THE FUTURE OF INFORMATION SCIENCES

FF press

5th International Conference
The Future of Information Sciences,
INFuture2015: e-Institutions – Openness, Accessibility, and Preservation
Zagreb, 11-13 November 2015

Organizer
Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

Publisher
Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

Printed by
Web 2 tisak d.o.o.

Impression
150 copies

All papers were reviewed by at least two reviewers. INFuture relies on the double-blind peer review process in which the identity of both reviewers and authors as well as their institutions are respectfully concealed from both parties.

ISSN 1847-8220

THE FUTURE OF INFORMATION SCIENCES

INFUTURE2015

E-INSTITUTIONS – OPENNESS, ACCESSIBILITY, AND PRESERVATION

Edited by

Karen Anderson, Luciana Duranti, Rafał Jaworski,
Hrvoje Stančić, Sanja Seljan, Vladimir Mateljan

Zagreb, November 2015

CONTENTS

Preface

This is the fifth publication in the series of biennial international conferences, *The Future of Information Sciences (INFuture)* organised by the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb. In 2015 the title of the conference is *INFuture2015: e-Institutions – Openness, Accessibility, and Preservation*, exploring issues surrounding the movement towards opening data and making the content globally accessible while at the same time recognizing the challenges of digital preservation.

The *INFuture2015* conference consists of 47 papers from 108 authors from all over the world – from the USA and Canada, Europe, Africa, all the way to Hong Kong. Through twelve chapters this publication follows the topics of the conference.

Keynote papers address the issues around open access initiatives, building trust and confidence in cyberspace, synergy between computer scientists and linguists, Croatian national strategy dealing with e-services and one of the results in that area – establishment of the national identification and authentication system in Croatia. The chapter on *e-Institutions* elaborates the issues of information and project management in the digital environment and social media. *Trust in the Digital Environment* analyses topics of data retention and disposition in the context of preservation of web-sites, records authenticity, legal aspects of digital preservation, hybrid curation and preservation of digitally signed records. The following chapter *Modern Information Systems* focuses on strategic planning and organisation of integrated business information systems, context-aware computer systems, and knowledge management systems. *Human-Computer Interaction*, *Language Technologies and Applications* is the chapter where authors analyse online summarising and translation technology as well as the accomplishment of interoperability between centuries-old dictionaries. SaaS quality management and data warehousing are investigated in the chapter *Big Data Analytics and Applications* while chapter *Multimedia Systems and Applications* address important aspects of multimedia instructional design, MMO games and virtual exhibitions. This is followed by the *Mobile Solutions* chapter where wireless network security and Twitter usage among startups in Europe are investigated. The *INFuture2015* conference has also opened a new topic – *Health Informatics*. It focuses on the development of e-Health services in the EU, usage of the NFC-solutions for secure medication, anonymization and pseudoanonymization of psycholinguistic research data as well as health-related life-long learning. The chapter *Technological Solutions for Society and Education* covers a broad range of topics including scientific communication, analysis of learning outcomes, development of reading skills, motivation and learning

processes using mobile phones, higher education institutions and social networks, all the way to the usage of online sources. It ends with the announcement of future scientometric and bibliometric analyses of the *INFuture* conferences. The following chapter provides information on the two *workshops* organised during the conference – one dealing with the issue of maintenance of authenticity and integrity of electronic records utilising digital timestamps, and the other analysing and presenting the initial maturity model for information governance arrangements in organizations, developed at the EU funded E-ARK project. The last chapter, *Innovative technologies,* provides information about PIQL – an innovative solution combining microfilm and digital records in one long-term preservation medium.

We believe that the results of the research presented in this publication will help us open our data, make content widely accessible and preserve authentic digital records. We also hope the research presented here will help you broaden your research, envision new projects, establish cooperation as well as inspire you to pursue new goals.

Editors

KEYNOTE PAPERS

University of Zagreb and Open Access Initiatives

Damir Boras
Rector
University of Zagreb
Trg maršala Tita 14, Zagreb, Croatia
rector@unizg.hr, dboras@ffzg.hr

Those who worked at universities in Croatia in the 90s, in the era before the World Wide Web, probably remember the internet application named *Gopher* which allowed access to hierarchically-organized text files that could be brought from university servers from all over the world to a text viewer application on a computer of a scientist, scholar or student. About 30 years ago higher education institutions relied on text-only information on the internet to communicate with the public and other scholarly institutions. The internet connection that was used by many teachers and researchers at Croatian colleges at that time was a 24k or 56k modem and a telephone line. The idea of open online access to published research papers, textbooks and teaching material was not shared even among the early adopters of internet technology at our universities. Books and journals were in paper format and either purchased or accessed through the National University Library (NSK) or specialized libraries at research institutes, colleges like the School of Medicine or the Faculty of Electrical Engineering and Computing, or small departments like Psychology or Information Science at our Faculty of Humanities and Social Sciences (traditionally known as Faculty of Philosophy).

Today, scholars, students and the public can use university web sites to access vast amounts of information about many areas of life and work at higher education institutions. Easily accessible web services like *Google Scholar* and *Google Books* complement the use of specialized libraries and online databases like *Scopus*, *Web of Science*, *Ebsco* or *Science Direct*. Electronic book readers like *Kindle* and tablet computers enable students to carry hundreds of books in their back-packs. Also, at any time students can use their smartphones to access college websites, university online services or e-courses from virtually any location.

It was 15 years ago that the scholars at the *Massachusetts Institute of Technology* (MIT) began the *OpenCourseware* project to fulfill the MIT's mission "to advance knowledge and educate students". In 2002 they had teaching material for 50 courses publicly available online on their website and in less than 5 years a total of 1250 MIT courses were sharing their teaching material with scholars and students all over the world.

In the area of information systems and services at the university level there is a lot of importance placed on the development, maintenance and protection of administrative, academic and research information resources. The concept of

academic *e-infrastructure* is sometimes extended to the idea of *e-university* which contributes to university vision and mission by advancement of ICT that supports institutional processes and enables academic analytics. Most of the related ICT and e-services are commonly provided by university departments for information systems or e-infrastructure. At the University of Zagreb this role is performed by the *University Computing Centre* (SRCE).

The principal "customer" of a university is the student. This customer is interested in acquiring a qualification for a career after graduation. For some of them it is teaching excellence of professors that motivates them to enroll a specific college, while for others the research reputation of the institution matters as well. The various aspects of college infrastructure and campus life are often used for rankings of universities which are available to prospective students. The information about the quality of teaching at the University of Zagreb is less available to our prospective students if course descriptions and teaching materials are not placed on our university web domain. The research papers of college departments are harder to access by the students and if they cannot be found on college websites. The scientific production of researchers can more easily be found by other scholars and perhaps cited if it can be accessed directly on the open web.

The research quality and quantity of universities is annually measured and universities are ranked on lists like *Academic Ranking of World Universities* (ARWU) of the Shanghai University, *CWTS Leiden Ranking*, and *QS World University Rankings*. In 2015 the ranking of the University of Zagreb has slightly improved regarding its research output. However, for the last 10 years universities have also been ranked by their *web presence* in form of the *Ranking Web* or *Webometrics* ranking of higher education institutions. On this measure of performance of universities from all over the world, based on their web presence and impact, the University of Zagreb is currently holding the 458 position.

The goals of the *Webometrics* ranking are to promote publication on the web and support Open Access initiatives. More importance is also placed on free and unrestricted internet access to scientific publications and other academic material, as well as to information about various activities of professors and researchers, including informal scholarly communication. Like the MIT *OpenCourseware* project, *Webometrics* ranking is intended to provide better access to scientific knowledge not only to students and researchers located in developing countries, but also to regional third parties like industry and government.

Croatian universities, our academics and researchers should contribute to the *Open Access* and *Open Educational Resources* initiatives. For instance, the Senate of the University of Zagreb has supported the *Croatian Declaration on Open Access*. However, greater engagement of higher education institutions, researchers and teachers is needed to make the contributions of our universities more available to the society and general public. This is one important aspect of the e-university we are trying to implement. This one possible contribution of

our higher education institutions to our society and students. By such initiatives our position among world universities may improve and our reputation enhance. The leap from *Gopher* application to *Webometrics* ranking in less than 30 years, the change from printed scholarly books and journals behind library walls to *Open Access* and *e-university*. We have an exciting academic INformation Future of INFuture before us and we just have to be more engaged to make it happen.

Building trust and confidence through sustainable information systems research: towards a common future

Karen Anderson
Department of Archives and Computer Science
Mid Sweden University
Universitetsbacken 1, 871-88 Härnösand, Sweden
Karen.Anderson@miun.se

Summary
The Information Systems research group at Mid Sweden University is the result of four separate groups, spread across three disciplines, artificially thrown together because we didn't seem to fit into other developed research groups. What we were seen to have in common was 'information'. The disciplines are: Archives and Information Science, Business Management and Organization, and Informatics.
Shortly after this uncertain start, the University launched a demanding internal process for Assessment of Research and Coproduction (ARC13) over the previous five years. We didn't have that much past together, so we set out to build ourselves a future, exploring our common interests as we gathered data for the evaluation. We pooled our research, hired a facilitator to help with a SWOT analysis, built trust within the group and achieved a remarkably good report from the external assessors. Participation in the multidisciplinary InterPARES Trust in Digital Records in an Increasingly Networked World Project has been an inspiring factor in the Archives and Information Science contribution to our new collaboration. Our Information Systems research group now focuses on five research challenges: Information value, quality, accessibility, sustainability and trustworthiness. Now we are having fun exploring the possibilities, confident that our InfoSys research group has its own sustainable future.

Keywords: sustainable information, information systems, social sustainability, archival science, trust, recordkeeping informatics

Introduction

Founding a research group is never an easy undertaking. It needs a group of likeminded, well-qualified researchers with similar or complementary interests. They should be willing, and preferably very keen to work together. Within a university, this encompasses not only the research they want to do, but also participation in the administration of the research group, and promotion of its interests within the life of the university. Successful researchers also teach from their

research, which in turn requires cooperation and coordination to review and continually update existing teaching programmes, or development of new courses and programmes. In order to be successful in these endeavours, research groups need to build a trusting environment in which to work and through that environment develop confidence in their own standing and ability to weather the challenges that they meet.

Purpose

The purpose of this paper is to share the experience of four small research groups within Mid Sweden University who became one, the InfoSys research group. It will reflect on why our melding was seen to be necessary and how the group has worked together to achieve our successes so far, identifying a theoretical ground that can frame both our teaching and research.

The background and our challenge

In 2010 three independent, discipline-based groups were challenged to combine to become one new research group by the then head of the Institute for Information Technology and Media (ITM) at Mid Sweden University. The groups were Archives and Information Science, Informatics, and Business Management and Organization. The reasoning was that all three were very much smaller than other, longer-established research groups within ITM, such that none were deemed to have sufficient critical mass to succeed alone and that all had a common interest in 'information'. Until this point each group had worked hard and made a sound start to developing their own cohesive research focus so this decision came as something of a shock. To add to the challenge, the three groups were geographically separate, spread across three campuses in the middle of Sweden: Härnösand, Sundsvall and Östersund.

The foundational groups

The disciplines within which the research groups had developed were:

- Archives and Information Science at the Härnösand campus, which in 2009 had formed the CEDIF research group – the Centre for Digital Information Management, bringing together professional and research expertise in recordkeeping (archives and records management), business process analysis, systems design, information architecture, and long-term digital preservation.
- Informatics at the Östersund campus had formed the CRIINFO research group - Critical Information. CRIINFO focused on design of systems for accessibility of critical information undertaking applied research in close cooperation with stakeholders in order to create benefit.
- Business Management and Organization at the Sundsvall campus at that stage was a single research group known as ValIT (Value creation with

> IT). ValIT focused on how value is translated, enacted and performed in IT-related processes and evaluation of these processes. At this stage they were still a single group, but by 2013 ValIT had defined their interests into two separate groups: CRIDIT – Critical Studies in Digital Information Technology and CIE – Computational Industrial Engineering.

Each of the partner groups had a unique position in a small and specialized niche: we had to start by getting to know each other and our various specializations. We instituted a series of monthly seminars, presenting two papers at each seminar, nominating two critical readers for each paper and circulating the chair. The university's videoconference facilities were used to ensure that as many as possible could participate. We also initiated a series of less frequent face-to-face meetings with the aim of trying to build a common identity, the first being held in October 2010, where we agreed on a name for the group that covered our common interests: Sustainable Information and Information Systems (SIIS). Since then we have held regular bi-annual meetings, which have become crucial to our development.

At this time ITM was in the process of developing a research strategy, a vision that could be presented to the public and research funding organizations and a development plan for our own measurable goals for evaluation purposes. Although its constituent campuses each had a considerably longer history, Mid Sweden University is celebrating 10 years as a university in 2015, so we were participating in building both ITM's and the new university's research identity. By the end of January 2011, we had our own vision statement:

> We shall be acknowledged and leading within sustainable information and information systems (SIIS). This will be achieved through innovative multi-disciplinary research which creates new perspectives and values. We use our double affiliation as both technologists and social scientists to nuance and challenge taken for granted ways of understanding information and information technologies, their development and use. We view technologies not only as separate artifacts but as contexts of information, people, technology, design, manufacturing and use, organizations and use practices. We also focus on how technologies achieve value and meaning in various use contexts, and contribute to conditions for sustainable ways of living and working for people, organizations and communities.

This statement was extrapolated into a position statement paper (Truex et al 2011), in which we made some first steps towards defining our understanding of the relation between the information systems field and the notion of sustainability. However, several members of the group expressed a need for caution, concerned that it was too early and we had not had time for each group to thoroughly explore the issues from their own perspectives before attempting to

write together. Nevertheless, we decided to go ahead and submit the paper as a first step towards developing our vision and stance.
From CEDIF's perspective, sustainability is central to archival and information science. In this field sustainable information is about recorded information (records) that can maintain authenticity, reliability, integrity, and still be usable over a long time, possibly forever. The records of road and rail transport systems provide an important example, for which maintenance of accurate accessible records is extremely important for the safety of future generations and the infrastructure bequeathed to them. Embedded in archival theory is the need to preserve the record for both current need and future use, underpinning business needs and an open and democratic society, which makes the social sustainability (e.g. McKenzie, 2004) construct relevant. The design of information systems in which records are created and managed are a particular challenge for sustainability: few information systems are designed to support the long-term preservation and accessibility of records. Indeed, Upward et al (2014, pp.38-39) describe the discipline as being in crisis and in crucial need of a re-think to develop what they call "a single-minded approach".

Building trust and confidence

Writing that first paper for the IFIP International Working Conference was strongly advocated in good faith, but in hindsight it was not well timed. It served best to make us realize that we had a lot more work to do to fully understand the potential contributions and standpoints of all our component subgroups. However, at the time, we simply did not have the maturity as a group and the confidence to decline the offer.
Shortly after this uncertain start, the University launched a demanding internal process for Assessment of Research and Coproduction (ARC13) over the previous five years. We didn't have that much past together, so we set out to build ourselves a future, exploring our common interests as we gathered data for the evaluation. We pooled the research we had done as individual groups and hired a facilitator to help with a SWOT analysis, which was a required part of the process. Working together on a frank analysis, starting with the strengths and weaknesses of the individual groups and then moving on to what we saw as the strengths and weakness of the combined group proved crucial to building trust within the group. It also helped us to identify common interests and most importantly, what we saw as shared challenges, and thus a way forward for tackling those challenges. When the assessment report was published in late 2014, we were very proud to have achieved an overall assessment of 'Very Good' from the panel of international assessors – the second highest on a 5-step grading scale. The report by the international Evaluation Panel commended our achievements in the areas of quality, productivity, co-production including external non-academic productivity, and impact on society (ARC13 2014, pp. 213-220). They also gave constructive advice and encouragement to continue

the strategic planning begun while working towards this evaluation and to consolidate and stabilize achievements (p.219). This was a very helpful factor in building our confidence in our own achievements and standing within the university. An important recommendation was to establish a clear leadership structure for the group. Planning our work to meet the requirements of the ARC13 process and writing the self-assessment for presentation to the Evaluation Panel had been entrusted to a group of three to ensure fair representation of all three of out component groups and all three campuses. The group consisted of Karen Anderson (Professor in Archives and Information Science), Katarina Lindblad-Gidlund (Associate Professor in Business Management and Organization) and Lena-Maria Öberg (Senior Lecturer in Informatics). At a meeting in 2014 the InfoSys Group discussed leadership possibilities and voted to maintain and confirm this representative leadership model. The advantage of this leadership troika is that it has good access and support across the whole group, which in turn facilitates not only leadership but also the group's capacity to implement and deliver upon agreed decisions.

Challenges facing the Information Professions

The digital turn in society brings both opportunities and problems. Information technology has made e-service and e-business development possible but raises also problems relating to democracy, the leadership of these new processes and new roles and working methods. In 2010-11 the CEDIF group undertook a project to benchmark information management practice in Sweden, using a web-based survey, focus groups and interviews (Anderson, Samuelsson & Morner Jansson, 2011). In one of the focus groups we conducted, information architects from large government departments noted the following competence deficits as being among the greatest challenges:

- Lack of relevant competence to create information and concept models in order to secure the quality in the information structures;
- Lack of relevant competence for work with enterprise business information modelling;
- Lack of alignment among different competencies and roles within the information management area leaves gaps and deficits in necessary competency levels.

Some archivists who were interviewed at the same time expressed a need for more knowledge and skills in the IT arena to enable them to better understand and participate in digital information management, particularly noting a lack of competence in enterprise-wide information modeling and development of strategies for ensuring quality in information structures, systems development and business process development. The survey also found that there was a lack of understanding of the complexity of managing information over the long-term, often because the focus was on IT solutions and under-using the professional

knowledge of the archivist. It is encouraging to note that there are now an increasing number of calls in Sweden for sustainable IT design. In our 2010 survey we asked specific questions about this, but found that at that time there was little action or awareness of sustainability issues in the field of information management.

Beginning in 2011 the National Archives of Sweden worked on a very large project (the e-ARD Project) to design and implement a set of common specifications for government agencies that would facilitate and standardise transfer of records to an e-archive. Although the project was successfully completed in 2014, it became clear as it progressed that there are very few archivists in Sweden who have sufficient competence to either design or to implement such standards in their organisations. Disturbingly, the National Archives does not have sufficient resources to provide training that will ensure implementation of the specifications across all public organizations.

In the Swedish job market, there are an increasing number of positions for so-called "IT-archivists", signalling employers' requirements for archivists capable of working with IT professionals and who have the appropriate competence to design, implement and maintain systems that capture and manage records and their flow into the archive, ensuring long-term preservation. It is also an indicator of increasing frustration with the lack of general technological competence among those who currently work as archivists, while information governance, quality recordkeeping strategies for current business, retrieval and re-use, the needs, opportunities and threats presented by our digital society receive less attention than preservation strategies. This should be a matter of deep concern for the profession, given that this leaves an opening for other professions to move into archivists' professional arena, a problem identified by Maria Kallberg in her research. Having identified a lack of recordkeeping awareness in both the political and managerial arenas, she noted: "As researchers within the field of archival science we should strive to communicate our results with other research communities, such as the information system technology field. Research findings indicate that IT professions are increasing their professional positions and influence within the strategic e-government work. … There is still a core of knowledge that is exclusive for this field … [that] needs to be nourished and explored in order to find new models, but furthermore, acknowledge and secure the societal function of archives in a broader perspective, as sources to serve the whole society and not only to serve organisations' business needs" (Kallberg 2013, pp. 123-4).

Reflecting on these issues we saw that education for information professionals needs a major overhaul. Our vision can be compared to that of Oliver et al (2009; 2010), who pointed out that records and archives management faces an imminent crisis and argued for an overhaul of the discipline, proposing what they call "recordkeeping informatics" and urging active engagement in reshaping and broadening philosophy and approach (2009a, p.18). They advocate

- Exploration of the technical and social environments facing organisations and their risk-related responses;
- Undertaking business analysis to understand the information architectures that intersect in sustainable ways with the enterprise architectures;
- Flexible and responsive frameworks to facilitate access to and rendering of authentic records and information in a secure and trustworthy manner;
- Understanding the information culture within organisations;
- Embracing a service oriented approach (Oliver et al 2010).

Furthermore, the InfoSys Groups shared research seminars also prompted us to begin exploring the issue of sustainability in the digital information society. Sustainability is usually seen as having three dimensions: environmental, economic and social. All are important, but it is the social aspects of sustainability that mesh best with our disciplinary interests. Yet Cuthill (2010, p. 363) notes that although there has been some exploration of concepts surrounding social sustainability, there does not appear to have been a broader academic response. Cuthill proposes a social sustainability framework "which describes an interdependent and self-reinforcing relationship between four key components: (1) social capital, (2) social infrastructure, (3) social justice and equity and (4) engaged governance".

Sustainable information systems and strategies for provision of sustainable information, i.e. information that is authentic, reliable and accessible over the long term, contributes to and underpins all four of Cuthill's components. Sound information governance and infrastructure, for example, are essential for supporting social justice and equity in a democratic society.

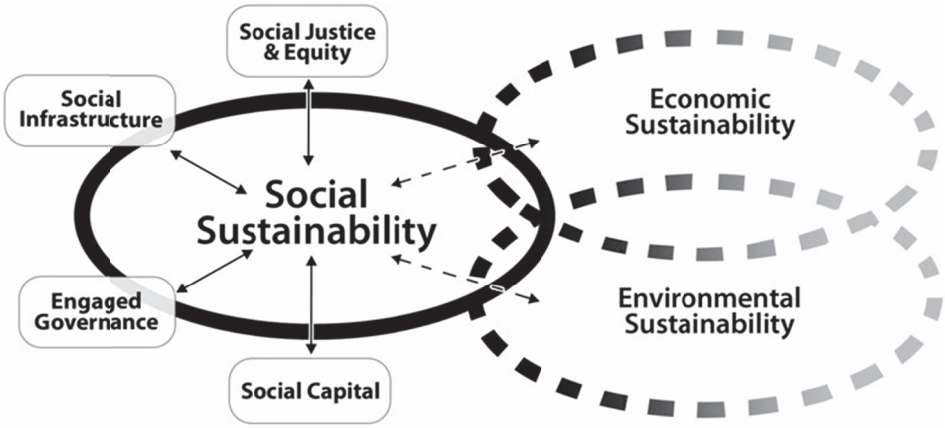


Figure 1: Conceptual framework for social sustainability (Cuthill 2010, p. 366)

Common ground

At a meeting in April 2015 the InfoSys Group affirmed five key research challenges: Information value, quality, accessibility, sustainability and trustworthiness. Each of our constituent groups presented their own theoretical and discipline-focused strategies for subscribing to these five challenges and archives and information science found no difficulty in matching them to our past and present research and teaching. As well as supporting the objectives of our regionally funded research projects, these key challenges also reflect the objectives of the multidisciplinary, international *InterPARES Trust in Digital Records in an Increasingly Networked World* Project, membership of which has been an inspiring factor in the CEDIF Group's contribution to our InfoSys collaboration. The *InterPARES Trust* Project's research goal focuses on conditions and strategies for ensuring public trust in digital records entrusted to the Internet grounded on evidence of good governance, a strong digital economy, and a persistent digital memory[1]. Thus we draw on our international research partnership as well as the combined academic disciplines assembled within the InfoSys group, well-placed to explore and research the relationship between social sustainability and sustainable information, and to educate professionals who will be capable of meeting these challenges.

A new Master degree in Sustainable Information Provision

All three of the discipline groups who are partners in the InfoSys group teach at undergraduate level, and all offer research education at doctoral level, but currently only Archives and Information Science offers a master program, which was identified as a structural problem in our education program. Better coverage of our disciplinary interests at master level is needed for several reasons. Our research and shared experience between 2010 and 2014 made us increasingly aware of a need for a new approach to master level education across our disciplines to provide graduates with the knowledge and competences urgently required in our increasingly networked information environment. We are committed to ensuring that our research should always have a strong connection to our education programs at all levels, since we see informing our teaching through our research as one of our most important tasks. And to complete the circle, we need to create a new pool of Master degree graduates who will be particularly well-qualified to research with us as doctoral candidates in the future.

This led us to propose a new Master degree in Sustainable Information Provision, with a special focus on management and design of information services, for which the first student intake will be in 2016. Students will graduate with one of three specialisations, however eight of the courses in the two-year pro-

[1] InterPARES Trust https://interparestrust.org (Accessed 12 October 2015)

gram will be common to all three streams. Of these, three are existing archival science courses covering Archival Theory, Information and Records Management and Current Issues in Archival Research. Three more will be new, jointly developed courses: Sustainable Information, Sustainable Architecture, and Design and Use of Information Systems. Courses contributed by the Informatics and the Business Management and Organization disciplines include Business Process Analysis, Change Management, Knowledge Management in Distributed Environments, Organisational Culture and Change, and Information Sharing. There will also be one project course in which students will work in groups comprising representatives of each specialisation, making nine common courses in all. In their final semester students will write a thesis oriented to their specialisation.

Our aim is to provide students with a deeper knowledge of information management that is fit for purpose in the evolving digital society. Students will also gain skills for managing projects across distributed work environments: essential for leading and designing information management solutions that are sustainable for multiple audiences and for achieving sustainable information over time. This will require specific knowledge of methods based on inclusive, democratic and participative principles.

Conclusion

The InfoSys Research Group's strategy of maintaining the identity and honouring the knowledge and expertise of each of our component subgroups, at the same time as we knit ourselves together has paid dividends. We have determined that strategic cooperative development will provide our best opportunities for future development of shared research and education, trusting and honouring the value that each discipline brings to the others rather than attempting to homogenise ourselves and water down our expertise to the lowest common denominator. In doing so, we believe we can make a real contribution to the future sustainability of our research group, to educating for sustainable information provision and in so doing, supporting a sustainable digital society.

Acknowledgements

I would like to thank my colleagues in the InfoSys group, particularly Katarina Lindblad-Gidlund and Lena-Maria Öberg. Working with them is always uplifting and they have provided the inspiration for this paper.

References

Anderson, Karen; Samuelsson, Göran; Morner Jansson, Marie. Benchmarking Information Management Practice and Competence in Swedish Organizations. // *5th European Conference on Information Management and Evaluation.* University of Insubria, Como, Italy. 8-9 September, 2011

ARC13: Assessment of Research and Coproduction at Mid Sweden University 2013. Sundsvall, Mid Sweden University: 2014. ISBN 978-91-87557-72-9 http://www.miun.se/siteassets/forskning/arc13/arc_bok_web_20140613.pdf

Cuthill, Michael. Strengthening the 'Social' in Sustainable Development: Developing a Conceptual Framework for Social Sustainability in a Rapid Urban Growth Region in Australia. *Sustainable Development.* 18, (2010). 362-373

Evans, Joanne; Reed, Barbara; Linger, Henry; Goss, Simon; Holmes, David; Drobik, Jan; Woodyat, Bruce; Henbest, Simon. Winds of Change. // *Records Management Journal*. Vol. 24 (2014), Iss 3; 205 – 223 http://dx.doi.org/10.1108/RMJ-01-2014-0006

Kallberg, Maria. 'The Emperor's New Clothes' Recordkeeping in a New Context. Doctoral Thesis. Sundsvall, Mid Sweden University (2013). ISBN 978-91-87557-21-7

McKenzie, Stephen. Social Sustainability: Towards Some Definitions. Magill, South Australia: Hawke Research Institute, University of South Australia: 2004

Oliver, Gillian; Evans, Joanne; Reed, Barbara; Upward, Frank. Achieving the Right Balance: Recordkeeping Informatics, Part 1. iQ, November (2009). 18-21

Oliver, Gillian; Evans, Joanne; Reed, Barbara; Upward, Frank. Achieving the Right Balance: Recordkeeping Informatics, Part 2. iQ, February (2010). 42-44, 52

Truex, Duane; Olsson, Leif; Lindblad-Gidlund, Katarina; Sefyrin, Johanna; Larsson; Aron; Nilsson, Olof; Anderson, Karen; Borglund, Erik; Asproth, Viveca. Position Statement: Sustainable Information and Information Systems (SIIS) // *Governance and Sustainability in Information Systems. Managing the Transfer and Diffusion of IT. IFIP WG 8.6 International Working Conference, Hamburg, Germany September 22-24, 2011, Proceedings*. Nüttgens, N.; Gadatsch, A.; Kautz, K.; Schirmer, I., Blinn, N. (eds). Heidelberg: Springer (2011); 306-309

Upward, Frank; Reed, Barbara; Oliver, Gillian; Evans, Joanne. Recordkeeping Informatics: Refiguring a Discipline in Crisis With a Single Minded Approach. // *Records Management Journal*. Vol. 23 (2013) Iss 1; 37 – 50 http://dx.doi.org/10.1108/09565691311325013

Cyberspace: A Communal Place or a Place of Separation?

Luciana Duranti
School of Library, Archival, and Information Studies
University of British Columbia
Vancouver, Canada
luciana.duranti@ubc.ca

Summary

Cyberspace, or the 5th dimension, is regarded by most as a virtual space, though – as P. J. Rey states –"real-virtual dualism is nothing more than a fiction." The idea, he contends, comes from a refusal to accept the physical extension of digital information in computer terminals and other machines. Cyberspace is a physical space of separation between those who provide information and those who access it. Yet, many perceive it as a communal space where documentary facts, actions and memories are shared. As a result, policies that attempt to regulate the activities carried out in cyberspace, take two opposite directions: on the one hand they support open access, open data, transparent and open government, redundancy, and permanent preservation, and on the other hand they proclaim the right to oblivion, or to be forgotten, the right to privacy, the duty to confidentiality and secrecy, expungement of data or elimination of links, jurisdictional control of storage locations, as well as destruction of records after they have served their usefulness for the purposes they were generated. This keynote will discuss these directions in light of the findings of international research.

InterPARES Trust (ITrust) is an interdisciplinary collaborative research project which explores issues concerning digital records entrusted to the Internet, with the goal of generating the theoretical and methodological frameworks to develop local, national and international policies, procedures, regulations, standards and legislation, in order to ensure public trust grounded on evidence of good governance, a strong digital economy, and a persistent digital memory. ITrust is a research partnership that comprises over fifty universities and organizations, national and multinational, public and private, in North America, Latin America, Europe, Africa, Australasia, and Asia. Its researchers are experts in archival science, records management, diplomatics, law, information technology, communication and media, journalism, e-commerce, health informatics, cybersecurity, information governance and assurance, digital forensics, computer engineering, and information policy.

The studies carried out by the ITrust project have clearly shown that, whereas government and corporate bodies' policies can be ambiguous in terms of open-

ness and restrictions, non-institutional/corporate users threat cyberspace as a true communal space, and this view of its function can be very powerful. This paper will discuss some of the 70 ITrust studies that reveal how users regard the Internet and identify some of the consequences.

Keywords: cyberspace, information providers, users, permanent preservation, right to be forgotten, InterPARES Trust

Approximate sentence matching and its applications in corpus-based research

Rafał Jaworski
Adam Mickiewicz University in Poznań, Poland
rjawor@amu.edu.pl

Summary

This article presents the technique of approximate sentence matching. It is known to perform well as an aid in linguistics-related tasks, among others in translation. The article presents two different applications of this technique in author's algorithms for concordance searching and sentence clustering.
Technical part of this article is succeeded by conclusions regarding the synergy between computer scientists and linguists. It is necessary for these two groups to overcome communication obstacles and work together to create linguistic tools of the future.

Keywords: approximate matching, corpus, searching, clustering, synergy

Introduction

Approximate sentence matching (ASM) is a technique of retrieving sentences from a large corpus that are similar to a given pattern sentence. One of the most important challenges in ASM is finding a good sentence similarity measure, which would reflect human intuition of sentence resemblance. Another challenge is designing a robust and scalable algorithm, capable of retrieving similar sentences from a large text corpus, using a given similarity measure. This article presents an author's algorithm for approximate sentence matching and sentence similarity computation. Its implementation is based on suffix arrays augmented by custom designed auxiliary data structures. In order to measure the sentence similarity it uses an improved version of Jaccard index (see [Jaccard, 1901]).

The second part of the article focuses on applying the sentence similarity measure in the task of building parallel corpora for under-resourced languages. This is achieved by selecting the most valuable sentences of a monolingual corpus by the means of cluster analysis.

Work on the development and testing of the above algorithms required extensive cooperation with linguists. The last part of this article lists observations and conclusions regarding this work.

Approximate sentence matching in CAT

As fully automatic translation often falls short of expectations, people turn to other methods of using computers in the translation process. One of such methods is Computer-Aided Translation (CAT). CAT systems use sets of previously translated sentences, called the translation memories. For a given sentence, CAT system searches for a similar sentence in the translation memory. If such a sentence is found, its translation is used to produce the output sentence. This output sentence is then used as a suggestion for translation. A human translator carries out the post-editing. This technique is used in many CAT platforms, such as SDL Trados [SDL, 2015] or Kilgray Memoq [Kilgray Translation Studios, 2015].

It is crucial for a CAT system that the translation memory contain translations of frequently used sentences. With such a translation memory, a CAT system generates good suggestions for translations, reducing the amount of human work on post-editing.

Author's approximate sentence matching algorithm

This section presents the author's solution for approximate sentence matching. It differs significantly from the approach described in the previous section. The most important difference is that author's solution allows to cover the searched sentence with more than one example sentence from the memory.

In order to carry out the search procedures efficiently, an offline index, based on the suffix array (see [Manber et. al., 1990], [Makinen et al., 2004]) and other auxiliary data structures was used.

Operations on index

Main operations performed on the index are the following:

- *void addToIndex(string sentence, int id)* This method is used to add a sentence to the index along with its unique id. The id is treated as additional information about the sentence and is then retrieved from the index by the search algorithm. This is useful in a standard scenario, where sentences are stored in a database or a text file, where the id is the line number. Within the *addToIndex* method the sentence is tokenized and from this point forward treated as a word sequence.
- *void generateIndex()* After adding all the sentences to the index the *generateIndex* method should be called in order to compute the suffix array for the needs of the fast lookup index. This operation may take some time depending on the number of sentences in the index. Nevertheless, its length rarely exceeds one minute (in reported experiments with 2 million sentences the index generation took 6-7 seconds).
- *simpleSearch(string pattern)* Basic search method takes a text fragment, tokenizes it and tries to locate all its occurrences in the index. The return

value is a list of matches, each holding information about the id of the sentence containing the fragment and an offset of this fragment in this sentence.

- *concordiaSearch(string pattern)* The main similarity search method returns the longest fragments from the index that cover the search pattern. This technique can be used for concordance searching.

Simple searching example

Functioning of the index generation and simple search algorithm is best illustrated by the following example. Suppose that the index contains the sentences presented in Table 1. Note that the id's of the sentences are not consecutive as there is no such requirement.

Table 1: Example index.

Sentence	ID
Novel methods were used to measure the system success rates.	23
Various statistics, including the school success rate, were reported.	12
The research is still ongoing	259

Let us now perform simple search for the fragment "success rate" in the example index. The expected results (in the form [sentence id, offset]) are *(23, 8)* and *(12, 5)*. Returned results allow for quick location of the contexts in which the phrase "success rate" appeared. Note also that the system is expected to return the result *(23,8)* ("the system **success rates**") even though the word "rate" was in plural in the index.

Index construction

Author's index incorporates the idea of a suffix array and is aided by two auxiliary data structures – the **hashed index** and **markers array**. During the operation of the system, i.e. when the searches are performed, all three structures are loaded in RAM. For performance reasons, hashed index and markers array are backed up on the hard disk.

When a new sentence is added to the index via the aforementioned *addToIndex* method, the following operations are performed:

- tokenize the sentence
- lemmatize each token
- convert each token to numeric value according to a dynamically created map (called dictionary)

Lemmatizing each word and replacing it with a code results in a situation, where even large text corpora require relatively few codes. For example, research of this phenomenon presented in [Jaworski, 2013] reported that a corpus of 3 593 227 tokens required only 17 001 codes. In this situation each word could be stored in just 2 bytes, which significantly reduces space complexity.

The following sections will explain in detail the data structures used by the index.

Index – hashed index

Hashed index is an array of sentences in the index. The sentences are stored as code lists. For example, let us compute the hashed index for the sentences of the example index shown in Table 1 ("Novel methods were used to measure the system success rates." "Various statistics, including the school success rate, were reported." and "The research is still ongoing"). First, the sentences are tokenized, then the tokens are lemmatized and mapped into integers. Results of this process for all three example sentences are shown in Table 2.

Table 2: Hashed sentences.

Novel	methods	were	used	to	measure	the	system	success	Rates
novel	method	be	use	to	measure	the	system	success	Rate
1	2	3	4	5	6	7	8	9	10

Various	statistics	including	the	school	success	rate	were	reported
Various	statistic	include	the	school	success	rate	be	report
11	12	13	7	14	9	10	3	15

The	research	is	still	ongoing
The	research	be	still	ongoing
7	16	3	17	18

The dictionary created during this process is shown in Table 3.

Table 3: The dictionary

Lemma	**code**	**lemma**	**code**	**lemma**	**code**
Be	3	rate	10	success	9
Include	13	report	15	system	8
measure	6	research	16	the	7
Method	2	school	14	to	5
Novel	1	statistic	12	use	4
ongoing	18	still	17	various	11

Code lists obtained from the sentences are then concatenated in order to form the hashed index. A special code (referred to as EOS – end of sentence) is used as sentence separator.

The hashed index is used as the "text" (denoted *T*) for the suffix array.

Index – markers array

When a fragment is retrieved from the index with the help of a suffix array, its position is returned as the search result. However, this position is relative to the "text", stored as the hashed index. For example, if we searched for the fragment "success rate", as in previous examples, we would obtain the information about

two hits: one at position 8, and the other at position 16 (mind that the positions are 0-based and EOS characters count as single text positions).
This result does not contain information about the id of the sentence where the fragment was found nor the offset of the fragment in this sentence. Naturally, this information is retrievable from the hash index alone. However, operation of that kind would require searching the hashed index in at least *O(n)* time in order to determine which sentence contains the given position. In addition, this would only return the ordinal number of the sentence, not its id, since this information is not stored in the hashed index.
In order to overcome these difficulties, a simple, yet effective data structure was introduced. Markers array is used to store information about the sentence id and offset of each word. Technically it is an array of integers of the length equal to the length of the hashed index. Each integer in the markers array stores both the sentence id and the offset of the corresponding word in the hashed index. Current implementation uses 4-byte integers, where 3 bytes are assigned to store the sentence id and 1 byte is used for the offset. This means the index can store up to 16 777 216 sentences, each no longer than 255 characters (one position is reserved for the EOS character). For example, the pair: $id = 342_{10} = 101010110_2$, $offset = 27_{10}=11011_2$ is stored as the integer: $10101011011011_2=10971_{10}$.
Even though the markers array is not free from redundancy, the cost space occupied by this data structure is affordable on modern computers. The benefits of its influence on speeding up the search process are much more significant.

Index – suffix array
The last element of the index is a generated suffix array. It is constructed after the hashed index is complete. Technically, this structure is a classic suffix array for the hashed index. As stated in Section *Index – hashed index*, hashed index plays the role of the "text" (*T*), whose "letters" are dictionary codes of words.
Algorithm used for construction of the suffix array is an implementation of classic construction algorithm proposed by Manber and Myers [Manber, Myers; 1990], running in *O (n log(n))* time. It differs significantly from the naive approach (generating suffixes, sorting them and reading their positions) which runs in $O (n^2 log(n))$.
Sorted suffixes for the example hashed index result in the following suffix array: [0, 1, 2, 18, 23, 3, 4, 5, 6, 14, 21, 7, 16, 8, 17, 9, 11, 12, 13, 15, 19, 22, 24, 25, 26, 20, 10].

Simple searching
Searching of the index is done according to the classic suffix array search procedure. In order to make this possible, an input search phrase must first undergo the same procedure as every sentence being added to the index (lemmatizing and coding).

Let us demonstrate the search on the same example search pattern presented in Section *Simple searching example*. We are searching for the pattern "success rate" in the example index. The pattern is tokenized and lemmatized, thus transformed into a sequence of lemmas: 'success' 'rate'. These lemmas are then encoded using the dictionary generated during the creation of the hashed index (see Table 3). As a result, the search pattern has the form '9 10'.
By searching in the suffix array with the help of the hashed index we know that the phrase "success rate" can be found in the source text at positions 16 and 8. In fact, we also know that this phrase is present only at these positions, as follows from suffix array properties. However, we expect that the final search results will be given in the form [sentence id, offset]. For that we need to check the markers array. In this example *M[8]=(23,8)* and *M[16]=(12,5)*. This corresponds to the expected final results.

Concordia searching
Concordia search internally uses the simple search procedure but serves for a more complicated purpose. It is aimed at finding the longest matches from the index that cover the search pattern. Such match is called "matched pattern fragment". Then, out of all matched pattern fragments, the best pattern overlay is computed.
Pattern overlay is a set of matched pattern fragments which do not intersect with each other. Best pattern overlay is an overlay that matches the most of the pattern with the fewest number of fragments.
Additionally, the score for this best overlay is computed. The score is a real number between 0 and 1, where 0 indicates, that the pattern is not covered at all (i.e. not a single word from this pattern is found in the index). The score 1 represents the perfect match - pattern is covered completely by just one fragment, which means that the pattern is found in the index as one of the examples. The formula used to compute the best overlay score is shown below:

$$score = \sum_{fragment \in overlay} \frac{length(fragment)}{length(pattern)} \cdot \frac{\log(length(fragment)+1)}{\log(length(pattern)+1)}$$

According to the above formula, each fragment covering the pattern is assigned base score equalling the relation of its length to the length of the whole pattern. This concept is taken from the classic Jaccard index (see [Jaccard, 1901]). However, this base score is modified by the second factor, which assumes the value 1 when the fragment covers the pattern completely, but decreases significantly, when the fragment is shorter. For that reason, if we consider a situation where the whole pattern is covered with two continuous fragments, such overlay is not given the score 1.

Let us consider an example illustrating the Concordia search procedure. Let the index contain the sentences from Table 4:

Table 4: Example sentences for Concordia searching:

Sentence	ID
Alice has a cat	56
Alice has a dog	23
New test product has a mistake	321
This is just testing and it has nothing to do with the above	14

Here are the results of Concordia searching for pattern: "Our new test product has nothing to do with computers":

Table 5: Concordia search results.

Pattern interval	Example id	Example offset
[4,9]	14	6
[1,5]	321	0
[5,9]	14	7
[2,5]	321	1
[6,9]	14	8
[3,5]	321	2
[7,9]	14	9
[8,9]	14	10

Best overlay: [1,5] [5,9], score = 0.53695

These results list all the longest matched pattern fragments. The longest is [4,9] (length 5, as the end index is exclusive) which corresponds to the pattern fragment "has nothing to do with", found in the sentence 14 at offset 7. However, this longest fragment was not chosen to the best overlay. The best overlay are two fragments of length 4: [1,5] "new test product has" and [5,9] "nothing to do with". Notice that if the fragment [4,9] was chosen to the overlay, it would eliminate the [1,5] fragment.
The score of such overlay is 0.53695, which can be considered as quite satisfactory to serve as an aid for a translator.

Sentence clustering

The idea of sentence clustering is applied to the task of preparing a specialized translation memory for a given purpose. It is based on the assumption that the most useful sentences that might appear in translation memory are those which occur most frequently in texts in the source language. A clustering algorithm classifies sentences from a monolingual corpus into clusters of similar sentences and selects one representative for each cluster. The translation for each representative is then produced manually by human specialists. The database prepared in such a manner forms a high-quality specialized translation memory,

covering a wide range of sentences intended for translation. This technique can be applied to prepare parallel corpora for under-resourced languages.

Clustering algorithm

Clustering algorithms work on sets of objects. They use a measure of distance between these objects in order to divide sets into smaller chunks containing objects which are "close" to each other (in terms of the sentence similarity measure). In our case, the set of objects is the monolingual corpus, the objects are sentences. In order to determine the distance between sentences we use two distance measures: "cheap" and "expensive" (the idea of using two distance measures is inspired by [McCallum et. al., 2000]). The "cheap" and "expensive" terms refer to complexity of calculations.

Cheap sentence distance measure

The cheap sentence distance measure is expected to work fast. It is based only on the lengths of sentences. The formula for computing the value of the measure is the following:

$$d_C(S_1, S_2) = 2^{-\frac{|len(S_1) - len(S_2)|}{10}}$$

where *len(S)* is the length of the sentence in characters.

Expensive sentence distance measure

The expensive sentence distance measure is expected to give more accurate assessment of sentence similarity. The first step of computing the distance between sentences S_1 and S_2 in this measure is to recognize named entities of the two sentences. Then, so called "penalties" are imposed for discrepancies between the sentences. The penalty values have been based on human translators' intuition, e.g. if one sentence contains a named entity and the other does not, then the sentences are not likely to be similar. (The values for penalties are bound to be calculated by self-learning techniques in future experiments).
Let us define:

- p – the sum of penalties imposed on sentences S_1 and S_2
- L_{S1} – the number of words and Named Entities in S_1
- L_{S2} – the number of words and Named Entities in S_2

$$d_E(S_1, S_2) = 1 - \frac{2p}{L_{S1} + L_{S2}}$$

The clustering procedure

The detailed clustering algorithm is:

IN: Set (S) of strings (representing sentences)
OUT: List (C) of clusters (representing clusters of sentences)
1. Divide set S into clusters using the QT algorithm with the "cheap" sentence distance measure (the measure is based only on sentence length).
2. Sort the clusters in the descending order by number of elements resulting in the sorted list of clusters, C.
3. For each cluster cL in list C:
(a) Apply the QT algorithm with the "expensive" distance measure (based on sentences contents) to cL, resulting in subclusters.
(b) Sort the subclusters in cL in descending order by number of elements.
(c) Copy the sentences from the cL into C

The QT algorithm is described in [Heyer et al., 1999]. The final step is performed by humans. They manually select the most valuable sentence from each clusters. The task is facilitated by a frequency prompt (the most frequent sentences always appear at the beginning of the cluster).

Evaluation

The clustering method presented above was applied in an experiment described in [Jaworski et al.; 2010]. The authors prepared a translation memory of about 500 sentences from Polish law texts and used it to augment existing memory of 20 000 sentences. The augmentation in size of the memory by merely 2.3% resulted in increasing the percentage of sentences found in the memory from 9% to 31% and decreasing the overall human effort in the process of translation by 12.5%.

Conclusion – need for synergy

The techniques presented above proved useful in experiments. Importantly, they were developed with the help of professional linguists. During this process, some subjects became clear. Firstly, linguists do not seem to know much about how computer software is created and which techniques are easy to implement and which are not. However, to be fair, computer scientists probably know even less about the translation process (see [Zetzsche, 2014]).

Moreover, the two groups are motivated differently – translators are primarily focused on the quality of their translation. They are, naturally, interested in optimizing their effort to achieve desired quality, but this is not their main concern. Computer scientists, on the other hand, are focused on the performance of their software and optimization of the efficiency of translator's work. Because of these differences, software developers and translators need to communicate with each other as often as possible.

The work model where a computer scientist comes up with ideas of improvement, implements them and presents to the linguist is actually one of the worst imaginable. Because of developer's lack of knowledge concerning the translation process and the different goals described above, solutions he or she sug-

gests are almost never perfectly useful for the linguist (which is what the developer hopes for). Instead, the linguist works hard to modify the developer's idea and the developer feels that his work has gone in vain, which leads to frustration on both sides.

Ideally, they should spend about 1-2 hours a week working together. They should exchange concepts and educate each other in their fields. The computer scientist should translate a document under supervision of the linguist. The translator should get accustomed with the architecture of the system he or she is using for their work. Ideas for new features in the software should be a result of their mutual thinking process. Only with this approach one can establish true synergy.

References

Heyer, Laurie; Kruglyak, Semyon; Yooseph, Shibu. Exploring Expression Data: Identification and Analysis of Coexpressed Genes. // *Genome Research*. 9 (1999); 1106-1115

Jaccard, Paul. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. // *Bulletin de la Société Vaudoise des Sciences Naturelles*. 57 (1901); 547-579

Jaworski, Rafał; Jassem, Krzysztof. Building high quality translation memories acquired from monolingual corpora. *// Proceedings of the IIS 2010 Conference/* (2010)

Jaworski, Rafał. Anubis – speeding up Computer-Aided Translation. // *Computational Linguistics – Applications, Studies in Computational Intelligence*. Springer-Verlag 458 (2013)

Kilgray Translation Studios. Memoq Translator Pro. https://www.memoq.com/memoq_translator_pro (8.05.2015)

Makinen, Veli; Navarro, Gonzalo. Compressed compact suffix arrays. // *Proc. 15th Annual Symposium on Combinatorial Pattern Matching (CPM) /: LNCS 3109 (2004), 420-433*

McCallum, Andrew; Nigam, Kamal; Ungar, Lyle. Efficient clustering of highdimensional data sets with application to reference matching. // *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (2000), 169-178*

Manber, Udi; Myers, Gene. Suffix arrays: a new method for on-line string searches. // *First Annual ACM-SIAM Symposium on Discrete Algorithms(1990)*. 319–327

SDL. Trados Studio. http://www.sdl.com/cxc/language/translation-productivity/trados-studio/ (8.05.2015)

Zetzsche, Jost. Encountering the Unknown. Part 2. // *Proceedings of the 17th Annual Conference of the European Association for Machine Translation EAMT2014*/ 2014, xvii

Croatian Strategy for the Development of Public Administration for the period from 2015 to 2020

Leda Lepri
Assistant Minister
Ministry of Public Administration
Maksimirska 63, Zagreb, Croatia
Leda.Lepri@uprava.hr

Introduction

The information and communications technology (hereinafter referred to as ICT) plays the leading role in the transformation of citizens' lifestyle, thus transforming the society as a whole. ICT as a generic term encompasses various types of information and communications technologies which enable the generation, storage, processing and exchange of information.

The potential of "the digital" for economic development and improvement of citizens' lifestyle is enormous, since everything is being recognized on an international level more and more quickly. Advanced public administrations have been exploring the ways of using this potential under their respective national circumstances.

Objectives

1. Improved business productivity of public administration through the use of ICT and new skills;
2. Enhanced quality of life through the use of public administration e-services;
3. Improved relationship between citizens and state administration through ICT use;
4. Insurance of secure environment for the provision of public administration e-services;
5. Increased competitiveness of economy through the use of public administration e-services
6. Opening up space for ICT-based innovations in public administration through the cooperation of public administration, scientific and business entities, and
7. Inclusion in the European Administrative Space.

Moreover, the quality and accessibility of public services in isolated and remote locations, especially on islands, has been lagging significantly behind other areas. There is a need to provide adequate services to the population, and the easiest way to do it is via the Internet. The problem of accessibility also concerns people with disabilities, vulnerable and socially endangered groups, including war veterans. Additional efforts are needed to adjust e-services to different

groups of people with disabilities and to adapt e-content in order to ensure the accessibility of the services for everyone. This will result in the creation of equal conditions and possibilities for the use of e-public services and the Internet for the entire population.

Pursuant to the Decision of the Croatian Government of 2 February 2012, the **Commission for the Coordination of Public Sector Informatisation** was established with the aim to rationalise the system and simultaneously enhance the quality of public services. The **National Information Infrastructure Act** (OG 92/2014) (hereinafter: the Act) was adopted by the Croatian Parliament in July 2014, as a response to the need to change the functioning of public administration and its relation towards citizens and businesses, and with the aim to ensure e-public services for citizens and businesses which are based on an integrated national information system.

In November 2014, the Croatian Government adopted the Decree Establishing the **Public Register for the Coordination of National Information Infrastructure Projects (Public Register ProNII).**

The Public Register ProNII was established for the purpose of rationalisation, steering of development and coordination of all activities and projects related to the application of national information infrastructure, simultaneously increasing the quality of public services and preventing the planning and implementation of the equal or similar projects in the public sector.

In order to solve problems in the state IT system management (in the sense of inadequate and insufficiently efficient cost and investment management in the IT sector), the Croatian Government has proposed that a single **Shared Services Centre** be established, thus setting up a single strategic location for the management and coordination of the development of the state IT; rationalising state IT expenses through supervision over the spending of budget resources; consolidating the integration of the IT systems of different state bodies, and enabling the use of a unified IT infrastructure. Consequently, public sector bodies could focus more on their basic activity and align their work with other public sector bodies by standardising common business processes through the unified IT system. Such a manner of work would lead to a new dimension in the transparency of the system and to creating new business possibilities for the private sector.

Snapshot

The current situation in Croatia is such that a vast majority of e-services is still at the maturity level 2, i.e. the level of one-way interaction. The reason for this is that until the summer 2014 there was no single point in the virtual world for interaction with citizens and businesses, so that every authority which wanted to provide personalised services needed to develop its own system of issuing mechanisms for the verification of eIdentity.

Almost all public sector bodies have developed e-services. On 10 June 2014 the eCitizens platform was launched with the three main components:

- Central Government Portal system,
- National Identification and Authentication System,
- Personal User Box system.

Problems of the dispersion of information and e-services, was resolved by introducing the system of a central government portal which integrates information and e-services in one place – the Central Government Portal which is realised within a single domain: **gov.hr.**

National Identification and Authentication System NIAS **is the single point of eIdentity verification for access to e-services.** The basic function of NIAS is secure eIdentification and authentication of e-service users.

Electronic credentials currently available through the NIAS system are as follows:

- ePass user name and password – a system developed for eCitizens;
- mToken – application developed for smartphones – a system developed for eCitizens;
- smart card of the Croatian Health Insurance Institute (CHII);
- user name and password from the AAI@EDU system – existing system of primary, secondary and higher education;
- HPB Token/mToken, and
- ePost (user name and password) – Croatian Post

The **Open Data Portal of the Republic of Croatia**, modelled on other similar projects in Europe and the world, **represents a catalogue of metadata** (data which describe sets of data more closely), and it helps users to get to desired open data easily. The portal was established at the address https://data.gov.hr.

The interoperability system in the Republic of Croatia is still in development. It should ensure interconnection of information systems developed on the principle of silos into a single system which will ensure the surpassing of legal, organisational, technical and semantic barriers in the development of national information infrastructure.

The **ProNII Public Register** serves to build interoperable solutions and to rationalise expenses in the development of public sector information systems in such a way that it guarantees control over projects, and on that basis decisions are made on joint project implementation, while preventing the planning and implementation of overlapping projects.

The **public register Metaregister** is also in place, functioning as a collaboration tool for the development of the system of interconnection of public registers. Apart from the entry of all the existing ways and possibilities of interconnection, the Metaregister ensures coordination during changes in the system of registers, and creates the so-called *referential integrity*.

Table 1: SWOT analysis of public administration ICT

STRENGTHS	WEAKNESSES
• Strong political commitment to building a modern, efficient and transparent public administration system, the aim of which is to ensure access for citizens and businesses to high quality services and coherent and reliable data through ICT • Existing strategic framework for the development of e-services and applications in sectoral strategies (and the Public Administration Development Strategy), with an additional advantage of the Digital Growth Strategy • Common rules for national information infrastructure development and implementation of projects related to the National Information Infrastructure Act will be defined through secondary legislation	• A relatively low level of e-services provision and use, and their low adaptability to the needs of individual users in relation to the EU average • Lack of a coordination mechanism for ICT projects in central government and local and regional self-government • Lack of a Chief Information Officer for strategic thinking and introduction of ICT and e-services in public administration bodies, who would answer directly to the head of a body
OPPORTUNITIES	**THREATS**
• Around 130m Euros from EU structural funds available in the period 2014-2020 for investments in public e-services development and public administration informatisation • Education and motivation of users for the use of e-services • Entry of new "digital natives", children raised with ICT	• Change in government policies and priorities related to the public administration modernisation and informatisation process, and consequent discontinuity in the implementation of government programmes and projects • Lack of IT experts in the labour market, especially in the public administration • Public fear over information security and personal data protection

e-Croatia 2020 – strategic objectives, activities and indicators

The goal of this Strategy is to ensure the development of the highest possible number of complex e-services oriented towards citizens and businesses in a standard manner. Complex, user-oriented e-services will be developed on an integrated national information system by respecting the basic principle of "only once" (the citizen delivers data only once, after that they are transferred), defined by the National Information Infrastructure Act, in cooperation of all central state administration bodies.

The strategic goal of the Strategy is to develop e-services required by citizens and businesses, and thus to increase the number of citizens using complex public administration e-services.

Activities of the Strategy with respect to external users – **The objective is** to increase the proportion of public e-services in the overall number of public administration services. The development of the e-services system will be based on the European Interoperability Framework. Special attention will be paid to

the political context as well as legal, organisational, semantic and technical interoperability.

Key areas and activities

The development of user e-services will go in the direction of the following key areas: rule of law and security; labour market and education; cultural and national heritage; health sector; war veterans; taxes and customs; agriculture and rural development.

The Republic of Croatia has chosen to establish a **Shared Services Centres** which will offer sharing of technical and business services. Shared resources will be available to a large number of users who will only have to take care of their own devices. This will enable dynamic scalability, operational speed and easy change of suppliers and services. The government cloud is a programme which will the ensure the use of a series of services in the cloud, and which represents a shift in the way we procure and work with information and communications technology.

Effects of the implementation of the Strategy

Implementation of the Strategy will primarily affect the life of citizens. Interaction with the public administration is not the purpose of the existence of citizens and businesses, so that it should be reduced to the necessary minimum, and their needs should be satisfied within the set time limit. Public administration informatisation will affect services by making them precise, legal, timely and transparent, and provided through the channel which is most convenient for citizens and businesses. By using e-services, the entire Croatian population is getting used to using new technologies and increasing their value in the labour market.

National Identification and Authentication System (NIAS)

Dario Belić
Director of IT Development Service
ICT Division
Financial Agency – FINA
Ulica grada Vukovara 70, Zagreb, Croatia
dario.belic@fina.hr

Summary

NIAS is an information-technological system for the central identification and authentication of e-services users at the national level with the ability to use multiple types of credentials of various issuers which are assigned different levels of security authentication.
All owners of credentials are allowed by NIAS safe use of NIAS connected electronic services.
At the same time, providers of electronic services are released from the issuing of credentials and user identity management.
During the authentication process all subjects or their information systems communicate with the NIAS system exchanging XML messages according to the SAML 2.0 standard.
When transferring SAML messages between subjects and NIAS, confidentiality and integrity are secured using the mechanisms of the electronic signature of SAML messages and SSL encryption.
NIAS provides Single Sign-On and Single-LogOut functionalities. NIAS is designed on the principles of the EU STORK project (Secure idenTity acrOss boRders linKed) respecting the already existing practices and accepted standards to ensure that in the moment of electronic connectivity with EU members the process is performed as simply as possible.
Cross-border connecting of NIAS system is currently in progress, based on the STORK standards, and also the expansion to e-business.

Keywords: NIAS, Credentials, STORK, SAML, Authentication, Identification, service Provider, identity provider, OIB

E-INSTITUTIONS

Effective Digital Information Management

Göran Samuelsson
Centre for Digital Information Management
Mid Sweden University
871 88 Härnösand, Sweden
goran.samuelsson@miun.se

Summary

Today, the increasing volume of digital information is used more frequently and over longer periods of time, increasing the need for structure. How can we improve the management of these digital information flows while ensuring the quality of the information over time? The status of information mangement in a number of public authorities has been analysed by conducting two studies. Effective information management has been defined and we asked the authorities in what areas they would consider help useful in order to become more completely digital authorities. Through interviews and questionnaires we have also identified future challenges. These challenges mainly involve generating information that communicates, that is to say, we need to increase interoperability between business systems and more long-term solutions for preservation (e-archives). This is increasingly important, as many authorities are becoming outsourcers, purchasing all services and products regardless of whether they are managing schools, railways or hospitals. This means that the benefit and value of the authorities is increasingly shifting towards managing and administering information – information that is often very long-lived. Another outcome of this review is identification of the need for a standardized, common information model shared by all information professions, regardless of whether one is a business architects, information security professional or an archivist. It is of utmost importance that we find techniques and methods to be able to work together to develop even more efficient sustainable information management.

Key words: information model, information arcitecture, information management, e-services, interoperability, e-archive

Introduction

In the last decade we have seen an increase and widespread re-use of information through various e-services. An important driver has been legal aspects in the finance sector, such as SOX and the European Basel III, collecting and gathering of information in the area of Environment and Health. Even the European PSI-directive, Public Sector Information from 1 July 2005 has actively influenced business to create more structured and complete digital information

flows for access (EU, 2015). There is also a more internal and business-driven reason for the growing amount of information. Business applications have been striving for full digital management and provision of online e-services 24 hours a day, seven days a week. The need for the business to support long-term customer and business relations requires information to be kept for many years. Using older data for statistics and analysis is a growing business, whether it is called Business Intelligence or Big Data (Digitaliser.dk 2015, E-delegationen 2013). In the near future, nearly all intrastructure and gadgets will include electronics, which in turn will be accessible online. This means that we are just at the beginning of an information avalanche. The storage industry expects the information to have grown by 50 times by 2020 all of which will require more storage space (EMC 21014)[1]. The only thing we can say about the future of everyday life with some certainty is that there is an increasing amount of information that will be used for longer by growing numbers of people. Therefore it is becoming increasingly important to order information well at an early stage. Previous studies have also shown that we need better coordination and new tools to be able to handle the new challenges posed by the greater and more complex amount of data and information (NAS 2010, Anderson 2011, Borglund 2010, Anderson 2010, Foscarini 2012). Not least as many authorities are moving towards becoming solely purchasers of services and products, and only manage and administer information about what should be or has already been done (Regeringskansliet 2011).

Purpose

An overall aim of this article is to compile the knowledge we in the Centre for Digital Information Management received through two studies carried out during the autumn of 2014 and the spring of 2015. Using this knowledge we aim to:

- provide an in-depth understanding of what public authorities consider as challenges in the scenario outlined above, where the amount of information is increasing and which, to a greater extent, is created by others, rather than the authority itself.
- compile this knowledge and define important research and development efforts to optimize and prioritize information management and administration over time.

Methods

We have carried out two different studies, one covers all public authorities in two Swedish counties – Västernorrland and Jämtland (Study I). The other study

[1] It is often the storage industry that produces this data. Studies that show the extent to which the large increase is due to more unique or widely copied or overlapping information is missing.

was conducted at The Swedish Transport Administration (Study II). Both studies were conducted in the same way. The work was divided into two stages: the first stage consisted of a description of the current situation and an analysis based on reading public authorities' internal and external reports and articles, in addition to a number of interviews and focus groups with the aim of gaining an understanding of important information flows and challenges. The second part consisted of an analysis where we, through literature and interviews, tried to get an idea of which information flows and challenges could benefit from a more in-depth analysis and structured management to contribute a significant social benefit and possible research tasks.

Study I began with workshops with key personnel, in which the term information management was defined and operationalized. These concepts and definitions have then been used as the basis of a questionnaire sent to all units of public authorities (state, counties and municipalities) in the region, a total of 50 authorities, out of which 44 responded. Not all responses were complete; on average, each question was answered by 29 authorities (16 governmental, 13 municipalities and county councils, giving a response rate of 58 percent). The web survey tool used was Netigate, which provided a summary and graphic representation of all responses.[2]

Study II was an in-depth analysis of one of the country's largest public authorities. A series of focus groups that included various professions – business architects, system administrators, controllers, people working with information security etc. (in total, approximately 30 people) gave an overview of the challenges the authorities face. A list of questions was used as a starting point, but the focus groups were conducted as conversations, lasting between 1.5–2 hours and sometimes longer.

Results

The two exploratory studies were carried out during the autumn of 2014 and the spring of 2015.

Study I: A regional questionnaire

The outcomes of this study were:

- An analysis of the concept of (effective) information management (presented in Figure 1) and its status in the region (Samuelsson 2015-1, pp 43 ff).
- Proposals to increase technical development skills, for example by an e-service development of "export modules" for information from business

[2] For a more detailed review of the questionnaire se the report and Appendix 1, http://www.miun.se/siteassets/forskning/center-och-institut/cedif/cedif-projekt/spif/slutrapport-for-strategisk-plattform-for-digital-informationsforvaltning-bil-1pdf.

systems and an adaption to the eGovernment Delegation's and the National Archives' Common Specifications for Government Agencies.[3]

The questionnaire sent to the authorities, structured according to these components revealed (in Figure 1) that the authorities need increased knowledge about how to analyse and identify quality requirements when creating and capturing information (Documentation, Registration and Records Management) in order to guarantee long-term preservation and reuse of information.

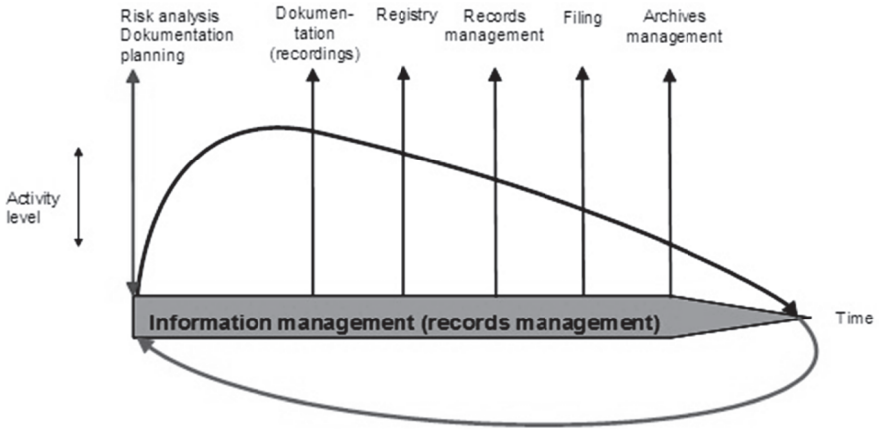


Figure 1. Components of digital information management

Based on the authorities' responses, the following are promising areas for research and development:

- an in-depth understanding of the importance of documentation
- means for a coherent registration of all metadata, whether in specific business systems or more traditional records registration systems.
- methods for improved decision-making about documentation requirements and selection of systems for documentation.

In each section of the questionnaire the aurthorities were asked whether they would like help to develop the relevant part of the information model. The number of 'No'-answers was never more than 20 percent (i.e. approximately 5/29). Thus there is great potential for joint development of solutions (Samuelsson 2015-1, pp 48 ff).

Our next step is to develop a more holistic model for information management, where a coordinated and structured management of the business' registration is

[3] Abbreviated in original language FGS, http://riksarkivet.se/Media/pdf-filer/Projekt/eARD_informationstext_eng.pdf

included. System management will be included, so that the proposed model will be able to handle and control the entire flow of business information. We also identified two additional functions to those usually included in definitions of information management: digital archiving and archive management, and the questionnaire revealed that these functions were handled very poorly. What is needed is a model to bridge the gap between the current business processes and archives. Such a model could be based on the initiative presented below – e-archive as a service. The following standards and tools will also be applied to the national and regional situation:

- ISO 30300, ISO15489, ISO 14721 as quality standards to guarantee information flows to the archive and for ingest compliant with the OAIS model. Inspiration and guidance will also be retrieved from Continuum Model (ISO 15489-1-2:2001, ISO 27001:2013, 30300:2011, ISO 14721, Cumming 2010, Foscarini 2012, McKemmish 2005).
- The National Archives' Common Specifications for Government Agencies (NAS, 2011).

Through discussion with the IT consultants and authorities involved, we have taken the first steps towards a more general tool for interoperability to handle digital archiving for much of the national public sector, using the working title "e-archive as a service". A web service could be developed to find an export scheme (xml) for different business systems – salaries, staff, etc, guiding the user step by step through the process until the information has been packaged for export to an e-archive solution (Samuelsson 2015-1, pp 67 ff).

Study II: A large national government infrastructure authority

In this study the overall aim was to develop an in-depth understanding of requirements to maintain the value of information over the entire life-cycle. A project directory was compiled indicating important research and development for the future, in order to prioritize and optimize the authority's information management over time. (Samuelsson 2015-2, pp 23 ff)).

Developing and managing durable physical infrastructure is important for this authority, making information highly valued and essential for operations in the long-term, although the use requirements for the information change with time. Interviewees mentioned difficulties preserving the value of information throughout the process for everyone involved. Connecting digital information and physical infrastructure is acknowledged to be a challenge. A uniform understanding of the value of information, within the authority as well as among external stakeholders is necessary in order to link information management with business processes and IT support (and the infrastructure) .

A particular challenge is ensuring more efficient reporting of information from various external actors. The authority is a purchaser of services and increasingly, information management tasks are handled by external actors. Thus the authority sets requirements rather than carrying out the information manage-

ment tasks itself. The challenge is to develop procedures to ensure that regulatory requirements are met. This also raises the question of appraisal at an early stage, where requirements for information creation and capture must be set as for example in large infrastructure investment projects.
With the transition to the role of purchaser, challenges in terms of trusting the authority and its activities emerge. An important question raised is how to maintain trust in the management of critical information when the information is produced and managed by external actors. The information must be auditable to ensure that the authority, as purchaser, knows that it has received what was requested, thus securing control for democratic purposes. These issues are crucial not only to ensure the authority is up to date, but also to be an authority citizens and decision-makers trust.
The results from the focus groups have been analysed (Samuelsson 2015-2, pp. 30 ff). Five areas of priority for research and development have been identified:

The role
The role of the authority as purchaser enhances the importance of information in processes of management and control, where it serves as an instrument of control. Because the authority is a public and a democratically controlled organisation, it is also a question of democracy to ensure the political control of its operation, in addition to ensuring that the authority has access to all information of relevance. The information is necessary to be able to maintain, control and manage the infrastructure over time.

Key information flows
Two key information flows have been identified as high priority are 1) the investment process, from planning to implementation and maintenance and, 2) management and governance processes. The latter is crucial for efficient information management to support the implementation of the authority's goals and strategic standpoints. It is also important given the transition to a role as purchaser, in which the authority becomes the client.

Business architecture – a common information model
The authority has begun to develop information and business architecture for a number of areas of work (E-delegationen 2013, Lindström, L. 2011, Nordström 2013). This is an important start, but developing a more uniform and joint information model must advance at a faster rate. Currently, a number of different information professionals are involved in information management and administration, and synchronization is needed. A second step could be to develop an information model integrating different approaches, also taking into account long-term preservation.

Information value and benefit
Different information professionals carry out different forms of appraisal of information, e.g. in archives, architecture, and information security. For example, in order to classify information security, the information must be identified and appraised. The question is how these appraisals relate to the appraisals carried out in the archive sector which aim to assess information based on what should be preserved or discarded. A prioritized research issue is to develop theoretical approaches, models and methods for appraisal of which information should be preserved which takes other perspectives into account. A first step is to identify how appraisal of information from different professionals (architects, archivists and heads of information security etc.) is currently carried out. The second step will be to develop models and methods for coherent information appraisal, taking all perspectives into account.

Quality
A related issue is how to guarantee information quality over time, where many activities with different information needs are involved. This also applies to information management fuctions handled by external actors, in a complex and ever-changing context. The authority seems to be extensively duplicating work because of a lack of trust in the quality of the information, resulting in increased costs. The extent of this has, however, not been assessed in this study. In archival science quality parameters such as authenticity, reliability and useability are defined. Whether these are sufficient and useful in this context and in this authority is also important to investigate.

Conclusions
In order for these authorities' information to be able to contribute to a wider societal perspective, rather than a narrower perspective that serves only the authorities' needs, a holistic approach to information management is required. Developing theoretical approaches, models and methods for appraisal of information and determining what should be preserved must be a high priority. This should be based on a wider societal perspective where the authorities' role as outsourcers and purchasers should be taken into account. The development of more uniform information models will be necessary, through which different information professionals are able to collaborate, be they information or systems architects, information security professionals, archivists, system administrators and others with an interest in high quality, authentic information, accessible and usable in future for as long as necessary.
The most important conclusion is that it must be possible to exchange information; without interoperability we will not be able to build functioning digital businesses or digital societies, Communicating here and now, between internal and external business systems, as well as over the long-term. To create a more

fully efficient digital information management, have we in this article underlined the importance of having:

- order and tidiness and a well thought out information management
- a common information model
- an enhanced interoperability where we suggests the an implementation of an e-archive can constitute "a common minimum level" of interoperable information. Here, we have proposed a development of one e-service, which among others consists of an "Export module" where the information can be exported from the business systems to the e-archive solutions – "e-archive as a service".

References

Anderson, K. , Samuelsson, G. & Morner Jansson, M. 2011: Benchmarking Information Management Practice and Competence in Swedish Organizations. I *Proceedings of the 5th European Conference on Information Management and Evaluation : University of Insubria, Como, Italy. 8-9 September*. Reading, UK, pp. 28-36.

Anderson, K., Borglund, E. A. M., & Samuelsson, G. 2010:.Strategies for high quality recordkeeping in public e-services. iRMA - Information & Records Management Annual, pp. 71-82.

Atherton, J. From life cycle to continuum: Some thoughts on the records management-archives relationship, Archivaria 21 (Winter, 1985-1986), pp. 43-51.

Borglund, E.A.M, & Anderson, K. Long-term Strategies for Electronic Recordkeeping in a Large Infrastructure Project. Paper presented at the Ninth Conference on Databases and Information Systems, 2010, RIGA, Latvia, pp. 441-454

Cumming, K. 2010: Ways of seeing: contextualising the continuum. Records Management Journal, 20(1), pp. 41-52.

Digitaliser.dk. Indgang til digitalisering af Danmark. Her samles centrale ressourcer, anbefalinger og vejledninger om it, formidling og it-udvikling. Målet er, at det kan styrke samarbejdet om at digitalisere Danmark mellem offentlige og private parter. See http://arkitekturguiden.digitaliser.dk/ .Accessed *2015-07-04*

E-delegationen: R 2013. Ramverk för digital samverkan. http://www.edelegationen.se/Documents/F%c3%b6rstudier%20och%20rapporter/Ramverk%20f%c3%b6r%20Digital%20Samverkan%20v1.0%20med%20bilagor%201-2.pdf?epslanguage=sv. Accessed 2015-07-01.

EMC, 2014: http://www.emc.com/leadership/digital-universe/2014iview/business-imperatives.htm. Accessed 2015-07-01.

EU 2015: Promoting a dynamic Public Sector Information (PSI) and Open Data re-use market across the European Union. http://www.epsiplatform.eu/content/about-us . Accessed 2015-07-01.

EU 2014: Delivering the European Advantage? 'How European governments can and should benefit from innovative public services'. A study prepared Delivering the European Advantage?

Foscarini, F. and Oliver, G. 2012: The information culture challenge: moving beyond OAIS: Workshop: Open Research Challenges in Digital Preservation – 9th International Conference on Preservation of Digital Objects (iPPRES). Toronto. October 1-5, 2012. Accessed 1 July, 2015 at: http://digitalpreservationchallenges.files.wordpress.com/2012/09/foscarini.pdf

International Standards Organization, ISO 15489-1:2001. Information and documentation -Records management – Part 1: General. 2001

International Standards Organization, ISO 15489-2:2001. Information and documentation -Records management – Part 2: Guidelines. 2001

International Standards Organization, ISO 27001:2013. Information technology – Security techniques – Information security management systems – Requirements, 2014.

International Standards Organization, ISO 30300:2011 Information and documentation – Management systems for records – Fundamentals and vocabulary,

International Standards Organisation, ISO 14721 Space data and information transfer systems – Open archival information system (OAIS) – Reference model. 2003.

Lindström, L. 2011: Informationsmodellering på IRM:s sätt. 2 uppl. Stockholm.

McKemmish,S. et al. Recordkeeping in Society, ed. Wagga Wagga N.S.W. 2005

National Archive of Sweden – NAS, 2010: Report on a survey – government departments' management of electronic records. Stockholm: Riksarkivet. In Swedish – Rapport rörande enkätundersöning – myndigheters hantering av elektroniska handlingar. 2010.

National Archive of Sweden – NAS, http://xml.ra.se/e-arkiv/. Accessed 2015-07-01

Nordström, M. 2013: pm3 evolutionen – en innovationsberättelse från systemförvaltning till IT-governance. Accessed 2015-07-01

Queensland Government Australia 2014: Queensland Government enterprise architecture 2.0. http://www.qgcio.qld.gov.au/products/about-the-qgea. Hämtad 2014-11-04

Reed, B. 2005: Reading the records continuum: interpretations and explorations. Archives and Manuscripts, Vol. 33(1), 18-43.

Regeringskansliet. IT i människans tjänst – en digital agenda för Sverige. 2011.

Samuelsson, G et al. 2015-1: Strategisk plattform för digital informationsförvaltning, Projektrapport. CEDIF, Mid Sweden University. Accessed 2015-06-15.

Samuelsson, G & Engvall, T 2015-2: Förstudien – En effektiv digital informationshantering. Forskningsportfölj – Trafikverket en modern myndighet, Trafikverket. Accessed 2015-06-16.

A project management approach to long-term preservation of optical media tasks

Arian Rajh
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
Agency for Medicinal Products and Medical Devices
Ksaverska c.4, Zagreb, Croatia
rarian_sp@yahoo.com

Summary

Optical media used in common practice are rarely suitable for long-term preservation (LTP) tasks, they are more a mean of transmission of records between the producer of records and the organisation/archive (any archive in broader sense). Migration is a LTP method which can be used after the act of receiving records, but its dynamics may vary and it can be implemented as a routine process or through project management approach. The author investigates migration process in Croatian National Competent Authority on the example of medicinal products records on CDs and DVDs. Studied projects belong to the same organisation and domain of records. The author investigates complexity of project management approach to migration of optical media and concludes with recommendations on setting up such projects.

Keywords: Documents and records management system, long-term preservation (LTP), migration of optical media, migration process, project management, XML

Introduction

Optical media are commonly used for transfer of digital content, among other media and channels like portals or USB drives. Producers of records are not always aware of archival requirements related to media and organisation of content of these media. Optical carriers can be damaged through usage in business processes. These are just some initial problems related to the preservation of digital records on original optical media. With migration process once established, organisations that receive records on these media may shift their focus from characteristics related to media to other qualities of records.[1] Then the em-

[1] INTERPARES 2 Project. Preservation Task Force Report. Pages 11-12. http://www.interpares.org/book/interpares_book_f_part3.pdf (accessed 29/4/2015)

phasis is given to preservation of content and evidential quality of records that represents that content. Of course, original media type should be mentioned in archival metadata. Overall tendencies in long-term preservation (LTP) moved away from keeping the original media in order to preserve contained information. Archival community takes widespread optical media (and non-archival storage media in general) for means of transfer of submission information packages (SIP), due to their volatility during manipulation and relatively short life expectancy. Optical media carriers are used for ingest of content from source storage to document and records management system, and for export from the system or for dissemination to other systems or users. Digital record is a set of data collected from the storage, dispatched to a memory, represented by review or edit software, and shown on user's screen – so authenticity of digital record is not necessarily tied in with the notion of digital media as a carrier. Consequently, long-term preservation attempt may go by without paying attention to the conservation of media. The question of preservation of optical media relates to the more general question of preservation of physical, logical and conceptual level of record, and the decision about what do we want and need to preserve in particular case.[2] This just means that the notion of authenticity differs from the one in world of paper and has to be ensured differently.

In general, preservation tasks related to the content transferred via optical media should be performed regularly. Records on optical media carriers that have not yet been migrated into document and records management system are in risk while awaiting copying. However, if creator or archive accumulated a large amount of compact discs over time and plans to migrate them in one or several efforts, preservation tasks should be set up as projects. Project is a unique set of activities with limited duration and resources and these activities must lead to planed and measurable result(s).[3] Beside results and activities, project should have clear purpose, defined scope (e.g. number of CDs/DVDs or particular records series), plan, schedule, budget, milestones for activities, verifiable indicators for results, analysed risks, and recognised issues that can be solved during implementation.

Organisation (creator and its archive) studied in this article is HALMED, Croatian National Competent Authority for regulation of medicinal products, medical devices and homeopathic medical products.[4] HALMED implemented docu-

[2] Stančić, H. Theoretical model of persistent preservation of authenticity of digital information objects, PhD dissertation. Zagreb: Faculty of humanities and social sciences, 2005, p. 15-16

[3] National competence baseline, Croatian version 3.0, Zagreb: Croatian association for project management, 2008, p. 15. http://capm.hr/preuzimanja/ (accessed 30/4/2015)

[4] Agency for Medicinal Products and Medical Devices (HALMED) is National Competent Authority for registration of medicinal products, medical devices and homeopathic medicinal products in accordance with legislation of the Republic of Croatia (http://www.halmed.hr/?ln=en&w=o_agenciji).

ment and records management system in 2014 (*Digital archival information system, DAIS*). System was developed on FileNet P8 platform. It consists of several modules (Content navigator, Enterprise records, Migration and ingest module and other) and two repositories. Repository called Records object store (ROS) is controlled by document management module (content navigator). File plan object store (FPOS) is controlled by records management module (Enterprise records) and by connected archival application *Pismohrana*.[5]

Migration

Migration motivator or drive in the case of submitted optical media is prevention of media decay.[6] There are several types of migration according to OAIS model/standard, from simple migration to complex migrations – refreshment, replication, repackaging and transformation.[7] Refreshment is migration of records from media to media of the same generation, without alterations of the information packages. Replication comes down to migration of packages to new media, repackaging[8] includes changes on the physical level, and transformation does not rule out changes on logical level. Transformation can be reversible or non-reversible (which is not recommended).

Migration of optical media in the National Competent Authority HALMED includes: (1) registration of SIP package on optical media into archival application with barcoding of the physical technical unit (disc), (2) copying folders and files into folder with the same barcode,[9] (3) migration of files by document and

[5] DAIS document and records management system was implemented by IPA 2009 TAIB project *Preparations for eCTD and implementation of DAIS*, proposed and managed by A. Rajh and implemented by Ericsson Tesla (business processes modelled by AAM Consulting). Archival application *Pismohrana* was developed by Omega Software in 2013. Please see: HALMED is implementing a one year IPA project "Preparation for eCTD and implementation of digital archive information system", http://www.halmed.hr/?ln=en&w=novosti&d=2014 &id=1053, 1/4/2014 (accessed 30/4/2015); Preparations for eCTD and implementation of digital archival information system, http://www.safu.hr/en/news/preparations-for-ectd-and-implementation-of-digital-archival-information-system, 17/9/2014 (accessed 30/4/2015).

[6] Reference model for an open archival information system (OAIS), Recommended practice CCSDS 650.0-M-2. June 2012. http://public.ccsds.org/publications/archive/650x0m2.pdf, p. 5-3 (accessed 30/4/2015)

[7] Reference model for an open archival information system (2012), ibid., p. 5-4 and further. For other definitions of the term related to file version please see Stančić (2005), ibid., page 82; Shepherd, Elizabeth and Yeo, Geoffrey, Managing records – a handbook of principles and practice, London: Facet Publishing, 2003, p. 198

[8] "The contents of the three files are moved to three new files on another media type, with a new directory and file implementation. Even if all the directory and file names have been preserved in the transfer, a Repackaging has taken place because the bits used to represent the Packaging Information have changed.", Reference model for an open archival information system (2012), ibid., p. 5-6

[9] See figure 1.

records management system DAIS into new digital technical unit, (4) automatic registration of new digital technical unit into archival application linked to DAIS, and (5) automatic file conversion into PDF/A. File conversion is made using *Aspose* software (www.aspose.com) integrated with DAIS system. Details of the whole migration process will be presented later in this article. The main characteristic of this process in HALMED are:

- content of the package are migrated to storage of different type
- folder structure is kept on conceptual level[10] (in document module of DAIS system), although storage itself is organised differently (metadata in database and objects in file storage until its replacement with archival storage, i.e. EMC Isilon or similar)[11]
- content is linked with original unit through barcode identifier provided by archival application
- metadata are re-used through this linking
- standardisation of file format (PDF/A) wherein the process is reversible (previous version is stored by document and record management system; system preserves all versions, but question is for how long previous versions in non-archival formats may be accessed and used)
- process is well-documented by DAIS and prescribed in formal internal act (standard operative procedure)

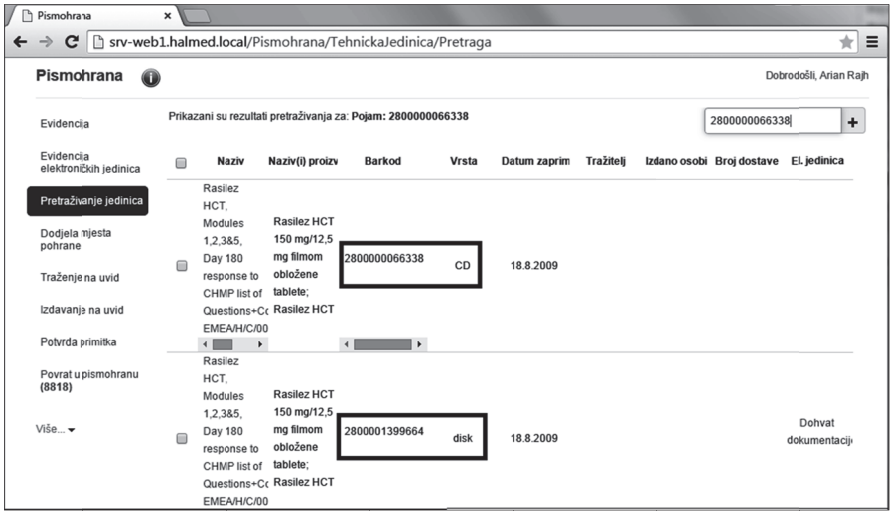


Figure 1: Usage of barcode identifier for linking original information package and digital package (Pismohrana archival application in HALMED, developed by Omega Software)

[10] See figure 2.

[11] See figure 3. For more about selection of storage please see: Critical Capabilities for Scale-Out File System Storage, 27 January 2015, http://www.gartner.com/technology/reprints.do?id =1-28XVMOC&ct=150130&st=sb (accessed 24/7/2015)

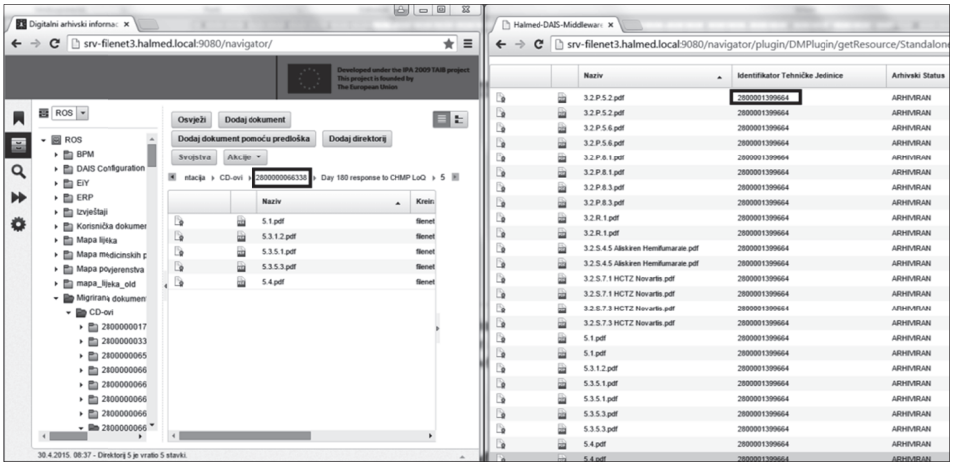


Figure 2: Presentation of structure in document and records management system (FileNet based DAIS system, developed by Ericsson Nikola Tesla)

Process has features of reversible transformation and it combines standardisation of format and OAIS approach, because DAIS system was developed following OAIS principles.[12]

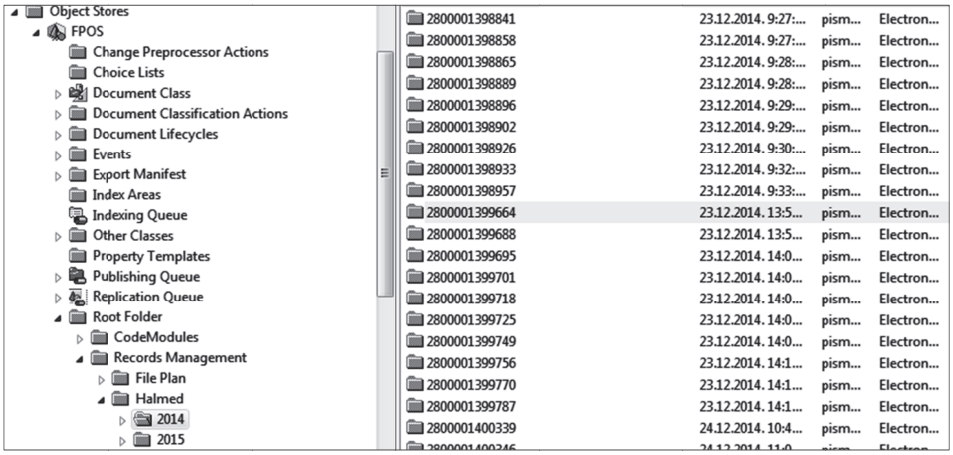


Figure 3: Content in FPOS store (FileNet based DAIS system, developed by Ericsson Nikola Tesla)

These were technical characteristics of optical discs long-term preservation process in HALMED. Below in the article process shall be reviewed from the perspective of organisation of job – (A) as regular use cases (or daily work) and (B) as projects dealing with backlog incurred before implementation of DAIS.

[12] Stančić (2005), ibid., p. 92, 102, 103, 104

The very need for similar preservation process as described in this article was recognised in HALMED in 2008, during pre-research or pilot study of optical media long-term preservation applicability.

Pre-research of suitability of optical media for LTP (pilot project)

Pilot study was designed by Stančić, Rajh and Tušek and conducted by Faculty of Humanities and Social Sciences students in HALMED in May 2008.[13] Students were checking the readability of optical media and defective media were detected. 279 media stored in hard CD cases or plastic sleeves were tested with CDR media code identifier, CD-R diagnostic, and DVDisaster applications. For 60 CDs, it was not possible to determine the date of recording. Some CDs were recorded in 2001, and majority of CDs were recorded between 2005 and 2006. Aim of testing with CDR media code identifier was to determine disc manufacturer and to see if ATIP (absolute time in pregroove) method of putting data is the same for all CDs (and it was not the case). CD-R diagnostic helped pilot project team to determine recording application, the recording date and organic dye layer (cyanine, azo, phthalocyanine).[14] Unreadable CDs were recorded using applications that CD-R diagnostic could not determine. DVDisaster showed us existence of unreadable sectors of media. 171 CDs were without any problems (61%), 75 CDs (27%) had problems[15] and 33 CDs (12%) were unreadable. Among CDs of branded manufacturers 8% were unreadable, and among CDs of unknown brand this rate was 14%. Pilot study was presented at 10/13th International information technology and journalism conference and in the European Medical Agency in London, on Telematic implementation group for electronic submission meeting, so it had some influence on national agencies' practices through group's message to agencies. Results of this pilot research indicated volatility of CDs and HALMED decided to store content of optical disks on other media, precisely on servers and archival storage. Immediate actions were

[13] Stančić, H.; Rajh, A.; Tušek, M. Are optical disks suitable for LTP?, 10/13th International Conference Information Technology and Journalism Collaborative Media and Content Management 15-23/5/2008. Pilot study was done as Digital preservation Europe Croatian Team's activity. Students on practice were: M. Borgudan, T. Pleše and T. Pavičić.

[14] Results of testing related to relation of dye layer and deterioration Stančić/Rajh/Tušek pilot study followed trend recognised in Slattery at al., Stability comparison of recordable optical discs, Journal of research of the NIST, Vol. 109, No. 5, 2004, p. 520 ("…phthalocyanine as the dye and silver and gold alloy as reflective layer, is far more stable than any of the other samples…"). Results in Stančić/Rajh/Tušek pilot study were 44% of errors for cyanine and azo dye (8/18) and 25% of errors with phthalocyanine (25/98). For 163 CDs CD-R Diagnostic application could not detect type of layer (with 25% or errors and 33% of unreadable CDs).

[15] "The end of life of a disc can be defined as the time when uncorrectable error occurs. Although the disc may still be readable after this point, some information has been lost.", Slattery at al. 2004 (ibid.), p. 518.

purchase of cabinets for optical media and publication of instructions for submitting documentation.[16]

Projects initiated for migration of accumulated optical discs and use cases for projects and regularly performed work

Realisation of that preservation planning decision from 2008 was supported by implementation of DAIS system in 2014. That is why HALMED included migration of accumulated CDs/DVDs during IPA project, initiated other projects aiming to migrate further quantities of media, and designed migration use case (new support process for migration of new accruals and for migration projects). Roll-out of these use cases happened in December 2014. Process (a set of use cases) was facilitated by migration module of DAIS system, based on FileNet P8 platform and customised by Ericsson Tesla during IPA 2009 TAIB project (figure 4).

The process consists of several use cases (figure 5): use cases for ingest of digitised SIP packages of paper technical units and use cases for migration of CDs. Use cases for CDs do not cover CD that have to be uploaded in business applications (review system for eCTD[17] dossiers, application for medical devices etc.). CDs for migration through DAIS must be previously recorded into archival application so links between original media and migrated content can be created. CDs/DVDs have to be copied on server mapped to DAIS in folder named after barcode of original media. Generic XML file with location and operation instructions have to be copied into folder named after barcode of original disc. After that, it is necessary to start migration module of DAIS and to run function that validates copied XML. Migration procedure starts after validation of XML. XML prescribes migration operations for CD/DVD as a SIP and its representation in logical organisation of AIPs in DAIS. Then XML prescribes migration operations performed sequentially on each file from CD/DVD. After migration, XML prescribes post-import operations. Post-import operations include linking new AIP (or migrated content) with recorded information about physical CD/DVD disc in archival application *Pismohrana*. Barcode identifier of original physical disc is used to link the disc and AIP. After these two entities are linked, archival application creates new description of migrated AIP and copies re-usable metadata from description of original physical disc to descrip-

[16] In which it was prescribed that optical media should be submitted in a hard cover, http://www.halmed.hr/?ln=en&w=lijekovi&d=dokumentacija (accessed 5/5/2015)

[17] eCTD or electronic Common Technical Document is electronic format of medicinal product's dossier with prescribed structure, files and folders naming conventions, records in PDF format and metadata in XML files. Dossiers consist of 5 modules: module for regional administrative records, module for summaries, module for quality related documentation, non-clinical and clinical module. For more information visit: http://estri.org/eCTD/.

tion of AIP. After migration is done, it is required to check access possibility to AIP through particular DAIS query or through query in archival application.

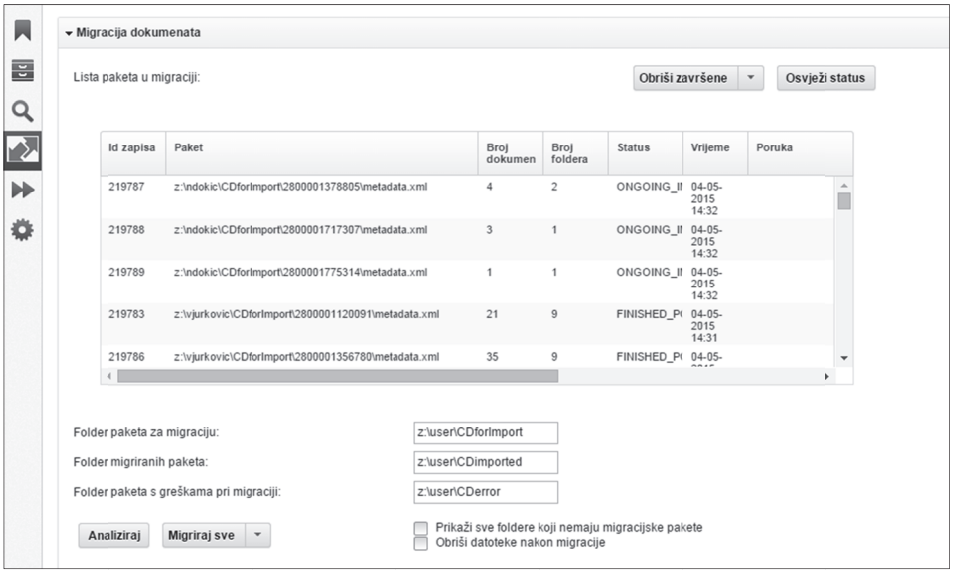


Figure 4: DAIS module for migration and ingest (developed by Ericsson Tesla for HALMED)

There were 4,567 CDs/DVDs migrated through this process during IPA 2009 TAIB project by Ericsson Tesla. Total capacity of migrated CDs/DVDs was 570 GB and they contained 376,501 files. There were also 79 unreadable CDs/DVDs (1.73%).[18] These 4.5 thousands CDs/DVDs were in project terms of reference and project scope. There were still 13.500 CDs left in HALMED after IPA 2009 TAIB project, without counting new CDs that HALMED received on daily basis. Although new CDs/DVDs are mostly in eCTD format and they are foreseen to be uploaded in business applications, so regular migration process should not be used for all of them, HALMED had to solve its backlog problem and migrate CDs before their deterioration.

That is why HALMED suggested a student practice proposal to Department of Information and Communication Sciences (Faculty of Humanities and Social Sciences, Zagreb). The proposal was adopted by the Faculty and practice was designed as 14-work-days project with the team of four students, practice mentor and project manager.[19] Expected project results were migrated optical media

[18] Final Report for the Contract "Development of the Digital Archive information system software in the Agency for Medicinal Products and Medical Devices, Republic of Croatia" EuropeAid/ 132507/D/SER/HR, Ericsson Tesla, 2014, p. 37.

[19] Students on the project - A. Vuk, K. Škrabo, N. Dokic, V. Jurkovic. Manager for this project was B. Romčević. Scope of the project was min. 3.500 up to 5.000 CDs and it was set up after initial migration tests. Project with students started on 4 May and ended on 22 May 2015. After

in the DAIS system and facilitation of records management and usage. The team that designed the project recognised several risks related to it, from scope related risks, technical and human errors, unreadability related risk, to delays associated with technological environment. Project was divided into three phases, the first one comprised preparation and training, second phase was migration process itself, and third phase dealt with final check-up before sending migrated discs for destruction. Project team used mapped server drive for copying, DAIS migration module for migration, and queries of archival application or DAIS Content navigator for verification. Activities of project team followed migration process shown in third lane of figure 5 (tasks 16, 17, 18, 19, 20). Physical destruction of the media had to be documented through archival application and each barcode had to be listed. Project was planned in ProjectLibre application (with Gantt chart, work breakdown structure and list of resources).[20] Students migrated 3,812 CDs/DVDs in May 2015, and error rate was 1.27% (49 CDs were unreadable and they were not migrated into DAIS). Project plan in ProjectLibre application was monitored by project manager on daily basis.

HALMED decided to perform migration of last backlog of CDs/DVDs with its internal team because any delay could result with further increase of number of unreadable CDs. This work was planned as a follow-up project that took place after student practice project. Results in this ongoing project are 5,668 CDs/DVDs migrated by internal employees by 23.7.2015. Current error rate is 2.49% or 141 CD/DVD.

HALMED started migrating new optical discs as they came to the Agency from spring 2015. Predictions based on project experience with students and project plan for follow-up project with internal employees indicate that migration of the backlog optical discs should be finished by the end of September 2015.

ending this project with students, HALMED initiated follow-up project with its own employees in June 2015.

[20] http://www.projectlibre.org/, downloaded and installed in April 2015.

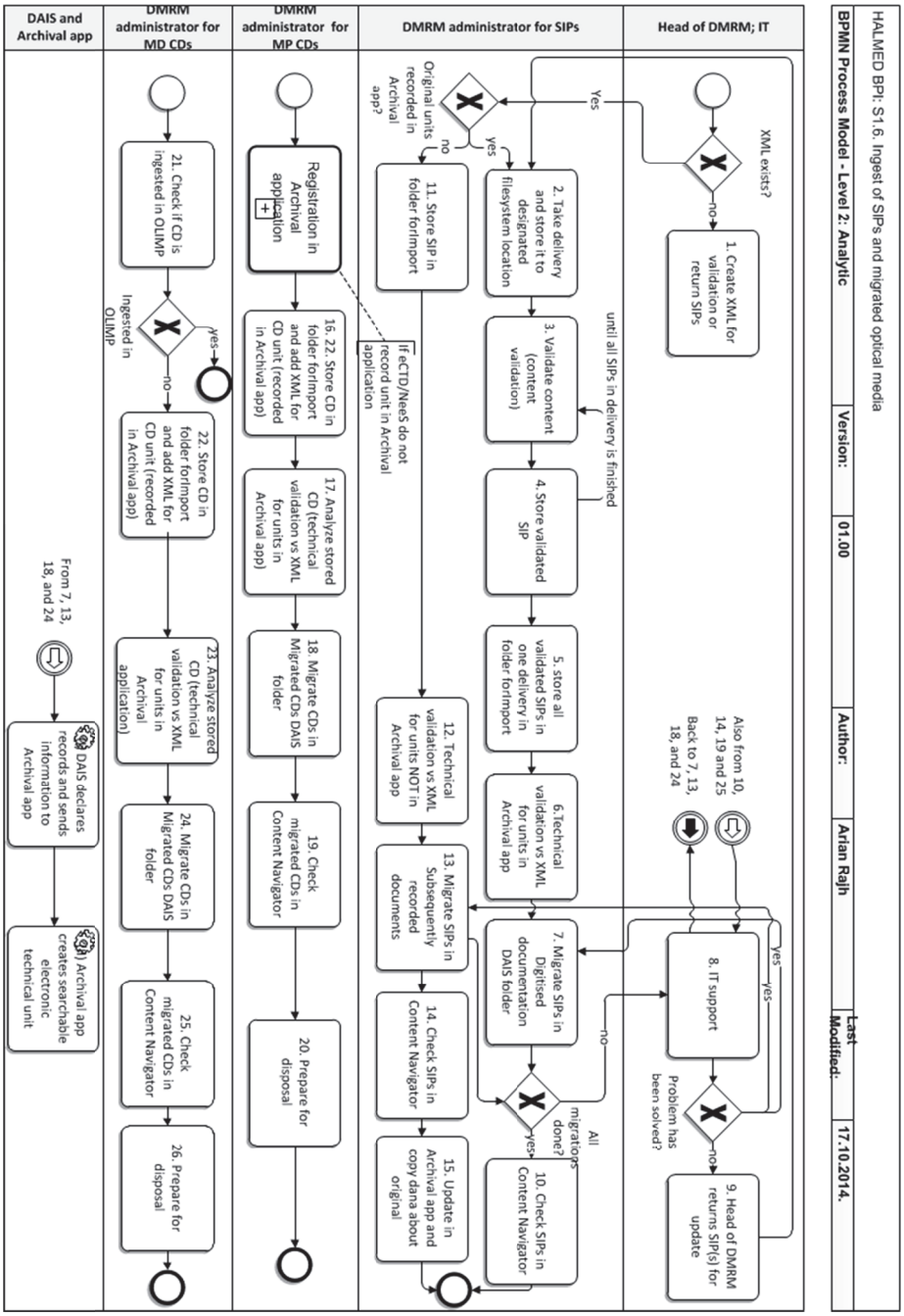


Figure 5: Migration and ingest process – tasks, process scheme created by A. Rajh in Croatian was included in his article *Sustavi za upravljanje digitalnom dokumentacijom – nove mogućnosti i novi izazovi za stvaratelje i arhive*, Arhivski vjesnik (Bulletin d'archives) 58/2015 (forthcoming 2015). English version of model was added to this article to show complexity of the process.

Quality, project risks, issues – project management

Quality in these projects is ensured by validation function in Migration and ingest module of DAIS and by checking the links in DAIS from archival application. For CDs with errors, HALMED will ask applicants (pharmaceutical representations and companies) for their copies after follow-up project ends. Risk related to unreadability occurred in much smaller ratio than during the pilot project from 2008. Sample in the pilot project was relatively small and that explains a large error rate. Second risk that occurred was risk related to DAIS system stress because several students (or employees) used to start migration at the same time. This issue was mitigated by sequencing of migration procedure. Beside those risks, another risk related to CDs occurred during migration of results of digitisation projects: some of the older CDs were inserted into binders with paper documents by HALMED's applicants. This was recognised earlier during the pilot project in 2008 as a bad practice and applicants were asked to submit CDs in separate hard covers.[21] Company that digitised paper documents marked those CDs with identifiers identical to identifiers of paper binders in which they were inserted. This risk can be handled by adding information about existence of original optical media in archival application and linking this information with migrated content. Although migration projects for CD/DVD discs are not categorised as large scale projects, and writing Lessons learned documents is not mandatory according to HALMED's project management methodology, recommendation based on projects experience would be to migrate content of optical media regularly or to establish very determinative technical environment for migration projects and very strict project monitoring mechanisms.

HALMED is not a project management oriented organisation, but it runs a large number of projects. In 2014 HALMED designed its project management methodology that can also be used for any internal migration project. Methodology was designed by A. Rajh, with an advisory help of R. Gospodnetić, team leader of AAM Consulting, a company engaged in IPA 2009 TAIB project for business process management. It consists of four broad phases[22] and twenty two steps (from initial agreement between project proposer and his or hers manager, to creation of project proposal document, creation of project environment in DAIS, creation of project charter, project plan and risk plan, all the way to reporting activities. HALMED methodology has some elements similar to Project Cycle Management, TenStep and PRINCE2. Projects governance and portfolio management include approving projects by unit managers, internal project management working group and Head of the Agency. Monthly monitoring is done

[21] See note 16.

[22] Project management methodology phases on Croatian: Formulacija, Inicijacija, Razrada i evaluacija, Upravljanje projektom

by Agency's project management working group. Exceptions during implementation of projects are handled in internal procedure of change management, e.g. changing the scope or handling project issues. For large projects (over 200.000 HRK and/or 12 months of implementation) HALMED's methodology prescribes using additional analytical tools and writing Lessons learned documents and other documents with extensive project related information. HALMED also engaged project consultancy Primakon for training in project management and it comprised project management principles and workflow of HALMED internal project management procedure.

Conclusion

Long term preservation of digital records is very demanding, and it can help a little if organisation cuts optical media unpredictability out of the equation. Of course, this does not mean that organisation or archive should not check regularly server storage or archival storage subsystem, do the backup frequently and according to the backup plans, and design its disaster recovery and business continuity process. Recommendation is to do migration from widespread recordable optical media to records management or file systems as daily process or to prescribe submission of more stable media. These media could be archival grade optical media, long-term backup DVDs or DTD discs with metal-ceramics inorganic layer. It is important to plan and execute migration projects according to the needs of the organisation. Using more stable optical carriers, like DTD discs with 160 years life expectancy stated after accelerated aging tests or other transfer carriers with longer life expectancy, will not solve all preservation problems associated with textual records – because it is not just the records' carriers archivists should be worried about. Relying on archival standards for file formats due to their known specifications that could be used for development of viewer and editor software in the future is much safer. Using embedded non-copyrighted fonts, typographically similar to common fonts for textual content, where they can be applied, is another proactive archival measure. Using standardised metadata for description is also recommended. Ensuring that migrated content is protected in document and records management system of some organisation, competent service provider or archival organisation is another level of archival custody. If such a system on top of preservation mechanisms ensures authenticity of its digital holding, this could provide public trust in the records creator, custodian or archive.

Documents and records management systems like DAIS have advantage before file systems because they can support migration process better and they can document it, which is essential. However, day-to-day migration of discs is easier (once established) so organisation or (its) archive can focus its LTP plans on file formats, reader applications, metadata, font fixing etc. If organisation or archive accumulated or inherited backlog, it should launch a migration project or a LTP programme of projects, but it has to be done with formal project man-

agement intellectual tools. Approach should strictly follow some internal or acquired methodology. Migration projects are easier to plan and execute if organisation has experience from similar projects or from migration tests, and if it owns appropriate technological environment to facilitate migration tasks.

References

Critical Capabilities for Scale-Out File System Storage, 27 January 2015, http://www.gartner.com/technology/reprints.do?id=1-28XVMOC&ct=150130&st=sb (accessed 24/7/2015)

Final Report for the Contract "Development of the Digital Archive information system software in the Agency for Medicinal Products and Medical Devices, Republic of Croatia" EuropeAid/132507/D/SER/HR, Ericsson Tesla, 2014

HALMED is implementing a one year IPA project "Preparation for eCTD and implementation of digital archival information system", 1/4/2014 http://www.halmed.hr/?ln=en&w=novosti&d=2014&id=1053 (accessed 30/4/2015)

Instructions for applicants for submitting the conventional dossier for medicinal products, 2009, http://www.halmed.hr/?ln=en&w=lijekovi&d=dokumentacija (accessed 5/5/2015)

INTERPARES Project. Storage Media (Characteristics and Composition) - Media. Draft appendix. October 2001. www.interpares.org/documents/media.pdf (accessed 30/4/2015)

INTERPARES 2 Project. Preservation Task Force Report. http://www.interpares.org/book/interpares_book_f_part3.pdf (accessed 29/4/2015)

National competence baseline, Croatian version 3.0, Zagreb: Croatian association for project management. 2008. Page 15. http://capm.hr/preuzimanja/ (accessed 30/4/2015)

Preparations for eCTD and implementation of digital archival information system. 17/9/2014, http://www.safu.hr/en/news/preparations-for-ectd-and-implementation-of-digital-archival-information-system (accessed 30/4/2015)

Rajh, A. Sustavi za upravljanje digitalnom dokumentacijom – nove mogućnosti i novi izazovi za stvaratelje i arhive, Arhivski vjesnik (Bulletin d'archives) 58/2015 (Manuscript submitted for publication, forthcoming 2015)

Reference model for an open archival information system (OAIS), Recommended practice CCSDS 650.0-M-2. June 2012, http://public.ccsds.org/publications/archive/650x0m2.pdf, p. 5-3 (accessed 30/4/2015)

Shahani, C. J., Manns, B., Youket, M. Longevity of CD media: Research at the Library of Congress, Preservation of Electronic Records: New Knowledge and Decision–Making: Post Prints of a Conference, Symposium 2003, 197–206. Ottawa: Canadian Conservation Institute, 2005. http://www.loc.gov/preservation/resources/rt/studyofCDlongevity.pdf, (accessed 23.7.2015)

Slattery, O.; Lu, R.; Zheng, J.; Byers, F.; Tang, X. Stability Comparison of Recordable Optical Discs – A Study of Error Rates in Harsh Conditions, Journal of Research of the National Institute of Standards and Technology, Volume 109, Number 5, September-October 2004

Standard operative procedure AG-ARR-0005 Ingest of SIPs and migration of optical media, 1st ed., HALMED, 2015

Standard operative procedure AG-OP-0053 Project management, 2nd ed., HALMED, 2015

Stančić, H.; Rajh, A.; Tušek, M. Are optical disks suitable for LTP?, 10/13th International Conference Information Technology and Journalism Collaborative Media and Content Management 15-23/5/2008

Stančić, H. Theoretical model of persistent preservation of authenticity of digital information objects, PhD dissertation. Zagreb: Faculty of humanities and social sciences, 2005

Shepherd, Elizabeth i Yeo, Geoffrey, Managing records – a handbook of principles and practice. London: Facet Publishing, 2003

Turner, J.R. (edt), Gower Handbook of Project Management, Aldershot: Gower, 2007

Lexicography and Encyclopaedistics in the Digital Environment

Nataša Jermen, Cvijeta Kraus, Irina Starčević Stančić
The Miroslav Krleža Institute of Lexicography
Frankopanska 26, Zagreb, Croatia
natasa.jermen@lzmk.hr, cvijeta.kraus@lzmk.hr, irinas@lzmk.hr

Summary

This paper presents the current as well as the planned future activities concerning the transformation of the Miroslav Krleža Institute of Lexicography into an e-institution. The authors outline the transformation process of the lexicography and encyclopaedistics from a traditional to a digital environment. The activities concerning the establishment of the publicly available repository of encyclopaedic knowledge are discussed. Furthermore, methods of digital communication with users are described, namely web pages and social media channels, in the context of the Institute's activities. The authors acknowledge that publishing of encyclopaedic editions faces challenges and changes in the digital environment. Digital technologies are transforming the encyclopaedic profession and ways of preparing and presenting encyclopaedic knowledge. The authors explain the changes that should be made in order to drive the Miroslav Krleža Institute of Lexicography towards becoming a fully blown e-institution.

Keywords: The Miroslav Krleža Institute of Lexicography, e-institution, lexicography, encyclopaedistics, digital environment, repository of encyclopaedic content, social media networks

Introduction

The Miroslav Krleža Institute of Lexicography is a public institution of special status for the Republic of Croatia.[1] It is the only Croatian institution that has been systematically engaged in lexicography and encyclopaedistics[2] since its foundation in 1950. As a publishing house, as well as a scientific institution, the Institute over the course of its 65-year history has published more than 250 dif-

[1] The Law on the Miroslav Krleža Institute of Lexicography (in Croatian), Official Gazette 96/03, 190/03.

[2] In Croatia, *Lexicography and encyclopaedistics* is a branch in the scientific field *Information and communication sciences*. Lexicography is engaged in systematic compiling of the words and their forms, according to certain rules and criteria which are then published in dictionaries, lexicons and encyclopaedias. Encyclopaedistics is a scientific discipline which is engaged in principles and practises of assembling encyclopaedias.

ferent encyclopaedias, dictionaries, lexicons and other editions that represent an important contribution to continuation and advancement of the Croatian intellectual standard.

The main role of the Institute is based on the activity of lexicography and encyclopaedistics which through systematising, synthesising, indexing, storing and distributing information gives the opportunity for a reliable access to global knowledge and enables a systemic orientation within the ever-increasing amount of data and information.[3] Furthermore, the role of the Institute is to research and understand Croatia's natural and historic heritage, as well as to preserve its cultural and social identity. The Institute's potential lies in the huge resources of created and stored knowledge, which are continuously being digitised and as such it is becoming accessible to a large number of users via the Institute's web pages. At present, the *Portal of Knowledge* offers free access to more than 60,000 articles from seven digitised encyclopaedias and lexicons[4]. Also, online and freely available editions currently are *Croatian Encyclopaedia* and *Proleksis Encyclopaedia* – the two general encyclopaedias that emphasise contents regarding Croatian national heritage, which together comprise of more than 120,000 articles. They are continuously updated and expanded[5]. Apart from the encyclopaedic content, the web pages of the Institute offer access to the digitised *Author Catalogue*[6] of the *Catalogue of Retrospective Bibliography of the Articles* which in total is comprised of 10 million leaflets containing lists of articles from periodicals published in the South-Slavonic area in the period from the end of the 18th century until the year 1945. Open access to the Institute's peer-reviewed scientific journal *Studia lexicographica* (launched in 2007) is also available.[7] By enabling the open access to the online encyclopaedic and scientific content, the Institute supports the open science initiative[8] and contributes to the dissemination of public knowledge in Croatia.

Certain aspects of the future development of the Institute as an e-institution have been recognised and incorporated in the Institute's vision statement.[9] One of the main goals of the Institute's activity in the years to come will be to up-

[3] Strategy of the Miroslav Krleža Institute of Lexicography 2015-2020 (in Croatian).

[4] Portal of Knowledge, http://enciklopedija.lzmk.hr (14.06.2015).

[5] Croatian Encyclopaedia, http://enciklopedija.hr; Proleksis Encyclopaedia, http://proleksis.lzmk.hr (14.06.2015).

[6] The Digitised Author Catalogue, http://katalog.lzmk.hr (14.06.2015).

[7] Studia lexicographica, http://www.lzmk.hr/hr/leksikografska-djelatnost/studia-lexicographica (14.06.2015).

[8] Digital Agenda for Europe – A Europe 2020 Initiative, http://ec.europa.eu/digital-agenda/en/open-science (14.06.2015).

[9] Strategy of the Miroslav Krleža Institute of Lexicography 2015-2020 (in Croatian).

grade its encyclopaedic online systems with the intention to develop public knowledge of a highest scientific relevance, as well as to create an integral, publicly available digital repository of encyclopaedic content.
This paper aims to give an insight into the current as well as planned activities of the Miroslav Krleža Institute of Lexicography concerning upgrade of the lexicographic and encyclopaedic system in the context of information and communication technologies (ICT). The challenge is how to utilise this technology successfully, and develop the Institute as an e-institution. The transformation process of the encyclopaedic profession from the traditional, i.e. paper-based environment to the digital environment, as well as the process of development of the repository of digitised and online encyclopaedic content will be shown. Moreover, the necessity of the interaction with users of the Institute's online content through the new methods of communication, such as social networks, is pointed out.

Towards an e-institution – basic concepts in the digital environment

The increase of ICT's influence has affected all spheres of human life and activities. This influence is highly felt in the institutional and business environment too. The spread of ICT's influence has challenged encyclopaedia publishers and brought changes to the institutions publishing encyclopaedias and lexicons, which nowadays are dominantly in the digital form. Although each institution is unique, any institution can be viewed as dynamic, open, efficient and undeterminable social system acting in an environment or at a market and has the task of achieving its goal, i.e. to transform inputs or resources to outputs, products or services intended for the users.[10]
When an institution is functioning in the digital environment and is developing a digital product distributable over the network, ICT is one of the foundation of its business processes. All information and data are in a digital form and usually stored in databases. Generally speaking, an e-institution should be looked upon as a complex business and information system consisting of five components: (1) hardware, (2) software, (3) orgware, (4) lifeware, and (5) e-content. In this context hardware consists of computers, network devices and other devices for data processing and transmission. Software consists of programs for conducting everyday business. Orgware consists of regulations and procedures aiming at keep the system functioning as a harmonious unit. Lifeware is referring to the employees, and e-content is the digital content that can be transmitted over a computer network such as the internet[11]. E-content could be any form of data such as text, graphic, picture, sound, video, animation or 3D object that can be stored in a database and be searched and retrieved.

[10] Ekonomski leksikon / ur. Z. Baletić. Zagreb, Leksikografski zavod Miroslav Krleža i Masmedia, 1995, 627.

[11] PC encyclopaedia, http://www.pcmag.com/encyclopedia/term/42225/e-content (14.06.2015).

Certain necessary preconditions need be satisfied in order for an institution to be considered as an e-institution. From the technological point of view, it is necessary that the ICT infrastructure has been implemented and is operational. From the business standpoint, new business model processes with accompanying standards, routines and procedures should be developed. Having the ICT infrastructure in itself is not enough. The new business model should enable full utilisation of ICT.

The transition of the Miroslav Krleža Institute of Lexicography to an e-institution will not happen overnight. The process will be a slow and complex, entailing installation and utilisation of the ICT equipment, introduction of the improved business model in the context of encyclopaedic profession, involvement of employees proficient with the concept of lexicography and encyclopaedistics in the digital environment, upgrade of the Institute's e-content and finally development of programs for conducting everyday business, such as content management systems (CMS), document management systems (DMS), records management systems (RMS) with the necessary rules and procedures for data management and preservation. A wider institution-level system, combining the mentioned systems, is called Enterprise Content Management System (ECMS). It denotes the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to institutional processes.[12]

Since institutions are dynamic systems that depend on the surrounding environment, changes in the environment in return influence institutional functions. Therefore, it is necessary that the Institute adapts the interaction with the users of its e-content, i.e. its environment, and develops an active social media presence.

Lexicographic work in the digital environment

The basic activity of the Miroslav Krleža Institute of Lexicography is lexicography and encyclopaedistics. This consists of the acquisition, processing and verification of data that, accompanied by fundamental scientific research in order to apply common lexicographic and encyclopaedic standards, are stored in a database and used for the creation and publishing of various printed or online editions.[13] Those editions offer scientifically verified information in a number of fields of human knowledge and contribute to the systematisation and public presentation of specific scientific disciplines, social and cultural phenomena.

Lexicographic work, which is carried out by *lexicographers*, is defined as the "performance of professional and scientific activities in the field of lexicogra-

[12] What is Enterprise Content Management (ECM)?, Association for Information and Image Management (AIIM), http://www.aiim.org/What-is-ECM-Enterprise-Content-Management (14.06.2015).

[13] The Law on the Miroslav Krleža Institute of Lexicography (in Croatian), Official Gazette 96/03, 190/03; Statute of the Miroslav Krleža Institute of Lexicography (in Croatian), Official Gazette 66/2013.

phy and encyclopaedistics".[14] There is a wide range of activities included in the process of lexicographic work, starting from the acquisition of data to the final step of their incorporation into the editions. The framework of lexicographic work is the composition (writing) of lexicographic and encyclopaedic units and the adjustment (editing) of those units to lexicographic and encyclopaedic as well as scientific standards. In addition lexicographic work consists of language, art, bibliographic and cartographic editing as well as creation of the web applications as the basis for the performance of lexicographic work.

There has been a step change from the traditional concept of lexicography and encyclopaedistics following the introduction of the ICT into the Institute's business process. Since the late 1990s the appropriate steps have been taken to transform the process of lexicographic work from the traditional (i.e. paper-based) to the digital environment. During this time, a computerised editorial system was created in order to support publishing of the paper-based *Croatian Encyclopaedia*. As only administrators had the full access to the system, all lexicographic processes were still done in the traditional way. In 2007 the Institute started developing web applications to support several printed editions. The first step of the Institute towards offering online editions was in 2009 when the *Portal of Knowledge* was made available online.[15]

At present, in common to many contemporary publishing houses, the working process on the majority of Institute's projects is now performed in the digital environment, using web applications, regardless of their planned future publishing be that printed or digital format. The *Croatian Encyclopaedia* and the *Proleksis Encyclopaedia*, two online encyclopaedias, are fully compiled and published in the digital form. The process of production of an online encyclopaedia differs from the production of its paper form, the new possibilities of content design, presentation, interconnection and usage leads to a paradigm shift. The switch to the digital editions enabled lexicographers to use the supporting ICT-based system to work more efficiently, to update content regularly, and to publish digital content faster. With that transition, "the very way in which encyclopaedias are produced has changed, at the same time as communication and use of encyclopaedic knowledge has become different, always avail-

[14] Regulation on the lexicographic work in the Miroslav Krleža Institute of Lexicography (in Croatian), 19 November 2013. The Law on the Miroslav Krleža Institute of Lexicography, by which the Regulation is governed, regulates that *lexicographers* are engaged both in lexicographic and encyclopaedic work. According to this, the authors use the term *lexicographer* equally in the context of lexicographic and encyclopaedic work.

[15] Even before 2009, the Institute had published several digital editions – in a co-publishing effort with the publishing house Masmedia, it has published *The Lexicon of Economy* on a CD-ROM (1995), a *Croatia Tourist Guide* on a CD-ROM (2000), and, with the co-publisher EURO-CAL, San Pedro, California, USA, a pocket device *World-traveler – 15 language translator* (1998).

able and in constant competition with other sources"[16] such as *Wikipedia*, the most commonly used encyclopaedia today. While encyclopaedias online strive to the total access to knowledge, allowing the exchange of ideas, they also protect cultural heritage and particularities, and constantly explore new possibilities for connecting (both existing and new) content. They assume a different organisation of information and interactivity in work.[17] The properties and functionalities of the digital media does not question the encyclopaedic concept, consisting of encyclopaedic principles,[18] but do question the profitability of printed editions. A redefinition of the role the lexicographers will play in the production and publishing of encyclopaedias in the digital form is needed.[19] They will be required to become proficient with the ever-changing digital concepts and technologies as well as understandable of the digital processes and user expectations.

The Institute's plan for the future is to upgrade existing web applications while creating new ones that will enable lexicographic work and presentation of its results entirely in the digital environment. The upgrading of the process of lexicographic work in the digital environment should facilitate the work of lexicographers as well as accelerate the publishing of new encyclopaedic content online. The process of improvement of the lexicographic work will consequently improve the development of the Institute's repository of encyclopaedic knowledge.

Repository of encyclopaedic knowledge

The necessary preconditions for the establishment of the Institute's e-content, i.e. digital repository of encyclopaedic content, include digitisation of published editions, development of an institution-level repository of all digitised and digital encyclopaedic content and its adequate presentation online. Digitisation of the Institute's published editions, as well as the development of the repository, currently only of the digitised encyclopaedic content, was initiated in 2009. It is an ongoing project, with the purpose of digitising all archival editions and making them available to the users. In time, digitally-born editions will be included as well. Digitisation is nowadays the Institute's dominant means for making the printed editions globally accessible and reusable. The repository of

[16] Sundin, Olof; Haider, Jutta. The networked life of professional encyclopaedias: Quantification, tradition, and trustworthiness. // First Monday, 8 (2013) 6.

[17] Prelog, Nenad; Bebić, Domagoj. From Who and What to How and Why – The Future of Online Encyclopaedias. // *INFuture2011: Information Sciences and e-Society* / Billenness, C. et al. (ed.). Zagreb: Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2011, p. 301.

[18] Jecić, Zdenko. Enciklopedički koncept u mrežnom okruženju. // *Studia lexicographica*, 7 (2013), 2 (13), p. 113.

[19] Starčević Stančić, Irina; Kraus, Cvijeta. Hrvatska enciklopedija – od tiskanoga do mrežnoga izdanja. // *Studia lexicographica*, 8 (2014) 1 (14), pp. 101-102.

digitised archival editions thus contributes to the dissemination and preservation of encyclopaedic heritage and facilitates the lexicographic and encyclopaedic activities. This highlights the need that all remaining published encyclopaedic content should be digitised as soon as possible. Once in the digital form, it will also facilitate the development of new encyclopaedic products.

The Institute's future publishing policy will be oriented to a large extent towards the more significant presence of the digitised and digital contents on the web as well as towards the improvement of the organisation of the existing repository. The plan is to link all digitised and online editions, both current as well as the future ones, in the repository of encyclopaedic content which will be available to the global community free of charge. The development of the fully-blown institution-level repository will need careful planning and adaptation of the Institute's business processes.

Establishment of such a repository of encyclopaedic knowledge, as a form of a public knowledge, would enable the possibilities of linking to digital repositories of other scientific and cultural institutions in Croatia, as well as abroad. On the international level, the Institute's plan is to initiate the project of online linking of the European general encyclopaedias, focusing on those from small to medium language populations, through cooperation with similar European institutions. The scope of the project, apart from the connecting of the encyclopaedically organised knowledge in Europe, would be preservation of the multicultural aspect of the European countries.

Further, creation and organisation of open repositories of knowledge would help reinforce research e-infrastructure, and would also serve as the didactic tools, thus supporting the development of e-learning systems and curriculum links.[20] To be precise, encyclopaedias form an important part of the infrastructure for learning in institutions like schools, libraries and universities and will, in all probability, continue to be indispensable in the future.[21] Online encyclopaedic content will also help popularise the results of scientific research and develop recognisability of the Croatian national heritage.

Digital communication with users

The growth of ICT has changed the ways how institutions communicate with their users. They use the internet in order to find new information and create new social interactions – make new friends, keep existing ones, observe interactions of other people and sometime copy their friends' behaviour. According to boyd and Ellison (2008) "we define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they

[20] Strategy of education, science and technology (in Croatian), 17 October 2014.

[21] Sundin, Haider, 2013.

share a connection, and (3) view and traverse their list of connections and those made by others within the system."[22] Social networks play an important role for individuals. Okazaki and Taylor (2015) think that "social media at the same time represent a powerful personalization tool as they enable individuals to both produce and distribute content by their own participation."[23]

Social networks and encyclopaedia publishers

Following the trend of increasing use of social networks it can be determined that various institutions are more and more involved with the social networks and by them attract a large number of users. Social networks are becoming a place where institutions can approach their users more easily and offer their content, exchange experiences, listen to their comments and use them to improve their products. Currently the most popular social networks are Facebook, Twitter, YouTube, LinkedIn, Instagram and Pinterest. Almost all encyclopaedia publishers have their own websites which are primary places for their content. This content is then used as a basis of interaction with the users via social networks.

Investigating the presence of encyclopaedia publishers on social networks, one can notice that *Encyclopaedia Britannica* updates its content frequently, the website is updated daily, publishes newsletters, develops mobile applications and in this way approaches end users and adapts to their ways of communication.

Swedish *Nationalencyklopedin* and Spanish *Enciclopèdia Catalana* are present on Facebook, Twitter, YouTube and Instagram. Italian *Encyclopaedia of Science, Letters, and Arts* (Treccani encyclopaedia) is present on Facebook, Twitter and YouTube. *Encyclopaedia Universalis* (in French, published by Encyclopaedia Britannica, Inc.) is present on Facebook and Twitter while Danish *Den Store Danske* and Croatian *Croatian Encyclopaedia* are present only on Facebook.

By following the trends of the previously mentioned encyclopaedia publishers, the Miroslav Krleža Institute of Lexicography has a nascent social media involvement. The website of the Institute is essential to the promotion and popularisation of the lexicographic and encyclopaedic activities[24]. Through it, the public is informed about current events and achievements related to the activities of the Institute and with its digital content. Further development of the Institute's website started in 2011, when a redesign of the previous web pages was

[22] boyd, danah m.; Ellison, Nicole B. Social Network Sites: Definition, History and Scholarship. // *Journal of Computer-Mediated Communication*. 13 (2008), 1; p. 211.

[23] Okazaki, Shintaro; Taylor, Charles R. Social media and international advertising: theoretical challenges and future directions. // *International Marketing Review*. 30 (2015), 1; p. 58.

[24] The Miroslav Krleža Institute of Lexicography, http://www.lzmk.hr (14.06.2015).

made. Part of the Institute's website content has been translated into English and a Facebook account was opened[25].

Having great content, knowing your audience and consistency[26] are the three main components when discussing what it takes for an encyclopaedia publisher to successfully run a social media channel.

The Institute needs to produce a plan for its social media approach that is interesting for users. When publishing content via social media channels it is necessary to include more people and set up a team that is responsible for social networks. The team should create a publication cycle (daily, weekly and monthly), create a calendar of important dates in order to facilitate pace of publication, define responsibilities among editors (monthly and weekly assignments), answer questions posted by users through social networks, provide feedback to the editors, moderate online discussions, monitor posts and develop new social media content accordingly.

Conclusion

The publications of the Miroslav Krleža Institute of Lexicography, thanks to its tradition and reputation, have played a significant role in creation, maintenance and distribution of the public knowledge in Croatia. In the context of information and communication technologies and ever-increasing amount of available information, they offer scientifically verified and systematised foundation for the orientation in the context of global information. In the field of new technologies, the digitisation of published editions has been going on since 2009, as well as the establishment of the online repository of digitised encyclopaedic content. Moreover, some of the fundamental ongoing editions (e.g. general encyclopaedias) are edited and presented to the public as online editions, too. In the forthcoming years the activity of the Institute will be focused towards the establishment of the freely available online repository of encyclopaedic knowledge, which, in part, will be continuously updated. For that purpose, there is a need to upgrade the process of lexicographic work from the traditional, i.e. paper-based environment towards the digital environment. This would facilitate the work of lexicographers, as well as accelerate the process of publishing online encyclopaedic content. Furthermore, the interaction with users should also be adapted to the digital environment and facilitated through the active social media presence.

As an organised system of the production of encyclopaedic content, the Institute has the possibility of participating in systematising, linking and building of

[25] In December 2011 the Miroslav Krleža Institute of Lexicography's website was voted No 1 website in Croatia in the category "Government organizations, government institutions and public services" by the VIDI Web Top 100 Contest.

[26] King, David Lee. Managing Your Library's Social Media Channels. Chicago : ALA TechSource, 2015, p. 22.

knowledge provided by various scientific, archival, documentation and information sources. By means of the possible future linking to the digital repositories of other scientific institutions, the development of the e-infrastructure of the Croatian educational and research area could be supported.

Finally, the development of the new ICT-based business model, entailing a system that would connect editing and publishing of the printed editions with the digital editions published online, is a complex and challenging task. Even more, one should also consider that the two processes, although having the same underlying lexicographic and encyclopaedic principles, are fundamentally different in their nature because of all the possibilities of platform portability, browsing, searching, text analysis, comparison, dynamic updating, interaction, long-term preservation of ever-changing digital media and formats etc. All those possibilities require professionals proficient with those concepts and understandable of the digital processes and user expectations. Nevertheless, there is no other way to go but to fine-tune lexicographic and encyclopaedic processes and practices to the digital realm, introduce the needed changes in the business processes and drive the Institute towards the concept of e-institution.

References

boyd, danah m.; Ellison, Nicole B. Social Network Sites: Definition, History and Scholarship. // *Journal of Computer-Mediated Communication*. 13 (2008), 1; pp. 210-230.

Digital Agenda for Europe – A Europe 2020 Initiative, http://ec.europa.eu/digital-agenda/en/open-science (14.06.2015)

Ekonomski leksikon / ur. Z. Baletić. Zagreb, Leksikografski zavod Miroslav Krleža i Masmedia, 1995.

Jecić, Zdenko. Enciklopedički koncept u mrežnom okruženju. // *Studia lexicographica*, 7 (2013), 2 (13), pp. 99-115.

King, David Lee. Managing Your Library's Social Media Channels. Chicago : ALA TechSource, 2015

Mullan, Eileen. What is Digital Content?, EContent, 19 December 2011. http://www.econtentmag.com/Articles/Resources/Defining-EContent/What-is-Digital-Content-79501.htm (14.06.2015)

O'Brien, James A.; Marakas, George M. Management Information Systems. New York : McGraw-Hill, 2011.

Okazaki, Shintaro; Taylor, Charles R. Social media and international advertising: theoretical challenges and future directions. // *International Marketing Review*. 30 (2015), 1; pp. 56-71.

PC encyclopaedia, http://www.pcmag.com/encyclopedia/term/42225/e-content (14.06.2015)

Prelog, Nenad; Bebić, Domagoj. From Who and What to How and Why – The Future of Online Encyclopaedias. // *INFuture2011: Information Sciences and e-Society* / Billenness, C. et al. (ed.). Zagreb: Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2011, pp. 299-308.

Starčević Stančić, Irina; Kraus, Cvijeta. Hrvatska enciklopedija – od tiskanoga do mrežnoga izdanja. // *Studia lexicographica*, 8 (2014) 1 (14), pp. 99-116.

Statute of the Miroslav Krleža Institute of Lexicography, Official Gazette 66/2013.

Strategy of education, science and technology, 17 October 2014.

Strategy of the Miroslav Krleža Institute of Lexicography 2015-2020.

Sundin, Olof; Haider, Jutta. The networked life of professional encyclopaedias: Quantification, tradition, and trustworthiness. // First Monday, 8 (2013) 6.

The Law on the Miroslav Krleža Institute of Lexicography, Official Gazette 96/2003, 190/2003.
What is Enterprise Content Management (ECM)?, Association for Information and Image Management (AIIM), http://www.aiim.org/What-is-ECM-Enterprise-Content-Management (14.06. 2015)

Links

Croatian Encyclopaedia, http://enciklopedija.hr
Den Store Danske, http://www.denstoredanske.dk
Enciclopèdia Catalana, http://www.enciclopedia.cat
Enciclopedia Treccani, http://www.treccani.it
Encyclopaedia Universalis, http://www.universalis.fr
Nationalencyklopedin, http://www.ne.se
Portal of Knowledge, http://enciklopedija.lzmk.hr
Proleksis encyclopaedia, http://proleksis.lzmk.hr
Studia lexicographica, http:// www lzmk.hr/hr/leksikografska-djelatnost/studia-lexicographica
The Digitised Author Catalogue, http://katalog.lzmk.hr
The Miroslav Krleža Institute of Lexicography, http://www.lzmk.hr

Original scientific paper

Communicating the Crisis: The European Commission on Social Media During the Economic Crisis in the European Union

Domagoj Bebić, Milica Vučković
Faculty of Political Science, University of Zagreb
Lepušićeva 6, Zagreb, Croatia
domagoj@edemokracija.hr, milica@edemokracija.hr

Marija Volarević
PhD student, Faculty of Social Sciences, University of Ljubljana
Kardeljeva pl. 5, Ljubljana, Slovenia
marija@edemokracija.hr

Summary

Looking at the official European Commission social media in the period of the economic crisis in Europe, the aim of this study is to test optimistic (Browning 2002, Coleman 2003; Hill & Hughes 1998; Ward, Gibson & Nixon 2003) and pessimistic (Kavanaugh 2002, Norris 2001) views about the role of the Internet in political communication. Relying on the optimistic ideas that social media can improve communication between representatives and the represented, we test to see if the European Commission used Facebook in order to improve communication with citizens. We use discourse analysis as a qualitative content analysis to look at status messages on the EC official Facebook page from 2010 until the end of 2012, focusing on which themes were encouraged by the EC and in which way did the EC communicate the crises to European citizens. Moreover, conducting a content analysis as a quantitative analysis we examine how often each theme was in the main focus of the status update.

Keywords: European Commission, social media, economic crisis

Introduction

In the first decade of the new millennium the world was faced with a new economic crisis. Starting in 2008 in only few months economic crisis defected all modern markets, including the European Union market too.

The role of the European Union and its institutions in resolving and communicating the crisis remained unclear. Some studies suggest that citizens were badly informed about the crisis and about the causes of the crisis. The Reuters Institute conducted extensive research on media coverage during the economic crisis in Europe. The study included 10 member states and more than ten thou-

sand articles in 40 national newspapers published from 2010 to 2012. The data showed that in most of the examined member states, European institutions, including the European Commission, were perceived as ineffectual and confusing in addressing the crisis. Moreover, the data revealed that the European Union and its institutions were negatively perceived in the national print media of the examined countries during the period of crisis.

While we could argue that European institutions are maybe underrepresented and framed as a scapegoat in national media coverage, we wanted to test if they used alternative media channels to communicate about the crisis. We analysed the European Commission because one of the main goals of this institution is "communication" with the citizens of the European Union. Moreover, in 2010 the European Commission proposed a 10-year strategy named "Europe 2020" that, among other important issues, underlines seven flagships initiatives to catalyse progress under each priority theme[1]. One of the flagships was "a digital agenda for Europe" whose aim is to deliver sustainable economic and social benefits from a Digital Single Market, based on fast and ultra-fast Internet and interoperable applications, with broadband access for all by 2013, access for all to much higher Internet speeds (30 Mbps or above) by 2020, and 50% or more of European households subscribing to Internet connections above 100 Mbp[2]. This regulation showed that the Internet and its development is considered an important and needed tool. Yet, it is not clear how the European Commission itself uses the Internet to empower its role among the citizens and member states of the European Union.

In order to reveal if and how the European Commission used Internet during the economic crisis we analyse 285 Facebook status messages on the official European Commission Facebook page from June 2010, when the European Commission opened its official Facebook page, to the end of 2012, when the intensity of the crisis in most of the EU member states was in decline. Using qualitative analysis we revealed the main themes in focus in the European Commission's communication with European citizens. Using quantitative analysis we examined how often each theme was in the main focus.

This paper is divided into two main sections – a theoretical and then empirical section. In the theoretical part we discuss new methods and tools of political communication in the 21st century with a focus on social media. In the empirical section we analyse the status updates of the European Commission during the economic crisis in Europe using both a qualitative and quantitative approach.

[1] Europe 2020 – 7 Flagship initiatives: (1) Innovation Union; (2) Youth on the move; (3) A digital agenda for Europe; (4) Resource efficient Europe; (5) An industrial policy for the globalization era; and (7) European platform against poverty

[2] Europe 2020 Flagship Initiative: "A Digital Agenda for Europe"

Theoretical part

Many citizens believe that governments make decisions like private clubs, without the engagement of citizens who can no longer find any connection between the process of governing and their own elective vote (Hain, 2003). Furthermore, many authors believe that contemporary trends in the field of political media coverage, as well as trends in political communication, sometimes regarded as "spin democracy" stressing high inclusion of media logic in the political process, have resulted in a general distrust of political institutions, citizens' skepticism and disengagement (Entman, 1989; Hart, 1994; Cappella/Jamieson, 1997). Optimistic viewpoints about possible role of the Internet and new communication technologies in politics strongly underline how the Internet could solve these problems, strengthening ties between representatives and the represented, enabling two way communications with citizens in the role of strengthening democracy (Barber, 1984; Coleman, 2004; Grossman, 1995).

Yet, the question is whether citizens will renew their faith in the politics now when they have a platform where they can be equal with their representatives? Coleman thinks that ICTs can help through the creation of a more transparent, interactive government that is engaged in a wide dialogue with an interactive citizenry (2003). Street finds that the Internet may offer solutions for problems that obstruct political participation – "time, size, knowledge and access" (2001: 217). Chadwick suggests that political websites should provide infrastructure for deliberation and that political participation will follow (2006: 26). The rise of social media across Western democracies raises important questions for political scientists (Vaccari et. al, 2013: 2). How parties, politicians and crucial political institutions now use these new communication platforms, and how they directly engage with the audience and voters have become a new research area for scholars of political communication. In particular, it is assumed that social media may be conducive to greater political engagement in various ways: by providing information about political issues, by offering social cues that motivate citizens to take action, and by reducing the costs of collective action (ibid). After a long period of the mass media serving as a mediator between politicians and citizens, today, social media have created possibilities for direct communication and interaction, for citizens' engagement and political promotion.

The growing relevance of the Internet in everyday life, especially applications for social networking, such as Facebook, and micro-blogging sites, such as Twitter, is undisputed. The revolution of information communication technologies (ITC) in last decade have completely changed traditional forms of communication and mediation. With 511,483,906 citizens in European Union (EU28), the number of Internet users in the Union is 391,395,602 and the number of Facebook subscribers is 194,342,680[3]. Hence, in March 2015, the European Com-

[3] Internet World Stat, http://www.internetworldstats.com/europa.htm

mission set out three main areas regarding the Internet: (1) Better access for consumers and businesses to digital goods and services; (2) Shaping the environment for digital networks and services to flourish, and (3) Creating a European Digital Economy and Society with long-term growth potential[4].
In this chapter we have so far discussed the importance and possibilities of Internet, and specifically social media, in improving communication between representatives and the represented. In the next section we examine how the European Commission, as one of the executive bodies of the European Union uses Facebook to communicate with citizens. We chose to examine the European Commission's communication during the economic crisis in Europe as it is assumed that during the crisis communication should be more intense.

Research design

Since the purpose of this study is to detect the most common themes and subthemes of the European Commission's status messages on Facebook, as well as the main focus of each message, our approach is qualitative and quantitative.
The aim of inductive qualitative analysis is to conclude a general theory or concept from less general parts of the text (Thomas, 2006: 237). That is why we use this method to extract main themes and sub-themes from the status messages.
On the other hand, content analysis is a method for data collection, whose aim is an objective, systematic and quantitative description of the manifest content of communication (Berelson, 1952). With this method the aim is to gain insight into the incidence of certain themes and sub-issues as well as detecting the change of the themes throughout the years.
In order to reveal the trends and changes in issues that were in the focus of the European Commission's status messages on Facebook during the economic crisis we propose following research questions:

- **RQ1:** Did European Commission recognize social media as an important channel in its communication with citizens?
- **RQ2:** Did the European Commission communicate about the economic crisis with its citizens on its official Facebook page.
- **RQ3:** What were the main topics of status messages on the official European Commission Facebook page?

We conducted an analysis on 285 status messages in the period from June 2010, when the EC opened its official fan page on Facebook, to the end of 2012. This time period is also the period in which all European countries were caught in the global economic crisis.
We divide the analysis into two parts. The first part involves an inductive qualitative approach where by analysing Facebook status messages we detect the

[4] Europe.eu: Digital Single Market Strategy: European Commission agrees areas for action (http://europa.eu/rapid/press-release_IP-15-4653_en.htm)

main themes and sub-themes that the European Commission published on its Facebook page in examined time period. This gives us an overview of the key issues and topics that the EC, as an important body of European Union, published on its official channel. The second part involves quantifying the data in the detected content categories. The purpose of this part of research is to obtain information concerning the representation of a specific issue and to detect differences in referring to that certain issue during the time of crisis in Europe.

Results and discussion

The qualitative analyses detected seven main themes: economy, European Union policy, democracy, member states' policy, references to history, international relations, and specific personas that are highlighted in the status messages (Table1).

Table 1. Themes and Sub-themes of European Commission status messages

THEMES	SUB-THEMES
Economy	Economic reconstruction; market economy; employment / unemployment; economic crisis; green policy and sustainable development
European Union policy	News form the Union; national development: European initiatives; policy for youth; regular European policy; collaboration with other European Union nations; the EU presidency; specific problems of the Union
Democracy	Pluralism, human and minority rights; criminal, corruption and nepotism; rule of law and justice
European Union member states policy	Specific initiatives; specific Member state problems; regional relations
Reference to the countries who are not member states of the European Union	Collaboration with the Union; accession countries; international associations; the European crisis
Reference to history	EU history; certain parts EU history; history before the EU; culture in the context of history
Reference to certain people in the Union	Individuals from EU; individuals not from EU

Quantitative part of the study deals with the presence of certain issues in the EC's status messages. For each status we coded every sub-theme in order to reveal the frequency with which specific issues were mentioned. The data show that in the examined period the EC's official Facebook channel dealt with the theme of the economy in 7% of the examined status messages, while the crisis was mentioned 14 times in the examined time period. Furthermore, the crisis was mentioned in the context of providing solutions or information to overcome it. Yet, most of the time it was mentioned in the context of neglecting the seriousness of the issue and promoting certain EU initiatives (see Chart 1).

Issues concerning regular European Union policy were mentioned in most of the status messages. For instance, the news from the Union was mentioned in 84

of the examined messages. News from the Union encompassed a summary of weekly news, information, new regulations, achievements, etc. During the examined time period the EC dealt with initiatives and EU policy in 67 and 34 of the examined status messages, respectively (Chart 2).

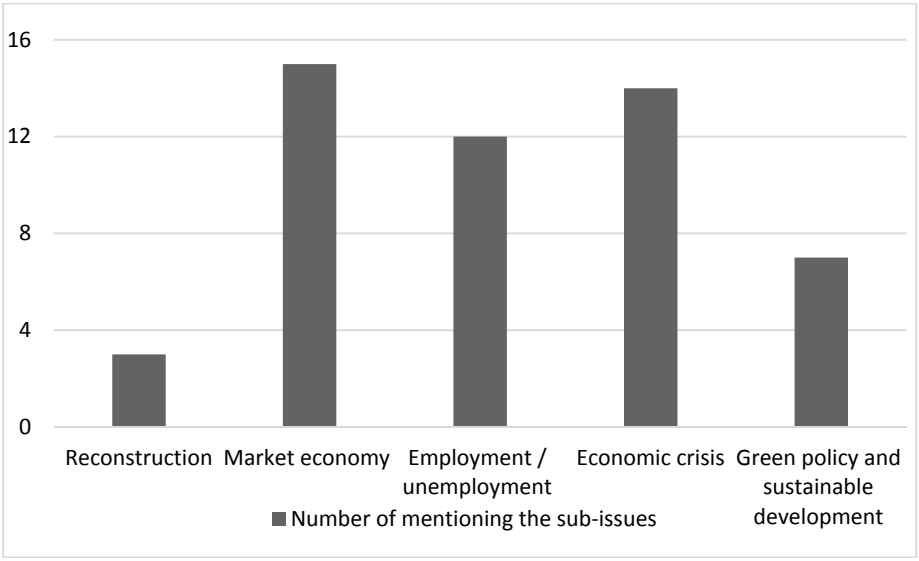


Chart 1. The frequency of mentioning economy issues

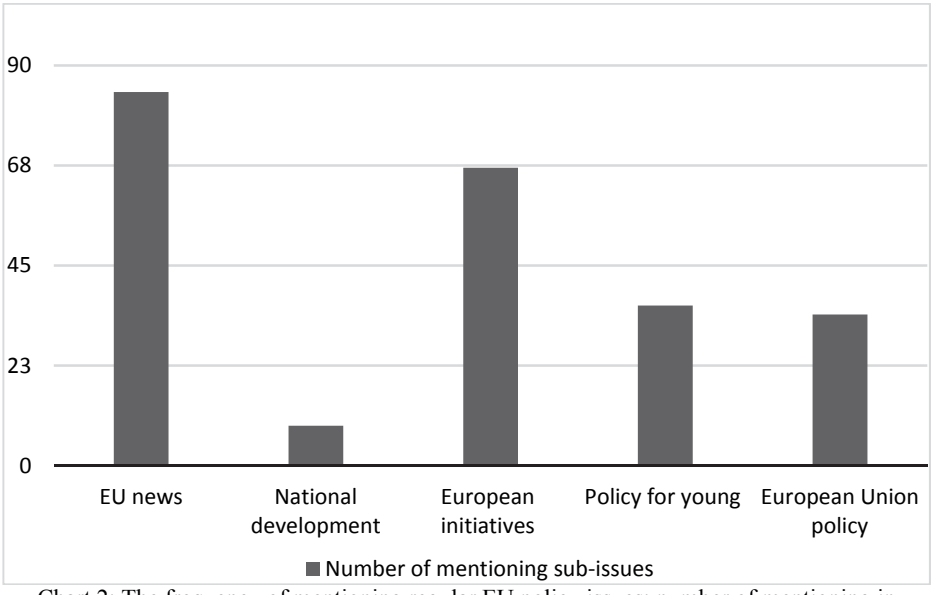


Chart 2: The frequency of mentioning regular EU policy issues; number of mentioning in Facebook status messages of European Commission

In 7% of the examined status messages the European Commission dealt with the theme of democracy. Issues such as human rights in the Union – which encompass regular citizens' rights, rights for more specific social groups – children and woman especially, were mentioned in 24 of the messages.

The data show that the European Commission in the period of economic crisis usually used its Facebook account to promote and communicate its projects, plans and achievements. During the period of research a great number of the status messages highlighted the initiatives of the European Union or the specific interests of certain social groups. Furthermore, our research shows that the EC was eager to stimulate discussion and action of users, inviting them to view some material, to comment, answer questions or to do some offline action (to vote, to watch some material on other media, or to get involved in a project). As a result, in the examined time period more than eight thousand comments were posted on the EC's official Facebook fan page, there were more than 18 thousand Likes on the 285 examined status messages and 10,228 shares.

This research showed that the EC was silent and ignorant about the economic crisis on its official Facebook page. Furthermore, the results suggest that the EC's Facebook page was used as a promotional tool for activities, policies and initiatives.

On the other hand, as already mentioned, the ongoing Reuter's research shows that the crisis and the role of the European Commission was an important part of the print media's coverage of European Union in member states. While on the Facebook account the EC was silent about problems and difficulties of the economic crisis the theme of the crisis had a greater presence in the national print media of certain member states. The Reuters research looked at the period from 2010 to 2012. So far the research has observed that the crisis was an important theme in the national print media of each analysed member state. The conducted research reveals that the stories about the European Commission were mostly written in a form of a news story in the first ten pages of the newspaper and usually in the section of business or finance. In most of the analysed countries, European Union institutions, as well as the European Commission, were perceived as ineffectual and confused in addressing the crisis. The data show that the European Union and its institutions were perceived more negatively than positively in the national media during the time of crisis. The conducted Reuters analysis shows that in most member states the root of the crisis refers to political roots, the structure of the Euro system, as well as national fiscal and social policies. In most member states the responsibility for the crisis refers to the members of the Eurozone as a group.

Following, it remains unclear why the European Commission did not take advantage of social media as a more transparent and journalistically unmediated platform to communicate its position, role and specific actions concerning and dealing with the economic crisis.

Conclusion

Due to its popularity and wide usage, social media has become an important and necessary tool in political communication. The main intention of this paper was to determine if the European Commission recognized Facebook as an important tool in connecting with citizens in the period of economic crisis in Europe. Our intention was to identify the main issues the EC presented in its status messages on its official Facebook page.

The research was conducted at two levels: first, a qualitative analysis of EC status messages identified the main themes and secondly, a quantitative analysis gave more specific insights into the presence of certain themes and sub-themes. Moreover, we examined the overall focus, technical details and response rate of the European Commission.

The Main themes we identified in the examined status messages during the economic crisis in European Union were the economy, European Union policy, democracy, and member states' policy, reference to history, international relations, and specific individuals. Our analyses showed that during the delicate period of crisis in the European Union, the European Commission mentioned economic crisis, which influenced the economy of entire Union, only 14 times on Facebook.

Our findings suggest that European Commission in general recognized Facebook only as a notice board while it did not recognize its Facebook page as an important platform for communicating with citizens. Bearing in mind that European institutions are often perceived as ineffectual and that their role remains vague and unknown to citizens, it is quite surprising that the EC did not use social media in a more communicative manner in order to clarify its role, policies and functions, especially during the economic crisis. Furthermore, social media is often a valuable source of information for journalists, which is why these findings are even more surprising.

Since the European Commission itself stresses the importance of the Internet and new communication platforms, as it is underlined in its strategic document *Europe 2020*, it remains to be seen if the European Commission will start using social media platforms more efficiently in the future. Moreover, a great number of citizens' comments, likes and shares on the status messages in the examined time period, suggest that citizens are willing to engage online if platform and content for engagement are provided.

References

Barber, Benjamin R. Strong Democracy: Participatory Politics for a New Age. Berekly: University of California Press, 1984

Barnhurst, Kevin G. The New "Media Affect" and the Crisis of Representation for Political Communication. // International Journal of Press/Politics. 16 (2011), 4; 573-593

Borchardt, Klaus-Dieter. Abeceda prava Europske Unije. Luxemburg: Ured za publikacije Europske Unije, 2011

Ćapeta, Tamara; Rodin, Siniša. Osnove prava Europske Unije na temelju Lisabonskog ugovora. Zagreb: Narodne novine, 2011

Cappella, Joseph N.; Jamieson, Kathleen Hall. Spiral of Cynicism : The Press and the Public Good. Pennsylvania: Oxford University Press, 1997

Chadwick, Andrew. Internet Politics: States, Citizens and New Communication Technologies. New York: Oxford University Press, 2006

Coleman, Stephen. Explorning New Media Effects on Representative Democracy. // Journal of Legislative Studies. 9 (2003), 3; 1-6

Coleman, Stephen. Explorning New Media Effects on Representative Democracy. // Journal of Legislative study. 9 (2004), 3; 1-16

Entman, Robert M. Democracy Without Citizens: Media and the Decay of American Politics. New York: Oxford University Press, 1989.

European Commission. Economic Crisis in Europe: Causes, Consequences and Responses. Luxembourg: European Economy 7, 2009.

Gibson, Rachel Kay; Nixon, Paul; Ward, Stephen. Political parties and the Internet: net gain? London: Routledge, 2003

Grossman, Lawrence K. The electronic republic. New York : Viking, Penguin, 1995

Hain, Peter. Why Bigger is Better: Enlargement and the Future of Europe. // New Economy. 10 (2003), 2; 95-100

Hart, Roderick. Seducing America: How Television Charms the American Voter. New York: Oxford University Press, 1994

Hill, Kevin A.; Hughes, John E. Cyberpolitics: Citizen Activism in the Age of the Internet. Lanham: Rowman & Littlefield Publishers, 1998

Kaldor, Mary; Selchow, Sabine; Deel, Sean; Murray-Leach, Tamsin. The 'bubbling up' of subterranean politics in Europe. London: London School of Economics and Political Science, Civil Society and Human Security Research Unit, 2012

Kavanaugh, Andrea; Patterson, Scott. The impact of community computer networks on social capital and community involvement. // American Behavioral Scientist. 45 (2001), 3; 496-509

Murray-Leach, Tamsin; Kouki, Hara; Mattes, Daniel; Mollenkamp, Daniel; Radice, Henry; Saltman, Erin Marie; Secondo, Donata; Sommer, Moritz. Crisis Discourses in Europe Media EU-phemisms and Alternative Narratives. London: London School of Economics and Political Science, Civil Society and Human Security Research Unit, 2014

Musiałkowska, Ida; Sapała, Magdalena; Wróblewski, Łukasz. The strengthening of the Single European Market vs. the crisis. // Poznań University of Economics. 12 (2012), 2; 74-105

Norris, Pippa. Political Communications and Democratic Politics, pp. 163-180 in Bartle, John and Griffiths, Dylan (eds) Political Communication Transformed: From Morrison to Mandelson. Basingstoke: Palgrave, 2001

Peter Dahlgren. The Internet, Public Spheres, and Political Communication: Dispersion and Deliberation. // Political Communication. 22 (2005), 2; 147-162

Savigny, Heather. Public Opinion, Political Communication and the Internet // Political science association, University of Birmingham. 22 (2002), 1; 1-8

Street, John. Mass Media, Politics and Democracy. London: Palgrave Macmillan, 2001

Vaccari, Cristian; Valeriani, Augusto; Barberá, Pablo; Bonneau, Richard; Jost, John T.; Nagler, Jonathan; Tucker, Joshua. Social media and political communication. A survey of Twitter us-

ers during the 2013 Italian general election. // Rivista italiana di scienza politica. 43 (2013), 3; 381-410

Wille, Anchrit. The politicization of the EU Commission: democratic control and the dynamics of executive selection. // International Review of Administrative Sciences. 78 (2012), 3; 383402

Wojcieszak, Magdalena. Transnational Connections Symposium: Challenges and Opportunities for Political Communication Research. // International Journal of Communication. 6(2012); 255265

Content Analysis Dataset, The Euro Crisis, Media Coverage, and Perceptions of Europe within the EU Project, Reuters Institute for the Study of Journalism, University of Oxfordm (2015): http://reutersinstitute.politics.ox.ac.uk/publication/euro-crisis-media-coverage-and-perceptions-europe-within-eu (Access date:15.4.2015.)

Europa.eu (2010): http://europa.eu/about-eu/institutions-bodies/european-commission/index_en.htm (Access date: 09.03.2015.)

Europe 2020. A strategy for smart, sustainable and inclusive growth, European Commission, COM (2010): http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:EN:PDF. (Access date: 15.3.2015)

Making sense of the impact of social media on crisis communication (2013): http://www.prconversations.com/index.php/2013/11/making-sense-of-the-impact-of-social-media-on-crisis-communication/ (Access date: 15.03.2015.)

The Euro Crisis in the Press : The Politics of Public Discourse in Europe – LSE blog (2015): http://blogs. lse.ac.uk/eurocrisispress/about/ (Access date: 27.04.2015.)

The European Parliament LGBT rights (2010): http://www.lgbt-ep.eu/work/priority-2/ (Access date: 09.03. 2015.)

Facebook as a Museum Content Sounding Board

Ileana Kurtović, Željka Miklošević
Department of Information and Communication Sciences
University of Zagreb Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
ilkurtovi@gmail.com, zmiklose@ffzg.hr

Summary

This paper gives an analysis of the ways in which three US museums with longstanding presence on social media and high online audience engagement use Facebook as a platform for distributing information about their programmes and topics relevant for their collections. Information technology influences the work environment in heritage institutions, changes their organization and calls for new responsibilities in the process of shaping online content in order to meet the needs of online users. The primary goal of the paper is therefore to reveal these activities and detect professional trends that can be of use in further research and help define online strategies for museums in general.

Keywords: Facebook, museums, communication, online information resources

Introduction

The development of new media has changed the way we find and exchange information, learn, establish and maintain relationships, or communicate in general. In order to follow these trends museums need to redefine and redesign their traditional cultural services into new forms of resources which engage audiences via new technologies. Online information resources play today a major role in determining museums' mission statements and communication strategies. Marty's research on the use of museum online or onsite information resources[1] shows that among 1065 participants 75.6% agreed[2] that museum websites should take advantage of the online environment to present unique experiences that cannot be duplicated in museums. In addition, 83.9% agreed that their needs and expectations of information on museum websites are different from their needs and expectations when visiting museums. Social media sites, for example, offer various ways of improving museum marketing and commu-

[1] Marty, Paul F. Museum websites and museum visitors: digital museum resources and their use. // *Museum Management and Curatorship*. 23 (2008), 1; 81-99.

[2] The percentages here represent both "agree" and "strongly agree" answers to the proposed statements.

nication practices[3] and establish physical or virtual relationships with their audiences. They are used by museums for "pushing" information, either to promote the institution and its programmes[4], or to distribute knowledge in the subject areas about which they build their collections. The former practices belong to what Kidd calls the marketing framework of social media use[5]. The analysis of three US museums' Facebook pages presented in this paper aims to illustrate how the selected museums use this particular social media site as a sort of sounding board for numerous institutional activities and information resources.

Research Questions and Methods

The museums selected for the study include New York's Museum of Modern Art, Museum of Science and Industry Chicago and the National Museum of American History. The selection is based on Edson and Dean's typology of museums according to collection type (art, social history and science)[6], on the number of Facebook likes provided by Museum Analytics[7] and the country of origin – the US as the country whose museums, in addition to UK ones, are leading institutions in social media use.[8] Their high social media presence and online value creation makes them representatives of online practices whose analysis can help in forming digital communication strategies for other museums. The research questions are therefore the following:

- What type of content do museums deliver through Facebook?
- What is the ratio between posts related to online and onsite content?
- What is the difference in museum online contents among different types of museums?
- What is the relationship between online content and engagement of Facebook users?

The first phase of the research included an analysis of posts published on the museums' Facebook pages from January to December 2014 and their categorization according to whether the information they contain relate to onsite programmes or online content. The formulation of categories and sub-categories followed a pilot study period during which the categories were tested and modified.

---

[3] Kelly, L. How Web 2.0 is Changing the Nature of Museum Work. // *Curator.* 53 (2010), 4; 405-410.

4 Dudareva, Natalia. Museums in Social Media. // *Museums and the Web 2013* / Proctor, N.; Cherr, R. (ed.). Silver Spring, MD: Museums and the Web, 2014. URL:http://mwf2014.museumsandtheweb.com/paper/museums-in-social-media/ (15.09.2014)

[5] Kidd, Jenny. Enacting engagement online: framing social media use for the museum. // *Information Technology & People*. 24 (2011), 1; 64-77.

[6] Edson, Gary; Dean, David. The Handbook for Museums. London, New York: Routledge, 1996

[7] Museum Analytics.URL: http://www.museum-analytics.org/museums/ (10.03.2015.)

[8] Padilla-Meléndez, Antonio; Águila-Obra, Ana Rosa del. Web and social media usage by museums: Online value creation. // *International Journal of Information Management.* 33 (2013); 892-898.

Posts related to onsite programmes include the following sub-categories:

1. Programmes in the museum (information related to current and upcoming events, workshops, lectures, special guided tours, film screenings, etc.)
2. Programmes in the museum shop (book and product presentations)
3. Programmes in other sites (events at other institutions)

Post related to online content include the following sub-categories:

1. Posts with links to museum sites[9] (museum collection with interpretation of objects, museum blogs with interpretation of objects, group of objects or exhibitions, blog wrote in collaboration with museum curators, museum web pages with games or educational resources)
2. Posts with links to non-museum sites (information published on news portals, online magazines, websites of other museums)
3. Posts with no links (information posted only on Facebook such as videos and photos with short introductory text)

The second phase of the research included a quantitative analysis of 1522 posts in the SPSS 17.0. A frequency table was used to describe the sample. The data were also evaluated by bivariate correlation analysis and the results displayed as pie and column charts in Tables 1 – 6. Considering the topic of the paper, the results shown here are more focused on online museum content and users' responses to it.

Table 1. Ratio between onsite programmes and online content

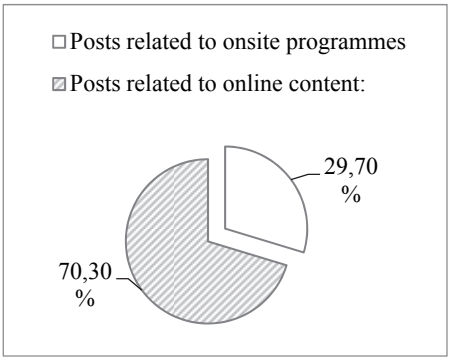
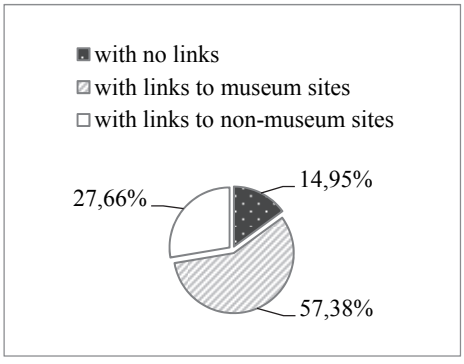


Table 2. Percentages of posts in three sub-categories of online content



[9] This sub-category does not include posts with links to the museum website content that solely provides short information about duration, participants, and venue of events or give a short description of an event and provide no interpretation of museum objects, exhibits or the like.

Table 3. Types of post related to online content on each museum's Facebook page

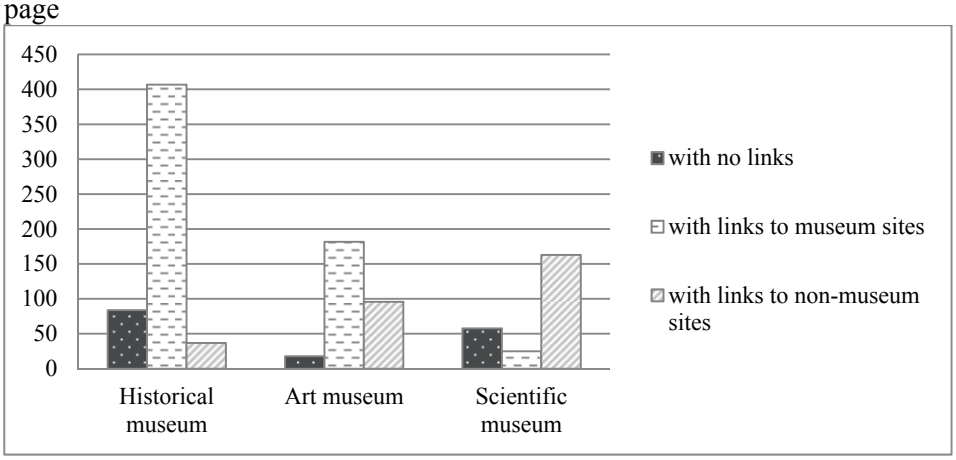


Table 4. Online content and number of likes

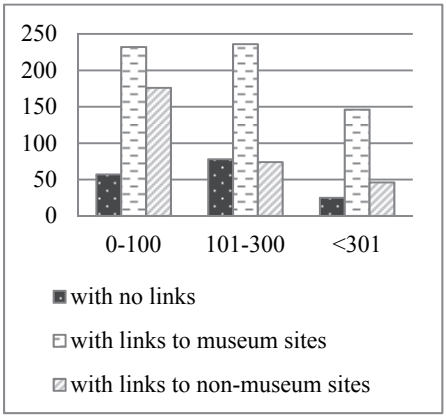


Table 5. Online content and number of comments

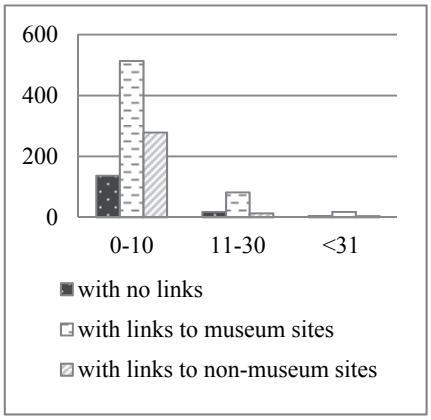
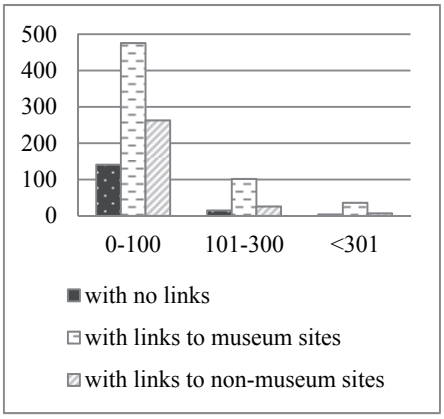


Table 6. Online content and number of shares



Discussion and conclusion

The results show that a considerable number of posts contain information related to online content (Table 1) proving that these three museums have moved away from using Facebook to promote the "face" of the institution through their onsite programmes. From the total number of posts to online content, the biggest number or 57.38% relates to contents on museum websites (Table 2). These resources can inform and prepare people for the physical visit, reinforce the experience after the visit, or create independent, online experiences that can be utterly different from onsite programmes. When examined individually (Table 3), the museums show a difference in the way they shape and contextualize different information. The Science and Industry Museum Chicago publish the biggest number of posts with links to non-museum sites such as online science magazines and portals that deal with science-related topics in general. By doing that, it provides much larger context for topics it deals with and objects it collects. This sort of activity supports the museum communication strategy that is mostly grounded in science phenomena more than objects that witness scientific developments. Conversely, MoMA's posts are more related to collections and exhibitions whose interpretation is given on the museum website. Its online content is more than in other two museums tied to the collection, which is especially evident in Facebook posts directly linked to the online catalogue of objects. Facebook posts of the National Museum of American History are largely dependent on the museum website where curators and educators develop interpretive packages for users. Very small number of links to other online sites may be explained by a wide range of approaches, views and interpretations of historical topics and the reluctance of the museum to refer to some of them. Further elaboration is needed on the relationship between the museums' subject area and digital resource formation.

The statistical analysis of likes, comments and shares, in relation to online content (Tables 4 – 6) shows that information developed on museums' web sites and distributed through Facebook engage users the most. On the one hand, it reflects the highest number of posts with links to online museum sites, but, on the other hand, it can be related to a specific manner of delivering information by museum professionals. However, better understanding of the relationship between digital content and user engagement requires a mixed-method research which could show correspondence between types of content and types and frequency of responses.

References

Dudareva, Natalia. Museums in Social Media. // *Museums and the Web 2013* / Proctor, N.; Cherr, R. (ed.). Silver Spring, MD: Museums and the Web, 2014
URL:http://mwf2014.museumsandtheweb.com/paper/museums-in-social-media/ (15.09.2014)

Edson, Gary; Dean, David. The Handbook for Museums. London, New York: Routledge, 1996

Holdgaard, N. The Use of Social Media in the Danish Museum Landscape. // *Museums and the Web 2011: Proceedings* / Trant, J.; Bearman, D. (ed.). Toronto: Archives & Museum Infor-

matics URL: http://conference.archimuse.com/mw2011/papers/the_use_of_social_media_in_the_danish_museum (24.05 2012.)

Kelly, L. How Web 2.0 is Changing the Nature of Museum Work. // *Curator.* 53 (2010), 4; 405-410

Kidd, Jenny. Enacting engagement online: framing social media use for the museum. // *Information Technology & People*. 24 (2011), 1; 64-77

Marty, Paul F. Museum websites and museum visitors: digital museum resources and their use. // *Museum Management and Curatorship*. 23 (2008), 1; 81-99

Museum Analytics.URL: http://www.museum-analytics.org/museums/ (10.03.2015.)

Padilla-Meléndez, Antonio; Águila-Obra, Ana Rosa del. Web and social media usage by museums: Online value creation. // *International Journal of Information Management.* 33 (2013); 892-898

TRUST IN THE DIGITAL ENVIRONMENT

Original scientific paper

A Comparison of Users Data in Retention and Disposition Processes in an Internet Website: The Ministry of Foreign Affairs of Israel as a Case Study*

Silvia Schenkolewski-Kroll
Bar Ilan University
Ramat Gan, Israel
sil_sch@netvision.net.il

Assaf Tractinsky
Israel State Archives
14 Hartom St., Har Hotzvim, Jerusalem, Israel
assaft@archives.gov.il

Summary

In our appraisal of the English website of Israel's Ministry of Foreign Affairs, we took into account the manner in which Israeli government ministries conduct their records management. However, due to the unique characteristics of internet websites, special parameters are needed, such as harvesting and user behavior. We conducted a functional analysis of user behavior on the MASHAV subsite and the "Foreign Policy" division of the ministry's website, using the "About Israel" division as a control unit for comparison. Since MASHAV provides information and services to users, it is possible to use the retention periods set in the Archives Law 5715–1955 for some of its parts. "Foreign Policy" is an informative division whose documents are replaced according to changing event.

In all three, we examined and analyzed the number of users and the length of time spent on the site during each access. The comparison was done for February and October 2014, and for the total year. The results show that in the "Foreign Policy" and "About Israel" divisions the number of users was greater than in MASHAV and fluctuated depending on the circumstances and events, while users remained in the MASHAV subsite for longer periods of time.

Keywords: appraisal; users; Ministry of Foreign Affairs – Israel

* This article is part of a research study of InterPARES Trust "Research of retention and disposition processes in an internet website of the Government of Israel: The Ministry of Foreign Affairs as case study" (Unpublished).

Archival Appraisal Processes in the Ministry of Foreign Affairs of Israel

As a short introduction, it should be noted that the Ministry of Foreign Affairs, as an integral part of the governmental system of the State of Israel, conducts its records management in accordance with the Archives Law 5715–1955. The State Archivist, through the Israel State Archives (ISA) and its administration, is responsible for enforcing the Law in all government ministries, state institutions, and local government. As such, archival appraisal, which is part of the Ministry of Foreign Affairs' records management, is executed in accordance with the system set forth by the ISA.[1]

A special department within the Ministry of Foreign Affairs, the Records Management and Databases Center, is responsible for records management. It has a long tradition of proper records management ensuring that all records of all the Ministry's units, including those in Israeli missions worldwide, undergo archival appraisal, including determination of retention schedules.[2] This pertains to both traditional and electronic records, so they do not differ in terms of appraisal criteria. Beginning in 2000, the Ministry integrated the retention schedules of electronic records. To date, no material from the Ministry's electronic systems has been disposed of. Retention schedules will serve the Ministry of Foreign Affairs for two purposes in the future: the first, to classify material in the electronic systems according to level of importance; and the second, to transfer material intended for permanent retention to the ISA. Regarding preservation of record authenticity, we will need to re-examine this issue within the framework of the abovementioned processes[3].

The Ministry of Foreign Affairs' websites, including the website serving as a case study, have yet to be addressed within the framework of the archival appraisal process.[4]

[1] Shoham, S and Schenkolewski-Kroll, S (2010) Israel: Libraries, Archives and Museums,' Encyclopedia of Library and Information Sciences, Third Edition, 1: 1, 3035 — 3040

[2] Prime Minister Office, State Archivist, Higher Archives Council, Protocols and Reports, Jerusalem, from years: 1972, 1982, 1985, 1986, 1987, 1988, 1991 (Hebrew), and Arie Zini, The Place of the Filing Unites in the Computerization Era, 2011(Hebrew) http://www.archives.gov.il/ArchiveGov/meyda/ElectronicRashuma/YomIyun2010.htm

[3] Schenkolewski-Kroll, S. and Tractinsky, A, (2015). "Archival Appraisal Processes in the Ministry of Foreign Affairs of Israel" in Ministry of Foreign Affairs website final Report, Submitted to the European team of InterPARES Trust (Unpublished)

[4] http://mfa.gov.il/MFA/Pages/default.aspx

Attempts to Find a Basis for Appraisal in Existing Retention Schedules

After examining the various sections we decided, at this stage, to deal with an appraisal of the new website, beginning with 2013, with a view to the future, rather than to conduct a retrospective appraisal. This decision was taken due to the different characteristics of the website that was created in 1993 and the difficulties that might arise in an attempt to analyze the two systems concurrently.
Two additional topics may affect website appraisal: first, harvesting and its application to the website or to portions of the site[5]; and secondly, the users – an issue which to the best of our knowledge has not been taken into consideration in any website appraisal framework. Can the ways in which the website is used – by user segmentation, by website sections and the way they are divided internally – be a factor sufficiently significant to serve as an appraisal and retention schedule criterion? Is it possible that a large number of views of a website section or of a certain website page points to their importance in the eyes of the users?
In addition, in light of the research findings, the Ministry of Foreign Affairs will have to conduct its website records according to records management procedures, in order to create and appraise authentic records.
This article focuses mainly on analysis of the statistical measurements of user behavior. The harvest method mentioned below serves only for background in presenting the main subject.

Appraisal of Archival Material – Websites

Archivists faced a new challenge with the emergence of websites in the 1990s: how to cope with electronic records that were organized in their entirety in a manner they had not done previously. The purpose of an organizational website is to make information accessible to the general public about the organization or institute, its objectives, activities, the content of its material, and any other information which it deems necessary, worthwhile, or suitable to be made public. The material posted on the site can be divided into two main groups: records intended to distribute information as a goal within itself; and records that constitute an alternative to administrative services, such as posting forms that are part of a service to the public and receiving them back on-line, in order to implement the public's rights. Due to the nature of the service and its operation, a website is defined as social media. The records constituting the site have different diplomatic forms: some conform to traditional and electronic records, which are the result of a service, and others are more similar to publications – these are

[5] Brown A. (2006), Archiving Websites: A Practical Guide for Information Management Professionals. Facet Publishing. pp. 24-41; TNA (2011), Web Archiving Guidance (2011). http://nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf; TNA (2014), Operational Selection Policy OPS 27, UK Central Government Web Estate, http://www.nationalarchives.gov.uk/documents/information-management/osp27.pdf

part of what was defined above as information. Everything depends on the nature of the site and its purposes. In light of the above, one of the special characteristics of a website is the constant updating of the records appearing on it, with the goal of providing current information in all the fields with which the site is concerned. Updates may be made as often as several times a day during critical periods, and up to every few weeks or more when dealing with permanent administrative material.[6] Because one of the objectives of the site is the distribution of current information, the number of users is one of the characteristic criteria that determine at any one time the measure of the site's success, as opposed to other records which are not made available to the public until they become archival material. This situation raises questions as to the nature of the appraisal and retention of records whose source is the Internet: is it possible to use the same rules of a particular appraisal method, which are applied to traditional and electronic records that deal with the same subjects in the administration of the same organization that operates the website, or is there a need to find appraisal methods that meet the specific needs of a website? Can the number of the users of the site's various sections serve as a criteria for determining the importance of a site, or one of its parts, and influence the determination of a retention schedule for specific materials?

With regard to the process of choosing records for retention on websites or on some of their pages, the system for doing so is harvesting. This refers to a type of software that makes it possible to locate the site or its sections; "to harvest" is to create copies in the computer on a specific date and at intervals of time that have been determined by those responsible for this process. This makes it possible to adapt harvested material to the structure of the site, its content, the time, and a specific situation. For example, on the Ministry of Foreign Affairs site, during times of war material will be harvested at very short periods of time, compared with harvesting information during periods of calm. Retention by harvest parameters is accepted practice, and is mentioned in the literature.[7] The issue of the number of users and its influence on appraisal has not yet been dealt with in the literature. Only a systematic study over a specific period of time of the number of users compared with the material presented on various parts of the site can provide an answer about this criterion (see below).

For the purpose of demonstration, in this case study of the Ministry of Foreign Affairs site, two of its sections were considered. Two different types were deliberately chosen, each having its own unique characteristics which serve different purposes.

[6] Brown A. (2006), Archiving Websites: A Practical Guide for Information Management Professionals. Facet Publishing. pp. 24-41

[7] See above note 4.

One was MASHAV that can be defined as a subsite, since it is operated by an administrative unit having a certain amount of autonomy when compared to other administrative units of the Ministry of Foreign Affairs. It is mainly operational, and serves as a means of fulfilling the functions of the unit responsible for it, such as presenting courses and everything related to them, MASHAV's activities on four continents, and more. The purpose of this subsite is partly informative and partly fulfillment of administrative functions, such as enabling registration for courses. The function of "Foreign Policy," as its name indicates is to present the past and current policy and activities of the Ministry through information from the past and current news. The format of the material is similar to traditional publications in the written and electronic media.[8]

MASHAV

MASHAV was founded in 1957 with the main purpose of circulating Israel's technological expertise among the developing countries. Today MASHAV is a unit at the Ministry of Foreign Affairs headquarters which works through the Ihud company to service Israeli organizations that deal with international training and consultation for developing countries, providing them with humanitarian as well as technical aid. It implements its training programs with the help of three training centers in Israel. Its fields of activity are those in which Israel has a comparative lead and expertise that have accumulated throughout the years. The subsite includes the structure and activities of the administrative unit. The documents are generally divided into two types: governance and management information, and research and innovative information. The division in which new documents are requested at the greatest frequency is the Newsletter, ranging from one to three documents a month. The site is managed independently by one of the MASHAV staff.[9]

As can be seen in Figure 1, the subsite is divided into seven divisions, some of which are divided into subdivisions.

With regard to divisions, we collected data that would help us with the appraisal according to the categories common to all records: the name of the division, number of documents, document type, date of creation or event, additional documents, internal or external links, and retention schedules. In addition, we examined whether there are links within the division to other places on the site, or off it.

This datum is important to know in order to plan the harvesting process of division and its various parts, or of the entire site. We marked the links with a + when they appeared, and a – when they didn't, as in the "Publications" division. Finally, we noted whether there are retention periods in the Archives Law for

[8] See above note 3.

[9] An Interviews with Avnit Rifkin, the MASHAV webmaster, 2/1/2014

material found in the division. We also noted the retention period and the clause in the Archives Law, as in the division "MASHAV Professional Courses," where the retention period is 10 years after file closure or end of issue.[10]

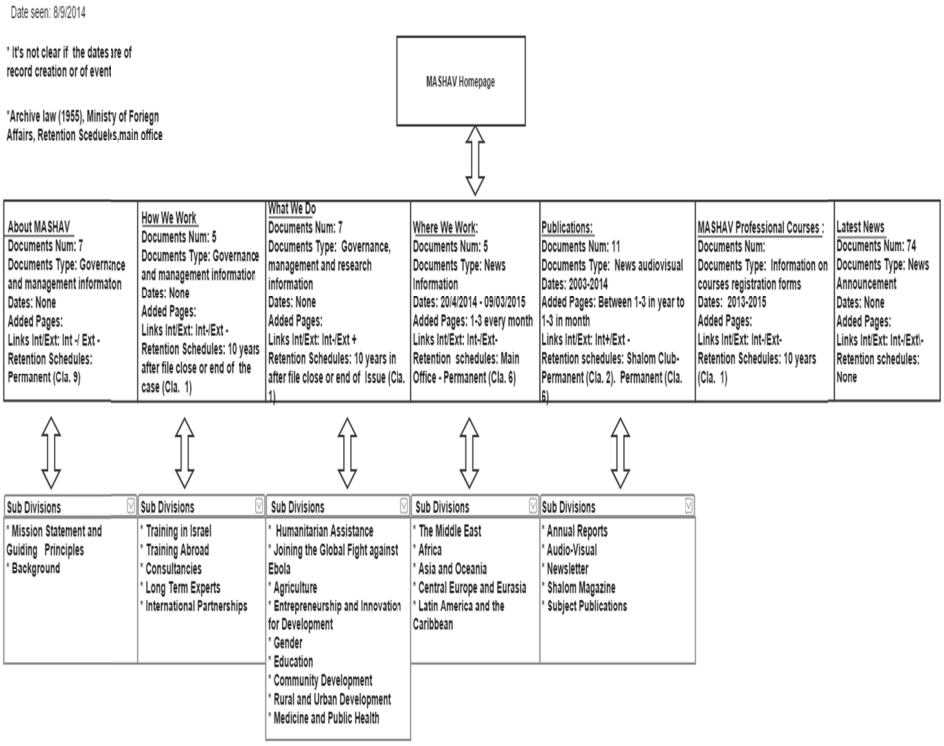


Figure 1 – MASHAV subsite and its seven divisions

Foreign Policy

The "Foreign Policy" division contains documents describing the foreign relations of the State of Israel. It changes and adds information according to the pace of events; in that, it is similar to traditional and electronic media. Various documents may appear on the site several times – in the divisions of the main page of the "Foreign Policy" division and in other divisions on the site, such as "International Organizations." The fact that the same documents appear in different places on the Ministry's site makes it difficult to determine time periods of updating. Obviously, when diplomatic activities are intensified, such as during Operation Protective Edge, material is updated more frequently, 2–3 times as often, depending on the judgment of the site manager. The main page of the

[10] Archives Law 5715–1955. Regulations and Instructions]p. 20 (Hebrew), http://www.archives.gov.il/ArchiveGov/meyda/TkufotBiur/taknot.htm

division then accumulates the latest major newsworthy items on Israel's foreign policy. In cases of such campaigns retention is permanent, especially if one takes into account the fact that the person responsible for the site has already applied a selection process as to which items to post (there are no clear rules on the issue, only personal judgment).[11]

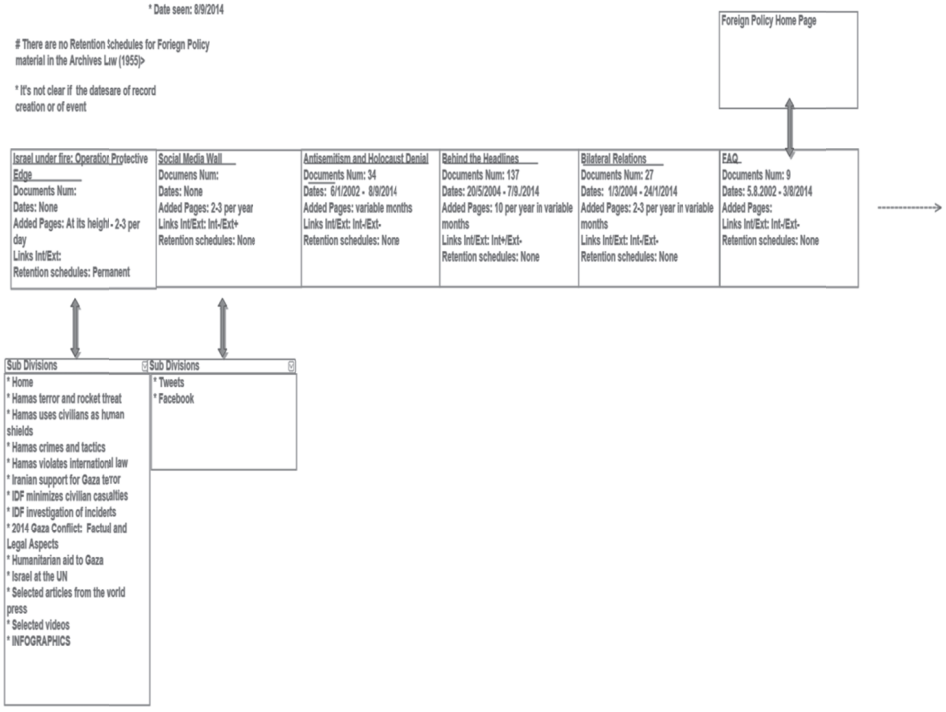


Figure 2A – Foreign Policy section and its twelve divisions

With regard to the divisions, we collected data according to the categories common to all: the name, number of documents, date of creation or event, additional documents, internal or external links, and retention schedules. The category "Document Type," which appears in the MASHAV data, was eliminated in this section because it was difficult to characterize the type of documents it presents. In addition, we examined whether there were links in the division to other places on the Ministry site or outside it. This item is important in order to know how to plan the harvesting process of the division and its various subdivisions or of the entire site. We marked + where links were found, and – where they were not, as was done in the "FAQ" division. Finally, we left in the category of retention periods, despite the fact that we did not find in the Archives

[11] An Interview with Sarah Lederhendler, the MFA webmaster, in 24/12/2013

Law periods for the material appearing in the section, and marked it "None." Only in "Israel under Fire: Operation Protective Edge" did we note the retention period, since this was a unique event.

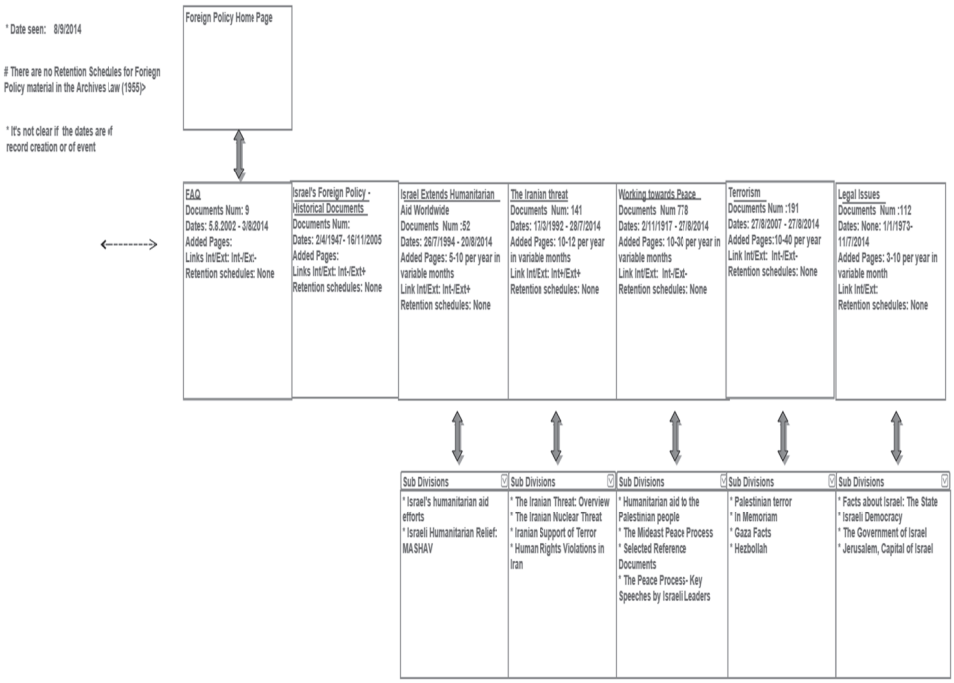


Figure 2B – Foreign Policy section and its twelve divisions – continue

Attempts to Determine Retention Periods for the Two Sections Examined

Because the Ministry of Foreign Affairs has updated retention periods for the entire administrative structure of the office,[12] an attempt was made to adapt the retention periods of the records for the two sections studied to the retention periods of corresponding administrative units in the structure of the Ministry. With regard to MASHAV, the parallels exist because that section represents a defined administrative unit. Therefore, clauses 1, 2, 5, 6 and 9 of the Archives Law guidelines, as applied to MASHAV,[13] compare with the divisions of the MASHAV section (see Fig. 1). The question remains whether this obligates MASHAV to apply its own retention period to its site.

[12] Archives Law,5715-1955, Regulations and Instructions, (Hebrew) http://www.archives.gov.il/ArchiveGov/meyda/TkufotBiur/taknot.htm

[13] See above note 11.

As for the "Foreign Policy" section, because of its publicist nature it has no parallel in the retention periods that were determined for the Ministry framework; therefore, it seems that only harvesting periods in accordance with the pace of events and/or the number of users can serve as criteria in this case. Since an advance selection of the material presented in this section is made by the person responsible for the site, and despite the fact that the criteria for his/her choices are unclear, the possibility of setting rules for selection has been discounted at this stage of the appraisal.[14] Therefore, it appears that there is no point in recommending setting retention periods for the "Foreign Policy" section.
From the above two examples, it may perhaps be possible to determine rules with regard to other sections on the site. Each one of them will require a process of study and analysis, both from the aspect of its internal organization and that of its content and the diplomatic form of the records.

A Comparison of User Data

The purpose of the experimental analysis is to try to integrate users' behavior on the website into the process of its appraisal. The goal of the data presented here is to present trends and estimates in the field, rather than results. The reasons for that are that the material on the website is mostly different than that of the internal organizational systems and, in contrast to those systems, the material on the site is accessible to the public at all times. Therefore, an external element may be required to assist in the site's appraisal, in addition to – or instead of – the criteria usually used in a standard appraisal, such as administrative use or legal, sociological, and research value.
The program we used was "Piwik,"[15] an open source system which can be downloaded from the Internet. It was installed on the Ministry of Foreign Affairs' website, in the government servers' farm, instead of the "Share Point"[16] system, which is used on the site and provides similar options.

Organization of the Data

The Piwik program data were collected from two sections, MASHAV and Foreign Policy, on which we conducted an in-depth study. In addition, a third section – About Israel – was chosen randomly to serve as a control group. The data on the three sections were recorded for the months of February and October 2014, as well as summary data for all of 2014. Comparative reports were produced for the three sections, which present the same data: the number of page views and average time on page during each of the periods studied, with page

[14] An Interview with Sarah Lederhendler, the MFA webmaster, in 24/12/2013.

[15] http://piwik.org/

[16] See for example: View Web Analytics Reports (SharePoint Server 2010). https://technet.microsoft.com/en-us/library/ee663487(v=office.14).aspx

views representing the number of times a page was visited. The average time on page represents the average amount of time visitors spent on a page.

February 2014

A. Number of Page Views in the Sections

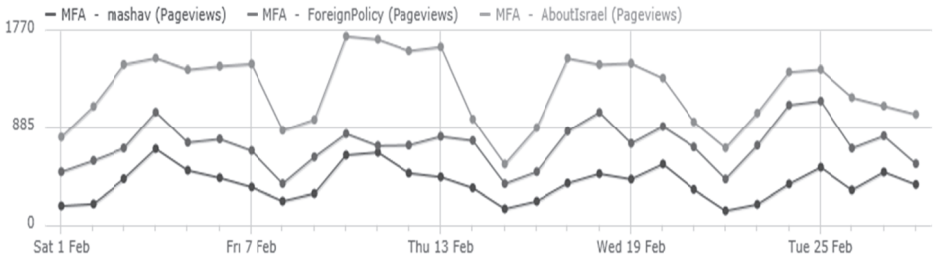


Figure 3 – The graph shows the number of views in the various pages of the three sections: MASHAV, Foreign Policy, and About Israel. (February 2014)

1. It can be seen that there are page views throughout the entire month in all the sections.
2. About Israel is the section most viewed, and the least viewed is that of MASHAV.

B. Time Spent on a Page

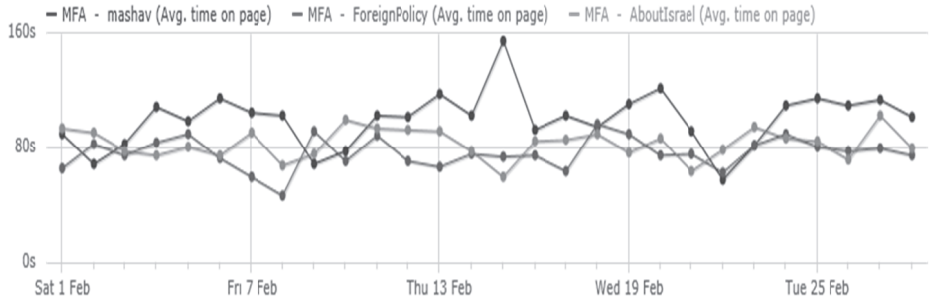


Figure 4 – The graph shows the average amount of time spent by users on the various pages in the three sections: MASHAV, Foreign Policy, and About Israel. The data are summarized in the system up to the section level (February 2014).

1. The average amounts of time spent in the three sections range around 80 seconds.
2. Most of the longest periods of time are found in MASHAV, with the peak of time spent on a page occurring on February 15, in that section, at a little less than 160 seconds.

October 2014

A. Number of Page Views in the Sections

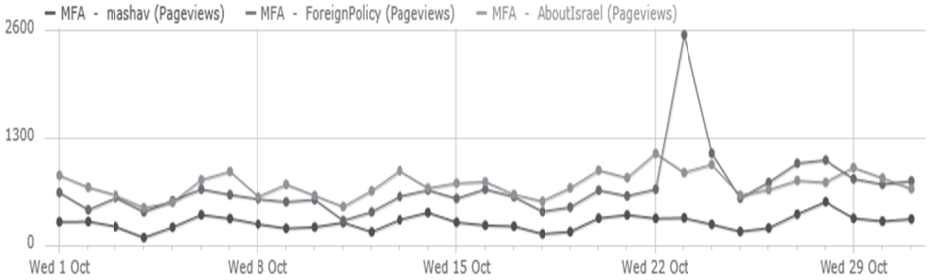


Figure 5 – The graph shows the number of views in the various pages of the three sections: MASHAV, Foreign Policy, and About Israel (October 2014).

1. All the sections were viewed during the month.
2. The number of page views – in general, the most page views occurred in "About Israel" and following that, "Foreign Policy." The fewest page views were in MASHAV.

A special peak can be seen on 23 October in "Foreign Policy." This may be related to a special incident which took place the previous day in Jerusalem: an Arab terror attack in which *a baby was killed* when it was deliberately run down by a car.

B. Time Spent on a Page

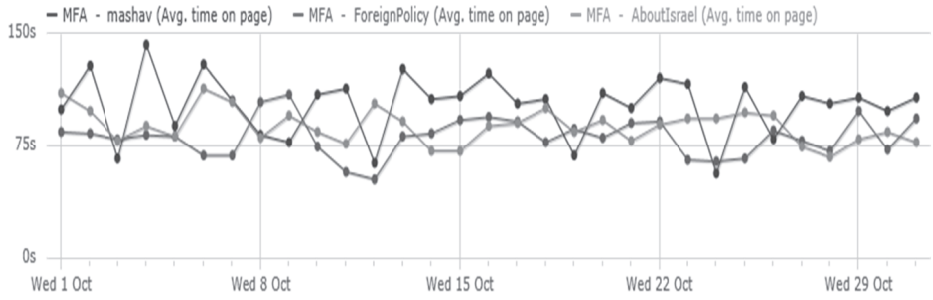


Figure 6 – The graph shows the average time spent by users on the various pages in the three sections: About Israel, Foreign Policy, and MASHAV (October 2014).

1. The average time spent in the sections, in most cases, is more than 75 seconds.
2. The longest average periods of time spent were in MASHAV, with the peak being a little less than 150 seconds.

Comparison of the February Data and October Data

In a comparison of the page views data for February and October, there are page views in all three sections that were studied. Most page views were in About Israel, while Foreign Policy was second. MASHAV had the least page views of the three. On the other hand, the average periods of time spent in the MASHAV section were the longest. In February the average periods of time spent in the MASHAV section were one minute and forty-one seconds, and in October they were one minute and forty-four seconds. The MASHAV section also has the longest average periods of time spent on a page, with a peak of slightly less than one hundred and sixty seconds in February and slightly less than one hundred and fifty seconds in October. Apparently, it is the operational nature of the section that causes the lengthy periods of time of page views.

Summary of 2014

A. Number of Page Views in the Sections

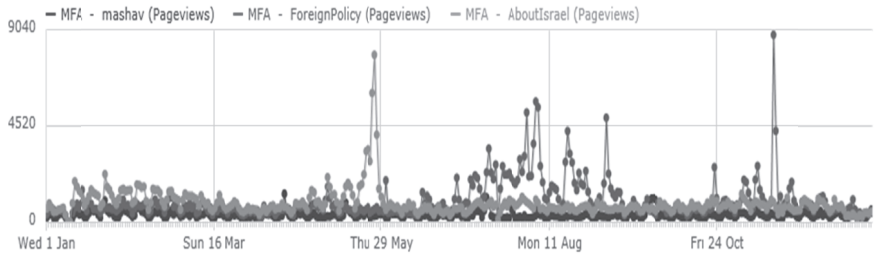


Figure 7 – The graph shows the number of views in the various pages in the three sections: MASHAV, About Israel, and Foreign Policy (Summary of 2014).

1. It is evident that throughout the first half of the year there were more page views in "About Israel" while in the second half there were more page views in "Foreign Policy."
2. We can see that there are some peak points of page views both in "Foreign Policy" and in "About Israel", with MASHAV having low page views data throughout the year.

B. Time Spent on a Page

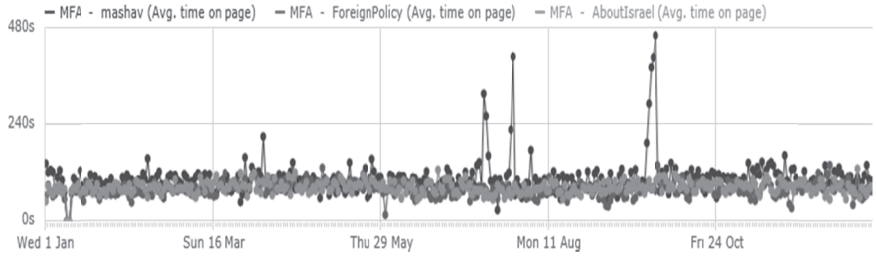


Figure 8 – Time Spent on a Page (Summary of 2014)

The graph shows the average periods of time spent by users throughout 2014 on the various pages of the three sections – About Israel, Foreign Policy and MASHAV.

The average periods of time spent on a page in 2014 were one minute and forty-five seconds in MASHAV, one minute and thirteen seconds in Foreign Policy, and one minute and twenty seconds in About Israel. The periods of time spent in MASHAV were lengthier during three periods of the year than the others. MASHAV is also characterized by four leaps in the data.

Conclusions/Trends

1. This type of system makes it possible to observe divisions and subdivisions throughout various periods of time. These data can serve as criteria that influence the appraisal and the determination of retention periods.
2. It can be seen that for the years and months studied the sections most viewed are About Israel and Foreign Policy. The section with the fewest page views is MASHAV. However, in MASHAV users spent the most time on pages, in accordance with the nature of the site and its operational nature.
3. According to the experience of the director of the website, the minimal significant amount of time spent on a page is at least 60 seconds. This datum requires additional study.
4. An additional element influencing the data presented here, which was not taken into account because it is presently beyond the scope of the research conducted, is the quality of the website from the aspect of its accessibility and links.
5. A study of the entry points of the users and their geographical origins may constitute parameters of significance in the evaluation.

This is the first attempt to examine the user behavior parameter to appraise internet websites. In order to know whether this parameter is valid beyond this specific case study, it is necessary to carry out additional studies which will either confirm the results or indicate the need of other types of examination.

References

Archives Law 5715–1955. Regulations and Instructions (Hebrew). http://www.archives.gov.il/ArchiveGov/meyda/TkufotBiur/taknot.htm

Brown A. (2006), Archiving websites: a practical guide for information management professionals. Facet Publishing. pp. 24-41

Prime Minister Office, State Archivist, Higher Archives Council Protocols and Report, Jerusalem (Hebrew)

Schenkolewski-Kroll, S. and Tractinsky, A, (2015). "Archival Appraisal Processes in the Ministry of Foreign Affairs of Israel" in Ministry of Foreign Affairs website final Report, Submitted to the European team of InterPARES Trust (Unpublished)

Shoham, S and Schenkolewski-Kroll, S (2010) Israel: Libraries, Archives and Museums, Encyclopedia of Library and Information Sciences, Third Edition, 1: 1, 3035-3040

TNA (2011), Web Archiving Guidance (2011). http://nationalarchives.gov.uk/documents/information-management/web-archiving-guidance.pdf
TNA (2014), Operational Selection Policy OPS 27, UK Central Government Web Estate, http://www.nationalarchives.gov.uk/documents/information-management/osp27.pdf
Zini, A. (2011). The Place of the Filing Unites in the Computerization Era, 2011(Hebrew). http://www.archives.gov.il/ArchiveGov/meyda/ElectronicRashuma/YomIyun2010.htm

Record Authenticity as a Measure of Trust: A View Across Records Professions, Sectors, and Legal Systems

Corinne Rogers
University of British Columbia
Vancouver, Canada
cmrogers@mail.ubc.ca

Summary

The ability to assess and protect the authenticity of records is axiomatic to archival science, and enjoys a centuries' long theoretical foundation. This paper presents a summary of the author's research into contemporary ideas about authenticity of digital records among records professionals. It found that the issue of authenticity is as much social as technical, with records professionals using social indicators over technical ones in managing authenticity, but relying on technical means to authenticate records. It presents the role of legal system as an independent variable in considerations of authenticity, opening up a line of inquiry for further study.

Keywords: Authenticity, Trustworthiness, Juridical system, Indicators of authenticity

Introduction

The ability to assess and protect the authenticity of records is axiomatic to archival science, and enjoys a centuries' long theoretical foundation. Authenticity is frequently identified by records professionals as a requirement, or a goal, when creating and preserving the records upon which modern society relies. Beyond that goal, however, identifying the attributes required to make an attestation of authenticity, or the conditions upon which authenticity can be assessed or presumed, is neither easy nor standardized. In this paper I will present a summary of research into contemporary ideas about authenticity of records and data, and practices employed in its service by records professionals – those entrusted both with managing current digital information, and preserving information no longer used by its original creator.

Background

Two aspects of records research that permeate archival discourse regardless of juridical system are the determination and maintenance of authenticity. The concept of documentary authenticity has ancient roots. The word authenticity

derives from the Anglo-Norman, Old and Middle French, with reference to a thing (as a noun, authenticum, originally and frequently a legal document), or a person (as an adjective, denoting trustworthy, credible, genuine, or legally or duly qualified). Its etymon is the Latin authenticus, referring to documents (2nd century a.d.), persons (3rd century a.d.), and later coming to mean something or someone who is authoritative (from 8th century in British sources), or a thing that is legally valid (12th century). In Hellenistic Greek, αὐθεντικός meant warranted, original, authoritative (Oxford English Dictionary 2014).

Record authenticity today is mandated in standards for records management and preservation frameworks such as ISO 15489:2001 and OAIS – Open Archival Information Systems (International Standards Organization 2001; CCSDS 2012), and the protection of authenticity is required by professional codes of conduct (cf. SAA 2011; ACA 1999). Although theoretical frameworks have been proposed for the protection of authenticity – some of them highly influential in practice, for example in the development of the DoD 5015.2 standard for records management systems (Duranti and MacNeil 1997) and in the development of Italian legislation (this research, Interview subject D073)), despite the wealth of past and current research, establishing and guaranteeing authenticity are still discussed as urgent problems yet to be solved. In the digital environment, research agendas in the information management communities focus on authenticity as an integral value that must be protected over time and across technological change through digital preservation (joining values of sustainability, accessibility, and understandability), broadening the scope of enquiry beyond records as defined by archival theory to include documents, data, and digital objects of all types.

Sir Hilary Jenkinson believed that archival documents (i.e. records) were "authenticated by the fact of their official preservation" (Jenkinson 1937, 4). To Jenkinson, records' history of legitimate custody alone, then, was a sufficient predictor and guarantor of the trustworthiness of the material. Michael Cook, writing 50 years later, dismissed Jenkinson's absolute faith in the documentary chain of custody (or perhaps the assumption that such chain of custody can be presumed or demonstrated) (Cook 1986, 129). Thus, archival institutions cannot trust the records they intend to acquire solely on the basis of their custodial history, but must test them for indications of their authenticity through studying their provenance and elements of their form (diplomatics) (Cook 1986, 7).

Today, digital technology has changed the way we communicate, conduct business, present our public face(s), and document our private lives. Digital culture is challenging the viability and legitimacy of many well-established social and cultural norms and their associated legal frameworks (Doueihi 2011, 12). One aspect of this evolution can be observed in our concepts of trust in digital information, our re-conception of what it means for a digital object to be authentic, and how we can assess its authenticity. "Virtual authenticity is not to be explained by a transfer of a well-known and ultimately problematic category from

one model to another; it is not to be restricted to a shift from the real to the virtual" (Doueihi 2011, 53).

Hypothesis and Research Questions

The hypothesis of this study is that, despite clear guidance offered by archival science on the means of ensuring, managing, and continuously assessing record authenticity, a guidance reflected in the products of several large-scale, significant and influential research projects on the topic of authenticity in the context of long-term preservation, the theoretical results of these projects are not being consistently applied in practice, and in fact records professionals are often unclear about how to define authenticity, how to protect it, and how to assess it (authenticate records). This research explores the practices, experience, and beliefs of records professionals. The primary purpose is to further our understanding of how these professionals, primarily but not exclusively archivists and records managers1, think about authenticity of the digital material for which they are responsible, and of what techniques or indicators of authenticity they use and rely on to ensure or continuously assess it. This study asks three broad research questions (the first two of which are discussed below) in order to investigate the convergence of theory and practice in the matter of establishing, maintaining, and assessing authenticity of digital records and data:

- Research Question 1: What elements of the context, content, and structure of digital records and data (indicators of authenticity) do records professionals use and rely on most to determine and manage authenticity?
- Research question 2: Do records professionals rely on the traditional archival model of record authenticity in the digital environment and if so, to what degree?
- Research question 3: Do records professionals consider the traditional archival model of authenticity sufficient to support a presumption of authenticity in the digital environment over time and across technological change?

Methodology

This research is grounded first and foremost in the theory and methodology of archival science and diplomatics, which provide a model of "record", and a means of understanding and defining record authenticity as well as the elements that comprise it. This model traces its origins to 17th century diplomatics, if not archival practice dating back as far as the Roman Empire and documented in the

[1] The distinction, perceived or real, between records managers and archivists, has been much discussed, and different national traditions view them as parts of one unified profession or two distinct ones. In formulating the research questions, I focused on the practice of records professional broadly – that is, any professional concerned with the nature of records and the establishment of their authenticity.

Justinian Code (Duranti 1998a, 36–40). It was further developed and tested for the digital environment through the research of the InterPARES Project. A record is defined as a document made or received in the course of practical activity and set aside for future action or reference. Authenticity is assessed by establishing its identity and demonstrating its integrity. In the digital environment a presumption of authenticity is an inference based on evidence about how the records have been created and maintained. Evidence may come from the records creator, or through further analysis to verify authenticity, such as comparison of the records with copies preserved elsewhere (redundancy), forensic analysis, testimony of a third party, or analysis of audit trails (MacNeil and Gilliland-Swetland 2005; Duranti 2005; Duranti and Thibodeau 2006; Duranti and Preston 2008).

This study used both quantitative and qualitative methods of data collection and analysis. Quantitative and qualitative data were collected through a questionnaire consisting of closed and open questions, circulated online through major English-speaking archival and records management listservs. Further qualitative data were collected through semi-structured interviews with selected respondents based on questions arising from the answers to the questionnaire.

The questionnaire was designed to measure practitioners' notions of authenticity of digital records and data and explore the relationship between practice and belief among records professionals – that is, what records professionals rely on in their work and whether their practice matches their belief or trust in specific authenticity indicators. Questions gathered basic information about the extent to which records professionals are concerned about the authenticity of digital records and data, how they ensure, assess, and/or protect it, and the level of importance that they place on specific indicators of authenticity.

Selected Results

The questionnaire was posted online from March 3-May 1, 2014 and reached professionals in 46 different countries or territories on six continents. Two hundred and ninety-three completed responses were received from archivists (n=134), records managers (n=97), and other (n=62), employed in government/industry (n=139) or cultural heritage institutions (n=158). Responses were received from civil law jurisdictions (n=64), common law jurisdictions (n=215), and pluralistic or religious law jurisdictions (n=14).

The main focus of the questionnaire was to explore the relationship among practice, experience, and belief with respect to the use of and value placed on a set of indicators of authenticity. The proposed indicators are based on the English-language archival literature, and categorized as social (S) or technological (T) (Table 1).

Table 1. Indicators of authenticity

Social indicators	
	Written policies and procedures governing the management of the records system
	Documentation about the record system (design, operation, management, etc.)
	Written policies and procedures governing digital records
	Classification scheme and/or file plan
	Retention and disposition schedules
	Archival description
Technical indicators	
	Information about the software used to create and manage the digital records
	Information about changes made to the digital records over time, (e.g. migration, normalization, etc.)
	Information about actions taken to preserve the digital records
	Access controls/security measures
	Audit logs
	Cryptographic validation techniques (e.g. digital signatures, hash digests, etc.)
	Standardized metadata

Social indicators are instruments developed by an organization to support the creation, management, or preservation of records (e.g. classification schemes, retention and disposition plans, policies and procedures documents). They are based on domain knowledge, and created and implemented by the intention of human actors (records professionals, management, legal counsel, etc.). They may or may not be present within a given organization; they may be mandatory or voluntary in their application or use, and even when mandatory, they may be circumvented or adapted, as Foscarini showed in her study of central banks (Foscarini 2009). They include the foundational instruments of archival and records management practice: policy instruments, classification schemes or file plans, retention and disposition schedules, and archival description or other descriptive measures (which may be captured in varieties of descriptive metadata).
Technical indicators are non-discretionary in their creation – that is, they are the result of a work process or state change in the records (e.g. system metadata capturing date created, and date modified), are algorithmically generated or implemented by the technological (e.g. computer, network) components of the overall record system (e.g. checksums, audit logs), are created to manage and control system access and security, or are created by a third party as specifications to a part of the technological system (e.g. documentation about software). Technological indicators may be used to control the records, but are more focused on controlling the system in which the records reside. They include audit logs, access controls and security measures, cryptographic validation techniques, and system metadata, as well as technical documentation.
Questions measured who used different indicators most or least in the course of their normal work to ensure authenticity, how these indicators were used in the event of a formal attestation of authenticity, what indicators practitioners be-

lieved to be most important in attesting to authenticity, and how frequently social versus technical indicators were invoked in use and in belief.
A key distinction between respondents is whether or not they have been required to make a formal attestation of authenticity in the course of their work. This allows a determination of whether work practice differs between those who have, and those who have not been required to authenticate a body of records. The differences between beliefs and practice is also explored by asking respondents who had not ever had to authenticate records what they think they would use if required to authenticate records.
Several themes emerged from the questionnaire that supported the research questions, and warranted further investigation through interviews. These were:

- the difference between practice and belief of records professionals regarding different indicators for ensuring or assessing authenticity,
- the relative weight of technical versus social factors in ensuring or assessing authenticity either in practice or belief, and
- the role of experience in making attestations of authenticity in practice and belief.

Furthermore, the interviews provided an opportunity to ask specific questions about the role, if any, of the system of law (civil or common law), and the extent to which traditional archival models of authenticity suffice in the digital environment. The results of the questionnaire were confirmed in 17 interviews (Table 2).

Table 2. Interviewees

	Government/Industry	**Cultural heritage**
Civil law (n=6)	Records managers – 4	Archivists – 2
Common law (n=11)	Records managers – 1	Records managers – 2
	Archivists – 2	Archivists – 3
	Other – 1	Other – 2

Role of legal system

It is interesting that in the questionnaire sample there was no statistical difference among professionals working in different legal systems based on the questions asked. In what way might a country's system of law influence the work of archivists and records managers in establishing and assessing record authenticity? The roots of archival theory and concepts of record authenticity are anchored in legal and administrative principles, first executed in centralized public repositories of written documents, then, with the spread of literacy, expanding into the regulated recordkeeping practices of public and private organizations, administrations, and homes (Eastwood 1994, 125; Duranti 1998b). Law and jurisprudence are the original pillars of influence that have guided the history and development of archival theory, reaching back through the centuries to Roman times. Principles from Roman law that have become part of the founda-

tion of archival knowledge include the idea that antiquity provides records with the highest legal authority, that deposit in a public place guarantees reliability of records as witnesses of actions, and that an unbroken chain of custody ensures records' continuing authenticity (Duranti 1996, 1).

The international records management standard, ISO 15489, is recognized in civil and common law systems. The idea of record authenticity is codified in ISO 15489 as follows: "An authentic record is one that can be proven to be what it purports to be, to have been created or sent by the person purported to have created or sent it, and to have been created or sent at the time purported" (International Standards Organization 2001 section 7.2.2). These concepts are explicated in domain distinctions among history, jurisprudence, and diplomatics – namely, the objects of authenticity inquiry (primary source material, works of art, legal instruments, records, etc.) that are important to historians, legal professionals, and archivists. For the purposes of understanding and analyzing documents and records, Duranti has differentiated among three types of authenticity: diplomatic, legal, and historical. Each is distinct and independent from the other, and reflects a specific purpose or focus of trust and the trust relationship in its discipline. Legally authentic documents bear witness on their own because of the intervention, during or after their creation, of a representative of a public authority guaranteeing their genuineness. Diplomatically authentic documents are those that were written according to the practice of the time and place indicated in the text, and signed with the name(s) of the person(s) competent to create them. Historically authentic documents attest to events that actually took place or to information that is true (Duranti 1998a, 45–46). An authentic document is conditioned by the discipline in which it is considered – and therefore the purpose the document serves. This holds regardless of legal system, although it may be expressed in different ways.

The system of law of many countries has developed out of one of two traditions – the English common law and the French civil law. These systems have spread through conquest and colonization throughout the world, common law transplanted to the United States, Canada, Australia, New Zealand, and parts of Asia and Africa and civil law to much of Europe, Latin America, and parts of Asia and Africa. Although there are differences in each tradition among different countries, in general the distinction can be summarized as follows: civil law systems utilize professional judges, legal codes and written records, while common law systems rely on lay judges, broader legal principles, and oral arguments (Glaeser and Shleifer 2002, 1196–1197).

In common law systems, authenticity is part of the foundation upon which the admissibility of documentary and real evidence is based, proven through the authentication of documents at the moment they are introduced as evidence. Authentication of documentary evidence can be proven through testimony, expert analysis, non-expert opinion, or, in the case of public documents or other special types, circumstances of record creation and preservation.

Traditionally, documentary evidence should be "best evidence", satisfied by production of an original. Authentication serves to establish the identity of the record and its relevance to the issues in the proceeding, while the best evidence rule demands proof of the integrity of the contents (MacNeil 2000, 46–48). However, the historical notion of authenticity, whereby an authentic record is one which attests "to events that actually took place or to information that is true" (Duranti 1998a, 46), focuses attention on the contents of a document and their truthfulness. Mapped to legal practice, historical authenticity is seen in moments when a document is considered "for the truth of its contents" (Sheppard and Duranti 2010, 27). This is tested not through authentication or establishment of best evidence, but in the consideration of the hearsay nature of the document, and is linked to reliability.

Some interviewees found solid legislative support for recordkeeping, while others noted a failure to keep up with technological advances. One interviewee in New Zealand noted that there is a high-level legislative regulatory apparatus for designing systems in the public sector, and guidance at a more granular level for recordkeeping systems in ISO 1617 Part 2. A template for writing records policy, and a metadata standard with technical requirements and a schema provide additional support. This provides a good basic framework to work from and test the systems against minimum requirements. In Australia, the UK, the US, and New Zealand standards assess and accredit recordkeeping systems from vendors, but throughout the 2000s this became increasingly untenable. The situation is more complicated in the UK and North America, where legislation regulating records generally does not address record authenticity specifically.

Civil law systems require a greater measure of procedural formalism than common law systems in regulating dispute resolution through bright line rules for collection and presentation of evidence (Djankov et al. 2002, 37). This can result in extremely detailed and prescriptive rules governing the creation of records of public administrations. For example, in Italy the attention of records professionals when designing systems for management and storage of records is focused on the implementation of national legislation which dictates a rich set of metadata establishing identity and integrity, supported by documentation about policies and procedures outlining clear responsibilities. Italian legislators have recently provided detailed guidelines to verify the quality of metadata and applications for records creation and (as a consequence) for their transfer to any kind of custodial environment (including cloud storage). Specific attention is paid to the documentation and metadata provided for transferring records from the creator environment to any other system (Interviewee D037).

Interviewees were asked if specific laws governed their work, based on the national system of law in which they worked. I wanted to understand from them whether there was a difference in the framework supporting authenticity of records between common law and civil law countries. I hypothesized that there would be a clear difference, but this was not the case. The legislative framework

in Italy has already been introduced. General legislation clarifies the minimum amount of metadata required for identity and integrity, including classification, reference code obligations, and specific rules for records management systems. Information related to provenance, identity, and context must be well defined and must be written and exchanged according to specific rules. The interviewee noted strong cooperation in Italy between archivists, records managers, and legislators. However, Italy appears to be the exception rather than the rule in civil law countries among the interviewees questioned. In another European civil law jurisdiction interviewees reported that there has been a change over the last six or so years, and today a common trend is to ignore the law and take the path of least resistance. An interviewee from a third civil jurisdiction noted that the regulations that do exist are at a very high level and do not provide any specific guidance.

Limitations and Conclusion

Web-based surveys conducted through professional listservs are convenient, but they do have limitations, primarily the inability of guaranteeing a representative sample. Individuals self-select to join a listserv and to participate in questionnaires. Results therefore may be practically significant, but as there can be no truly random sample, results may be difficult to generalize, nor can validity be objectively measured. However, as an indicator of general practice, such surveys provide useful information.

This research explored the relationship between practice and belief in establishing and assessing authenticity of digital records among a sample of records professionals. Overall, respondents adopt a pragmatic approach to authenticity based on resources, sensitivity of the records, and organizational or legislative framework. In response to the first two research questions, across variables of profession, sector, and legal system these professionals rely most heavily on social indicators of authenticity in daily work, but exercise greater trust and reliance on technical indicators in the process or expectation of authenticating records. Experience of authenticating records affected belief in the value of indicators, with those who had never had to authenticate records placing higher faith in technical indicators. Interviewees generally agreed that the traditional archival model of record authenticity still held in the digital environment (RQ3), but required adaptation. Although the research did not originally intend to explore in depth the function of legal system as an independent variable, the findings that did emerge open up an interesting line of inquiry for further study.

References

ACA. 1999. “Code of Ethics | The Association of Canadian Archivists.” http://archivists.ca/content/code-ethics.

CCSDS. 2012. “Reference Model for an Open Archival Information System (OAIS): Recommended Practice Issue 2.” Consultative Committee for Space Data Systems. http://public.ccsds.org/publications/archive/650x0m2.pdf.

Cook, Michael. 1986. *The Management of Information from Archives*. Aldershot, Hants, England ; Brookfield, Vt., U.S.A: Gower.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silane, and Andrei Shleifer. 2002. “Courts: The Lex Mundi Project.” Working Paper 8890. National Bureau of Economic Research. http://www.nber.org/papers/w8890.

Doueihi, Milad. 2011. *Digital Cultures*. Cambridge, Massachusetts: Harvard University Press.

Duranti, Luciana. 1996. “Archival Science.” In *Encyclopedia of Library and Information Science*, 59:1–19. New York, Basel, Hong Kong: Marcel Dekker.

———. 1998a. *Diplomatics: New Uses for an Old Science*. Lanham: Scarecrow Press.

———. 1998b. “The Odyssey of Records Management, Part 2.” *Records Management Quarterly* 23 (3).

———. 2005. *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. San Miniato: Archilab.

Duranti, Luciana, and Heather MacNeil. 1997. “The Preservation of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project.” *Archivaria* 42 (Spring): 46–67.

Duranti, Luciana, and Randy Preston. 2008. *Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential. Interactive and Dynamic Records*. Padova: Associazione Nazionale Archivistica Italiana.

Duranti, Luciana, and Kenneth Thibodeau. 2006. “The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES.” *Archival Science* 6 (1): 13–68.

Eastwood, Terry. 1994. “What Is Archival Theory and Why Is It Important?” *Archivaria* 37 (Spring): 122–30.

Glaeser, Edward L., and Andrei Shleifer. 2002. “Legal Origins.” *The Quarterly Journal of Economics* 117 (4): 1193–1229. doi:10.1162/003355302320935016.

International Standards Organization. 2001. “ISO-15489 Information and Documentation-Records Management, Part 1, 2.” Text. Geneva: ISO. http://www.iso.org/iso/catalogue_detail?csnumber=31908.

Jenkinson, Hilary. 1937. *A Manual of Archive Administration*. New and Revised. London: Percy Lund, Humphries & Co. http://www.archive.org/details/manualofarchivea00iljenk.

MacNeil, Heather. 2000. *Trusting Records: Legal, Historical, and Diplomatic Perspectives*. Dordrecht: Kluwer Academic.

MacNeil, Heather, and Anne Gilliland-Swetland. 2005. “Authenticity Task Force Report.” In *The Long-Term Preservation of Authentic Electronic Records: Finding of the InterPARES Project*, edited by Luciana Duranti. San Miniato, Italy: Archilab.

Oxford English Dictionary. 2014. “Authentic, Adj. and N.” *OED Online*. Oxford University Press. http://www.oed.com.ezproxy.library.ubc.ca/view/Entry/13314.

SAA. 2011. “SAA Core Values Statement and Code of Ethics.” *Society of American Archivists*. http://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.

Scanlan, Daniel M. 2011. *Digital Evidence in Criminal Law*. Aurora, Ont.: Canada Law Book.

Sheppard, A. F., and Luciana Duranti. 2010. “The Canadian Legal Framework for Evidence and the Digital Economy: A Disjunction?” SSHRC Knowledge Synthesis and the Digital Economy. University of British Columbia.

E-ARK Legal Issues Report: European Cultural Preservation in a Changing Legislative Landscape

David Anderson
University of Portsmouth
2 Winston Churchill Avenue, Portsmouth, Hampshire, UK
cdpa@btinternet.com

Summary

This paper summarizes the main findings of a report on the legal issues relating to digital preservation within the European Union that arise as a result of the many items of national and EC legislation, but principally from the proposed changes in data protection currently being discussed by the European Parliament, the Commission, and the Council. This report has been prepared as a Public Deliverable (D2.2) by the EC Policy Support Programme Project "E-ARK (European Archival Records and Knowledge Preservation) number 620998 (1 February 2014 – 31 January 2017). The full report is available for download free of charge from the E-ARK Project website – http://www.eark-project.com/resources/project-deliverables.

Keywords: legislation, data protection, data governance, European Community, digital preservation, E-ARK

Introduction

Archives provide an indispensable component of the digital ecosystem by safeguarding information and enabling access to it. Harmonisation of currently fragmented archival approaches is required to provide the economies of scale necessary for general adoption of end-to-end solutions. There is a critical need for an overarching methodology addressing business and operational issues, and technical solutions for ingest, preservation and re-use.

In co-operation with commercial systems providers, the E-ARK consortium aims to create and pilot a pan-European methodology for electronic document archiving, synthesising existing national and international best practices, that will keep records and databases authentic and usable over time. Our objective is to provide a single, scalable, robust approach capable of meeting the needs of diverse organisations, public and private, large and small, and able to support complex data types.

The practices developed within the project will reduce the risk of information loss due to unsuitable approaches to keeping and archiving of records. The project will be public facing, providing a fully operational archival service, and access to information for its users. The project results will be generic and scalable

in order to build an archival infrastructure across the EU and in environments where different legal systems and records management traditions apply. E-ARK will provide new types of access for business users.

At present, no comprehensive survey of the legal and organisational framework under which European recordkeeping, preservation and access take place is available to practitioners in the field.

Facilitated by the DLM Forum with its broad EC-wide membership comprising public bodies, service providers, technology providers and national archives, we aim to provide an overview report in relatively plain language dealing with the legal and regulatory requirements for data protection, the reuse of public sector information, and copyright legislation. In particular, this report provides coverage and an analysis of the following EC Directives and Regulatory Instruments:

- Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.
- Directive 92/100/EEC of 19 November 1992 on rental right and lending right and on certain rights related to copyright in the field of intellectual property
- Directive 95/46/EC of the European parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
- Directive 96/9/EC of 11 March 1996 on the legal protection of databases (the "Database Directive")
- Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001, on the harmonisation of certain aspects of copyright and related rights in the information society
- Directive 2001/84/EC of the European Parliament and of the Council of 27 September 2001 on the resale right for the benefit of the author of an original work of art
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)
- Directive 2003/98/EC of the European parliament and of the Council of 17 November 2003 on the re-use of public sector information
- Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
- Directive 2009/24/EC of 23 April 2009 on the legal protection of computer programs (Codified version replacing the abrogated Directive 91/250/ EEC of 14 May 1991, known as the "Computer Programs Directive")

- "Proposal for a Regulation of the European parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)", Brussels 25th Jan 2012
- Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information

The findings presented in this report are intended to provide a greater understanding of the legal framework as it impacts on cross-border co-operation. This report will be used to inform the other Work Packages within E-ARK as it is essential to ensure the project aligns with EU Directives as implemented by Member States.

Three broad areas are examined:

Data Protection
At the time of writing, it is not possible to say exactly what regulatory provisions for Data Protection will be put in place by the EC, as discussions are still taking place within (and between) the European Parliament, the Council of Ministers, and the Commission about exactly what changes should be made to the provisions of Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. However, the broad brush strokes of the new regulations are reasonably clear, and some 17 key areas, where change seems more or less certain, are examined reasonably closely.
The approach has been to present and analyse the current requirements as set out in Directive 95/46/EC, followed by presenting and examining the regulations which are expected to replace them, finally, some concluding remarks are offered.

Re-use of Public Sector Information.
The general approach here is broadly similar to that taken with Data Protection. The background to the regulatory framework is discussed, and placed in context. The obligations placed on Member States by Directive 2003/98/EC on the re-use of public sector information, are explained, and then compared against Directive 2013/37/EU, which was introduced to amend it.

Copyright Legislation
Copyright protection is an area of European regulation that is both more diffuse than the other areas considered, in that there is not a single over-arching Directive to consider, and the Directives are more stable in the sense that they have not been subject to major revision over recent years.

In addition to providing analysis and commentary on matters of law, this report also provides some introductory material, which examines the broad legal context within which modern legislators are operating. To this end, there is discussion of a number of conventions such as:

- The Paris Convention for the Protection of Industrial Property (1883)
- Berne Convention for the Protection of Literary and Artistic Works (1886)
- Universal Declaration of Human Rights (1948)
- European Convention on Human Rights (1950)
- Council of Europe Convention 108 (1981)

as well as influential state and national legislation such as:

- Hessisches Datenschutzgesetz (1970)
- Datalag (1973)

Extensive free-standing appendices will be produced to accompany the full report, and will include full copies of the principal legislation under discussion, together with related material such as the Malmö Ministerial Declaration on eGovernment that sets out eGovernment practices up to 2015.

The intention is to provide in a single location many of the resources which practitioners may need to have available to navigate these three key areas. Somewhat against normal academic practice, extensive use is made of in-line quotation of the text of Directives and other regulatory instruments. These are generally placed directly alongside explanation and analysis. The purpose behind this approach is to simplify the process of using this report in practice, and to avoid the need to engage in "footnote hunting", a task often made particularly difficult for readers for whom English is not their first language.

The adoption by the EU of the Data Protection Directive (95/46/EC) marked a pivotal moment in the history of European personal data protection. Two decades later, the fundamental principles around which the Directive was structured continue to be relevant, but the ever-increasing pace of technological change, and globalisation have undoubtedly presented challenges for data protection that the original Directive is ill-equipped to address. The world of the early 21st Century is the world of social networking, apps, cloud computing, location-based services and smart cards. It is almost impossible for individual citizens to go about their daily business, or to buy goods and services without leaving digital footprints. Without effective control over how this information is stored and used, the potential for adverse consequences is obvious.

So it is that the European Commission is currently engaged in a process of modernising the EU legal system for the protection of personal data. One of the key policy objectives behind the revisions is to make more consistent the implementation and application of the protection of personal data in all areas of the Union's activities. Anticipated benefits include the strengthening of the

rights of individuals, reduced administrative overhead, and an improved flow of personal data within the EU and beyond.

The main part of this report covers the requirements of Directive 95/46/EC, which have been implemented by Member States in a variety of legislative instruments since the adoption of the Directive in 1995. These are set alongside the General Data Protection Regulation (GDPR) proposals currently under discussion between the Commission, the Council of Ministers and the European Parliament. As a final form of the text has not been agreed at the time of writing some of the conclusions reached in this report are necessarily tentative in nature.

Individual citizens (or data subjects) are not the only stakeholders on the digital playing field. Within the context of this report, we will also pay attention to institutional stakeholders, particularly in the cultural heritage sphere. Memory institutions such as galleries, libraries, archives and museums are both custodians of our common digital heritage, and aggregators and generators of large quantities of born digital and newly digital information. Many of the leading organisations such as national archives and libraries have a legal deposit responsibility which obliges them to collect and retain vast quantities of digital information, and to make this, as far as possible, available to the public today and in the future.

The law, even within a single national jurisdiction, is often complex in character, and legislation is generally drafted in a form that lay readers struggle to comprehend. The situation is made even more difficult when many pieces of legislation may potentially apply to an activity, and where the law makes competing demands. Thus, a national archive may have a general obligation under the Directive(s) on the Re-use of Public Sector Information to ensure that information held by them is made available to the public, while the Data Protection Directive, may oblige them to protect the privacy of individual data subjects by keeping some information undisclosed.

Preserving files intact is a natural activity for memory organisations, yet there is increasing pressure for data subjects to be given the right to have data concerning them purged altogether. In some cases this may not even be technically feasible. Even the act of preservation, which constitutes much of the raison d'etre of galleries, libraries, archives and museums, may in the digital context, involve techniques and processes which conflict with EU Directives, while simultaneously being required under national legislation.

The legal landscape is thus far from clear, even to experts in the field, and while discerning the overall legal requirements in every case may not be an intractable problem, it does provide an on-going, and ever more complex challenge to those charged with preserving our digital records.

The Commission's proposals amount to a fundamental modernisation of Europe's data protection rules, establishing a number of new rights for citizens of which the right to be forgotten is only one.

General jurisdictional scope

The new regulatory arrangements both simplify the existing arrangements, and extend significantly the reach of EU legislation, taking it beyond Europe's borders. Under the new regime, processors of personal data will fall under the regulations. The existing old "means" and "equipment" tests are abandoned in favour of concentrating on whether non-EU controllers are providing goods/services to data subjects in the EU, or are monitoring their behaviour. However, some potential remains for legal uncertainty arising from a lack of clarity about the meaning and scope of key terms in the new proposals.

Scope of personal data

Under Directive 95/46/EC there is some divergence of opinion between Member States as to what constitutes 'personal data' The new proposals are expected to establish a single broad definition of personal data for the whole of the EU. Henceforward, 'identification' will depend on the likelihood of 'singling out' an individual directly or indirectly, rather than being limited to the possibility of knowing details such as their name and address.

It will be prudent to take a very conservative approach to the collection, processing, and retention of personal data. Only the minimum data should be handled; data should be assumed to be personal unless there are clear grounds for believing otherwise; personal data should be held only for the minimum time required mindful of the purpose for which it is being held and processed; organisations should be able to demonstrate an audit trail showing that data no longer held has been securely deleted; where possible data should be anonymised.

The Obligations and Liabilities of Data Controllers

It is something of a truism to assert that the notion of 'data controller' is key in data protection regulation. The new proposals introduce a modify somewhat the definition of 'controller' used in Directive 95/46/EC, and having done so, then pay considerable attention to delineating obligations and liabilities which controllers must respect.

Echoing the provisions of Directive 95/46/EC, the general principles which govern personal data processing are stated and may be understood as stipulating "the less the better". Thus, data should be retained no longer than absolutely necessary, and processing should be kept to a minimum.

Controllers will be held responsible for ensuring the existence of transparent and easily accessible policies with regard to the processing of personal data, and for the exercise of data subjects' rights, as well as ensuring that any information or communication concerning the processing of personal data uses clear and plain language. They will also be required to provide the means for data subjects to exercise their rights.

The new regulations assert the right of data subject's right to data 'portability', that is to say, they will have the right to both obtain those data from the controller, and to have them provided in a structured and commonly used electronic format.
Controllers will have to respect the 'principle of accountability' and be able to demonstrate their compliance. Typically this would mean being able to show internal policies and mechanisms for ensuring such compliance. There is also a requirement for controllers (and processors) to carry out a data protection impact assessment prior to risky processing operations.
The new proposals introduce 'joint controllers', who are understood to be processors working beyond the controller's instructions, and clarify the obligations of the controller and the processor for co-operation with the supervisory authority.
Building on the personal data breach notification in Article 4(3) of the e-privacy Directive 2002/58/EC, the new proposals place an obligation on controllers to notify supervisory authorities of personal data breaches, and to notify personal data breaches to data subjects.
Finally, the new rules build on Article 23 of Directive 95/46/EC to extend the rights of data subjects to damages resulting from the action of processors and clarify the liability of joint controllers and joint processors.

Lawfulness of processing

The new regulations follow closely the existing requirements under Directive 95/46/EC. However, one area where a significant tightening of the rules will take place is the regime for obtaining valid consent.
Controllers will be required to bear the burden of proof for the data subject's unambiguous consent to the processing of their personal data for specified purposes. Data subjects will have the right to withdraw their consent at any time.
In cases where there is a significant imbalance between the position of the data subject and the controller, consent will not be regarded as providing a legal basis for processing.
The impact these amendments will have in individual Member States, will naturally depend on the extent to which their current national legislation takes a stricter or more lenient position on consent.

The Right to be Forgotten

It is clear that while, under the new regulations, data subjects are set to enjoy the right to be forgotten, this right will be by no means unrestrained. Data controllers will be required to attenuate the right to be forgotten against, particularly, the right to freedom of expression when determining whether to accede to removal requests. Controllers will also have the option to 'restrict processing' of contested data rather than to remove it completely, but, in practice, the burden imposed on data controllers by expecting them to balance the right to be forgotten against the right to freedom of expression, and deciding whether it is

more appropriate to restrict processing or to complete erase data, is likely to be severe. This is, if anything, exacerbated by cascading this responsibility down to secondary controllers.
Either way, many data controllers are likely to find themselves acting as both judge and jury when considering requests. The right to be forgotten has been the subject of much discussion at Council level, particularly in the light of the decision of the Court of Justice of the European Union in Google Spain.

Data Portability

There remains considerable debate over the provisions for data portability, whether they would not sit more appropriately under competition law, and what limitations may apply. Undoubtedly, compliance with the regulations in their current form would impose on businesses a significance cost burden. The extent to which this is justifiable, particularly in the absence of any real evidence of 'customer lock-in', is questionable. While we may be reasonably confident that data portability, in some form, will feature in the final version of the new regulation, it is far from clear what that form will be.

Automated Individual Decisions / Profiling

It is not yet possible to have any clear idea what the final shape of the new regulations will be with respect to profiling. However, a balance needs to be struck between providing, on the one hand, rights for data subjects to object to automated profiling, and on the other the interests of businesses who depend for the viability on being able to 'target' audiences, or discriminate between potential customers. What that balance will look like is by no means clear.
Data protection officials/officers
The appointment of a Data Protection Officer represents a significant administrative and cost overhead on businesses, in consequence of which there has been a robust debate as to whether the new regulations should require them to be employed, or to permit organisations to continue with the current voluntary arrangements. Counter-proposals include limiting the mandatory appointment of a Data Protection Officer to cases where a certain threshold of data processing activity has been crossed in addition to limiting the requirement to public bodies and larger enterprises. It is simply not clear at this point how this particular aspect of the proposed new regulations will be resolved in the final text.

Data protection by design and by default

Privacy by Design (PbD) is an approach to systems engineering which promotes privacy and data protection compliance from the outset and involves the whole engineering process. The gold standard for PbD is encapsulated in the seven 'foundational principles of privacy by design' produced by The Canadian Privacy by Design Centre of Excellence. The proposals put forward by the Commission fall some way short of incorporating all seven of foundational princi-

ples, and reflect to some extent the debate which has been going on between the European Commission, Parliament and Council as to the scope and detail of the PbD requirements.
Nevertheless, it is clear that the new regulatory framework will require organisations to take full account of developments in technology and solutions for privacy by design and data protection by default and will no longer be satisfied to see privacy and security as something of a post hoc addition to products and processes.

Jurisdictional scope: Controllers not established in the Union

Proposals are still under discussion about bringing non-EU processors conducting business within the EU, and processing EU data subjects' personal data under the scope of the new law.
However desirable this may be, it will not be clear for some time after the introduction of the new rules whether it is possible, in practice, to enforce the rules. Some commentators have questioned whether sufficient resources will be available to enforcement agencies to bring to a successful conclusion prosecutions outside the geographical boundaries of the European Union.

Security of Processing

Measures to ensure the security of data processing are implemented differently in the various Member States. Directive 95/46/EC gives relatively little guidance on how to handle security. The new proposals while broadly repeating the approach of Directive 95/46/EC do make some movement in the direction of providing indicative compliance benchmarks

Personal Data Breach Notification

While there may be some amendment of the precise time periods within which notification to the competent authority, and the data subject must take place, there is little doubt that the new regulations will require controllers and processors to make notification of breaches within a relatively short time. Mindful of the sanctions proposed for non-compliance these deadlines will need to be respected.
It will take some time after the new regulation comes into effect before it is clear whether this aspect of the new rules will be workable in practice. On the one hand, notification within 24-72 hours may prove to be too challenging, while on the other, concern over the possible consequences of being found in breach of an obligation to notify may lead controllers/processors to err on the side of caution and notify so frequently that the system fails in practice.

Transfer of personal data to a third country

At present there are marked differences in how Member States treat the transfers of personal data to third countries in those cases where neither the Commission nor their national authorities have determined the adequacy of the arrangements in place.

Overall, the intention under the new proposals appears to be to build on the current framework. Organisations who are acting solely in the capacity of data processors will need to be mindful of the rules which govern international data transfers, as significant penalties may be incurred for breaches of the regulations.

It should be noted that under the new proposals the Commission will have sole authority to determine which countries are deemed to provide adequate safeguards for personal data, and that decisions once taken will continue to be subject to being overturned or revised. There is general approval for the idea of a European Data Protection Seal, and this will be only one of a number of new mechanisms for certifying data processing as adequately safeguarded. An important distinction has been drawn in the new proposals between safeguards (such as one-off contractual clauses) which will continue to require authorization from a data protection authority, and those (such as legally binding and enforceable instruments between public authorities) which will not. It is also worth highlighting that data transfer may, if the Council has its way, henceforward require explicit consent to count as valid.

Legal enforcement & Penalties

Final decisions have not yet been reached about the sanctions and penalties that will be available under the new regulatory scheme. However, it is already clear that sanctions will in the future be much onerous than those in place today. Originally, the commission proposed fines amounting to 2% of annual global turnover be imposed in the most serious cases, but that figure seems to have been abandoned in favour of even more severe penalties. We can expect that sanctions will be set at a level that compels data holders to take very seriously the potential legal consequences of paying insufficient attention to (particularly) their corporate data protection responsibilities. Whereas the relatively modest sanctions scheme provided under Directive 95/46/EC meant that organisations could, if they chose, afford to risk infringing data protection requirements, this course will no longer be open under the new scheme.

Reproduction Rights

With respect to reproduction rights, Community law does not provide an appropriately accommodating legal framework. Articles 5.2 (c) and 5.3 (n) of the Information Society Directive of 22 May 2001, appear to provide libraries with public access, educational establishments, museums and archival services, lim-

ited exceptions to the general restrictions placed on unauthorised reproduction and communication. However these do not cover computer programs or databases and therefore transfers of this kind of material remain problematic.
More and more, digital objects are multimedia in nature. Problematically, no definition of multimedia exists in Community law. Therefore it is necessary to look to national interpretations to determine their legal nature. The legislative frameworks examined for this report (France, Germany, The Netherlands) regard multimedia works as 'complex works' and take a distributive, fragmented approach in which each component part of a multimedia work: audio, graphics, software, database, etc., is considered separately. Since multimedia works are not, in general, made available on computer platforms in such a way that individual elements can be removed from the whole, this means that, in practice, a multimedia work will enjoy, as a whole, the strongest protection under law that is available for any of its constituent parts.

Technological Measures of Protection (TMP)

Many works are made available in a form to which technical measures have been applied to prevent or restrict the use that may be made of them. This might take the form of a simple password protection scheme or may involve considerable technical sophistication.
The Information Society Directive (2001/29/EC) recognises the "need to provide for harmonised legal protection against circumvention of effective technological measures and against provision of devices and products or services to this effect". Article 6 [2], stipulates that "Member States shall provide adequate legal protection against the circumvention of any effective technological measures, which the person concerned carries out in the knowledge, or with reasonable grounds to know, that he or she is pursuing that objective." However it also permits Member States to be given the option of "providing for certain exceptions or limitations for cases such as educational and scientific purposes, for the benefit of public institutions such as libraries and archives".
The potential for exemptions is quite limited, and does not extend to permitting the creation or use of tools by individuals to bypass TMP generally.

Legal Compliance and Technical Capability for Privacy-sensitive Data Protection in the Cloud

Eun G. Park
McGill University, Canada
eun.park@mcgill.ca

Summary

This study aims to investigate if both the legal compliance and technical capability are sufficient to protect privacy-sensitive data and records at privacy-sensitive institutions in the cloud. The study examines the applicability of current legal requirements in Canada, the United States and Europe to the cloud environments. We also conduct a case study to test the implementation of current privacy-preserving techniques to protect privacy information at selected government agencies or health institutions.

Keywords: privacy, security, cloud computing, security assessment

Introduction

Many privacy-preserving techniques have been developed to address the privacy issues in different data sharing scenarios that are used in the field related to records and data management (Fung et al., 2010; Mohammed et al., 2013; Park, 2014). The emergence of cloud computing has significantly improved the potential of sharing records and data. However, the major obstacle to adopting this technology in the public sector is a lack of trust in sufficient security and privacy protection. Research has shown that simply removing explicit identification information of patients, participants or citizens, including the person's name, social insurance number, telephone number, and address, is insufficient for privacy protection. Therefore, there is a need to assess the privacy-preserving techniques and tools that are available and popular in the field and test how well these techniques can actually help protect privacy and security to records managers at the real settings (e.g. privacy and security sensitive institutions, such as government agencies or health institutions). Government agencies hold a large amount of citizens' data with their privacy information on the data (e.g. social insurance number, etc.). Health agencies also create, use and house a large amount of patients' data with their health information, such as blood type, disease, drug, etc. These institutions tend to be more sensitive with privacy information in their data sets. It is important to examine whether legal requirements as minimum requirements that are based on government privacy guidelines are satisfactory from a viewpoint of records management in implementing privacy-preserving techniques and tools at different institutional settings in re-

ality. In addition, there is a pressing need to examine at what extent these techniques and tools can actually help reduce security and privacy risks in records and data management practices in the cloud at the institutional contexts.

Objectives of the Study

The objectives of this study are: (1) to identify the security and privacy challenges of hosting person-specific information on cloud; (2) to evaluate the state-of-the-arts privacy-preserving techniques and their applicability to the cloud environment; and (3) to study the readiness of the health and government agencies of the technological shift to the cloud.

Method

To answer the first objective, this study examines the current legal guidelines of privacy management in Canada, the United States and Europe, especially, new laws and guidelines on the cloud and draw the privacy preserving and security criteria that are driven by literature review. This study takes a close look at the (e.g. Personal Information Protection and Electronic Document Act (PIPEDA), Privacy Act of 1974, Health Insurance Portability and Accountability Act (HIPAA) and Privacy Act of 1974 and does a comparative analysis of the three acts' requirements on privacy protection. In addition, new acts, notices or polices related to the cloud are examined to identify how to apply to the cloud environments in Canada, the United States, and Europe (e.g. Privacy Impact Assessments for Personal Information Banks, Information Sharing Agreements for personal information sharing, Office of the Privacy Commissioner of Canada, Information Breach Protocol, etc.).

To answer the second objective, this study examines whether the available security and privacy-preserving techniques and tools can meet the criteria at the practical sites, especially government and health agencies in Canada or the United States. The security and privacy-preserving techniques and tools to review include the following four approaches: single provider and single release, sequential release, collaborative data integration, RFID trajectory data release, etc. In addition, we developed one security technique to be applicable to the cloud environment.

Further, we plan to conduct a case study at government agencies or health agencies or government to examine whether the legal requirements and technical capability are met at the chosen sites. Based on the findings at the chosen agencies, this study plans to make suggestions on how to manage security and privacy risks in records and data management at government agencies or health institutions in Canada.

Progress to Date and Expectations

This study is in progress. To date, laws and cases on the applicability to the cloud have been reviewed. Literature review on the applicability of laws and regulations to the cloud are in progress. Further, based on the literature review, the privacy preserving and security criteria will be developed and the popular privacy-preserving and security techniques have being in progress of review. To test the privacy preserving and security criteria, we are contacting a couple of sites for having access to their data and plan to conduct a case study at the chosen sites. We plan to complete this study by the end of 2015.

Acknowledgements

This study is part of InterPARES Trust project.

References

Fung, B. C. M., Wang, K., Fu, A. W.-C., and Yu, P. S.. Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques, ser. Data Mining and Knowledge Discovery. 376 pages, Chapman & Hall/CRC, August 2010.

Mohammed, N., Jiang, X., Chen, R., Fung, B. C. M., and Ohno-Machado. L. Privacy-preserving heterogeneous health data sharing. Journal of the American Medical Informatics Association (JAMIA), 20(3):462-469, May 2013. BMJ.

Park, E. G. 2014. InterPARES Trust 2nd International Symposium, October 17, 2014: A Privacy-preserving Approach for Records Management in Cloud Computing. Victoria, BC: University of Victoria Library.

Managing Digital Records in a South African Public Sector Institution

Shadrack Katuu[1]
International Atomic Energy Agency
Vienna, Austria
skatuu@gmail.com

Mpho Ngoepe
Department of Information Science, University of South Africa
Pretoria, South Africa
ngoepms@unisa.ac.za

Summary

South Africa's public institutions generate public records that need to be managed in compliance with the legislative framework. However, there are challenges in managing digital records brought about by the complexity within the legal terrain as well as the disparate nature of information systems in institutions. This article provides a brief review of literature that informs the background to a South African research study conducted by the InterPARES Trust Project's Africa Team.

Keywords: Electronic Document and Records Management Systems, Enterprise Architecture, InterPARES Trust Project, Legal framework, National Archives, South Africa

Introduction

South Africa's public sector institutions generate public records that need to be managed in compliance with the country's legislative and regulatory instruments. For instance, national archival legislation dictates that public records that are older than 20 years and that have enduring value should be physically transferred into the custody of the National Archives. Recent technological developments that have impacted how public records are created and managed such as mobile computing and cloud computing have raised a number of challenges to the traditional means of recordkeeping. The identification and transfer of digital records from public institutions into archival custody has not happened in any systematic manner because the national archival system has struggled to effectively manage such records and facilitate their long-term preservation (Ngoepe and Keakopa, 2011). This is because the national archival system is

[1] The views expressed herein are those of the author and should not be attributed to the International Atomic Energy Agency.

plagued by poor infrastructure and the lack of skilled capacity to ingest any digital records for permanent preservation (The Archival Platform, 2015, p. 102). As a result, these records are left to the creating agencies to manage and preserve them even though they lack the infrastructure as well as knowledge and skills to preserve digital records in the long-term. It is against this background that a research study exploring the management of digital records is being coordinated by Team Africa that is part of the InterPARES Trust research project (InterPARES Trust, 2015). The particular research study is entitled *Managing records in networked environments using EDRMS applications*. The aim of the research study is to assess the recordkeeping environment in a public enterprise known as Rand Water which is the largest bulk water utility institution in South Africa. Rand Water supplies water to 12 million people in several metropolitan and local municipalities as well as large industries. The research study is assessing both the legal terrain as well as the technological requirements for the management of digital records in the custody of the institution. The research study began in 2015 and will be conducted over a period of 18-24 months.

An initial part of the research study is a review of literature that provides contextual background on the current situation within the public sector in South Africa. This article outlines two aspects related to the literature reviewed as part of the research study: an exploration of the legal and regulatory framework as well as the technological terrain in South Africa. The article provides an outline of the emerging issues that inform the ongoing research study.

Literature review on the legal terrain

South Africa has a 'hybrid' or 'mixed' legal system consisting of three distinct legal traditions (Du Bois, 2004, p. 9-16). The first legal tradition is a civil law system inherited from the Dutch and commonly referred to as Roman Dutch law that draws from two sources: “judicial decisions and the writing of the old Dutch jurists” (Madhuku, 2010, p. 50). The second legal tradition is a common law system inherited from the British, and the third a customary law system inherited from indigenous cultures and is often termed as African Customary Law (Alberts and Mollema, 2014). These traditions have had complex interrelationships with each other sometimes causing areas of strain (Toufayan, 2014).

South Africa’s legal and regulatory system has a tremendous impact on how records are managed in the country’s public sector. Allan (2009, p. 174) argues that there are two major categories of legislative instruments that relate to information management, those that control information across all public institutions and those that “relate to specific information held in specific sectors or structures”. Three of legislative instruments that “control information” across all public institutions are:

- The National Archives Act (NAA) that was promulgated in 1996 and whose key mandate is the proper management and care of records of governmental bodies (South Africa, 1996).
- The Promotion of Access to Information Act (PAIA) that was promulgated in 2000 and facilitates public access to records from both public and private sector institutions within a stipulated time (South Africa, 2000).
- The Protection of Personal Information (POPI) Act that was assented in 2013 and whose key objective is the protection of personal information processed by public and private bodies (South Africa, 2013).

In order to facilitate the management of digital information there several additional legislatives instruments, two of which are:

- The Electronic Communications and Transactions (ECT) Act that was promulgated in 2002 and that facilitates electronic communications and transactions by promoting legal certainty whenever public administration and private business activities need to be conducted in digital form (South Africa, 2002a, p. 16-18).
- The Regulation of Interception of Communications Act (RICA) that was promulgated in 2002 and that regulates the interception of certain telephonic as well as internet communication (South Africa, 2002b).

While these legal instruments all play significant roles in how records are managed in the public sector, the NAA provides the anchor for the management of records, including digital records. The Act defines a public record as "a record created or received by a governmental body in pursuance of its activities" (South Africa, 1996, p. Sec 1). Additional guidance on the management of these public records is found in a number of advisory pamphlets and policy documents and those related to digital records include:

1. *Advisory Pamphlet Number 1 - Managing public records and the law* (National Archives and Records Service of South Africa, 2007c).
2. *Advisory Pamphlet Number 2 – Electronic Records and the Law* (National Archives and Records Service of South Africa, 2007a).
3. *Advisory Pamphlet Number 5 – Managing email and the law* (National Archives and Records Service of South Africa, 2007b).
4. *Guidelines for Managing Electronic Records* (National Archives and Records Service of South Africa, 2006, p. i).

One of the main legislative challenges that public institutions in South Africa face is distinguishing between the original digital record from a copy. The South African Law Reform Commission (2010, p. 52-53) advocates for a review of the definition of data message as well as the inclusion of definitions of 'electronic', 'copy' and 'original' in the ECT Act. Clarifying these definitions is critical when dealing with electronic evidence of transactions and communication in South African legal proceedings (Mostert, 2005, p. 7). Force (2013, p.

127-128) discusses how judges in Canada base their decisions on the admissibility of a record in court proceedings and argues that records professionals should be able to identify original records and be able to provide information on whether they are originals or not. For digital records, this is largely dependent on the "metadata from and electronic recordkeeping system" that gives account of versions of records, when they are destroyed and as well as an audit trail (Force, 2013, p. 127-128). Unfortunately records professionals in South Africa are not able to make this determination because of the gaps in the current legislative and regulatory instruments. Therefore, a review effort such as the one called upon by the South African Law Reform Commission should include examining the law of evidence and other legislative instruments as well as explore the implications to the recordkeeping practices in the country's public institutions.

Literature review on the technological terrain

The technological challenges of managing digital records in South have been discussed in research projects since the late 1990s and throughout the 2000s (Abbott, 1999, Kwatsha, 2010). Keakopa (2010) and Kemoni (2009) have argued that South Africa is the most advanced African country in the implementation of software applications to manage digital records. A survey of South African institutions aimed at investigating their implementation of software applications used to manage digital records revealed that, by 2010 when the research was conducted, more than 40% of the institutions had five or more years of practical experience (Katuu, 2012, p. 48-49). The survey also revealed that these institutions were using different modules such as document management, records management, imaging and workflow within their applications either as open source or proprietary software (Katuu, 2012, p. 50-51). A common term used for applications that include both document and records management is Electronic Document and Records Management Systems (EDRMS) (Nguyen et al., 2009, Wilkins et al., 2009).

However, records professionals often overlook the wider technological environment within which EDRMS applications are utilised. The enterprise architecture of such an environment often includes many other information systems. Lappin (2010, p. 254) noted that organizations "have an information archaeology, not an information architecture. New applications are brought into the organizations' information estate, but old applications persist rather than disappear." Hepsø et. al (2009) demonstrated this information archaeology in an oil and gas company in Norway that sought to introduce MS SharePoint but had to acknowledge and deal with the presence previous information management systems such as shared file drives and Lotus Notes.

A South African example can be drawn from the public healthcare sector. The National Health Act (61 of 2003) makes specific reference of the need for a National Health Information System (NHIS) with each of the nine provincial gov-

ernments responsible for contributing in the management and consolidation of health information. However, the presumption of a single system is not the reality and the Department of Health (2012, p. 5) acknowledges in its eHealth Strategy that the existing information systems are fragmented, lack coordination and are not interoperable. This is mostly aptly demonstrated in the secondary and tertiary health institutions where there are at least 15 different patient management or hospital information systems in use within different provinces of the country as shown in the table below (Department of Health [South Africa], 2012, p. 14).

Table 1. The Patient Management/Hospital Information Systems in South African provinces

Province	Patient Management/Hospital Information System in Use
Eastern Cape	Delta 9
Free State	Meditech; PADS
Gauteng	Medicom; PAAD; PharmAssist; Soarian MedSuite
KwaZulu-Natal	Meditech; Medicom; PALS; Pro-Clin; ReMed
Limpopo	Medicom
Mpumalanga	PAAB
North West	PAAB
Northern Cape	Nootroclin
Western Cape	Clinicom; Delta 9; PHCIS; JAC Pharmacy

As shown in the table above, four of the provinces use more than one information system demonstrating the need for integration within each province before a nation-wide effort in the single NHIS (Katuu, 2015, p. 160). The Department of Health (2012, p. 26) identifies a number of aspects that require implementation towards reaching the goal of the NHIS including:

- Implement the foundation of the Electronic Health Record (EHR) and particularly a national patient registry and Patient Master Index (PMI).
- Implementation of primary health care patient management and Electronic Medical Record (EMR) system/s at clinics.
- Implementation of Pregnancy and Neonatal EMR system to record clinical details with link to EHR.
- Implementation of EMR system/s to monitor anti-retroviral treatment (ART) and tuberculosis treatment.
- Pharmacy systems interface to EMR systems.
- Implement a uniform Integrated Document and Records Management System (EDRMS) at all levels.

This health sector example demonstrates that efforts in the identification and management of digital records within the public institutions would have to acknowledge that EDRMS applications are only part of an archaeology of information systems within health institutions. Therefore South African records professionals have to be cognisant of the whole enterprise architecture of sys-

tems and not just the EDRMS applications when engaging on efforts to manage digital records.

Conclusion

This article has provided an outline of the literature review being carried out as part of a research study entitled *Managing records in networked environments using EDRMS applications*. The aim of the research study is to assess the recordkeeping environment in a South African public enterprise known as Rand Water. The purpose of this article was to outline emerging issues that provide contextual background on the current situation within the public sector in South Africa. The article has demonstrated that the country has a complex legal and regulatory environment where determining original digital records from copies is an ongoing challenge. In addition it has outlined, using the example of the health sector, that public institutions often have an extensive information system archeology. Therefore the management of digital records should not be just of those originating from EDRMS applications but from other disparate information systems in the enterprise architecture.

References

Abbott, B. S. (1999), Preserving Electronic Memory: An Investigation into the Role played by the National Archives of South Africa in the Management of Electronic Records of Central Government. Masters thesis, University of KwaZulu-Natal.

Alberts, M. and Mollema, N. (2014), "Developing Legal Terminology in African Languages as Aid to the Court Interpreter: A South African Perspective", Lexikos, Vol. 23 No. 1, pp. 29-58.

Allan, K. (Ed.) (2009), Paper Wars: Access to Information in South Africa, Wits University Press, Johannesburg.

Department of Health [South Africa]. (2012), "eHealth Strategy South Africa 2012-2016", Pretoria: Department of Health [South Africa], available at: http://www.health.gov.za/docs/strategic/2012/eHealth_Strategy_South_Africa_2012-2016.pdf (accessed 19th September 2015).

Du Bois, F. (2004), "Introduction: History, System and Sources", in Van der Merwe, C. G. and Du Plessis, J. E. (Eds.) Introduction to the Law of South Africa, Kluwer law international, The Hague, pp. 1-54.

Force, D. C. (2013), Pursuing the “usual and ordinary course of business”: An Exploratory Study of the role of Recordkeeping Standards in the Use of Records as Evidence in Canada. Phd thesis, University of British Columbia.

Hepsø, V., Monteiro, E. and Rolland, K. H. (2009), "Ecologies of e-Infrastructures", Journal of the Association for Information Systems, Vol. 10, pp. 430-446.

InterPARES Trust. (2015), "Team Africa", Vancouver, available at: http://interparestrust.org/trust/aboutus/africa (accessed 31st May 2015).

Katuu, S. (2012), "Enterprise Content Management (ECM) implementation in South Africa", Records Management Journal, Vol. 22 No. 1, pp. 37-56.

Katuu, S. (2015), Managing Records in South African Public Health Care Institutions - A Critical Analysis. Phd thesis, University of South Africa.

Keakopa, S. M. (2010), Management of Electronic Records, Lap Lambert Academic Publishing.

Kemoni, H. N. (2009), "Management of Electronic Records - Review of Empirical Studies from the Eastern, Southern Africa Regional Branch of the International Council on Archives (ESARBICA) region", Records Management Journal, Vol. 19 No. 3, pp. 190-203.

Kwatsha, N. (2010), Factors affecting the Implementation of Electronic Document and Records Management Systems. Masters thesis, University of Stellenbosch.

Lappin, J. (2010), "What will be the next Records Management Orthodoxy?", Records Management Journal, Vol. 20 No. 3, pp. 252-264.

Madhuku, L. (2010), An Introduction to Zimbabwean law, Weaver Press, Harare.

Mostert, W. (2005), "The Electronic Communications and Transactions Act: Guide for Consumers, Businesses and Public Bodies", Cape Town, Mostert Opperman Incorporated.

National Archives and Records Service of South Africa. (2006), "Managing Electronic Records in Governmental Bodies: Policy, Principles and Requirements", Pretoria: National Archives and Records Service of South Africa, available at: http://www.national.archives.gov.za/rms/NARS_DMLIB-4878-v1-Managing_electronic_records_Policy__principles_and_Requirements_April_2006.PDF (accessed 23rd February 2014).

National Archives and Records Service of South Africa. (2007a), "Advisory Pamphlet No 2 - Electronic records and the law: What governmental bodies need to know", Pretoria: National Archives and Records Service of South Africa,, available at: http://www.national.archives.gov.za/Advisory%20Pamplet%20No%202%20April%202012.pdf (accessed 27th August 2015).

National Archives and Records Service of South Africa. (2007b), "Advisory Pamphlet No 5 - Managing email and the law: What governmental bodies need to know ", Pretoria: National Archives and Records Service of South Africa,, available at: http://www.national.archives.gov.za/rms/ad_pamphlet5.PDF (accessed 4th December 2013).

National Archives and Records Service of South Africa. (2007c), "Advisory Pamphlet No. 1 - Managing Public Records and the Law: What governmental bodies need to know", Pretoria: National Archives and Records Service of South Africa,, available at: http://www.national.archives.gov.za/rms/ad_pamphlet1.PDF (accessed 27th August 2015).

Ngoepe, M. and Keakopa, S. M. (2011), "An Assessment of the State of National Archival and Records Systems in the ESARBICA region: A South Africa-Botswana Comparison", Records Management Journal, Vol. 21 No. 2, pp. 145-160.

Nguyen, L. T., Swatman, P. M. C., Fraunholz, B. and Salzman, S. (2009), "EDRMS Implementations in the Australian Public Sector", available at: http://aisel.aisnet.org/acis2009/16/ (accessed 27th March 2012).

South Africa (1996), "National Archives Act (Act 43 of 1996)", Pretoria, Government Printer.

South Africa (2000), "Promotion of Access to Information Act (Act 2 of 2000)", Pretoria, Government Printer.

South Africa (2002a), "Electronic Communications and Transactions Act (Act 25 of 2002)", Pretoria, Government Printer.

South Africa (2002b), "Regulation of Interception of Communication and Provision of Communication-related Information Act (Act 70 of 2002)", Pretoria, Government Printer.

South Africa (2013), "Protection of Personal Information Act (Act 4 of 2013)", Pretoria, Government Printer.

South African Law Reform Commission. (2010), "Media Statement by the South African Law Reform Commission concerning its investigation into the Review of the Law of Evidence (Project 126)", Johannesburg, available at: http://www.justice.gov.za/salrc/media/2009_ip27_pr126_media.pdf (accessed 2nd February 2014).

The Archival Platform. (2015), "State of the Archives: An Analysis of South Africa's National Archival System", Cape Town: University of Cape Town, available at: http://www.archivalplatform.org/images/resources/State_of_the_Archive_FOR_WEB.pdf (accessed 19th May 2015).

Toufayan, M. (2014), "When British Justice (in African Colonies) Points Two Ways: On Dualism, Hybridity, and the Genealogy of Juridical Negritude in Taslim Olawale Elias", in Onazi, O. (Ed.) African Legal Theory and Contemporary Problems, Springer, Dordrecht, pp. 31-70.

Wilkins, L., Swatman, P. M. C. and Holt, D. (2009), "Achieved and Tangible Benefits: Lessons Learned from a Landmark EDRMS Implementation", Records Management Journal, Vol. 19 No. 1, pp. 37-53.

Trustful hybrid curation, transformation and organization of records in Belgium

Fiona Aranguren Celorrio
National State archives
Brussels, Belgium
Fiona.arangurencelorrio@arch.be

Laurence Maroye
Université Libre de Bruxelles
Brussels, Belgium
lmaroye@ulb.ac.be

Marie Demoulin
University of Montreal
Montreal, Canada
marie.demoulin@umontreal.ca

Sébastien Soyez
National State Archives
Brussels, Belgium
sebastien.soyez@arch.be

Cécile De Terwangne
University of Namur
Namur, Belgium
Cecile.deterwangne@unamur.be

Seth Van Hooland
Université Libre de Bruxelles
Brussels, Belgium
seth.van.hooland@ulb.ac.be

Bénédicte Losdyck
University of Namur
Namur, Belgium
benedicte.losdyck@unamur.be

Odile Vanreck
University of Namur
Namur, Belgium
odile.vanreck@unamur.be

Summary

This paper presents the objectives, methodology and preliminary conclusions of the HECTOR research project (Hybrid Electronic Curation, Transformation and Organization of Records). This interdisciplinary project aims at developing a model for the management and preservation of hybrid records (i.e., paper-based, digital-born or digitized) in the Belgian federal administrations.

Keywords: Hybrid records management, authenticity, evidence, digital signature, public service, original record, copy

Introduction

The coexistence of paper-based and digital-based environments has been an inconvenient situation for a long while in public and private administrations worldwide. The paperless office is still a myth despite major and fast spreading technological improvements in the fields of electronic records management and archiving (Sellen & Harper, 2003). This uneven transition and its resulting situation have created a hybrid environment of records: paper-based records that are digitized and sometimes printed again; digital-born records that are printed and

then scanned. As a consequence, tracking the authenticity and integrity of the record along the chain of custody (Duranti, Eastwood & MacNeil, 2002) is increasingly difficult. This heterogeneity results in information loss and redundancy, waste of valuable resources, legal conflicts and a lack of efficiency of public services.

To tackle these and other challenges and to boost the cultural and scientific research, the Belgian Federal Science Policy Office (Belspo) launched the BRAIN-be program in 2012 (Belgian Research Action through Interdisciplinary Networks)1as an answer to the need for scientific knowledge of the federal departments and to support the scientific potential of the Federal Scientific Institutions2. During its first phase and until 2017, the program creates a framework for the development of research projects within 6 thematic areas related to ecosystems, geo-systems, cultural heritage, public strategies, societal challenges and the management of collections.

The HECTOR3 research project was selected as part of this program, within the thematic area "Federal public strategies", and will carry out its research work throughout four years.

The HECTOR research project

The main goal of the HECTOR research project is to clear the path for Belgian federal administrations and guide them through an often flurry hybrid environment. To fulfill this objective, the project team will offer hybrid records management and archiving advice in order to enhance trust, transparency and security as well as to minimize the use of unnecessary paper and countless digital copies and to improve the current conditions for the long-term preservation of the information heritage of public authorities. This advice will be accompanied by two transversal models for a trustful, secure and efficient records management.

This guidance will be achieved through an interdisciplinary collaboration between members of the following institutions, each offering expertise in its own fields: the Research Center of Information, Law and Society of the University of Namur for the field of legal sciences; the Information and Communication Science department of the Université Libre de Bruxelles for the field of information sciences; and the Digital Preservation and Access Division of the State Archives Belgium for archival sciences. To complete the team and in pursuance of a more international approach, the University of Montreal and its École de

[1] http://www.belspo.be/belspo/brain-be/index_en.stm

[2] Within these eleven institutions, there can be mentioned the State Archives, the Royal Museum of Art and History or the Royal Observatory of Belgium and Planetarium among others.

[3] http://www.hector-project.be

Bibliothéconomie et des Sciences de l'Information will provide a cross-domain expertise in both archival sciences and law.
Concerning the methodology chosen by the research team and in addition to the aforementioned interdisciplinary approach, the project adopts a bottom-up approach. The starting point of the research are the particularities of hybrid records management found through the analysis of a selection of case studies within the Belgian federal administrations. These case studies reflect the two levels of analysis of the project: 1) the hybrid management of records is analyzed at the Federal and Local Police, the Federal Ministry of Employment and at the Courts; 2) the hybrid management of files is analyzed at the Federal Agency for the Safety of the Food Chain, the Federal Agency for Nuclear Control, the State Archives in Belgium, the Federal Ministry of Finances, and the Federal Ministry of ICT.
The research team will also seek the advice of a Follow-up Committee which is composed of representatives from international and national organizations, federal and local Belgian administrations, as well as of some independent experts and other potential users. All these stakeholders are encouraged to give the team their critical input on the progress of the project in order to directly benefit them in their daily records management and archiving practices.

Preliminary results

After the first year of research, the first level of analysis (the hybrid management of records) is coming to an end. An initial theoretical analysis of the chosen case studies and more than a dozen interviews with the main stakeholders have enabled the research team to draw some preliminary observations.
On the one hand, there exists a tangible legal insecurity regarding key issues such as the digital signature, the probative value of records, the difference between the many versions of the same record in different formats, the difference between an original record and its copy, or the status of digital records. Such legal insecurity constitutes an important obstacle for on-going and future digitization projects, records management policy implementations and digital archiving strategy developments.
On the other hand, the participating stakeholders have shown a widespread interest in the concerns arisen by the research project, an interest clearly linked to a feeling of urgency engendered by never-stopping technological progress and continuously decreasing budgets in the public services. These concerns address technological, legal, organizational, societal, political and economic matters, all of them affecting directly the records management and archiving practices of these same public institutions.
Besides these observations, the field analysis has also shown a worrying lack of proper records management and archiving systems, also called EDRMS4 (Ste-

[4] Electronic Document and Records Management System (EDRMS).

phens, 2007). Usually due to reduced financial, technical and human resources, the solutions in place are limited to mere storage and back-up systems. Furthermore, the acceptance and use of this kind of information management solutions depend on the implication of every member of the organization in the development and implementation project of the solution from its early stages. This fact is often understood and tackled through various trainings and yet, these efforts are not enough. Hence, further change management initiatives should be encouraged.

Conclusion

In the context of an ever-changing records management environment, the Belgian public administrative structures are confronted with new challenges due to the existence of hybrid records and files. For the sake of the efficiency in the processing and the exchange of records within these institutions, as well as the appropriate preservation of their information heritage, this research project will continue to analyze different case studies and to assess the subsequent results during its second year, in order to build and share the foreseen models for a hybrid records management and archiving during the third year. Eventually, the fourth and last year of the project will be dedicated to develop e-learning tools on top of an international conference so that the recommended guidelines and models can be shared with different stakeholders.

References

Duranti, Luciana ; Eastwood, Terry ; MacNeil, Heather. Preservation of the Integrity of Electronic Records. Dordrecht: Kluwer Academic Publishers, 2002

Stephens, David O. Records Management: Making the Transition from Paper to Electronic. Lenexa: ARMA International, 2007

Sellen, A. J.,; Harper, R. H. R.: The myth of the paperless office. Cambridge, MA: MIT Press, 2003.

Digitally Signed Records – Friend or Foe?

Boris Herceg, Hrvoje Brzica
Financial Agency – FINA
Ulica grada Vukovara 70, Zagreb, Croatia
boris.herceg@fina.hr, hrvoje.brzica@fina.hr

Hrvoje Stančić
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb, Croatia
hstancic@ffzg.hr

Summary

Long-term preservation of digitally signed records may be a challenging task. Digital signatures expire and digital certificates may be revoked, thus influencing the trustworthiness of archived digital records. The authors firstly explain the problems and then proceed to a description of the planned research. The research will be conducted on a sample of digitally signed and archived electronic forms stored in the PDF file format, originating from the period from 2006 to 2009.

Keywords: electronic forms, digital records, digital signatures, PDF, file format, long-term preservation

Introduction

Thibodeau (2002) said that "the preservation of digital objects involves a variety of challenges, including policy questions, institutional roles and relationships, legal issues, intellectual property rights, and metadata" but also that "the variety and complexity of digital information objects engender a basic criterion for the evaluation of possible digital preservation methods, namely that they must address this variety and complexity." Becker, Kulovits, and Rauber (2010) further explain that since "a digital object needs the correct environment in order to function, we can either recreate the original environment (emulation) or transform the object to work in different environments (migration)". However, the digital object could also be converted to a newer version of the same file format, or to a different file format (conversion). All those changes could impact the trustworthiness of a digital record, i.e. influence its authenticity, reliability, accuracy, integrity, and/or usability. In order to prevent any of the possible unwanted changes, Duranti (1999) explicated that "irrespective of the long-term solution for the preservation of authentic electronic records, it is quite clear that there will not be much worth preserving for the future if serious measures

are not taken by records creators to guarantee the trustworthiness of electronic records (in both meanings – trustworthiness of content and trustworthiness of the record as a record) since the moment of creation." This could be a complex task if the records being preserved are digitally born records signed with (advanced) digital signatures which depend on the (qualified) digital certificates and (trusted) timestamping process, and are entrusted to the cloud (Stančić, Rajh, & Brzica, 2015). This article will focus on the issues dealing with the long-term preservation of digitally signed records, specifically on the records in the .PDF file format, and explain the planned research .

The problem

Digitally signed records, although well preserved over the long-term period, could lose their legal validity if the digital signature cannot be validated, or if it loses its characteristic of non-repudiation. Adobe confirms that "the mere existence of a digital signature is not an adequate assurance that a document is what it appears to be". If the validity check of a digital signature results in an error, e.g. no proof of existence, the record's trustworthiness may be compromised.

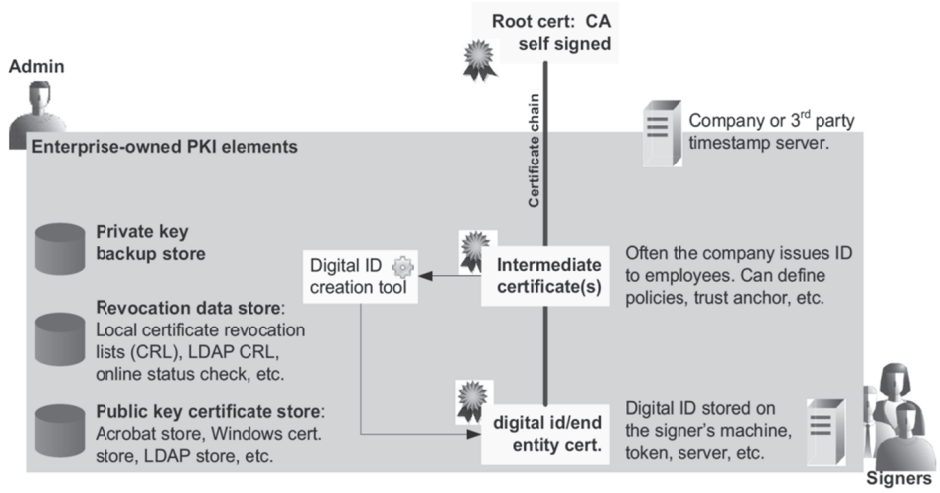


Figure 1. Common PKI elements in signature workflows (Adobe, p. 3)

The problem is that digital signatures have a valid time span, and that their validation requires a connection with the certification authority relying on the Public Key Infrastructure (PKI) (Figure 1). If any of the elements in this chain fails, the validity check will fail too. This is even more important if the records with advanced digital signatures, i.e. digital signatures using digital certificates, are being preserved. Digital certificates expire, but they also may be revoked. Should the historic information on the revocation lists be preserved along with the digitally signed records? Is that technically possible? However, this problem

should be appropriately addressed before it appears. In the case of this planned research, the digitally signed records are already archived, and it yet remains to be investigated if the passage of time will influce/influences their trustworthiness. The results of this preliminary research will provide some answers relevant to the more crucial questions recognised above, and will help prepare a later stage of the research in which those questions will be addressed in full.

Research

The research will be conducted on electronic forms of the public administration services created from 2006 up until 2009. The electronic forms subject of research will be the electronic form of the Croatian Pension Insurance and the electronic form of online registration to the Court Register, both created as PDF documents and digitally signed. The working hypothesis is that the technological progress has no effect on the long-term preservation of the content and the key elements of electronic records, and that both the content and the key elements of electronic records are fully preserved. Of course, this hypothesis can prove to be true, partially true or false.

Methodology

The methods of collection, sampling, parametrisation, testing, comparison, analysis, synthetisation, and abstraction will be used in the course of the study.

The plan is to organise the research in several steps as follows:

1. Definition of the research parameters and the testing environment.
2. Collection of samples of electronic records (.PDF files) for analysis from the production system of each electronic service, i.e. from its archives.
3. Organisation of all samples of electronic records of one electronic service into a single directory in the test environment.
4. Duplication – a copy of the sample will be made before the testing in order to ensure that the testing has no impact on the original sample.
5. Testing will be done by opening the sample PDFs in the version of the reader from the time of record creation, following by opening the same sample in the consecutive, newer versions of the reader. During the testing, it will be necessary to examine whether the individual Reader version contains a documented bug. In that case it will be necessary to register these findings and take the next stable version of the Reader in which a particular bug was corrected. The characteristics of the records will be investigated.
6. Data analysis.
7. Synthetisation of the findings.
8. Writing of the final report and recommendations.

For the purpose of this research the following characteristics of the archived electronic forms of the public administration services in the .PDF format will be investigated:

1. Readability of the content.
2. Validity of digital signatures.
3. Functionality of digital signatures' visualization.
4. Display of digital signatures' elements.
5. Size of digital signatures.
6. Size of electronic records.
7. Legal usability of the electronic records.

Other characteristics may prove to be relevant as well.

The testing environment should include:

1. Possibility of installation of all versions of the reader using the latest OS environment.
2. Ability to recognize extensions and attributes of electronic records.
3. The possibility to duplicate the testing sample without losing the attributes of the sample.

Expected results

The researchers hope to prove the hypothesis, i.e. to find out that the technological progress has no effect on the long-term preservation of the digitally signed PDFs, and that both the content and the key elements relevant to the concept of trustworthiness of electronic records are fully preserved. These expectations are based on the facts that the investigated records are stored in the stable and widely accepted file format and that the vendor takes into the consideration the backward compatibility of their products.

Results of this research will be used in the next stage in which the preservation of the historic information on the revocation lists, along with the digitally signed records, as well as the application of timestamps to the records with the expiring certificates, will be investigated.

References

Adobe. (n.d.). Digital Signatures in a PDF. Retrieved September 15, 2015, from https://www.adobe.com/devnet-docs/acrobatetk/tools/DigSig/Acrobat_DigitalSignatures_in_PDF.pdf

Becker, C., Kulovits, H., & Rauber, A. (2010, 1). Trustworthy Preservation Planning with Plato. (P. Kunz, Ed.) *ERCIM News - Special theme: Digital Preservation*, pp. 24-25.

Duranti, L. (1999). Concepts and principles for the management of electronic records, or records management theory is archival diplomatics. *Records Management Journal, 9*(3), 149-171.

Stančić, H., Rajh, A., & Brzica, H. (2015, June). Archival Cloud Services: Portability, Continuity, and Sustainability Aspects of Long-term Preservation of Electronically Signed Records. *The Canadian Journal of Information and Library Science (CJILS), 39*(2), 210-227.

Thibodeau, K. (2002). Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In *The State of Digital Preservation: An International Perspective* (pp. 4-31). Washington, D.C.: Council on Library and Information Resources (CLIR).

MODERN INFORMATION SYSTEMS

Nucleus of the Strategic Planning of an Integrated University Information System

Ljerka Luić
B4B d.o.o.
Ulica grada Vukovara 271, Zagreb, Croatia
ljerka.luic@b4b.hr

Ana Fudurić
Karlovac University of Applied Sciences
Trg J. J. Strossmayera 9, Karlovac, Croatia
ana.fuduric@yahoo.com

Summary

Nucleus of the strategic planning of the integrated information system (SP-IIS) is determined by implication of mutual impacts of its four principal elements: information strategy, info management strategy, info technology strategy and change management strategy, based on which the subject of research is described in the paper, viewed from the aspect of social dimension alignment of the academic (elected) and administrative (permanent) university management with the aim of drafting a simulation model of strategic planning of the integrated university information system (SP-IIS-Uni). An analysis of the research results provided evidence indicating that, with respect to the strategic planning of an integrated business information system (SP-IBIS-Uni), university top management expresses the greatest interest in the planning of alignment between the business information system (BIS) and the academic information system (AIS). They believe the impact of using ICT is crucial for their future work and operation of the university, but also necessary for more efficient integration with university institutions in EU countries. In the context of investigating this domain, the nucleus of the strategic planning of an integrated university information system is viewed as an implication of mutual impacts of information strategy, info management strategy, info technology strategy and change management strategy, but also the social implications by which the SP-IIS-Uni nucleus strategy affects the information & organizational environment of the university. Elaborated premises, hypotheses set and a comparative analysis of results from similar research resulted in a conclusion on an inseparable synergic impact of nucleus particles on the strategic planning of integrated university information system development.

Keywords: nucleus, strategic planning, information system, university

Introduction

In the concept, strategic planning is deceptively simple. The analysis includes the present and expected future states and determines the direction and development of resources required to accomplish the mission. In reality, strategic planning is an extremely complex process requiring a systematic approach to determining and analyzing factors within and beyond the business system. A business system's information system is of great importance for its viability and operation, which is why its strategic planning is as important as the strategic planning of the business system. Based on this premise, the development of modern integrated information systems is a foreseeably demanding and comprehensive organizational and financial venture for any business system, including an academic one, which needs to be strategically planned in accordance with the strategic business plan.

Background

From the scientific research viewpoint, investigating strategic planning of information systems (SPIS) is not easy, especially considering the fact that this discipline of information science is very young. In initial research of information systems where works of Burrell and Morgan[1], Wood-Harper[2], Klein and Lyytinen[3] and Iivari[4] dominate, the researchers agreed upon the identification of four basic paradigms defined as functionalism, interpretation, radical humanism and radical structuralism. These paradigms may be analyzed by investigating four basic dimensions: ontology, human nature, epistemology and methodology. It is unquestionable that the two dominant paradigms in the studying of this area are the interpretative and functional paradigms. Among others, Han and Jones[5] additionally broaden the view and draw attention to the fact that a new research body is being developed that assumes a more pragmatic position because, according to Orlikowski[6], early work considered technology to be an objective external force that would have relatively deterministic effects on organizational characteristics such as structure. Conversely, a group of researchers later focused on human activity with respect to technology, perceiving it more as a product of joint interpretations or interventions.

[1] Burrell, G., Morgan, G., *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life*, Ashgate Publishers, 1979

[2] Wood-Harper, T., *Research Methods in Information Systems*, Elsever Science Publishers, 1985

[3] Klein, H.K., Lyytinen, K., *The Poverty of Scientism in Information Systems,* Manchester, 1984

[4] Iivari, J., *A paradigmatic analysis of contemporary schools of IS development*, European Journal of Information Systems, 1(4), 1991, pp. 249-272

[5] Han, P., Jones, M., *The dialectics of Information Systems*. University of Cambridge Research Paper: Judge Institute of Management Studies, 1993

[6] Orlikowski, W. J, *The duality of technology: rethinking the concept of technology in organisations*, Organisation Science, 3(3), 1992, pp. 398-427

The third group reverted to technology, however, as "soft" determinism where technology is defined as an external force having its effects which are modeled by the human factor and organizational contexts. The research described in this paper integrated the abovementioned concepts through paradigmatic frameworks and is in this context classified in the third group, based on the theory of structuralism and focused on investigating the social dimension of alignment. Except for several statements concerning the domain of the social dimension of alignment, defined by Reich i Benbasat[7] as the state in which business and IT executives within an organization understand and are committed to business and IT objectives, mission and plans, and several papers published by Nelson and Cooprider[8], Taylor-Cummings[9], Subramani, Henderson and Cooprider[10], and Reich and Benbasat[11], there are no systematic scientific papers investigating methods of the social dimension of alignment viewed through the time dimension and focused on measuring the state in which different groups of executives within a line of business understand and take part in the strategic planning of an integrated information system contemplated in the research, the results of which are provided below.

Problem Statement

Academic circles show a high level of disagreement on how to position the making of a decision to initiate the strategic planning of information systems, who should be responsible and which persons from which domains should be actively involved. This disagreement is often associated with positioning information in the center of the success of university organizations. The way in which information is used (automatically or manually integrated) is a key factor that determines how much each department and university as whole is efficient and competitive. It is unquestionable that, as regards the relevance, possibility of dissemination and support to integral academic functions, information is the principal element of primary university activities such as teaching and research activities. It is also important to note that a university is often expected to generate information relevant to specific stakeholders, from the founder to the greater social community. Providing all this information is very often a difficult and major challenge to a university and the academic community as a whole.

---

[7] Reich, B.H., Benbasat, I., *Measuring the linkage between business and information technology objectives, MIS Quarterly,* 20(1), 1996, pp. 55-81

[8] Nelson, K.M., Cooprider, J.G., *The contribution of shared knowledge to IT Group performance*, *MIS Quarterly,* 20(4), 1996, pp. 409-432

[9] Taylor-Cummings, A., *Bridging the user-IS Gap, A study of major information systems projects*, Journal of Information Technology, (13), 1998, pp. 29-54

[10] Subramani, M., Henderson, J.C., Cooprider, J., *Linking IS-user partnerships to IS performance: a socio-cognitive perspective*. *MISRC Working Paper Series WP 99-01*. Boston, 1999

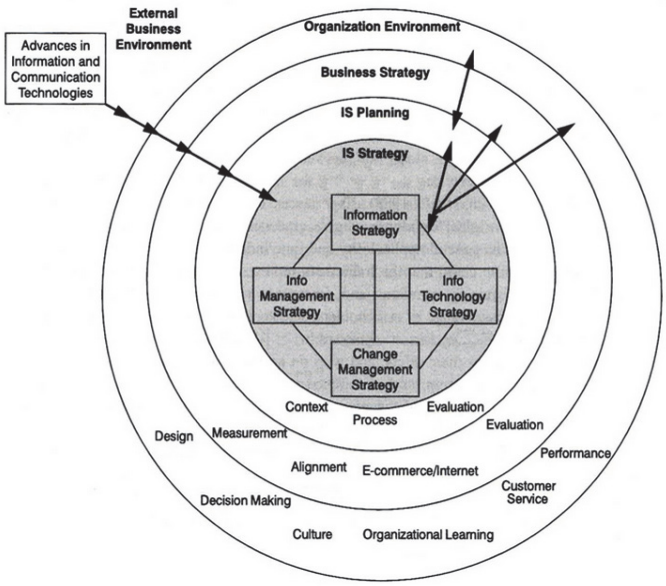
[11] Reich, B.H., Benbasat, I., *Measuring the Information Systems - Business Strategy Relationship*. In: Galliers, R.D., Leidner, D.E., *Strategic Information Management: Challenges and Strategies in Managing Information Systems*. Oxford: Elsevier Butterworth-Heinemann, 2003, pp. 265-310

After the Bologna process was implemented in the higher education system of the Republic of Croatia, an acute problem arose in the optimization of using available resources: budget funds, internal revenue, staff, space and equipment, in particular the optimization of allocating funds by introducing lump sum funding. It is therefore necessary for the development and management of a complex system such as an integrated information system to design a uniform information strategy as a link between the academic strategy, the information system strategy and the business strategy. The complexity of this process requires specific forms of strategic thinking, planning and activity management, and ultimately strategic planning of the university.

Nucleus of the Strategic Planning of an Integrated Information System

According to Ward and Peppard, the nucleus of the strategic planning of an integrated information system is determined by the implications of the mutual impacts between its particles: *Information Strategy, Info Management Strategy, Info Technology Strategy, and Change Management Strategy*, based on which the subject of research described in this paper is observed from the viewpoint of the social dimension of the short-term and long-term alignment of academic (elected) and administrative (permanent) management of the university and its constituents.

Figure 1: Particles of the Information System Strategic Planning Nucleus



Source: Ward J., Peppard J. *Strategic planning for Information systems*, Wiley series in information systems, 2010, p. 137

Strategic planning of information systems remains the most important issue in many organizations. SPIS cannot be understood by taking into consideration only formal methods. Planning processes and plan implementation are just as important, so Galliers and Leidner accordingly propose five different approaches to SPIS: business-led, method-driven, administrative, technological and organizational approaches.[12] Each approach has different characteristics and, consequently, different chances to be successful. The results of completed research show that the organizational approach is the most efficient one. The taxonomy of these five approaches also offers a diagnostic tool for analyzing and evaluating an organization's experience with SPIS. It is recommended for SPIS to focus on the following areas:[13]

- investing in IT strategy according to business objectives
- using IT to gain competitive advantage
- directing efficient management of information strategy resources
- developing technological policies and architectures.

The above graphically and textually elaborated statements show that the strategy nucleus particles form a synergic, almost indivisible unit, which points to the conclusion that their interactive relations are highly relevant to the strategic planning of integrated information systems (SP-IIS).[14]

Material and Methodology

Research Sample

The research on which this paper is based was conducted on all 7 universities in the Republic of Croatia represented by their Rectorates, and their 66 constituents (faculties) represented by their Dean's Offices. The final selection of 73 higher-education institutions of a university nature was ultimately viewed as a reflection of Yin's[15] strategy of "literal replication" where all cases are theoretically the same, which achieves the principal objective of the research as it was conducted on 68.5% of all institutions and this allowed for statistically significant results to be provided, as defined by Reich and Benbasat[16]. The final sample based on said stratification was organized according to Sumsion's[17] typol-

[12] Galliers R.D. Leidner D.E., *Strategic Information Management: Challenges and Strategies in Managing Information Systems*. Oxford: Elsevier Butterworth-Heinemann, 2003, p. 81
[13] Ibid
[14] Fudurić, A, *The nucleus of the business and information strategy*, Graduate thesis, Karlovac University of Applied Sciences, 2014, p. 35
[15] Yin, R. K., *Case Study Research: Design and Methods*, London: Sage Publications, 2003
[16] Reich, B.H., Benbasat, I., *Measuring the Information Systems - Business Strategy Relationship*. In: Galliers, R.D., Leidner, D.E., *Strategic Information Management: Challenges and Strategies in Managing Information Systems*. Oxford: Elsevier Butterworth-Heinemann, 2000
[17] Sumsion, J., *Survey of resources and uses in Higher Education Libraries*. Loughborough, Library and Information Statistics Unit, Loughborough University of Technology, 1994

ogy and observed dually: (1) from the viewpoint of legal entities, 73 public institutions of a university nature, and (2) 365 individuals (executives).

Research Methodology

Based on the fact that a survey questionnaire is a fundamental measuring instrument for data gathering in social sciences, including information science, this measuring instrument was used for the purposes of the research concerned. A total of 24 questions were grouped in three thematic units: (1) Alignment of the Business and Academic Information Systems, (2) Planning of Integrated Information System Development, and (3) Using ICT. The subjects completed the questionnaire personally, all subjects within an institution at the same time in the presence of a researcher, a co-author of this paper.

Hypotheses

An analysis of earlier similar research indicated that there are no generally accepted models for investigating the social dimension of alignment. It was therefore proceeded with identifying the categories of factors that may affect alignment according to theoretical and empirical literature.

By using relations studied in literature and defining new ones, these factors were organized in a research model containing four constructs distributed across two levels. These constructs were identified as KNOWLEDGE, PLANNING, ATTITUDE AND FUNDING and allocated by taking into account the time dimension to short-term and long-term alignment levels. The main hypothesis and sub-hypotheses were set based on these statements, defining each construct of the model as provided below in a logical order.

$\underline{H0_1}$: The strategic planning of an integrated information system (SP-IIS) largely depends on the long-term alignment of a university institution's management, on which the KNOWLEDGE and FUNDING constructs have a direct impact.

$H0_2$: The common KNOWLEDGE of the domain (SP-IIS) has the greatest weighting factor and is the only one distinguishing between management on the short-term and long-term alignment levels.

$H0_3$: The perception of the importance of the strategic PLANNING of an integrated information system largely depends on the type of executive function (permanent vs. elected).

$H0_4$: The ATTITUDE of management directly affects the level of their perception regarding the significance of the strategic planning of an integrated information system.

$H0_5$: The mechanism and sources of university's FUNDING have a direct impact on the strategic PLANNING of an integrated information system.

$H0_6$: The long-term alignment of management with SP-IIS is defined on a causal basis: KNOWLEDGE → ATTITUDE → FUNDING → PLANNING.[18]

Results

The main hypothesis was fully confirmed. Namely, the results of the linear regression model with PLANNING as a dependent variable and ATTITUDE, FUNDING and PLANNING as independent variables show that the independent variables explain a statistically significant part of the variability of the dependent variable PLANNING, while a PATH analysis[19] demonstrated that KNOWLEDGE and FUNDING reflect the state where permanent and elected executives share a common vision with respect to the strategic planning of an integrated business information system.[20]

Discussion

The Main Scientific Evidence Obtained by the Research

The main evidence indicates the fact that, as regards the strategic planning of an integrated business information system, university management shows the greatest interest in the domain concerning the alignment of the business and academic information systems, followed by the use of ICT, while their interest is the least focused on the planning of integrated information system development. Although not the largest, a sufficiently substantial share of management believes the success of their institution's current operation is not dependent upon the impact of using ICT, but that it is crucial for their future work and operation. The second main evidence confirms that elected management places much more importance on a systematic approach to the building of IS-Uni which includes project planning and is primarily based on organizing the existing processes and information flows, compared to permanent management that hardly notices these activities.

By observing this evidence resulting from a descriptive analysis on an integral basis, it is possible to establish that strategic thinking of the use of information sources is substantially more present in large universities compared to small universities, which implies the assumption that a large university's management is more inclined to the strategic planning of integrated business information systems. In addition, the above explanations indicate four key impacts: knowledge, planning, attitude and funding, which were tentatively assumed to affect the level of management's perception of the domain concerning the stra-

[18] Luić, Lj., *Strategic Planning of Integrated Business-Information Systems – Model Design Example of Higher Education*, Doctoral thesis, Faculty of Humanities and Social Sciences of the University of Zagreb, Croatia, 2009, p. 31

[19] Lleras, C., *Path analysis*, Encyclopedia of Social Measurement, Vol. 3, No. 1, 2005, pp. 25-30

[20] Luić, Lj., *Strategic Planning of Integrated Business-Information Systems – Model Design Example of Higher Education*, Doctoral thesis, Faculty of Faculty of Humanities and Social Sciences of the University of Zagreb, Croatia, 2009, pp. 166-169

tegic planning of an integrated business information system, as demonstrated by using scientific methods.[21]

A Comparison with Evidence Obtained in Similar Research

Ljerka Luić vs. Allen David (1995) *Information Systems Strategy Formation in Higher Education Institutions*. Information Research, 1(1) Both research papers confirmed the existence of a certain degree of subject's awareness of the need for strategic planning of information resources and making them the center of interest for the institutions represented by them. Compared to Allen's research, the qualitative contribution of the research evidence obtained by the co-author of this paper is reflected in the quantitative assessment of the degree of awareness of the need for strategic planning of integrated information systems, which in relation to the determining whether or not it exists reflects a certain contribution to the development of said research.

Ljerka Luić vs. Francis Greene, Brendan Loughridge i Tom Wilson (1996) *The Management Information Needs of Academic Heads of Departments in Universities: a Critical Success Factors Approach*. British Library Research and Development Department. A research paper by Greene, Loughridge and Wilson contemplating the information needs of department heads provided evidence that the balance between the internal and external critical success factors for a university depends on four key factors: the political & economic environment, the institutional environment, the relative position of the department, and the culture of the department. If the strategic planning of an integrated business information system is perceived as part of university management's needs for management information, the evidence obtained within the research conducted by the co-author of this paper may be compared with the abovementioned evidence in the context of achieving a balance between different management groups, which in relation to the comparable research also depends on four key influential factors which are, however, very different. In this research paper, they are: knowledge, funding, attitude and planning.

Ljerka Luić vs. Marcella Rita i Knox Karl (2004) *Systems for the management of information in a university context: an investigation of user need*. Information Research. Marcella and Knox investigated the efficiency of the existing data and information systems supporting the work of management and academic, research and administrative staff of higher education institutions. The evidence obtained by them indicate the fact that actual flaws exist in the implementation of the analyzed university's information strategy and that such flaws need to be dealt with as part of the development which needs to focus on improving the future strategic efficiency. The solutions proposed for improving the existing information systems indicate the fact that senior executives are aware that the main priority in the implementation of such improvements is the

[21] Ibid, pp. 136-138

development of a single integrated system that is flexible enough to meet specific demands while ensuring maximum accuracy and reliability of system data which needs to be quick, user-friendly and designed to support the actual operating functions, in particular those of administrative staff.
Similar evidence was obtained by the co-author of this paper. Namely, her results indicate the fact that actual flaws exist in the university management's strategic thinking regarding the development and implementation of an integrated business information system. Both research papers show a clear attitude of management with respect to the development of a single integrated system flexible enough to meet specific demands. The fact that the research studies were conducted independently, with a delay and in different cultural and business environments additionally confirms the need to contemplate and use SP-IPIS-Uni.[22]

Recommendations for Further Research
An extension of the *Strategic Planning of an Integrated Information System* is justified in the context of contributing scientific evidence such as discovering new algorithms and variables that may be used to describe more precisely the changes occurring within each construct of the model and their interactions, but also in the context of its actual use as the efficiency of using new technologies in the operation of a university lies in the synergic use of strategic planning nucleus particles in university management's strategic thinking.

Conclusion

As a result of the implication of the mutual impacts between its main particles (Information Strategy, Info Management Strategy, Info Technology Strategy and Change Management Strategy), the IS strategic planning nucleus affects the overall information environment of a university: academic, business and organizational. In this process, nucleus particles are almost inseparable from each other and their mutual link is of great importance for the strategic planning of information systems and their entire lifecycles. [23]
An analysis of the research results provided evidence indicating the fact that, as regards the strategic planning of an integrated university information system, management of a higher education institution expresses the greatest interest in the domain concerning the alignment of the business and academic information systems. In addition, the results obtained indicate the fact that the opinions of elected executives are much more represented in decisions concerning the informatization domain compared to permanent executives, however, neither group recognizes IT professionals as adequate collocutors on the subject of an

[22] Ibid, pp. 170-172
[23] Fudurić, A, *The nucleus of the business and information strategy*, Graduate thesis, Karlovac University of Applied Sciences, 2014, p. 64

integrated business information system, which foreseeably also excludes the participants in the process of the strategic planning thereof. [24]

Acknowledgments

The publication of this paper was supported by the B4B, to which the authors of this article are grateful for its substantial support.

References

Burrell, Gibson; Morgan, Gareth. Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life. Farnham: Ashgate Publishing, 2007

Fudurić, Ana. The nucleus of the business and information strategy, *Graduate thesis*, Karlovac: Karlovac University of Applied Sciences, 2014

Galliers, Robert; Leidner, Dorothy. Strategic Information Management: Challenges and Strategies in Managing Information Systems. Oxford: Elsevier Butterworth-Heinemann, 2003,

Han, P., Jones, Matthew. The dialectics of Information Systems. *Research Paper.* Cambridge: University of Cambridge, Judge Institute of Management Studies, 1993

Iivari, Juhani. A paradigmatic analysis of contemporary schools of IS development // *European Journal of Information Systems*, volume 1 (1991), number 4; pp. 249-272

Klein, Heinz; Lyytinen, Kalle. The Poverty of Scientism in Information Systems*,* In Allen, David. Information systems strategy formation in Higher Education Institutions// *Information Research*, 1(1) Available at: http://InformationR.net/ir/1-1/paper3.html (7 September 2015)

Lleras, Christy. Path analysis // *Encyclopedia of Social Measurement*, volume 3 (2005), number 1, pp. 25-30

Luić, Ljerka. Strategic Planning of Integrated Business-Information Systems – Model Design Example of Higher Education, *Doctoral thesis*. Zagreb: Faculty of Humanities and Social Sciences of the University of Zagreb, 2009

Nelson, Kay, Cooprider, Jay. The contribution of shared knowledge to IT Group performance // *MIS Quarterly,* volume 20 (1996) number 4, pp. 409-432

Orlikowski, Wanda. The duality of technology: rethinking the concept of technology in organisations // *Organisation Science*, volume 3 (1992), number 3, pp. 398-427

Reich, Blaize Horner, Benbasat, Izak. Measuring the Information Systems - Business Strategy Relationship. In: Galliers, Robert; Leidner, Dorothy (eds.) Strategic Information Management: Challenges and Strategies in Managing Information Systems. Oxford: Elsevier Butterworth-Heinemann, 2003, pp. 265-310

Reich, Blaize Horner; Benbasat, Izak. Measuring the linkage between business and information technology objectives *// MIS Quarterly*, volume 20 (1996) number 1, pp. 55-81

Subramani, Mani; Henderson, John; Cooprider, Jay. Linking IS-user partnerships to IS performance: a socio-cognitive perspective, *MISRC Working Paper Series WP 99-01*. Boston: Systems Research Center at Boston University, 1999

Sumsion, John. Survey of resources and uses in Higher Education Libraries. Loughborough, Library and Information Statistics Unit, Loughborough University of Technology, 1994

Taylor-Cummings, Andrea. Bridging the user-IS Gap, A study of major information systems projects // *Journal of Information Technology*, volume 13 (1998), pp. 29-54

Wood-Harper, Trevor. Research methods in information systems: using action research. In Mumford, E., Hirschheim, R., Fitzgerald, G. and Wood-Harper, T. (eds.) Research Methods in Information Systems. Amsterdam: Elsever Science Publishers B.V North Holland, 1985

Yin, Robert. Case Study Research: Design and Methods, London: SAGE Publications, 2003

[24] Luić, Lj., *Strategic Planning of Integrated Business-Information Systems – Model Design Example of Higher Education*, Doctoral thesis, Faculty of Faculty of Humanities and Social Sciences of the University of Zagreb, 2009, p. 176

Patterns-based Information Systems Organization

Sergej Lugović
Polytechnic of Zagreb
Vrbik 8, Zagreb, Croatia
slugovic@tvz.hr

Ivan Dunđer
Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
idundjer@ffzg.hr

Marko Horvat
Polytechnic of Zagreb
Vrbik 8, Zagreb, Croatia
mhorvat1@tvz.hr

Summary

The socio-technical systems research paradigm is about the complexity of real situations. It confronts us with the quest for variables that could provide us with insight into the behavior of such systems. Their behavior emerges according to internal system properties and adaptation of the system to external conditions.

In our view, behavioral patterns are one of those particular variables since machines can recognize them and their dynamics. Based on the synthesis of three different theoretical frameworks, this paper proposes a concept of patterns-based information system organization. The authors built the concept on the Deacon discussion of theory of information, Hofkirchner's unified information theory and related system behavior, and Kelso's explanation of pattern creation processes in self-organizing systems. All three researchers have included patterns in their theoretical proposal.

According to this analysis of the existing theories and their synthesis, we conclude that in order to design machines that can automatically support new behavior, we have to analyze humans and machines as a complex whole with dynamic relationships and emerging patterns as a dependent variable of behavior. By developing this theoretical concept, we establish a departure point for future research and search for different variables that correlate with pattern formation.

Keywords: information, patterns, self-organizing, information behavior, information systems

A pattern is a message, and may be transmitted as a message. (Wiener, 1950)
The transformation of noise back into signal is part of the game. (Cohen, 2006)
Messages can be studied according to their form, content, goal, producers, and recipients. (Capurro, 2003)

Introduction

If patterns are messages and could transform as messages, first they have to be recognizable. In a recent massive scale experiment of Facebook users, that's exactly what happened (Kramer et al 2014). The pattern of user behavior changed according to the amount of positive or negative content in their feed. When users were exposed to more positive content, users posted more positive status updates. In this experiment, researchers manipulated the users' content and measured the amount of positive or negative words in their status updates. So by recognizing patterns of agent behavior and by intervening correlated variables, we can transmit patterns of behavior. It's interesting to look where user behavior comes from. Because the agent (FB user) didn't know about the intervention, it's fair to say this behavior is self-organized and emerged from the internal agent behavior and not from the surrounding environment.

According to (Camazine, 2003, p.7), in self-organizing biological systems, patterns emerge from internal interactions using only information stored locally, while in systems lacking self-organization, a supervisor, directives, or already existing patterns in the environment can impose order. Is this human(s) who changed emotional behavior self-organizing or lacking self-organization? In our view, it's the position of the observer observing phenomena. If we look it from outside the system, we can recognize reference to the environment and feedback, but if we look at it from inside (from the position of the user) not knowing about intervention, then we could recognize reference to internal interactions and information stored in the user cognitive system.

If we're analyzing dynamic patterns formation, there are two problems we have to deal with. One is how very large material components construct the pattern, and the other relates to how given patterns are constructed to accommodate different circumstances. The former could be "the problem of complexity of substance" and the latest "the problem of pattern complexity" (Kelso, 1997, p. 5). Or how Batson (1972, p. 20) writes, "[M]ental process, ideas, communication, organization, differentiation, pattern, and so on, are matters of form rather than substance." Pattern as a phenomenon to observe could provide us with a lens to analyze how substance is formed. And it's in the domain of information science since it's concerned with the form and organization of information and the underlying structures, a fundamental methodological stance Bates (1999) described as socio-technical.

There are two modes of apprehending our experience's structure, state description and process description. The former characterizes the world as sensed and later characterizes the word as acted upon. "The organism must develop correlations between goals in the sensed world and actions in the world of process"

(Simon, 1962). The former provides criteria for identifying objects or process of relating dynamic pattern formations with circumstances, while it later provides the means for producing objects or describes how a dynamic pattern of behavior is constructed.

In the Facebook research, we could see limitations of analyzing social agents and technology agents separately. Do we observe the social agent or tools he uses? The research on socio-technical systems is dealing with the complexity of real situations rather than analyzing separated aspects (Ropohl, 1999). When a system is complex and dynamic, to understand the system, it's not enough to observe its parts because knowing properties of each doesn't give complete information about the system. Therefore, we must treat such a system as a whole. To understand the principles of such a system, feedback, or interaction of the parts isn't enough. What is important is the assumption that such a dynamic system has complex behaviors and that those behaviors can be goal-seeking forming dynamic patterns (Ashby, 1956, p. 55).

In our view, by analyzing patterns, which relate to complex behavior and dynamic interactions in a system composed of social and technical agents, we can get insight into how they emerge according to the reflection of the environment and internal properties of such a system. So, patterns become dependent variables of the socio-technical system behavior, including internal properties and environment states. By understanding how behavior is constructed according to the environment and internal norms stored in a system, we can comprehend the complexity of real situations.

This aligns with the socio-technical practice in information science, which calls for research efforts to analyze how communities use tools to access documents and create knowledge (Tuominen et al., 2005). We believe that in this approach we overcome the divide underlined in two cybernetics approaches, one of the first order and another of the second order. According to Geyer (1995), there are four main differences. The first one is dealing with the observer and the system where in the second-order cybernetics relationship between the observer and the system should be included in the analysis. The second one is that the second-order cybernetics is about constantly changing properties of the system, not its stability. In other words, it's about self-organization that isn't about forecasting but about understanding. The third is about the second-order cybernetics recognizing that all living organisms are self-steering and that their behavior can be steered from outside up to certain limits. The fourth is recognizing the continuous emergence of complexity resulting in new behavior and new interaction with the environment.

By observing patterns, we can take a step back from the system and observer relationship, and we can observe patterns in stable processes and in the changing ones. Different patterns will depend on steering from the outside and/or on internal system capabilities. By applying independent and objective machine observation, we can precisely observe the dynamics of those pattern emer-

gences. What's important, in our view, is that patterns can be easily described by machine quantitatively, making them different from behavior, which is qualitative and hard to recognize by machines.
For the purpose of further discussion and in regard to conceptual development, we define patterns as a time sequence happening between the moment when an agent experiences information (i.e., when structured data become information) and the information-searching and seeking process starts until the moment when the agent stops interaction with the structured data as his information needs are satisfied. Such a process can occur in social and technical domains. This is a working definition describing the scope of the analysis.

Self-organization and socio-technical systems

Can we call socio-technical systems self-organizing? Humans live in the technotope, surrounded by technologies and material context, including technical artifacts, which are not only neutral, but they shape our perceptions, behavioral patterns, and activities (Geels, 2004). As the information society develops in terms of blurring the difference between online and offline, we're becoming inforgs[1], socially connected information organisms, which will compete with other artificial inforgs. And this isn't happening because of the transformation of our bodies, but through re-ontologization (a radical form of reengineering) of our environment (Floridi, 2009).
So, to respond to changes in the environment we have to design and use technologies to adapt better, and by doing so, we have to move from our point of observing social or technical entities separately to observing socio-technical systems. Socio-technical systems are complex and behave in the new environment by developing new properties based on internal information processing capabilities. Such systems are self-organized since they don't have an internal agent doing the organization, but they form dynamic patterns of its behavior (Kelso 1997, p. 8). There's no agent defining the exact pattern of how author uses this machine while writing this text, but the usage pattern emerges according to cognitive, affective, social, and technical variables relationships.
We can discuss four specific properties of complex systems in terms of socio-technical systems. The first is non-determinism since it's impossible to fully understand the system's behavior by knowing only its constituting functions. The second is limited functional decomposability because it's a dynamic system, and it isn't possible to study the system's properties by decomposing it into functional parts. Its interaction with the environment and the self-organizing property allow it to functionally restructure itself. The third is the distributed nature of information and representation; some of the system functions cannot

[1] Inforgs can be part of a hybrid agent that is, for example, a family with digital devices such as digital cameras, cell phones, tablets, and laptops. (Floridi, L. Ed. (2010). The Cambridge handbook of information and computer ethics. Cambridge University Press).

be precisely localized. This is the cause of socio-technical system properties comparable to distributed systems. The fourth is the self-organization related to the emergent properties that we can't comprehend by understanding the components (Pavard and Dugdale, 2006).

We cannot know how a human and a mobile phone will behave by knowing only the phone's functions. Also, we can't understand what functions a user can execute on a mobile phone by knowing only the phone's functions. By adding more users with their mobile phones, a bigger system emerges, which is more difficult to understand regarding its structure, functions, behavior, and goals. Recent examples of self-organization of socio-technical systems based on users and mobile phones include the Arab Spring movements or in the rising popularity of mobile banking in Africa.

Another example of self-organizing socio-technical systems is the rise of the popularity of Bitcoin. Such a self-organization has many technical capabilities to process data, but a socio-cognitive and affective aspect will give meaning to that data and affect how patterns will form. Also, such a self-organization must expand in terms of selection and understanding how such "self-ordered properties permit, enable and limit the efficacy of natural selection" (Kauffman, 1993, p. xiv). By understanding selection along with self-organization, we have better insight into the process of new function emergence and what the correlations with self-organizing properties are.

Also, it's important to make a distinction between determinism and causality. When systems are far from an equilibrium and far from a deterministic clockwork state, they start to show self-organizing capabilities by following three propositions. The first one is there is no transformation mechanism that unambiguously turns cause to effect. Cause and effect relations allow a different cause to have the same effect and the same cause to have a different effect. The second one is that little changes in the cause may lead to a big change in effect. The third is that the more complex a system is the less probable is the return of the certain state in the future (Hofkirchner, 1998). Or to cite Hofkirchner (2011), "self-organization may be looked upon as the way evolutionary systems come into existence or change their structure, state or behavior and the way they maintain themselves (their structure, state or behavior)."

As stated by Kelso (1997, p. 32), patterns arise spontaneously as the result of interacting components, and the nature of such interaction is nonlinear and in such systems the whole is more and/or different from the sum of its parts. We can trace these ideas back to Immanuel Kant who pointed out that on one hand, organisms will always resist complete mechanistic and mathematical explanations since they have certain goal directedness (purpose), which humans can describe only in functional terms. On the other hand, biology as a science requires a mechanistic approach using forward causal explanation that needs to be expressed in mathematical terms. Kant stressed that the necessity of this dual ap-

proach for studying organisms is not a property of nature but rather our limited faculties (Roth, 2011).

If we're about to design machines to help us process or compute information, this is a very important point. If we're trying to describe machines by these functional limitations, then machines by themselves will have functional disabilities. And if such machines are essential parts of the self-organizing socio-technical systems, then, this dysfunction is a dysfunction of the socio-technical systems. If we could better describe the reality of nature, we could design machines that are more intelligent, and such machines would be part of the socio-technical systems making them more intelligent too.

In our view, the social "half" of such a system is extremely important because the self-steering capacity of humans provides essential non-deterministic properties of such a system built on social and technical components. So, to have a more intelligent system, we have to overcome the problem of dualism in explaining nature.

Can we overcome this problem of dualism in explaining nature? We think that by observing dynamics of patterns as dependent variables of socio-technical systems, which emerge in self-organizing processes of such systems, we could open up new perspectives in different disciplines in information and computer science. The emphasis is on the dynamic patterns emergence, not static patterns. As they emerge, they aren't reducible to a substance from which they emerge because they have their own properties. These properties relate to the purpose of the higher order level, while the mechanics of their formations are at the level below (Wilson, 2002). And they emerge as a result of the process of organizing the observed system's internal properties and as a reflection of the environment's circumstances.

So, patterns as phenomena observed in the behavior of self-organizing socio-technical systems can lead us to know about how these systems behave, including mechanical principles and principles related to the purpose of such systems. Patterns, as phenomena that correlate a mechanical view and a purpose view, can present behavior that's, we believe, qualitative and can create a mechanical structure that's quantitative. With development of computational technologies, we can observe such patterns emergence processes with extensive support of the machines, obtaining objective insights about how they form and what their dynamic is, according to internal or external factors influencing their formation.

So, we could use machines to help us observe dynamic pattern emergence, use them as a dependent variable in the research of socio-technical systems. By using those findings, we could design better machines conditioned by social self-steering forces based on information needs. Such a force is implemented through the feedback, creating dynamic adaptation of those machines in a dialectic process (thesis, antithesis, and synthesis).

In addition to the above-mentioned authors, we can find further socio-technical systems concept elaboration in the work of one of the founders of this paradigm Trist (1981) and a PhD thesis written by Shields (2007) analyzing different theories and practices used to study technology systems. An important contribution to the socio-technical system paradigm are the books *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (Bijker et al., 1981) and *Science in Action: How to Follow Scientists and Engineers Through Society* (Latour, 1987). Current research approaches the socio-technical system from the evolutionary perspective and is in the work of Heylighen (2007).

Information, Systems, Organization, and Patterns

In this part of the paper, we select and propose existing theoretical frameworks that support our views on patterns-based information system organization. We combined three existing theoretical frameworks related to socio-technical systems and information, offering theoretical fundamentals for proposal of a new synthesis, one that will support the design of machines that could automatically support new behavior of the socio-technical systems.

Information and patterns

Deacon (2010) discussed the information from the semiotics perspective, pointing out that information could refer to the sign or signal functioning without any meaning or reference, such as information content in bits. Such information is syntactic information. Information could also represent meaning of the bits from syntactic information. Such information could be semantic information. Information also could refer to the aspects of those ideas, which we call pragmatic information (Deacon, 2010). Syntactic information's basis is the Shannon theory of information; the semantic is from the Boltzmann and Shannon entropy, while the pragmatic one expands it with the theory of evolution.

In the same work, Deacon pointed out that "both the patterns that we deliberately create in order to convey an idea and those we discover in nature can convey information. Ultimately, this demonstrates that almost anything can qualify as information in the syntactic sense, because this is only an assessment of the potential to inform." And such a basic notion of information corresponds to the Shannon theory of information. So, a pattern could be looked as a sign, which when recognized, could provide a basic component of the message transferred through the communication channel from the receiver to the sender. Here, we see that a pattern could be looked as a fundamental property of the information. To remove uncertainty we have to recognize the pattern and send it as a message to the receiver. Those three types of information are in Table 1 (adapted from Dodig-Crnkovic, 2012).

Table 1. Deacon's three types of information

Type of information	Underlying theories	Focus on	Semiotics perspectives
Information 1	Shannon	Data, pattern, signal, data communications	Syntax (what it exhibits)
Information 2	Shannon + Boltzmann	Intentionality, "aboutness," reference, representation, relation to object or referent	Semantics (what it conveys)
Information 3	Shannon + Boltzmann + Evolution	Function, interpretation, use, pragmatic consequence	Pragmatics (what it is for)

Systems and patterns

Hofkirchner (1999) used semiotics as a basis for discussion about systems and information behavior, pointing to the concept of Peirce's triad of syntactic, semantic, and paradigmatic sign aspects. The information process consists of drawing distinctions in terms of system organizing based on those distinctions, whether it's in the new system that emerges or the novelty in the existing system.

Table 2. Three different types of system formation (based on Hofkirchner, 1999)

Type of system	Evolution of sign processes	System formations: from patterns to new goals
Self-restructuring systems	Reflective pattern formation in dissipative systems – emergence of signs	Forming patterns is the way self-restructuring systems reflect some changes in the conditions in the system's environment.
Self-reproducing systems	Intelligent symbolization in autopoietic dissipative systems – emergence of symbols	Self-reproduction requires structures to functionalize for survival. Functionalized structures are not plain patterns any more but something that contains meaning.
Self-determining systems	Conscious formation of ideas in recreative autopoietic systems – emergence of goals (ideas)	Systems that do not merely (re)produce themselves and strive for survival, but in doing so, seek additional goals, which they're committed to and which they chose on their own. There are three steps of idea formation: 1. Perception of signals from outside the system resulting in modification of the system structure; 2. Interpretation of the perceptions by which a system is modified; 3. Evaluation of the interpretations modifying a system by affecting behavior.

In realizing those novelties, a distinction is drawn and information produced. So, there exists a relationship between the old and the new sign; the new sign isn't reducible to the old sign but is dependent on it. Since we have a relationship of signs, we can refer to this as syntactic relations of signs. As a new state of the system also relates to the system's environment, a new relationship develops between the sign and the environment, and the sign gains significance,

forming a semantic relation of signs (having meaning). As the sign gains significance, it can also relate to the aims of the system. We can call such a relationship a pragmatic relationship of signs (Hofkirchner, 1999). Such an analysis of different sign relations in self-organizing systems could give insight into how systems behave.
In Table 2, we present a brief summary of different system properties, the evolution of sign processes, and how a system forms itself in the process of self-organization.

Organization and patterns
Kelso proposes The Tripartite Scheme with the aim to theoretically summarize laws constituting the order parameter dynamics. The idea proposed is straightforward, proposing that understanding any level of organization starts with knowledge about three main things: parameters acting on the system, interacting elements of the system, and emerging patterns. There's also a clear point that this model is not rigid in terms of micro and macro (Kelso, 1997, p. 18). We present the Tripartite Scheme in Table 3.

Table 3. The Tripartite Scheme (based on Kelso, 1997)

Knowledge about	Description
Cooperativities	The emerging patterns or modes to which they give rise
Set of primitives	The interacting elements themselves
Boundary conditions	The parameters acting on the system

Synthesis
We could clearly see that three theoretical models presented above have patterns in common. Deacon's model is dealing with information, the Hofkirchner's with systems, and the Kelso's with organization. We could synthesize those theoretical proposals into the concept of **patterns-based information systems organization.** Such a concept could be presented through three layers of self-organisaion:

- Structure level – the most simple biological self-organizing system (patterns formation according to internal properties) or simple self-restructuring mechanical systems (such as a PC that self-restructures data on a hard disk) making it applicable to socio-technical system analysis. Such a system is self-restructuring when the patterns emergence forms new structures, information explained in terms of the Shannon information theory and it organized around emerging dynamic patterns, the essence of cooperation.
- Function level – the autopoietic biological self-organizing system, which can reproduce itself, and to the intelligent mechanical system, which can reproduce its components to maintain a predefined goals execution process, making it applicable for socio-technical system analysis. Such a

system is self-reproducing in terms of functional structures; information is defined in terms of Shannon and Boltzmann entropy, which explains intentionality, "aboutness," and reference to the information in the communication channel. And it organizes around interactions of elements for which prerequisite are formations of cooperatives.

- Goal level – the conscious socio-technical system, which can intervene in the environment and create new goals, according to its ideas. Such a system is self-determining, when perception determines new goals, interpretation of those perceptions, and evaluation of the interpretations results in the new behavior. Information is defined in terms of evolutionary principles describing use, functions, and pragmatic consequences. And it's organized according to the boundary conditions producing parameters acting on the system.

We can see that the last application of patterns-based information systems organization cannot separate into biological and mechanical systems. This leads us to the proposition that if we want to have machines to support new behavior of social structures by developing new structures, functions and new goals, we cannot analyze machines and humans separately, but as a complex whole with dynamic relationships and emerging patterns as a dependent variable of behavior.

Another condition to achieve such properties of the system is that the system should not have one central agent to control this process because it's self-regulated. It's also important to underline ethical implications related to such a process as an important dimension in further research.

If we want to gain the insight into how the system behavior emerges and how such a system performs in reshaping the environment according to its goals, what is observable are the dynamic patterns, with an emphasis on the dynamic. Therefore, if we take a point of observation from outside the system, we can see its performance, but we cannot gain insight into the emergence of particular behavior, which caused such a performance. By observing the system from the inside, we can understand the system properties and the emergence of behavior, but we can't see the result of such behavior.

Stepping back and observing the process of patterns formation as a dependent variable of system behavior and the system environment, we can learn how they emerge and how they relate to system performance. And such a position of observation is objective and considers the internal system behavior and reflection to the system environment.

Hjørland pointed out the "tendency to try to measure the users' information need by questioning them or by studying their behavior, seems to be mistaken. What information is needed to solve a given problem is not primarily a psychological question, but a theoretical/philosophical one" (Hjørland, 2002).

In our view, the concept of patterns-based information systems organization, is overcoming such pitfalls since patterns are more objective and quantitative then psychological, theoretical, or philosophical questions.

Conclusion

In this paper, we provided an overview and analysis of some of the existing theories about self-organizing systems, how they interpret information and how they organize it. It's important to state that in this work we did not focus exclusively on the fields of patterns recognition and analysis from the perspective of computer science. Our analysis aims at searching for the explanations of dynamic patterns formation in different types of self-organizing, socio-technical systems from the perspective of information science.

Consequently, we synthesized theoretical frameworks into the concept of patterns-based information systems organization. We select those theoretical frameworks in terms that all three of them address patterns. We did not base selection on systematic literature reviews but with the aim to support concept construction. Such a conceptualization provides the point of departure toward future research on additional variables, which correlate with the patterns as behavior-dependent variables of socio-technical systems.

As a part of future research, our next task is to investigate existing empirical studies concerning patterns in socio-technical systems, including theoretical foundations. We believe that an aggregate investigation of the information behavior, information systems with pattern recognition, and application of cybernetics principles of feedback and self-organization could lead to new, encouraging and interdisciplinary research perspectives.

References

Ashby, W. R. (1956). *An introduction to cybernetics* (Vol. 2). London: Chapman & Hall.

Bates, M. J. (1999). The invisible substrate of information science. *Journal of the American Society for Information Science*, 50(12), (1999). 1043-1050.

Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. University of Chicago Press.

Bijker, W. E., Hughes, T. P., Pinch, T., & Douglas, D. G. (2012). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT press.

Camazine, S. (2003). *Self-organization in biological systems*. Princeton University Press.

Capurro, R. (2003). Angeletics – A Message Theory. Hierarchies of Communication. Karlsruhe: ZKM-Center for Art and Media, 58-71.

Cohen, I. R. (2006). Informational landscapes in art, science, and evolution. Bulletin of Mathematical Biology, 68(5), 1213-1229.

Deacon, T. W. (2010). What is missing from theories of information? *Information and the Nature of Reality*, 146, chapter 8.

Dodig-Crnkovic, G. (2012). Physical computation as dynamics of form that glues everything together. *Information*, 3(2), 204-218, citing Deacon, T. W. (2011). *Incomplete nature: How mind emerged from matter*. WW Norton & Co.

Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research policy*, 33(6), 897-920.

Geyer, F. (1995). The challenge of sociocybernetics. *Kybernetes: The International Journal of Systems & Cybernetics*, 24(4), 6-32.

Floridi, L. (2009). Web 2.0 vs. the semantic web: A philosophical assessment. *Episteme*, 6(01), 25-37.

Heylighen F. (2007): Accelerating Socio-Technological Evolution: from ephemeralization and stigmergy to the global brain, *Globalization as an evolutionary process: modeling global change*, eds. Modelski, G., D. Tessaleno, & Thompson, W. London: Routledge.

Hjørland, B. (2002). Domain analysis in information science: eleven approaches–traditional as well as innovative. *Journal of documentation*, 58(4), 422-462.

Hofkirchner, W. (1998). Emergence and the logic of explanation. An argument for the unity of science. Acta *Polytechnica Scandinavica*, Mathematics, Computing, and Management in Engineering Series, 91, 23-30.

Hofkirchner, W. (1999). Towards a unified theory of information – The merging of second-order cybernetics and semiotics into a single and comprehensive information science. In 15e *Congrès International de Cybernétique*, Namur 1998, Namur 1999, 175-180.

Hofkirchner, W. (2011). Does computing embrace self-organisation. *Information and computation*: *Essays on scientific and philosophical understanding of foundations of information and computation*. Vol. 2. (edited by Gordana Dodig-Crnkovic and Mark Burgin).

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press.

Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior*. MIT press, Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Pavard, B. and Dugdale, J. (2006). The contribution of complexity theory to the study of socio-technical cooperative systems. *Unifying Themes in Complex Systems*. Springer Berlin, Heidelberg.

Ropohl, G. (1999). Philosophy of socio-technical systems. *Techné: Research in Philosophy and Technology*, 4(3), 186-194.

Roth, S. (2011). Mathematics and biology: a Kantian view on the history of pattern formation theory. *Development genes and evolution*, 221(5-6), 255-279.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6), 467-482.

Shields, W. M. (2007). Theory and practice in the study of technological systems. Doctoral dissertation, Retrieved from Virginia Polytechnic Institute and State University.

Trist, E. (1981). The evolution of socio-technical systems. *Ontario Quality of Work Life Centre*, (2), Ontario, Canada.

Tuominen, K., Savolainen, R. and Talja, S. (2005). Information literacy as a sociotechnical practice, *The Library Quarterly* 75(3), p329-345.

Wiener, N. (1988). The Human Use of Human Beings: Cybernetics and Society (No. 320). Da Capo Press.

Wilson, T. (2002). "Information science" and research methods. Knižnicná a informacná veda [Library and Information Science], XIX, 63-71. (Published by the Department of Library and Information Science, Comenius University, Bratislava, Slovak Republic.

Modelling states of a computing system aware of an aspect of context

Krunoslav Peter
Andrija Stampar Teaching Institute of Public Health
Mirogojska 16, Zagreb, Croatia
kruno.peter@stampar.hr

Summary

The paper describes an approach to the modelling adaptive behaviour of a computing system aware of a single aspect of context. It senses states of an aspect of context and changes its state accordingly. Its adaptation to context can be modelled by using a finite-state automaton. It is possible to implement its decision-making subsystem by translating the state-transition table of the finite automaton to the set of decision rules of the subsystem.

Keywords: context, sensor, context-aware system, adaptable system, decision-making subsystem, finite-state automaton, state-transition table, decision rule

Introduction

Modern computing systems are "responsive to their contexts and environments" (Ornellas, 2014). For example, a system aware of light senses light conditions of its environment and adjusts its display's backlight for optimum visibility and lower power consumption. The system adapts to context automatically "without distracting from the user's task" (Chalmers, 2011, 3) and "reduces the burden of excessive user involvement" (Loke, 2007, 7). Any adaptation is a change of system's state. Such behaviour can be modelled by using a general computation model called the finite automaton or finite-state automaton. Additionally, the state-transition table of a finite automaton may be useful for implementation of the decision-making subsystem of a context-aware system. It can be translated to the set of decision rules of such subsystem.

The goal of this article is to present a simple and technology-dependent approach to the problem of implementing the thinking subsystem. Other approaches are mentioned in the article. They are more complex than rule-based approaches but effective in certain domains.

Context

Context of an adaptable system includes information such as network connectivity, communication costs, light conditions, location and time (Schilit et al, 1994). It is information about system's environment and system itself. Such a piece of information is called the *aspect of context* (Chalmers, 2011, 68).

"*Sensors*, biological or nonbiological, provide a means to acquire" such information (Loke, 2007, 15).
Some aspects of context are *discrete* (Loke, 2007, 20), for example, network connectivity. A system is connected or disconnected to a wireless network during a particular period of time. A connected system adapts to context by executing its firewall (it does not bother a user to execute "a strange program called Firewall"); a disconnected system will automatically stop and unload the firewall to free main memory. The states of network connectivity are "connected" and "disconnected"; they are also 1 (true) and 0 (false) – the Boolean values. Those states can be thought as elements of an *alphabet of states*. They change arbitrarily during the time. A sequence of those discrete states forms a *word* of some *regular language* (Abiteboul et al, 1995, 14), for example disconnected, connected and disconnected (010). It is a *formal language* that can be described by using a *regular expression* and accepted by a particular finite-state automaton. The previous language is specified by the regular expression *disconnected (connected disconnected)**.
Noise is an example of a *continuous* aspect of context (Loke, 2007, 20). A noise-aware system might set louder sound in a noisy environment. To implement such system, its developer might set the *thresholds* for sound pressure. For example, a noise-aware system senses "normal" and "loud" noise. Sound pressure in a quiet office is about 40 dB (Wikipedia, 2015); a system considers it "normal". Traffic on a busy street exceeds 80 dB (Wikipedia, 2015); it is sensed as "loud" noise. A noise-aware system recognizes a regular language of the "normal" state (< 80 dB) and the "loud" state (≥ 80 dB): *(normal + loud)**. They are the elements of an alphabet of noise states and they change arbitrarily.

Context-aware system

Besides main functionality, a *context-aware system* is able to utilize "contextual information about the physical world" (Loke, 2007, 5). It senses context and processes contextual information (Ferreira et al, 2015). Such system "reacts to changing context" (Schilit et al, 1994) and behaves adaptable; it does so automatically. Using of perceptual information about the environment distinguishes *context-aware computing* from traditional computing (Loke, 2007, 7; Chalmers, 2011, 67).
The principle of "separating the acquisition and representation of context from the use of context" is important for developing context-aware systems (Ferreira et al, 2015). Therefore, there are three additional functionalities of a context-aware system: *sensing*, *reasoning* and *acting* (Loke, 2007, 15). These functionalities can be realized in a centralized or a distributed architecture (Schilit et al, 1994; Loke, 2007, 15). According to these functionalities, an abstract architecture of context-aware system has three subsystems (Loke, 2007, 25): the *sensing subsystem*, the *thinking subsystem* and the *acting subsystem*.

The first subsystem acquires data about system's environment; the second subsystem performs reasoning about these data. Finally, the third subsystem performs an adaptation to the environment.
The thinking or decision-making subsystem "uses input from software and hardware sensors to decide how, when, and where to adapt the system" (McKinley et al, 2004). This subsystem might be a simple rule-based system or an arbitrary complex reasoning system (Loke, 2007, 23; McKinley et al, 2004). If the thinking subsystem is a rule-based system whose states changes according to the states of context, then it is possible to model and implement it like a finite-state automaton.
The paper presents two examples of general approaches to implementing software adaptation. Setting system's sound loudness in a noisy context is an example of *parameter adaptation*; it modifies system's parameters that determine adaptive behaviour. On the other hand, *compositional adaptation* exchanges system's components with others that improve system's adaptability to its context or add new behaviour (McKinley et al, 2004), e.g., to run the firewall of a connected system. Thinking subsystems for both software adaptations might be modelled by using finite automata.

Finite-state automata

As is previously stated, an appropriate model for a system aware of an aspect of context is the *finite-state automaton* (FSA). An FSA (Sipser, 2013, 35; Abiteboul et al, 1995, 13) is a 5-tuple $(S, \Sigma, \delta, s_0, F)$, where

S – a finite set of *states*;
Σ – an *alphabet*;
δ – the *state-transition function* $\delta: S \times \Sigma \rightarrow S$;
s_0 – the *start state*;
F – a set of *accepting states*.

The state-transition function is often represented as a *state-transition table* (Dovedan, 2003, 65).
The FSA M accepts the language $L(M)$; it is a *regular language*. It can be specified by writing an FSA accepting it (Abiteboul et al, 1995, 14) or by using the regular expression (Dovedan, 2003, 19).
"A regular expression over Σ is written using the symbols in Σ and the operations concatenation, * and +" (Abiteboul et al, 1995, 14). * stands for set of all words over Σ and + stands for union.

Specifying adaptive behaviour

Adaptive behaviour of a system aware of the aspect C of context can be specified like an FSA $(S, \Sigma, \delta, s_0, F)$:

S – a finite set of system's states;
Σ – a finite set of states of the aspect C of context;
δ – the state-transition function $\delta: S \times \Sigma \rightarrow S$;
s_0 – the initial state
F – a set of accepting states.

Here is an example of a simple context-aware system (Example 1):

Example 1
The system M_1 that is aware of noise conditions. It senses noise conditions every second; if noise is louder than a specified threshold, the system's sound will also be louder.
The system recognizes language $L(M_1)$ – an arbitrary sequence of conditions *normal* and *loud*[1]:

$(normal + loud)^*$

The system is defined as the FSA M_1 ($\{ volume_1, volume_2 \}, \{ normal, loud \}, \delta, \{ volume_1 \}, \{ volume_1, volume_2 \}$), with a state-transition table (Table 1)

Table 1: The state-transition table of the FSA M_1

δ	*normal*	*loud*
$volume_1$	$volume_1$	$volume_2$
$volume_2$	$volume_1$	$volume_2$

Example 2 shows a bit complicated system:

Example 2
A system M_2 is aware of noise conditions. It has three states. It is not allowed the change of system's state from the "normal" state to the loudest of all states. A user of system must not be shocked by the loudness change.
The system recognizes language L (M_2) – an arbitrary sequence of conditions *normal*, $loud_1$ and $loud_2$:

$(normal + loud_1 + loud_2)^*$

It is defined as the FSA M_2 ($\{ volume_1, volume_2, volume_3 \}, \{ normal, loud_1, loud_2 \}, \delta, \{ volume_1 \}, \{ volume_1, volume_2, volume_3 \}$), with a state-transition table (Table 2)

[1] One can use characters or numbers instead *normal* and *loud*.

Table 2: The state-transition table[2] of the FSA M_2

δ	***normal***	***loud*$_1$**	***loud*$_2$**
volume$_1$	*volume*$_1$	*volume*$_2$	*volume*$_2$
volume$_2$	*volume*$_1$	*volume*$_2$	*volume*$_3$
volume$_3$	*volume*$_2$	*volume*$_2$	*volume*$_3$

Implementing a system aware of single aspect of context

In the case of a continuous aspect of context, it is possible that the sensing subsystem delivers discrete values according to defined thresholds. For instance, if sound pressure exceeds 80 dB, the sensing subsystem delivers the "loud" value. The thinking subsystem may store this value in a *context variable*. Additionally, a value of the *state variable* represents system's state; it is an input of the acting subsystem. The state-transition table is implemented by the set of *if-then* rules (Schilit et al, 1994) of the thinking subsystem. The syntax for the decision rule ***if*** *(State and Context)* ***then*** *State* corresponds to the state-transition function $\delta: S \times \Sigma \rightarrow S$. Example 3 shows decisions rules for the thinking subsystem defined by the state-transition table of the FSA M_1.

Example 3
The state-transition table of the FSA M_1 might be directly translated to the following decision rules[3]:

Input:
state
context

if *(state == "volume*$_1$*" and context == "normal")* ***then*** *state = "volume*$_1$*".*
if *(state == "volume*$_1$*" and context == "loud")* ***then*** *state = "volume*$_2$*".*
if *(state == "volume*$_2$*" and context == "normal")* ***then*** *state = "volume*$_1$*".*
if *(state == "volume*$_2$*" and context == "loud")* ***then*** *state = "volume*$_2$*".*

Output:
state

Assigning a value to the state variable that is different from the previous value triggers the acting subsystem in this implementation of the noise-aware system.

In the system aware of many aspects of context, there could be the state-transition tables for any aspects of context.

[2] Shades of grey in the state-transition table visually emphasise distinct values.

[3] In the syntax of decision rules, == is the equality operator and = is the assignment operator.

Other approaches

Developers of the decision-making subsystems also apply "first-order logic-based formalisms to represent context and situations, and rules that map situations to required actions" (Loke, 2007, 23). There is an approach that is "inspired by biological processes, such as the human nervous system" (McKinley et al, 2004). Decision makers might even learn about and adapt to dynamical context and user behaviour (Schilit et al, 1994; McKinley et al, 2004).

Conclusion

Modelling adaptive behaviour of a general computing system aware of an aspect of context like a finite-state automaton helps a developer of a system to define context's and system's states. Adaptation to an aspect of context is specified in the form of the state-transition table of a finite automaton; this table might be translated to the set of decision rules of the thinking subsystem.

Literature

Abiteboul, Serge; Hull, Richard; Vianu, Victor. *Foundation of Databases*. Upper Saddle River: Addison-Wesley, 1995.

Chalmers, Dan. Sensing and Systems in Pervasive Computing: Engineering Context Aware Systems. London: Springer Verlag, 2011.

Dovedan, Zdravko. *Formalni jezici – sintaksna analiza*. Zagreb: Zavod za informacijske studije, 2003.

Ferreira, Denzil; Kostakos, Vassilis; Dey, Anind K. *AWARE: mobile context instrumentation framework*. 20 April 2015. http://journal.frontiersin.org/article/10.3389/fict.2015.00006/full (15 May 2015)

Loke, Seng. *Context-Aware Pervasive Systems: Architecture for a New Breed of Applications*. Boca Raton: Auerbach Publications, 2006.

McKinley, Philip; Sadjadi Seyed, Masoud; Kasten, Eric; Cheng, Betty. *Composing adaptive software*. // *IEEE Computer* 37 (7), 2004. p. 56-64.

Ornellas, Abigail. *2015 is about IoT, Smart Washing Machines and Context-Aware Technology*. 2 December 2014. http://info.groveis.com/blog/2015-is-about-iot-smart-washing-machines-and-context-aware-technology (26 May 2015)

Schilit, Bill N.; Adams, Norman; Want, Roy. *Context-Aware Computing Applications*. // *IEEE Workshop on Mobile Computing Systems and Applications*, December 8-9 1994.

Sipser, Michael. *Introduction to the Theory of Computation, 3rd Edition*. Boston: Cengage Learning, 2013.

Sound pressure. 15 May 2015. http://en.wikipedia.org/wiki/Sound_pressure (15 May 2015)

Preliminary communication

A project of developing a knowledge management system

Ante Martinić*, Petra Bago**, Nives Mikelić Preradović**,
Goran Glavaš***, Bojana Dalbelo Bašić***, Jan Šnajder***

*KONČAR – Power Plant and Electric Traction Engineering Inc.
Fallerovo šetalište 22, 10000 Zagreb, Croatia
**Faculty of Humanies and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
*** Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, 10000 Zagreb, Croatia
ante.martinic@koncar-ket.hr, {pbago, nmikelic}@ffzg.hr, {goran.glavas,
bojana.dalbelo, jan.snajder}@fer.hr

Summary

Knowledge is an essential element of modern business and increasing attention is given to its acquisition, distribution and exploitation in everyday business activities. Therefore, KONČAR launched the development of a knowledge management system for its own demands and initiated a collaboration with the academic community for scientific research purposes and potential broader social significance of the project. With regard to the multidisciplinary nature of knowledge management, an agreement was reached with the University of Zagreb, the Faculty of Humanities and Social Sciences and the Faculty of Electrical Engineering and Computing. The knowledge management system will enable an effective management of all segments of intellectual capital of an organization, resulting in increase in productivity and higher market competitiveness, as well as an increased capability for generating new values for all parties to the agreement.

Keywords: knowledge management system

The purpose of knowledge management

It is estimated that 80% of today's economic value comes from intangible resources (Petrick et al., 1999). The knowledge relates to its intangible resources, consisting of resources recorded on information storage media and knowledge possessed by the organizations' employees. The entire knowledge of an organization constitutes its intellectual capital.

Knowledge management can be defined as a set of activities practiced within an organization with a goal to identify, record and apply knowledge, thereby stimulating also the creation of new knowledge. There are two major forms of

knowledge: (1) explicit – the knowledge that is recorded on some permanent medium, and (2) tacit – the knowledge that can be found in people's minds (Dalikr, 2005). To enable an effective distribution of knowledge, it is necessary to convert as much tacit knowledge as possible into explicit knowledge.
The intellectual capital, i.e., the knowledge relevant to organizations, can be broken down to human, structural and relational capital (Pearse, 2007). The human capital is represented by knowledge and skills of organization's employees. The structural capital includes systems and processes necessary for an efficient performance of organization's goals. The relational capital is represented by connections between the organization and other business entities. The main elements of the intellectual capital are given in Figure 1.

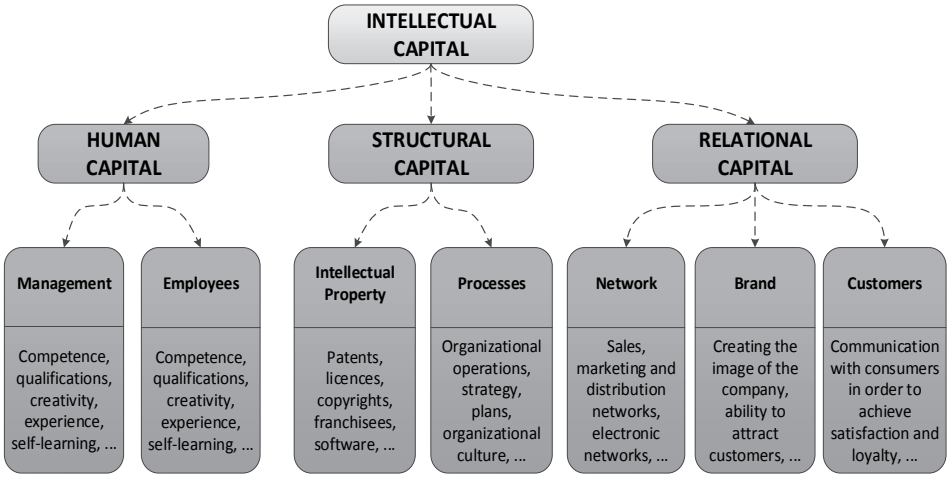


Figure 1. The structure of the intellectual capital (http://autopoiesis.foi.hr/wiki.php?name=KM+-+Tim+50&parent=NULL&page=Intelektualni%20kapital)

Knowledge management at KONČAR Group

The business activity of the KONČAR Group lies mostly in power generation, distribution and transmission of electrical energy, industry and electric traction. Founded in 1921, KONČAR Group nowadays consists of 18 dependent companies and 1 affiliated company, and has around 3,800 employees. The KONČAR Group builds its business on research activities and production of its products, alongside employing a number of highly educated employees and practicing an individual approach to clients. The companies of the KONČAR Group implement their projects using products that are primarily the result of organization's knowledge and development. In order to manage its own knowledge effectively, KONČAR initiated a project of development and implementation of a

knowledge management system. The primary objective of this project is to increase the productivity of the KONČAR Group, which can eventually lead to increased market competitiveness and contracting and more effective bidding and implementation of projects. To support a more successful implementation and application of modern technical and scientific achievements, the KONČAR Group initiated an agreement with the University of Zagreb and its constituents - the Faculty of Humanities and Social Sciences and the Faculty of Electrical Engineering and Computing. It is expected that current and future business demands will be met through joint activities, increasing the capability to generate new values for all parties.

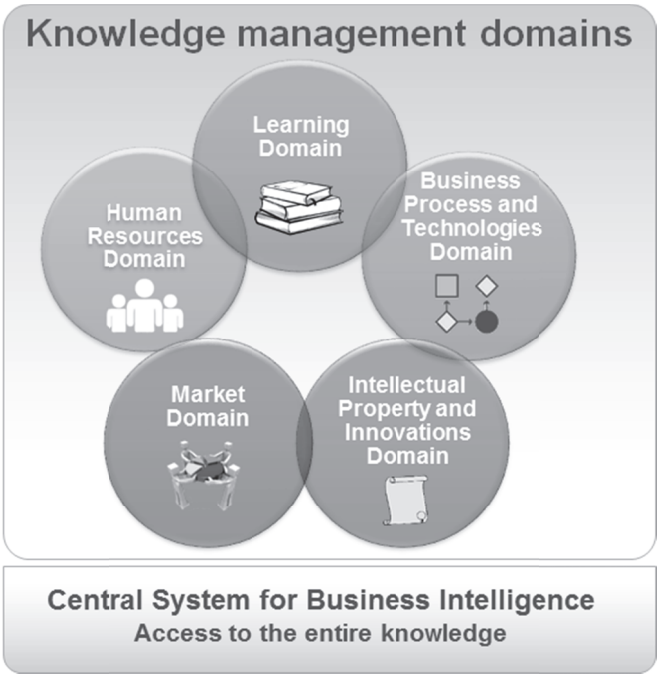


Figure 2. Knowledge management domains

Developing a knowledge management system for the KONČAR Group

The knowledge management system will encompass all three elements of the intellectual capital: (1) the human capital, (2) the structural capital, and (3) the relational capital. The elements of the intellectual capital of KONČAR's knowledge management system are divided into knowledge management areas as follows:

- Human Resources,
- Learning,
- Natural Language Processing,

- Business Process and Technologies,
- Intellectual Property and Innovations,
- Market.

In addition to developing knowledge management solutions for all of the abovementioned areas, the project includes construction of the Central System for Business Intelligence which will enable a single-point access to the entire knowledge of the company, i.e., to the knowledge from all six knowledge management areas mentioned above. Each knowledge area encompasses business processes that are implemented through related subprojects.

The relations between elements of the intellectual capital and knowledge management areas are shown in Figure 3.

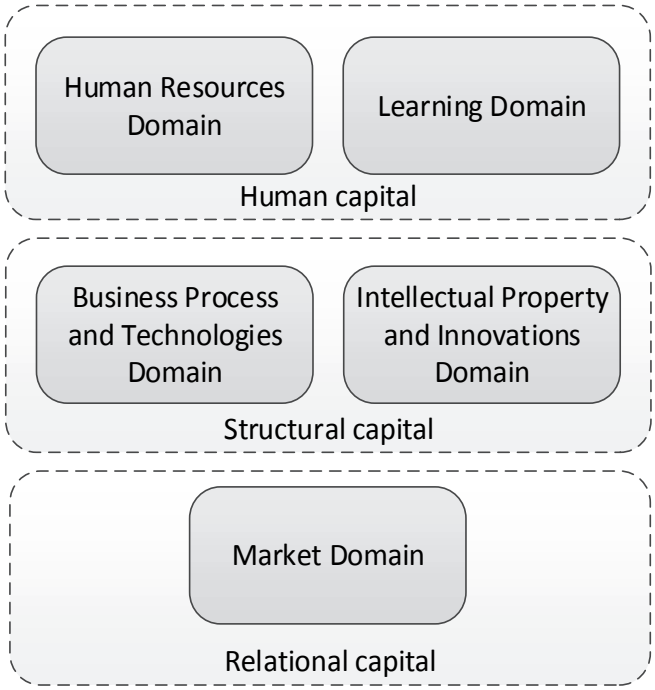


Figure 3. The relation of the elements of the intellectual capital to knowledge management areas

Considering the large scale of the knowledge management system, the iterative approach will be practiced for its development and implementation, which will reduce risks and achieve greater adoption by end users.

Business processes that will be introduced initially are those considered to increase productivity and result in storing a considerable amount of knowledge. Business processes with different stakeholders will be selected for simultaneous

introduction to various domains of knowledge management. That way the employee workload will be reduced, increasing the opportunity for their active participation, faster adoption of new business processes and related information systems.
In the next section, we examine parts of the project that illustrate the complexity of a large-scale knowledge management system such as the one envisioned by KONČAR. This is an open type project, which means that (apart from the elements listed in the following section), it can likewise include other topics, participants and subprojects.

Collaboration on the development of the knowledge management system
The scientific community participating in this project includes the University of Zagreb and its two constituents - the Faculty of Humanities and Social Sciences and the Faculty of Electrical Engineering and Computing. These two constituent units represent the leading research institutions in the Republic of Croatia in the areas of social sciences, humanities and technical sciences. These institutions participated in more than 150 international and almost 300 national research projects. The development of challenging and comprehensive company's knowledge management models may additionally position these research institutions as leaders in the higher education domain when it comes to knowledge management, which will further strengthen their status as leading research institutions in Croatia, as well as in the region.
One of the main tasks of the project includes automation of the following processes: collection, transformation and induction of knowledge. Knowledge is derived from information that is precisely formulated in advance, based on the data. The acquisition of existing knowledge is related to the automated (1) collection of information from structured and unstructured (primarily textual) data that the KONČAR Group and similar companies have in abundance, and (2) derivation of knowledge from information by automated reasoning (i.e. machine learning or expert systems). Furthermore, it is necessary to develop a central business intelligence system able to make conclusions based on the information incorporated from structured and unstructured resources.
KONČAR Group and companies of similar technical profile have abundance of unstructured textual data containing crucial business information. Thus, from the technical point of view, it is especially important to focus on the development of methods for efficient text analysis and information extraction from text. Tools for text analysis will significantly contribute to:

1. the management of human resources (automatic analysis of resumes and competencies; linking employees' competencies and projects' requirements);
2. the management of intellectual property and innovations (detecting patentable innovations based on the technical documentation; efficient search of patent databases);

3. the management of business processes and technology (computer-aided development of company's knowledge base);
4. the knowledge management of the market (automated analysis of tenders; analysis of competitor activity and extraction of relations between entities relevant for company's business activities).

The management of digital resumes of the employees and their competencies is one of the basic tasks of the human resource management. It is necessary to structure and standardize the dynamic digital resumes and automate the process of their loading into the central database. Additionally, we need to identify and analyze the preferences of the employees regarding the specific methods of acquiring knowledge (Coffield et al., 2004) in order to plan and organize the system of educational and training activities, and to develop a taxonomy of the employees' competencies (Jacobs and Washington, 2003). Analyzing employees' competencies and characteristics, it will be possible to optimize job positions, as well as find employees with an optimal profile for a specific job position.
In their study, Hilbert & López (2011) presented the world's technological capacity to store information on analog and digital media in 2007, estimating that 94% of the stored data was in digital form, while the analog form was represented by only 0.007%. We can assume this ratio in KONČAR Group to be more in favor of the analog form, given that official documents (although mostly digitally born) still need to be printed. However, since the KONČAR Group was founded in 1921, there are also large quantities of documents in paper form. In order to make information from these documents available, these documents must first be digitized (scanned or photographed), i.e., converted from paper form to a form suitable for computer processing (machine-readable text). Next, it is necessary to structure and standardize the digitized documents as well as to annotate them with metadata. Depending on the language contained in the documents, normalization of the text to standard Croatian will be required. For each step in the digitization process, it is necessary to apply existing or develop new (semi)automated methods.
An e-document management system, i.e., a digital archival information system has to be created in order to search and manipulate through large collections of both digitally born and digitized documents. It is necessary to enable the management of e-documents and e-records, to ensure their longevity and to ensure the trust of users in this type of records. Digital signatures will enable the verification of authenticity of digitally born and digitized documents. Furthermore, it is necessary to develop mechanisms of authenticating a printed record which has been digitally signed. All digitized documents will be archived in the central business intelligence system, and thus available for the business intelligence analyses. This will allow historical (mostly unstructured) and new (equally structured and unstructured) data to be interconnected, which will enable derivation of new knowledge and efficient decision-making.

Conclusion

The knowledge management project launched by KONČAR aims to develop a knowledge management system for the KONČAR Group's own demands, through collaboration with the academic community with the potential broader social significance of the project. The expertize of the researchers from the Faculty of Humanities and Social Sciences and the Faculty of Electrical Engineering and Computing at the University of Zagreb in the relevant areas, as well as the experience of these institutions in participation and management of international and national projects, will contribute to improving the market position of the KONČAR Group. Finally, this collaboration can help to advance KONČAR Group's own business processes, contribute to joint appearance and to the implementation of the knowledge management solutions in other companies.

References

Coffield, Frank; Moseley, David; Hall, Elaine; Ecclestone, Kathryn. Learning styles and pedagogy in post 16 learning: a systematic and critical review. The Learning and Skills Research Centre, 2004.

Hilbert, Martin; López, Priscila. The World's Technological Capacity to Store, Communicate, and Compute Information. // *Science*. 332 (2011), 6025; 60-65.

Jacobs, Ronald; Washington, Christopher. Employee development and organizational performance: a review of literature and directions for future research. // *Human Resource Development International*. 6(3) (2003), 343-354).KONČAR – Inženjering za energetiku i transport, Program uspostave sustava upravljanja znanjem u Grupi KONČAR, September 2014.

Petrick, Joseph A.; Scherer, Robert F.; Brodzinski, James D.; Quinn, John F.; Ainina, M. Fall. Global leadership skills and reputational capital: Intangible resources for sustainable competitive advantage. // *Journal of the Academy of Management Executive*. 13(1) (1999), 58–69.

Dalikr, Kimiz. Knowledge Management in Theory and Practice. Burlington: Elsevier, 2005

Pearse, Noel. The Role of Experience in the Creation of Intellectual Capital // *Proceedings of ICICKM 2007 The 4th International Conference on Intellectual Capital, Knowledge Management and Organisational Learning* / Remenyi Dan (ed.). Cape Town, 2007, 329-335.

Bitrix24 – integrated business-information system for sales process support

Marta Alić
University of Applied Sciences, Zagreb
Vrbik 8, Zagreb, Croatia
marta.alic@zg.htnet.hr

Summary

With the development of information technology, its increased application in the organization of the business process has emerged. In order to manage and keep track with shifting modern markets, companies started to automatize its processes and improve internal and external communication, as a first postulate of agility to keep track with new trends. The implementation of complex, integrated business and information systems is the first step. This paper gives the overview of Bitrix24 system basic functions as a cloud network solution that integrates the customer relationship management (CRM) with the information systems of the entire company.

Keywords: Bitrix24, information system, business process, project management, customer relationship system, document management, human resources

Introduction

Bitrix24 is an integrated business system, realized as an intranet network in the cloud environment, that connects multiple information systems of various business processes in the company.

The base system is a CRM (Customer Relationship Management) system, which requires joint efforts of all departments of the company, supporting marketing and sales business functions [1].

Regarding the purpose and use [2] of such systems, Bitrix24 is a compound system of customer relationship management which:

1. supports collaboration and communication with customers, partners and suppliers,
2. allows the construction of the employee communication and collaboration along the organizational structure of the company and
3. supports business process automation.

Bitrix24 is a business network based on the principles of the social network paradigm. At the base of the system is activity feed, as a basic form of interaction among employees. Every employee can see the latest news and track changes that took place within the company and interact with them. Through this interface one can easily schedule an event, organize a poll, upload files and/or make

a post as a part of a project management IS. Functionalities of requests (for business trip approval, leave approval, purchase requests, etc.) made from customizing forms in the interface can serve as integral part of the human resources (HR) and accounting information systems.

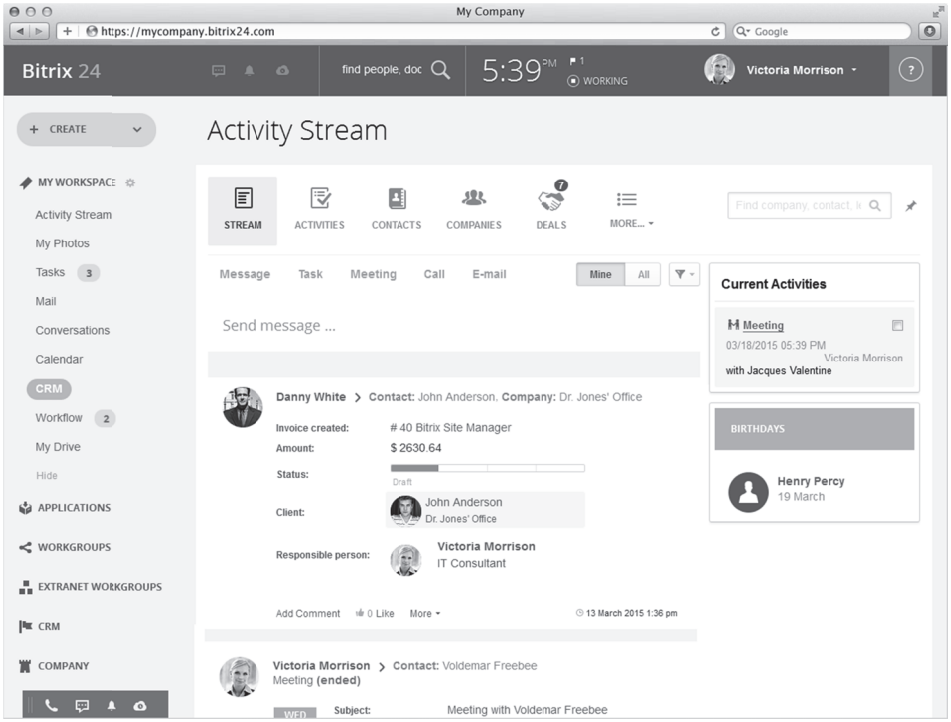


Figure 1. Activity stream workspace (source: https://www.bitrix24.com/features/crm.php)

CRM system

As a CRM system, Bitrix24 offers easy automation of business processes through the default workflow or by building customized ones. Business processes can include various actions on elements of the system; such as sending e-mails, assigning responsible persons, making tasks; and operations of the overall document management system (DMS): from record management to archiving. Based on the lead management methodology, the system serves as a bridge between marketing and sales information systems.

Lead is a CRM object that represents a buyer's interest in a product or service and it is usually a result of the organization's marketing strategy and actions through various sales channels (website, social networks, direct sales channels, etc.). By passing through system leads are being qualified and prioritized according to their level of the realization possibility. At the end of a process, leads can be qualified as junk or quality leads. First ones are then being removed from

system or sent back for remarketing, while second ones are being converted to deal.
Deals are system objects of sales business processes that contain the interaction with a buyer pertaining to a transaction. Deals proceed through the system by a number of customizable statuses given by the system or customized by a workflow and they result with a closed sale (won or lost). At this stage of the sales process, system offers the possibility of quote generation, accompanied with referring invoice, as two new objects in system, connected with accounting operations IS and DMS.

Project and task management

Through the system for the organization of operational processes in sales, Bitrix24 in its structure clearly follows the mentioned stages of lead management and sales process, integrating them with the system for project management. When setting clear organizational structure of departments and functional units in system adjustment phase, same structure is easy to follow for adjustment of the organization internal communication system. Every unit can be set as one workgroup, sub-system network where employees within have every functionality of the entire system and where all collaboration and work is documented on activity feed. Projects are also structured through workgroups, giving the possibility of organizing team of employees from different units and on different functions with the same document management panel. When a new version of the document is loaded, the previous is not deleted, but is stored in order to be available at all times and records history provides information about the time when the version created and the user who created it.
Documents can be edited in a cloud with providing Google Docs and Microsoft Office Online tools for text documents, spreadsheets, and presentations on cloud or locally on the computer and by saving they are easily indexed and subject to search, as well as all records of activities, messages or system objects.
By creating a project task, time management option is enabled that keeps track of task duration and enable transparent performance reports of the project at any time. Employees can independently input and record absence days through the system. Reports can be made for each department and for each employee on weekly, monthly or annual terms, supporting also HR management.

Product catalog

Bitrix24 system comprises of the product catalog as an important part of customers relations and sales process. Products can be categorized in sections and have a range of settings options in the system (tax rates, units of measure, currency) as well as customizing options for fields and attributes. They are easily connected with lead and deal objects in the system, giving more detailed overview of realized sales.

Management support

One of the major benefits of CRM system implementation is its support to the management. With evolved reporting IS CRMs are often used for planning and forecasting of future sales activities.

Bitrix24 offers a detailed statistical report of sales business processes by individual components - objects, their status and the persons responsible. There are also embedded reports by products, activities and invoices sent and revenue made. Bitrix 24 in its purpose is not analytical system and in-depth analyzes and forecasts cannot be made within the system, so the segmented data are used primarily for optimization of business processes and as raw data for redistribution.

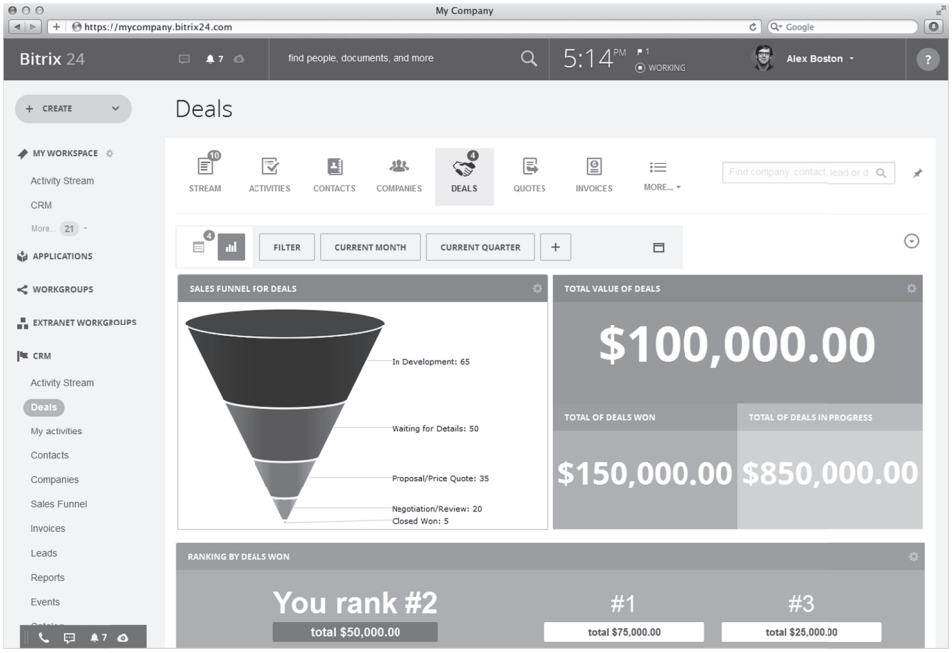


Figure 2. CRM report dashboard (source: https://www.bitrix24.com/features/crm.php)

Conclusion

By introducing elements of social networks in the communication and collaboration processes, Bitrix24 system sets different approach to their organization. Although primarily designed to support the sales processes and customer relations, the system sets wider functionalities of business management, incorporating information systems of document management and project management, human resources, accounting and product catalog functionalities; providing automatization of related activities. Guided by the principles of project management, system analyzes complex processes to simple operations that, accompa-

nied with a social networking component, gives a unique approach to organization of business, making Bitrix24 more than just CRM system, but complex business platform.

References

Bibiano, Luis; Mayol, Enric; Pastor, Joan Role and importance of business processes in the implementation of CRM systems. : Zaragoza: JISBD, 2007.

Harewood, Alton ERP, CRM and BPM vs. Business Process Optimization. May 13th, 2013 http://blogs.aspect.com/2013/05/13/erp-crm-and-bpm-vs-business-process-optimization/ (May 20th, 2015)

Nguyen, Bang The dark side of customer relationship management: Exploring the underlying reasons for pitfalls, exploitation and unfairness, Journal of Database Marketing & Customer Strategy Management. 19 (2012), 56–70

Srića, Velimir; Muller, Josip. Upravljanje odnosima s klijentima. Zagreb: Delfin – razvoj managementa, 2005.

HUMAN-COMPUTER INTERACTION, LANGUAGE TECHNOLOGIES AND APPLICATIONS

Original scientific paper

Information Transfer through Online Summarizing and Translation Technology

Sanja Seljan*, Ksenija Klasnić**,
Mara Stojanac*, Barbara Pešorda*, Nives Mikelić Preradović*
*Department of Information and Communication Sciences,
**Department of Sociology
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
sseljan@ffzg.hr, kklasnic@ffzg.hr,
mstojana@ffzg.hr, bpesorda@ffzg.hr, nmikelic@ffzg.hr

Summary

Information access – presented in proper language, in understandable way, at the right time and right place can be of considerable importance. Information and communication technology, wrapping also human language technologies, can play important role in information transfer to the specific user. Translation technology along with summarizing technology has opened new possibilities and perspectives, requiring in the same time the critical opinion in information analysis. The main purpose of this research is to present the impact of text summarization and online machine translation tools on information transfer.
The research was performed on texts taken from online newspapers in five domains (politics, news, sport, film and gastronomy) in English, German and Russian languages. The total of N=240 evaluations were analysed, performed by the same three evaluators.
In the research three types of assignments were made. The first assignment was to evaluate machine-translated sentences at the sentence level for the three language pairs (English-Croatian, German-Croatian and Russian-Croatan). In the second task, the similar evaluation was performed, but at the whole text level. In the third assignment, which was related to information transfer, the evaluators were asked to evaluate the overall quality of the texts process in the pipelined process (online summarization and online machine translation) for English and German. Assessment was based on the finding the answers to the following questions – who, what, when, where, and how? The results were analysed by ANOVA, t-test and binary logistic regression.

Keywords: information transfer, text summarization, online translation tools, evaluation, statistical analysis, ANOVA, t-test

Introduction

Information access – presented in proper language, in understandable way, at the right time and right place can be of crucial importance. Information and communication technology, wrapping also human language technologies, can play important role in information transfer to the specific user. One of tasks stated in Tongia et al. (2005) was to make ICT universally available, accessible, and affordable, usable and applicable.

In the situation when acquiring up-to-date information is important, automatic processes and human language technologies can play an important role. Huge amount of data cannot be managed and analysed by humans. Moreover, access is often limited, presented in various forms, of different length and in various languages, often not easily retrieved. That is the situation where automatic or semi-automatic processes of summarization and machine translation can provide the basic insight. Being piled up with huge amount of data, users have to discern valid from invalid information and to decide whether to read or not the whole documents. Because of growth of non-English speakers, large amount of information is often needed on target languages. Therefore, information access, cross-language information retrieval and information transfer represent one step further in global communication.

There are various projects dealing with information access and information retrieval, including human language technologies including online summarization and machine translation. But those free online tools are mainly created for widely spoken languages, while for less-resourced languages (e.g. Croatian) they exist mainly as standalone applications. Online summarization tool for Croatian – CroWebSum was presented by Mikelic Preradovic et al. (2010) yet not integrated in the broader process for information transfer. Free online machine translation services for Croatian still obtain lower grades than for other languages (Seljan et al. 2015a, 2015b), due to scare language resources, differences between language pairs, specific domains, etc. and depending on user's expectations.

In the paper, the pilot research was performed in order to evaluate information transfer by use of free online summarization tool and by two online machine translation tools. Information transfer was evaluated by questions: who, what, when, where, and how? Information quality, completeness, understanding and, finally, reader's possibility to recognize information was analysed. This pilot research was made in five different domains: politics, news, sport, film and gastronomy. All texts were firstly summarized and then machine translated from English, German and Russian into Croatian language. Evaluation was performed by three different evaluators, using t-test between two types of evaluations and between the two systems. The main limitation of the pilot study was small number of summaries included in the pilot research.

Related work

Summarization and machine translation tools are often integrated as modules into various systems dealing with information access or cross-language information retrieval. An example of automatic public service is *Europe Media Monitor* (EMM) developed by Joint Research Centre (JRC), Directorate General of the European commission performing real-time monitoring and analysis in order to detect threats (natural disasters, diseases, etc.). The system is developed for more than 60 languages, performing gathering, clustering, information extraction, summarisation, machine translation and generation (Steinberger et al., 2009).

The paper presented by Damianos et al. (2003) describes the MiTAP system using human language technologies created for monitoring of infectious disease outbreaks providing multilingual information access to various resources in 8 languages. *MITRE* is the system of Text and Audio Processing attempting to solve these problems using natural language technology and careful focus on the end user. The system is designed to perform, among other, machine translation and summarization.

Afantenos et al. (2005) bring survey on summarization of medical documents used in order to quickly determine the main points of a document, elaborating on types of summaries, input factors (single or multi-document, languages, text, speech or multimedia), purpose factors (informative or indicative summaries, generic or user-oriented, general or domain-specific), output factors (quality, extracts or abstracts), evaluation methods (intrinsic or extrinsic) and various summarization techniques. It focuses especially on summarization in medical domain. Possible evaluations could be based on utility values, relevance, information inclusion, reading comprehension, or other.

Chin-Yew (1999) describes the design and implementation of MuST – a multilingual information retrieval, summarization and translation system which enables to perform cross-language information retrieval, summarization and machine translation.

Xiaojun et al. (2010) present cross-language document summarization aiming to produce summary in a target language, focusing on English-Chinese, with purpose to understand the major content. The future plans include manual translation of reference summaries and automatic ROUGE metric to perform automatic evaluation.

Chieze et al. (2010) present information system created for legal professionals which integrates natural language processing technologies, like text classification and summarization. Text is then submitted for bidirectional statistical machine translation (SMT) between English and French.

Online text summarization tools

Text summaries represent a necessity in today's information age. Due to huge savings in time, they represent a useful tool for managing the vast available online texts. The process of summarization reduces the complexity and length of the original document, providing the visibility of the subject matter and key ideas of the work. (Mikelic Preradovic, Vlainic, 2013)

„Text summarization represents a method of extracting relevant portions of the input document, presenting the main ideas of the original text. It is a process of condensing a source document into its shorter version preserving the information content."(Mikelic Preradovic, Vlainic, 2013) Automatic text summarizations are being used for summarizing news to SMS or WAP-format, for mobile phones of PDAs or for TTS systems. It is also used for compressed descriptions of the search results in the search engines as well as in keyword directed subscriptions of news which are summarized and sent to the user. Summaries are useful for orientation (google maps) and in a decision making process (e.g. TV guide).

There are various summarization systems which use statistical or linguistic approach or the combination of these two. Although these summarization systems mainly rely „on the shallow features of the text, they all generate informative extracts satisfying quality expectations of the human users" (Mikelic Preradovic et al., 2014, 9). There are several basic types of summaries: informative and indicative (based on the aim of the summary). Indicative summaries give the review of the summarized information from the most relevant topics in the document. Informative reduce the quantity of the information, but still keep the relevant information (Mikelic Preradovic et al, 2007). Summarization techniques can be classified into three levels (Vlainic, Mikelic Preradovic, 2013):

- surface methods – relying on the frequency of words, sentence position, words in the title or to the presence of cue phrases in text
- entity level – model text entities and the semantic relationships between these entities; the relations between entities are based on similarity, proximity and cohesion
- discourse level methods – model the document's global structure and its relation to the communicative goals, taking into account the rhetorical structure of the text.

The summarized text should give the answers to the following questions: *who, what, when, where, and how*? This pilot research will present the extent to which the information obtained by text summarization changed the quality, completeness and the original meaning of the text which is important while evaluating the quality of content. Attempts will be made in order to evaluate the final information transfer after using the results of three respondents who gave answers to the mentioned questions of the text which was created after the summarizing and translating.

Online translation tools

In recent years, machine translation technology develops to the extent that it causes great interest in every possible area. In the education market, the international institutions and many other places, machine translation technology is searching for the opportunity and place for its integration. Machine translation makes quick and easy translation from one natural language into another. The aim of this paper is to show the impact of online machine translation tools to information transfer. Certain online machine translation tools will show the quality of translation and whether the information is lost or not. Summarized text was machine-translated by use of freely available online translation technology. In the research the role of both technologies is analysed in the process of conveying and understanding text from one natural language into another – in this case, from English, German and Russian into Croatian language by use of two types of technology – well-known Google Translate and Yandex Translate.

Nowadays when almost everyone can produce and share various content at the Internet, it is highly important to know sources and tools that are of good quality, precision and accuracy. However, the question is how can one say with certainty that one automatic translation system is more "suitable" or "better" than the other? Evaluation of machine translation can answer this question, but it is a complicated task due to difficulty and complexity of the process itself. Human and automatic evaluations are two types of evaluation of machine translation. As the human evaluation requires more costs and time, automatic evaluation is requisite in the process of machine translation.

Evaluation of machine translated text for Croatian, still obtain lower scores by various online tools when using different methods – 1) human evaluation with one or more reference translations or 2) various automatic metrics (Seljan, 2015a; Seljan and Dunđer, 2015b).

In this paper, human evaluation was used. When texts from five different categories of each language were translated, respondents were given the text for evaluation. The respondents needed to perform quality evaluation of translated text comparing sentence by sentence individually in the target and source language, after the text was firstly summarized. Then they evaluated the quality of the whole text after it was only translated by each tool, not having to compare it individually sentence by sentence as they did in the previous assignment.

Research Methodology

The research has been conducted in order to receive required conclusions about information transfer by dint of questions posed to three respondents, native Croatian speakers. The corpus that was used in this research contained texts from three different languages – English, German and Russian. Each language comprised texts from five different categories:

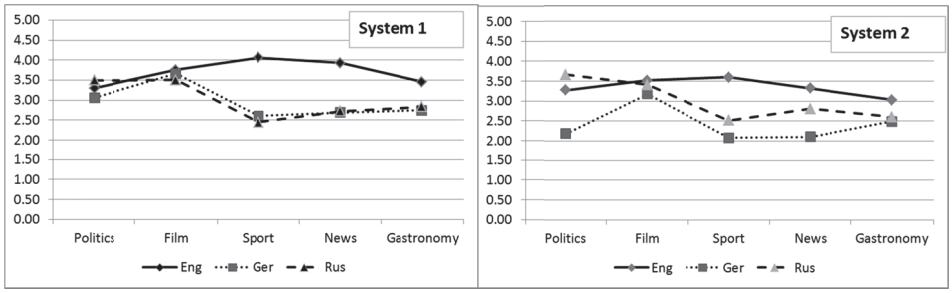
- politics,
- news,
- sport,
- film and
- gastronomy.

In the evaluation process, the same three native Croatians speakers participated. The total of N=240 evaluations were analysed:
In the first task 90 evaluations for machine translated texts were made (3 language pairs, 5 domains, 3 evaluators, 2 online tools) where 60 evaluations for the pipelined process of summarization and machine translation (2 language pairs, 5 domains, 3 evaluators, 2 online tools) and 30 evaluations for the machine-translation process (1 language pair, 5 domains, 3 evaluators, 2 online tools). This evaluation was made at the sentence level.
In the second task, the similar evaluation was performed, but at the whole text level: 90 evaluations (3 languages, 5 domains, 3 evaluators, 2 online tools).
In the third task, which was related to information transfer, 60 evaluations were done (2 languages, 5 domains, 3 evaluators, 2 pipelined processes)
Each evaluator had three assignments.
The first assignment was to evaluate machine-translated sentences at the sentence level for the three language pairs (English-Croatian, German-Croatian and Russian-Croatan). Machine translation was made by two online services: Google Translate and Yandex Translate for all three languages. Texts on English and German were firstly summarized and then machine translated into Croatian language. The summarization was made from 108 sentences to 47 sentences in English and from 103 sentences into 49 sentences for German. As the summarization for Russian language was not possible, evaluation of text accuracy for Russian-Croatian was made only on the basis of machine translation services, without summarization step. Texts on English and German were previously summarized by online summarization tool Swesum. on the scale from 1 to 5 to each sentence (partial values as 1.0, 1.5, 2.0 etc. were also allowed in the evaluating process) using the following scale: 1=not understandable, not enough information, 2=hardly understandable, necessary to repair almost everything, 3=understandable, but many changes needed, 4=very good, but with some mistakes, 5=excellent. The evaluators could compare the sentences from the source and target languages at the same time. The number of sentences depended on the extent of the text. Therefore data analysis was performed on the basis of average score of all sentences for the specific domain.
The second assignment consisted in quality evaluation of the translated text as a whole. Moreover, the respondents did not have the text from the source language to compare it to the translated text. Also, likewise in the first assignment, they gave quality score to the text ranging from 1 to 5.

In the third assignment, which was related to information transfer, the respondents were asked to evaluate the overall quality of the texts process in the pipelined process (online summarization and online machine translation) for English and German. Assessment was based on the finding the answers to the following questions – *who, what, when, where, and how*? If the respondents could give the answer to the asked question, they gave score 1, and if they could not give the answer, they marked 0 as an answer. Furthermore, the respondents did not have the source language to compare with.

Results

Description



Figures 1, 2. MT system 1 and system 2 evaluation – mean accuracy scores

Diagrams 1 and 2 present mean scores of machine-translated texts in five domains. Both systems obtained the best grades for the English language. The system 1 (Google Translate) obtained the second best grades for the German-Croatian language pair and the system 2 (Yandex Translate) obtained the second best grades for the Russian-Croatian. The system 1 received close scores for the domains of film and politics for all three languages, while for other domains differences were more considerable. The system 2 received similar scores for the domain of film and gastronomy, while the politics was best represented for Russian language. The second system obtained generally better results for the Russian language as the second language and the first system for the German language as the second language.

Evaluation at the sentence level

The first assignment was to evaluate the accuracy of each sentence wherein evaluators could compare the sentences from the source and target languages at the same time.

A one-way between subjects ANOVA was conducted to compare the average level of accuracy of transmitted information when using two different free online tools for machine translation (Google Translate and Yandex Translate)

for texts translated from English, German and Russian into Croatian at the sentence level. Normal distribution of accuracy variable across all three languages and both services was verified by Kolmogorov-Smirnov test (p=.057).
There was a significant effect of using different online tools for different languages in accuracy of transmitted information at the p<.05 level for the six conditions [$F_{(5,84)}$=4.78, p=.001].

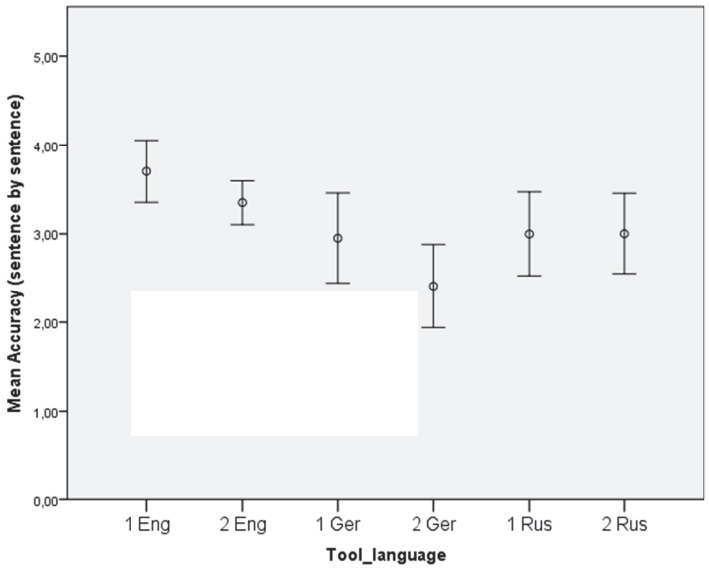


Figure 3. Error bars (mean and 95% CI for means): accuracy by tool (1-Google Translate, 2-Yandex Translate) and language at sentence level

Post hoc comparisons using the Tukey HSD test indicated that there was no statistically significant difference among tools compared by the same language pair (e.g. English-Croatian for both tools) when transmitting information. However, two statistically significant differences in average accuracy scores were determined.
There is a difference between the use of pipelined online tools for summarization and machine translation by Google Translate from English to Croatian (M=3.70, SD=0.64) and pipelined tools for summarization and online Yandex Translate from German to Croatian (M=2.41, SD=0.85). Google Translate from English to Croatian resulted in higher mean accuracy than Yandex Translate from German to Croatian (p<.001).
Second statistically significant difference is a difference in mean accuracy scores of information between Yandex Translate from English to Croatian (M=3.35, SD=0.44) and Yandex Translate from German to Croatian (M=2.41, SD=0.85). Yandex Translate from English to Croatian resulted in higher mean accuracy than Yandex Translate from German to Croatian (p<.001).

Evaluation at the text level

The second assignment consisted in quality evaluation of the translated text as a whole in which respondents did not have the text from the source language to compare it to the translated text. This task is similar to the previous one, but the evaluation is performed at the whole text level.

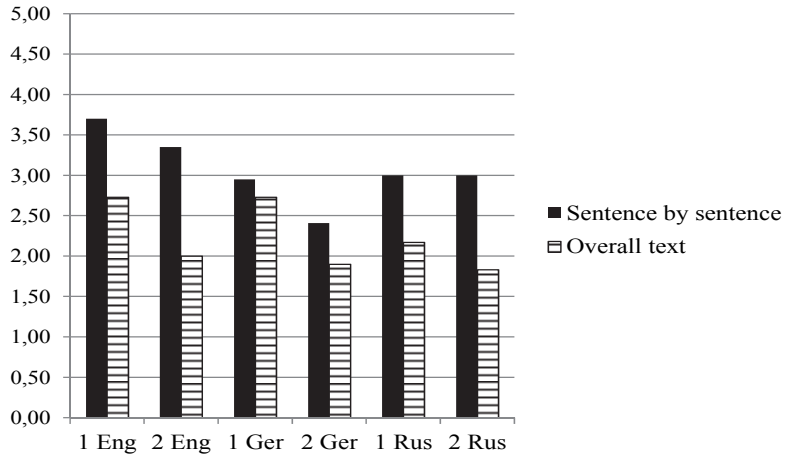


Figure 4. Comparison of sentence by sentence mean scores and text evaluation mean scores(1-Google Translate, 2-Yandex Translate)

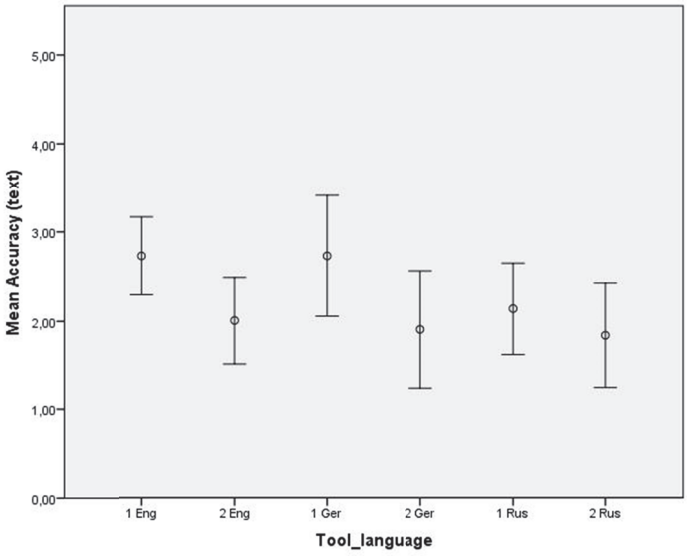


Figure 5. Error bars (mean and 95% CI for means): accuracy by tool and language on text

When comparing sentence by sentence translation and the overall evaluation, sentence by sentence evaluation received overall mean score of 3.07 (SD=.85), while overall evaluation received 2.22 (SD=1.07) score. This difference was statistically significant [$t_{(89)}$=7.20, p<.001].

Reason for this probably lied in the fact that respondents did not have the text from the source language in order to make a comparison like they could in the first assignment. Furthermore, respondents explained that in this assignment they did not have to be as concentrated as in the previous task which was detailed due to working on each sentence individually, and not with the whole text like it was the case here.

A one-way between subjects ANOVA was conducted to compare the average level of accuracy of transmitted information when using two different free online tools for machine translation (Google Translate and Yandex Translate) for texts translated from English, German and Russian into Croatian on overall text level.

There was a significant effect of using different online tools for different languages in accuracy of transmitted information at the p<.05 level for the six conditions [F(5,84)=2.35, p=.048].

Post hoc comparisons using the LSD test indicated that there was one statistically significant difference among tools compared by the same language and three statistically significant differences among tools compared by different languages.

The only statistically significant difference among tools compared by the same language is for German language, between the use of machine translation by Google Translate from German to Croatian (M=2.73, SD=1.24) and online machine translation Yandex Translate from German to Croatian (M=1.90, SD=1.20). Google Translate from German to Croatian resulted in higher mean accuracy than Yandex Translate from German to Croatian (p=.030).

Second statistically significant difference is a difference in mean accuracy scores of information transfer between Google Translate from English to Croatian (M=2.73, SD=0.80) and Yandex Translate from German to Croatian (M=1.90, SD=1.20). Google Translate from English to Croatian resulted in higher mean accuracy than Yandex Translate from German to Croatian (p=.030).

Third statistically significant difference is a difference in mean accuracy scores of information between Google Translate from English to Croatian (M=2.73, SD=0.80) and Yandex Translate from Russian to Croatian (M=1.83, SD=1.06). Google Translate from English to Croatian resulted in higher mean accuracy than Yandex Translate from Russian to Croatian (p=.019).

Fourth statistically significant difference is a difference in mean accuracy scores of information transfer between Google Translate from German to Croatian (M=2.73, SD=1.23) and Yandex Translate from Russian to Croatian (M=1.83,

SD=1.06). Google Translate from German to Croatian resulted in higher mean accuracy than Yandex Translate from Russian to Croatian (p=.019).

Information transfer

In the last assignment the respondents were asked to evaluate the overall quality of the texts based on the finding the answers to the following questions – *who, what, when, where,* and *how*?

The average information score for German was 3.8 and for English 4.4. English summaries missed mainly *where?* and *when?* information. German summaries also answered *who?* and *what?* questions, more *when?* information, but missed more *where?* information. The best score was obtained for *who? (0.95),* followed by *what? (0.87), how?* (0.83), *where? (0.72)* and *when? (0.60).*

Binary logistic regression analyses was used to test whether accuracy evaluations for English-Croatian and German-Croatian translations of both systems can predict the odds of giving the answers to five listed questions. This analysis was performed using sentence level because of higher accuracy scores.

Four out of five analyses (dependent variables: *who*?, *what?*, *where?* and *when?*) did not show any statistical significance of the effect of accuracy. Accuracy has shown to be statistically significant predictor only for the odds of giving the answers to *how?* question. Analysis showed that for a one-unit increase in accuracy on sentence by sentence level the odds of giving the answer to the question *how?* for transmitted information increases 6.3 times (95% C.I.: 2.1 – 18.5) (p=.001).

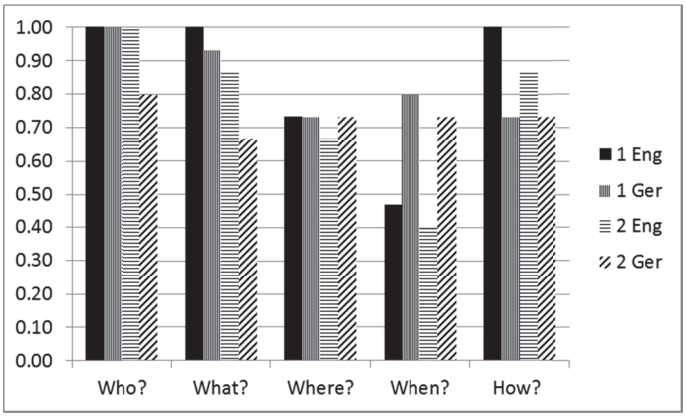


Figure 6. Information transfer in summaries across all domains

Conclusion

The paper presents information transfer in five domains (politics, news, sport, film and gastronomy) for text taken from online newspapers for 3 languages (English, German and Russian). In the research three types of assignments were made.

The first assignment was to evaluate machine-translated sentences at the sentence level for the three language pairs (English-Croatian, German-Croatian and Russian-Croatan). A one-way between subjects ANOVA was conducted to compare the average level of accuracy of transmitted information when using two different free online tools for machine translation (Google Translate and Yandex Translate).

In both systems, the best grade was obtained for the English language. The second best grades for Google Translate were obtained for the German-Croatian language pair and for the Yandex Translate it was Russian-Croatian. Overall scores for the German-Croatian were better for Google Translate, and for Russian-Croatian the overall scores were better, but not statistically significant, for Yandex Translate.

When comparing different language pairs and different tools, there was a significant effect in accuracy of transmitted information at the $p<.05$ level for the six conditions [$F_{(5,84)}=4.78$, $p=.001$].

Post hoc comparisons using the Tukey HSD test indicated that there was no statistically significant difference among tools compared by the same language pair (e.g. English-Croatian for both tools) when transmitting information. However, two statistically significant differences in average accuracy scores were determined: between machine-translated texts by Google Translate for English-Croatian (M=3.70, SD=0.64) and Yandex Translate for German-Croatian (M=2.41, SD=0.85). Google Translate from English to Croatian resulted in higher mean accuracy than Yandex Translate from German to Croatian ($p<.001$).

Second statistically significant difference is a difference in mean accuracy scores between Yandex Translate for English-Croatian (M=3.35, SD=0.44) and Yandex Translate for German-Croatian (M=2.41, SD=0.85). Yandex Translate from English to Croatian resulted in higher mean accuracy than Yandex Translate from German to Croatian ($p<.001$).

The second assignment consisted in quality evaluation of the whole translated text, without possibility to compare with source language text.

When comparing sentence by sentence translation and the overall evaluation, sentence by sentence evaluation received overall mean score of 3.07 (SD=.85), while overall evaluation received 2.22 (SD=1.07) score. This difference was statistically significant [$t_{(89)}=7.20$, $p<.001$].

A one-way between subjects ANOVA was conducted to compare the average level of accuracy of transmitted information when using two different free

online tools for machine translation for texts translated from English, German and Russian into Croatian on overall text level.

There was a significant effect of using different online tools for different languages in accuracy of transmitted information at the $p<.05$ level for the six conditions [$F(5,84)=2.35$, $p=.048$].

The only statistically significant difference among tools compared by the same language is for German language, between the use of machine translation by Google Translate from German to Croatian ($M=2.73$, $SD=1.24$) and online machine translation Yandex Translate from German to Croatian ($M=1.90$, $SD=1.20$). Google Translate from German to Croatian resulted in higher mean accuracy than Yandex Translate from German to Croatian ($p=.030$).

In the third assignment, which was related to information transfer, the evaluators were asked to evaluate the overall quality of the texts processed in the pipelined process (online summarization and online machine translation) for English and German. Assessment was based on the finding the answers to the following questions – *who, what, when, where, and how*? The average information score for German was 3.8 and for English 4.4.

Binary logistic regression analyses was used to test whether accuracy evaluations for English-Croatian and German-Croatian translations at the sentence level of both systems can predict the odds of giving the answers to five listed questions. Analysis showed that for a one-unit increase in accuracy on sentence by sentence level the odds of giving the answer to the question *how?* for transmitted information increases 6.3 times (95% C.I.: 2.1 – 18.5) ($p=.001$).

Taken together, results suggest significant differences in information transfer when using different online tools. Although they work best for the English language, there are significant differences among other languages and online tools. The second aspect is the user information perception who gave significantly higher scores in sentence by sentence evaluation, than on the whole text evaluation. Information transfer in the pipelined process of summarization and online translation shows significant connection between accuracy and the question *how?* However, these results are to be taken as preliminary due to small number of test data analysed in this pilot research. The following pilot research would include finding methods for improving results for the specific language or by adding online terminology/ontology resources.

References

Afantenos, Stergos D.; Karkaletsis, Vangelis; Stamatopoulos, Panagiotis. // Summarization from Medical Documents: Survey. *Journal of Artificial Intelligence in Medicine*. 33(2005);157-177

Chieze, Emmanuel; Farzindar, Atefeh; Lapalme, Guy. An Automatic System for Summarization and Information Extraction of Legal Information. // *Semantic Processing of Legal Texts. Lecture Notes in Computer Science.* 6036 (2010); 216-234

Chin-Yew, Lin. Machine Translation for Information Access across the Language Barrier: te MuST System. *Machine Translation Summit* VII, 1999

Damianos, Laurie E.; Wohlever, Steve; Kozierok, Robyn; Ponte, Jay M. MiTAP: A Case Study of Integrated Knowledge Discovery Tools. // *36th Hawaii Int. Conf. on System Sciences (HICSS-36)*; 2003; 69

Hofman, Hans. How to keep digital records understandable and usable through time? // *Proceedings Long-Term Preservation of Electronic Records* / Paris, 2001

Lee, Alice Y.L.; So, Clement Y.K. Media Literacy and Informtaion Literacy: Similarities and Diferences. // *Comunicar.* 21 (2014), 42; 137-145

Mikelic Preradovic, Nives; Ljubesic, Nikola; Boras, Damir. Croatian web text summarizer (CroWebSum). // *32nd Int. Conf. on Information Technology Interfaces (ITI)*. 2010; 109-114

Mikelic Preradovic, Nives; Boras, Damir ; Vlainic, Marta. Importance of Surface Methods in Human and Automatic Text Summarization. // *Journal of Computers*. 8 (2014); 9-16

Tongia, R.; Subrahmanian, E.; Arunachalam, V. S. Information and Communications Technology. // *Information and Communication Technology for Sustainable Development: Defining a Golobar Research Agenda.* / Bangalore: Allied Publishers, 2005; 19-42

Seljan, Sanja; Tucaković, Marko; Dunđer, Ivan. Human Evaluation of Online Machine Translation Services for English/Russian-Croatian. // *Advances in Intelligent Systems and Computing – New Contributions in Information Systems and Technologies.* Springer. 353 (2015a); 1089-1098

Seljan, Sanja; Dunđer, Ivan. Machine Translation and Automatic Evaluation of English/Russian-Croatian. // *Proceedings of Corpus Linguistics*. St. Petersburg State Univ., 2015; 72-79

Steinberger, Ralf; Pouliquen, Bruno; van der Goot, Erik.n introduction to the Europe Media Monitor family of applications. // *Information Access in a Multilingual World – Proceedings of the SIGIR 2009 Workshop (SIGIR-CLIR'2009)*. 2009; 1-8

Vlainic, Marta; Mikelić Preradovic, Nives. Comparative Study of Automatic Text Summarization System Performance. *// Recent Advances in Information Science*. 13 (2013); 222-227.

Xiaojun, Wan; Huiying, Li; Jianguo, Xiao. Cross-Language Document Summarization Based on Machine Translation Quality Prediction. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL).* 2010; 917-926

Original scientific paper

Interoperability of an 18th century Italian-Latin-Croatian dictionary

Petra Bago, Damir Boras
Department of information and communication sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
{pbago, dboras}@ffzg.hr

Summary

The digitization process of historical texts is increasing in the last several decades, thereby developing the field of digital humanities. However, the digitization projects are usually isolated to the project teams, universities and institutes, but also to individuals developing the project. There was a lack of communication between the community members, which resulted in resources, tools and systems that are not able to exchange information.
Consequently, recently it was demonstrated that there is a demand for standardization of technologies, but also of all processes of development. Interoperability emerges as the key concept at this phase of digital humanities, which aims to facilitate communication of data.
To enable semantic interoperability of historical dictionaries, they have to be encoded using some standard. In this paper we present the encoding of della Bella's trilingual 18th century dictionary entries using a TEI (Text Encoding Initiative) encoding scheme for dictionaries. The dictionary is a trilingual dictionary containing Italian, Latin and Croatian language.

Keywords: historical dictionaries, interoperability, Text Encoding Initiative, encoding, digital humanities

Introduction

The digitization process of historical texts is increasing in the last several decades, thereby developing the field of digital humanities. However, the digitization projects are usually isolated to the project teams, universities and institutes, but also to individuals developing the projects. There was a lack of communication between the community members, which resulted in resources, tools, and systems that are not able to exchange information.
Consequently, recently it was demonstrated that there is a demand for standardization of technologies, but also of all processes of development. Interoperability emerges as the key concept at this phase of digital humanities, which aims to facilitate communication of data.

Interoperability defined

In order to overcome the lack of communication between language data, and achieve interoperability of language resources, two big international projects joined forces. One of the projects is a USA project called *Sustainable Interoperability for Language Technology* (SILT). Its goal is to turn existing, fragmented technology and resources developed to support language processing technology into accessible, stable, and interoperable resources that can be readily reused across several fields (SILT, 2014). The second project comes from Europe and is called *Fostering Language Resources Network* (FLaReNet). Its goal is to develop a common vision of the language resources and language technologies, and foster a European strategy for consolidating the sector, thus enhancing competitiveness at EU level and worldwide (FLaReNet, 2014).

The aim of this international cooperation is to involve community members in the field of language technology, but also those working in related fields, to collaborate. The collaboration should create a consensus related to sharing data and technologies for language resources and applications, working towards the interoperability of existing data and, where possible, to promote standards for markup and resource creation (Ide et al., 2009).

One of the main achievements of this international cooperation is a definition of the term "interoperability" within the field of language technologies (Ide et al., 2009).

According to Ide et al. (2009) interoperability can broadly be defined as a measure of the degree to which diverse systems, organizations, and/or individuals are able to work together to achieve a common goal. For computer systems, interoperability is generally divided in two types: syntactic interoperability and semantic interoperability. Syntactic interoperability aims at enabling communication and data exchange, relying on specific data formats, communication protocols and the like. It is important that information is exchanged, but there is no guarantee that interpretation of this information will be the same. On the other hand, with semantic interoperability, systems have the ability to automatically interpret exchanged information meaningfully and accurately in order to produce useful results via compliance to a common information exchange reference model. It is important that interpretation of exchanged information is the same on both sides of the communication (Ide et al., 2009).

To enable interoperability of language resources, the focus is to specify an abstract data model for structuring linguistic data to which syntactic realizations can be mapped, together with a mapping to a set of linguistic data categories that communicate the information (linguistic) content. Hence, syntactic interoperability in the context of language resources can be defined as the ability of different systems to process exchanged data either directly or via trivial conversion. On the other hand, semantic interoperability in the context of language resources can be defined as the ability of systems to interpret exchanged information in meaningful and consistent way. In that sense, language resources

have to focus more on semantic and not on syntactic interoperability (Ide et al., 2009).

Semantic interoperability is the reason why we decided to use TEI (Text Encoding Initiative) encoding scheme for dictionaries ("Dictionaries", 2013). The TEI Consortium develops and maintains the Guidelines (TEI P5, 2014), intended for everyone who is generating and/or processing textual resources in digital form. The Guidelines are recommendations of the Consortium on how to encode implicit features of textual resources, thereby making these features explicit. Once the structure or some other feature of a text is explicitly encoded, later processing by computers is easier. Text encoding with a predefined scheme enables data exchange without or with minimal information loss, and correct interpretation of the information. Since the Guidelines are based on XML, they are independent of computer programs and operation systems. The Guidelines can be applied to all natural language texts, without any limitations to textual form or content, as well as digitized and digitally born texts ("About these Guidelines", 2014).

The 9th chapter of the Guidelines defines tags that can be used to encode dictionary entries, as well as other lexical resources, such as glossaries ("Dictionaries", 2014). We consulted this chapter when encoding entries from an old Croatian printed dictionary with a very complex entry structure. Our research is conducted on the first volume of a second edition of a trilingual dictionary "Dizionario italiano-latino-illirico" (Italian-Latin-Croatian dictionary) compiled by Ardelio della Bella and printed in Dubrovnik in 1785 (della Bella, 1785). This dictionary, among several other important Croatian historical dictionaries, was digitized as part of the scientific projects "Croatian dictionary heritage and dictionary knowledge representation"[1] and "Croatian dictionary heritage and Croatian European identity"[2] funded by the Croatian Ministry of Science and Technology. The digitization process of the dictionaries is not the scope of this paper.

The process of encoding digitized dictionary entries is performed manually, which is time consuming and tedious. Since the process is a manual task, it paves a way for an encoder to introduce inconsistencies. If the process of encoding could be (semi)automated, it would reduce cognitive load in encoders and time spent on the task. However, to (semi)automate the encoding process, first we have to analyze the structure of the entries of a chosen dictionary, and encode the content, preferably using a standard. The scope of this paper is to

[1] Boras, Damir (principal researcher). Hrvatska rječnička baština i prikaz rječničkoga znanja; Znanstveni projekt, Zagreb, 2003-2006. http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451 (25.8. 2014.)

[2] Boras, Damir (principal researcher). Hrvatska rječnička baština i hrvatski europski identitet; Znanstveni projekt, Zagreb, 2007-2013. http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451 (25.8. 2014.).

describe this first step i.e. to present the results of the manual encoding of della Bella's dictionary entries.

About della Bella's dictionary

The dictionary was intended for Italian Jesuit missionaries to help them spread the faith in a national language i.e. Croatian language, but also other Slavic languages. For this reason a Croatian grammar can be found inside the dictionary preamble. The dictionary contains 899 pages and consists of two parts. The first part is a preamble written in Italian language on 54 pages. The second part is the dictionary, containing around 19,000 headwords. The dictionary is printed in two volumes: the first volume contains the preamble and the dictionary part from letters A to H, while the second volume contains the dictionary part from letters I to Z. For the first time in Croatian lexicography, della Bella's dictionary contains examples of uses of headwords in various literary works and oral literature (Bago & Boras, 2012).

Encoding dictionaries

Two main problems arise while encoding dictionaries. One problem is related to the structure of the entry, while the other problem is related to the information found within the entry. The structure of entries varies widely among and within dictionaries. Since the encoding scheme is supposed to be suitable for various entry structures, it allows every element of the scheme to appear anywhere in a dictionary entry. However, there are dictionaries that have consistent structure, thus the need for the scheme to support such cases. Since della Bella's dictionary has a complex, but still consistent structure, we decided to encode the entries with the element <entry>, which is intended for consistent and regular dictionary entries[3].

The second problem arising while encoding dictionaries is the problem of representation of information within an entry. Most information in dictionaries is implicit or compressed. Therefore, the encoder must decide whether to capture the precise typographic form of the source text or the underlying structure of the information it presents. Some encoders may find it important to be faithful to an original printed version. On the other hand, others may find only lexical information important, thereby changing, adding or deleting the information contained in the printed form. Both approaches are also possible to combine[4].

We decided to encode old Croatian dictionaries by keeping the information contained in the printed form as content of the elements, while additional information that cannot be explicitly found in the original is encoded as attributes

[3] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html

[4] TEI Consortium (ur.). "9. Dictionaries".TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Verzija 2.6.0). TEI Consortium. http://www.tei-c.org/Guidelines/P5/ (10. 7.2014.).

of the elements. That way we separate the original content from the new, added content.

Scheme of della Bella's dictionary entries

In this section we present the scheme structure of entries found in della Bella's dictionary.

The root element for an entry is the element <entry>. The element <form> groups all the information on the written and spoken forms of one headword[5]. The attribute @type classifies what form is encoded, i.e. simple, compound, etc. The value for the headword itself is "lemma", which denotes that the headword is in its elementary form. Additional attribute is used to define the language of the headword as important information for bilingual and multilingual dictionaries. Since della Bella's dictionary has only an orthographic form of the headwords, the element <orth> is used[6].

All punctuation characters are explicitly encoded with the element <pc> (punctuation character)[7]. There is a space before every punctuation character that is not part of an abbreviation. There is no space before a punctuation character that is part of an abbreviation.

According to the description given above, the headword "Abate" is encoded in the following way:

```
<form type="lemma" xml:lang="it">
<orth>Abate</orth>
<pc>.</pc>
</form>
```

Within a dictionary entry, a headword can additionally be found in a form other than the usual dictionary form. Generally, not whole words can be found but its suffixes. The reason for this is saving space. These suffixes are encoded with the value "inflected" of the @type attribute. A suffix "ta", which represents a suffix of the singular noun in genitive of the translation "Opat", is encoded as following:

```
<form type="inflected">
<gramGrp>
<case value="genitive"/>
<number value="singular"/>
  ta
<pc>.</pc>
</gramGrp>
</form>
```

[5] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html

[6] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-orth.html

[7] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pc.html

Three new elements can be found above. The element <gramGrp> (grammatical information group) groups morpho-syntactic information about a lexical item[8]. The element <case> contains grammatical case information for a given form, and its value of the case in an attribute @value[9]. In this case the value of the attribute @value is "genitive". The element <number> indicated grammatical number associated with a form and its value of the attribute @value[10]. In this case the value is "singular". The elements <case> and <number> are not explicitly given, but are implicitly given via the suffix. Therefore these elements are added as empty elements. As we mentioned before, the content of the elements are only those parts found in the original printed version.
Furthermore, with the element <form> a compound can be encoded that is formed from simple lexical items, one being the headword. It is encoded with the value "compound" of the @type attribute. The element <oRef/> (orthographic-form reference) is used to indicate a reference to the orthographic form(s) of the headword[11]. A compound "Dignita d'Abate" of the headword "Abate" is encoded as following:

```
<form type="compound">
<orth>Dignità d'<oRef/></orth>
<pc>.</pc>
</form>
```

In the printed dictionary following a suffix for a singular noun in genitive is an abbreviation for the morphological gender of a lexical item, which we encode with the element <gen> and the attribute @value[12]. Apart from an abbreviation for the gender, we added an element for the part of speech (<pos>) and the value of the assigned part of speech is recorded with the attribute @norm[13]. The element <pos> is an empty element, since it is not explicitly given in the printed dictionary. An abbreviation for a masculine noun is encoded as following:

```
<gramGrp>
<pos norm="noun"/>
<gen value="m">m<pc>.</pc></gen>
</gramGrp>
```

The element <sense> groups together all information relating to one word sense in a dictionary entry, i.e. definitions, examples and translations. With the attrib-

[8] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html

[9] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-case.html

[10] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-number.html

[11] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html

[12] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gen.html

[13] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pos.html

ute @n one can encode the number of different senses found within an entry[14]. Inside this element, the element <cit> (cited quotation) with a value "translation" of attribute @type contains a translation of the headword, while the value "example" of the same attribute contains an example of the headword[15]. A mandatory element within the <cit> is <quote>, which contains a phrase or passage attributed by author of the dictionary to some agency external to the text[16]. A translation of the entry "Abate" on Latin is encoded as following:

```
<cit type="translation" xml:lang="la">
<quote>Abbas
<pc>,</pc></quote>
</cit>
```

An example of Croatian translations are encoded as following:

```
<cit type="translation" xml:lang="hr">
```

while in examples there is a source of the example given, tagged with the element <bibl>[17]. Since there is no example in the entry "Abate", here we present an example found in the entry "Anno":

```
<cit type="example" xml:lang="hr">
<quote>Evo gre pet godin', dàsam gne sluga ja <pc>;</pc></quote>
<bibl>Sciſc<pc>.</pc></bibl>
</cit>
```

Sometimes dictionaries contain unclassifiable piece of information to guide sense choice, which is encoded with the element <usg> and attribute @type with value "hint"[18]. In the following example the lexicographer describes the headword "Affrica" more precisely by stating that it is "one of the four parts of the World":

```
<usg type="hint">
una delle quattro parti del<lb/>Mondo <pc>.</pc>
</usg>
```

Dictionaries have many references to other entries within the same dictionary, which are encoded with the element <xr> (cross-reference phrase) and attribute @type with value "see" 19. Within this element, there is the element <ref> (reference) that defines a reference to another location, in this case another headword20. To define the location of the referenced entry, we use the attribute

[14] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html

[15] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html

[16] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-quote.html

[17] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-bibl.html

[18] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html

[19] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html

[20] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html

@target with the value of the location of the entry, as in the following example of the entry “Abbarbicare“:

```
<xr type="see">V<pc>.</pc>
<ref target="#barbare">Barba-re</ref>
<pc>.</pc>
<ref target="#radicare">Radicare</ref>
<pc>.</pc>
</xr>
```

In this case, the target entries have to have an @xml:id with a unique value within the XML document in the <form> element. An example of the entry “Barbare” referenced in the entry “Abbarbicare“ is encoded as following:

```
<form type="lemma" xml:lang="it" xml:id="barbare">
<orth>Abate</orth>
<pc>.</pc>
</form>
```

Finally, we encode line breaks (<lb/>), column breaks (<cb/>) and page breaks (<pb/>).

Below we present the whole content of the entry “Abate” as found in the della Bella’s dictionary:

Abate . *Abbas , tis* . *m.* Opat , ta . m. Igu-|men , ena . m. Dignità d’Abate . Opat-|ſtvo , va . n. Igumenſtvo , tva . m.

and the encoding of the corresponding digitized dictionary entry:

```
<entry>
<form type="lemma" xml:lang="it">
<orth>Abate</orth>
<pc>.</pc>
</form>
<sense n="1">
<cit type="translation" xml:lang="la">
<quote>Abbas
<pc>,</pc></quote>
<form type="inflected">
<gramGrp>
<case value="genitive"/>
<number value="singular"/>
   tis
<pc>.</pc>
</gramGrp>
</form>
<gramGrp>
<pos norm="noun"/>
<gen value="m">m<pc>.</pc></gen>
</gramGrp>
```

```
</cit>
<cit type="translation" xml:lang="hr">
<quote>Opat
<pc>,</pc></quote>
<form type="inflected">
<gramGrp>
<case value="genitive"/>
<number value="singular"/>
    ta
<pc>.</pc>
</gramGrp>
</form>
<gramGrp>
<pos norm="noun"/>
<gen value="m">m<pc>.</pc></gen>
</gramGrp>
</cit>
<cit type="translation" xml:lang="hr">
<quote>Igu-<lb/>men
<pc>,</pc></quote>
<form type="inflected">
<gramGrp>
<case value="genitive"/>
<number value="singular"/>
   ena
<pc>.</pc>
</gramGrp>
</form>
<gramGrp>
<pos norm="noun"/>
<gen value="m">m<pc>.</pc></gen>
</gramGrp>
</cit>
</sense>
<sense n="2">
<form type="compound">
<orth>Dignità d'<oRef/></orth>
<pc>.</pc>
</form>
<cit type="translation" xml:lang="hr">
<quote>Opat-<lb/>ſtvo
<pc>,</pc></quote>
<form type="inflected">
```

```
<gramGrp>
<case value="genitive"/>
<number value="singular"/>
    va
<pc>.</pc>
</gramGrp>
</form>
<gramGrp>
<pos norm="noun"/>
<gen value="n">n<pc>.</pc></gen>
</gramGrp>
</cit>
<cit type="translation" xml:lang="hr">
<quote>Igumenſtvo
<pc>,</pc></quote>
<form type="inflected">
<gramGrp>
<case value="genitive"/>
<number value="singular"/>
    tva
<pc>.</pc>
</gramGrp>
</form>
<gramGrp>
<pos norm="noun"/>
<gen value="m">m<pc>.</pc></gen>
</gramGrp>
</cit>
</sense>
</entry>
```

Conclusion

To enable semantic interoperability of digitized historical dictionaries, the dictionaries have to be encoded using some standard. In this paper we present the successful encoding of della Bella's 18th century dictionary entry using a TEI (Text Encoding Initiative) encoding scheme for dictionaries. We consider the TEI scheme an adequate method for encoding dictionaries, since it is a *de facto* standard for encoding textual resources in digital form, spread within the field of digital humanities. Additionally, we consider it adequate, considering it enables us to encode all the necessary information found in dictionaries, regardless of its complicated structure.

The process of encoding digitized dictionary entries is a manual task, which is time consuming and tedious, and paves a way for an encoder to introduce in-

consistencies. If the process of encoding could be (semi)automated, it would reduce cognitive load in encoders and time spent on the task. We could further apply natural language processing methods (such as machine learning) to try to speed up the process of encoding the structure of the dictionary entries.
By using the TEI scheme, we could additionally link the entries and concepts to external resources, such as online encyclopaedias (i.e. the Croatian Encyclopaedia of the Miroslav Krleža Institute of Lexicography[21] or the free-content Wikipedia[22]), online dictionaries (i.e. Croatian language portal[23]), Linked Data[24] or DBpedia[25].

References

"About These Guidelines" in TEI P5: Guidelines for Electronic Text Encoding and Interchange. TEI Consortium (ed.).. http://www.tei-c.org/Guidelines/P5/ (10. 7.2014.).

Bago, Petra; Damir, Boras. A framework for consolidating most important digitized Croatian dictionaries from 1595 to 1945 // 17. International Information Technology and Journalism conference. Dubrovnik, 2012.

"bibl" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-bibl.html (25.09.2014.).

Boras, Damir (principal researcher). Hrvatska rječnička baština i prikaz rječničkoga znanja; Znanstveni projekt, Zagreb, 2003-2006. http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451 (25.8. 2014.).

Boras, Damir (principal researcher). Hrvatska rječnička baština i hrvatski europski identitet; Znanstveni projekt, Zagreb, 2007-2013. http://zprojekti.mzos.hr/public/c2prikaz_det.asp?cid=1&psid=25&ID=451 (25.8. 2014.).

"case" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-case.html (16.09.2014.).

"cit" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cit.html (21.09.2014.).

The Croatian Encyclopaedia of the Miroslav Krleža Institute of Lexicography. http://www.enciklopedija.hr/ (05.09.2014.).

The Croatian Language Portal. http://hjp.novi-liber.hr/ (05.09.2014.).

DBpedia. http://wiki.dbpedia.org/ (05.09.2014.).

della Bella, Ardelio. Dizionario italiano-latino-ilirico. Nella Stamparia Privilegiata, Prima edizione Ragusea, 1785.

"Dictionaries" in TEI P5: Guidelines for Electronic Text Encoding and Interchange ed. The TEI Consortium (ed.). 2014. V. 2.6.0. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html.

"entry" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-entry.html (15.09.2014.).

eXtensible Markup Language (XML). http://www.w3.org/XML/ (10.7.2014.).

"form" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-form.html (15.09.2014.).

Fostering Language Resources Network (FLaReNet). https://www.flarenet.eu/ (5. 9. 2014.).

"gen" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gen.html (18.09.2014.).

[21] http://www.enciklopedija.hr/

[22] https://en.wikipedia.org/wiki/Main_Page

[23] http://hjp.novi-liber.hr/

[24] http://linkeddata.org/

[25] http://wiki.dbpedia.org/

"gramGrp" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gramGrp.html (16. 09.2014.).

Ide, Nancy; Pustejovsky, James; Calzolari, Nicoletta; Soria, Claudia. The SILT and FLaReNet International Collaboration for Interoperability. Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009, str. 178-181.

Linked Data – Connect Distributed Data across the Web. http://linkeddata.org/ (05.09.2014.).

"number" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-number.html (18.09. 2014.).

"oRef" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-oRef.html (18.09. 2014.).

"orth" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-orth.html (15.09.2014.).

"pc" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pc.html (15.09.2014.).

"pos" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pos.html (18.09.2014.).

Pustejovsky, James; Ide, Nancy. "What Does Interoperability Mean, Anyway? Toward an Operational Definition of Interoperability," 2010. http://www.cs.vassar.edu/~ide/papers/ ICGL10.pdf (5. 9. 2014.).

"quote" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-quote.html (25.09. 2014.).

"ref" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html (28.09.2014.).

"sense" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sense.html (18.09. 2014.).

Sustainable Interoperability for Language Technology. http://www.anc.org/LAPPS/SILT/index. html (5.9.2014.).

TEI Consortium (ed.). TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Version 2.6.0). TEI Consortium. http://www.tei-c.org/Guidelines/P5/ (10. 7.2014.).

"usg" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-usg.html (28.09.2014.).

Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Main_Page (05.09.2014.).

"xr" in TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-xr.html (28.09.2014.).

BIG DATA ANALYTICS AND APPLICATIONS

Software as a Service (SaaS) Quality Management and Service Level Agreement

Weimei Pan
University of British Columbia
Vancouver, Canada
weimei.pan@ubc.ca

Grant Mitchell
International Federation of Red Cross and Red Crescent Societies
Geneva, Switzerland
grant.mitchell@ifrc.org

Summary

This paper discusses the finding of a case study concerning the service quality of a Software as a Service (SaaS) application. It was found that while the service level agreement could help guarantee the level of service promised, the service customer would resort to a reactive approach rather than a proactive approach in terms of monitoring the service delivered. In addition, trust and a partnership relationship between the service customer and the service provider are also critical in maintaining the service quality.

Keywords: Software as a Service; SaaS; Service Level Agreement; Cloud Quality Management; Service Quality

Introduction

Software as a Service (Hereafter SaaS) is an emerging software delivery model based on cloud computing. It is the capability "to use the provider's applications running on a cloud infrastructure" via a network, primarily, Internet, through a web browser or a program interface on a pay-per-use basis (Mell and Grance 2009, p.2). SaaS shares the characteristics of every cloud computing delivery model, namely flexibility, scalability, ease of use, and redundancy. As SaaS applications increasingly become an integral part of an organization's IT structure, one important issue is how to manage and guarantee the quality of the service delivered. The aim of this paper is to contribute to this area through presenting the finding of an InterPARES Trust[1] project case study. As one of many child projects of the InterPARES Trust project, this case study examines how the is-

[1] The InterPARES Trust is an on-going multi-national, interdisciplinary research project, which explores issues concerning digital records and data entrusted to the Internet. To learn more about the project, please go to https://interparestrust.org/

sues raised (e.g., the management of SaaS application project, contract terms, cloud contract negotiation, retention and disposal scheduling of electronic records, and privacy protection of personal data) by cloud-based services are being addressed in practice. Case study methodology ensures a contextualized and in-depth understanding of the use of SaaS application within an organization. The chosen case is an e-recruitment SaaS application (*E-Recruitment*) recently launched by the International Federation of Red Cross and Red Crescent Societies (the International Federation) located in Switzerland.

We will present the finding of our analysis of the service level agreement between the International Federation and the service provider of *E-Recruitment* in relation to the service quality: a couple of other factors (i.e. trust, and partnership) that influence the quality of SaaS application will be discussed as well. The remainder of this paper is organized as follows: first, a brief review of literature related to IT service quality management and SaaS quality management is presented; next, the methods for data collection and data analysis and the chosen SaaS application (i.e., *E-Recruitment*) in terms of its functionality and the data it collected and stored are introduced; then, the research findings on service level agreement, trust, and partnership are discussed; the article will conclude by summarizing the main points of this paper and identifying areas for future research.

Literature review

From IT service quality management to Software as a Service (SaaS) quality management

The concept of IT service quality management belongs to the much broader research of quality management. According to the international standard ISO 8402, quality is "the total characteristic of a product or service concerning its suitability to fulfill predefined requirements" (ISO8402). Accordingly, quality management is quality-oriented management aiming to continuously improve the quality of products or services; it "defines quality policies, quality targets, quality controlling and continuous quality improvement within an organization" (Praeg and Spath 2009, n.p.).

IT service quality management is an extension of quality management with a focus on IT service. There are a variety of views on IT services, e.g., components of applications, part of IT organization, set of utilities, and business assets (Spath, Bauer and Praeg 2010). In this paper, IT services are understood as "a set of utilities used by business processes", which can be provided by either an internal or external IT provider (Spath, Bauer and Praeg 2010, p.7). IT service quality management uses a quality management framework to accommodate the different dimensions and views on quality (i.e., both internal and external perspectives {e.g. market developments, customer demands}) and incorporate the different levels of management and managing processes (e.g. strategic, process and infrastructure) (Spath, Bauer and Praeg 2010). It builds on IT service man-

agement frameworks (e.g. IT Infrastructure Library {ITIL}, ISO/IEC 20000, and Control Objectives for Information and Related Technology {COBIT}) yet differs by its comprehensive and thorough view of the quality improvement process.

Quality management process is comprised of four steps: plan, execute, control, and improve (Spath, Bauer and Praeg 2010). In the plan phase, the IT service quality targets will be defined based on an understanding of the business strategy, customer's expectation, and others. Next, in the execution phase, the IT service management process will be evaluated and assessed against the defined targets. Then, in the control phase, possible gaps between the IT service management reality and the targets, and reasons for such gaps will be identified. Finally, proper procedures will be implemented to address these gaps and to improve the overall quality level. In addition to the quality management process, in the IT service quality framework that Spath, Bauer and Praeg (2010) introduced, they mapped different quality methodologies along the IT service lifecycle phase, including requirement engineering, sourcing and procurement, design/test and orchestration, operational maintenance/support, and replacement.

As SaaS applications increasingly become an integral part of an organization's IT service, it is necessary to consider the implications of using SaaS applications for an organization's IT service quality management. Existing studies on the quality of SaaS applications are mostly focused on identifying the SaaS application attributes for quality measurement (e.g., Burkon 2013; Wen and Dong 2013; Khanjani et al. 2014; Benlian et al. 2012). Some may regard SaaS as one type of cloud service, and investigate the attributes of cloud computing as a whole (e.g., Zheng et al. 2014; Thoss et al. 2014). Regardless, recognizing the service-oriented feature of SaaS, a considerable number of these studies draw on the conceptual frameworks of service quality, including the set of attributes concerning quality in a service-oriented environment designed by O'Brien et al. (2007), and SERVQUAL, a multiple-item scale for measuring customer perceptions of service quality developed by Parasuraman et al. (1988).

Though all these studies acknowledge the set of characteristics of cloud computing in general, and SaaS applications in particular to be the foundation to generate the attributes for quality measurement, the way these characteristics is conceptualized differs greatly one from the other. As a result, the same concept may mean different things in different frameworks, and seldom are two sets of attributes exactly the same. In addition, most of these frameworks focus on the non-functional requirements, with functional requirements often unaddressed. What's more, thus far, most studies concentrate on the "control" step of the whole quality management lifecycle (i.e., plan, execute, control, and improve), seldom are there studies that investigate SaaS quality from the whole lifecycle point of view.

Service level agreement

A Service Level Agreement (Hereafter SLA) is "a binding contract between the service provider and the service customer, which outlines the responsibilities, describes the service to be provisioned, defines the service commitment guarantees, penalties for non-compliance, and emergence contacts" (Qiu et al. 2013, p.730). Considering the service-oriented feature of cloud-based service, the establishment of a commonly agreed SLA and the service provider's commitment to the SLA can assure the service customer of the expected levels of service (Sun et al. 2012). It is also a vital instrument helping the service customer define and measure the service level in quality management.

Some of existing studies have identified the main components of SLA. For instance, Baset (2012) states that a typical SLA has the following components: *service guarantee*, *service guarantee time period*, *service guarantee granularity*, *service guarantee exclusions*, *service credit*, and *service violation measurement* and *reporting*. Gulia and Sood (2013) define that the main sections found in most SLA are *definitions*, *performance management* (the service commitments), *problem management* (SLA exclusions), and *remedies* (service credit, and credit request/claim). Despite the differences in the terms used, the two lists are consistent in their identification of several major components: *definition*, *performance* (service guarantee period, service guarantee granularity), *limitation of liability*, *service credit*, and *remedies*. Bon et al. (2007), also propose the minimum contents that a SLA should include *description of service*, *service targets*, *communications and reporting*, *authorization details and validity period*, *financial management details*, *service provider liability and obligations*, *customer responsibilities*, *supporting and related services*, *impact*, *urgency and priority guidelines*, *service hours*, *date exceptions*, *critical business periods and out-of-hours cover*, *workload limits*, *contact details of people authorized to act in case of emergencies*, *actions to be taken in the event of a service interruption*, *escalation and notification process*, *scheduled and agreed interruptions*, *complaints procedures*, *housekeeping procedures*, *exceptions to the terms given in the SLA*, and *glossary of terms*.

Other studies focus on conceptualizing cloud SLA through systematically analyzing publicly available SLA (e.g., Qiu et al. 2013; Gulia and Sood 2013; Baset 2012). These studies usually discuss what attributes are present in these SLAs, what attributes are missing from these SLAs, what attributes should be included in SLAs, and others. For instance, Qiu et al. (2013) summarized that cloud SLA attributes include *responsibilities of parties*, *service definition*, *service guarantee*, *availability definition*, *calculation method*, *monitoring mechanism*, *security policy*, *privacy policy*, *data protection policy*, etc. Qiu et al. (2013) found that *security*, *privacy*, *protection*, and *backup* policies are generally missing in most SLAs.

These studies help reveal the status quo of cloud-based service in terms of what service levels are agreed upon and what are the relationships between service

provider and customer. However, in order that SLAs serve as effective assurance for the expected services, more research is required to both refine the pre-SLA risk assessment and ensure post-SLA monitoring and prediction (Sun et al. 2012). In addition, proactive SLA assurance should be a mutual achievement between the service provider and the service customer, and therefore should be examined from these two perspectives (Sun et al. 2012). This requires us to not only understand SLA in the context of SaaS adoption, implementation, and maintenance, but also in relation to the whole lifecycle of quality management.

Methods

Data collection and data analysis

In order to gain an in-depth understanding of the adoption of a SaaS application and how the issues raised by the use of a SaaS application are addressed in practice, multiple methods have been used to collect data, including policy analysis (e.g., information security classification standard, ICT security policy, and cloud services request form), contract analysis (e.g., terms and conditions, and service level agreement), system analysis, and semi-structured interviews. Data collection started in December 2014 and ended in February 2015.

Ten interviews were conducted with staff from the legal department (2), library and archives unit (1), IT department (1), the human resources department (5), and risk and audit department (1). Five interview guides were developed prior to the interview for staff from each department. While in general the interview questions prompted interviewees to describe how the issues raised by the use of this SaaS application were mitigated, depending on the role and responsibilities of each interviewee in the adoption, implementation and use process, interview questions were geared towards the area with which each person is familiar, to ensure a comprehensive understanding of the adoption, implementation, and use of this SaaS application. For instance, the staff from the IT department, who is also the project manager of the *E-Recruitment* project, was asked to discuss more about the management of this SaaS application project, contract negotiation process, and the organization's cloud strategy.

Each interview lasted between 25 minutes to 90 minutes. Interviews were audio recorded and transcribed afterwards. All personal identifying information was removed; each interviewee was assigned a unique identification number, e.g., interviewee1, interviewee2. A thematic analysis was performed on the data collected. Reports on the research has been sent back to the International Federation staff to do an accuracy check.

The SaaS application

Formally launched in January 2014, *E-Recruitment* is one of several SaaS applications that the International Federation has adopted. It is a public-cloud based e-recruitment application aiming to streamline, standardize, and improve the recruitment process. Consistent with the recruitment process, the core func-

tionalities of *E-Recruitment* include creating a new job opening, approving the new job opening, advertising the job opening on internal and external career pages, selecting the preferred candidate, and making an offer. Types of information collected and stored by *E-Recruitment* include candidate data (e.g., name, address, CVs, references, and application data), contract data, notes on the interviewing and evaluation of candidates, policies and guidelines, job description, and career information on the public website.

Finding

Service level agreement

The contract between the International Federation and the service provider of *E-Recruitment* is comprised of the following documents: order form, terms and conditions, terms and conditions for other services, service level agreement, and documentation of the functionalities provided by the service provider. To understand the service level agreed upon between the International Federation and the service provider, we conducted an analysis of the SLA against the conceptual SLA framework identified by previous research studies. More specifically, two analyses were conducted: one on the components of SLA and one on SaaS quality of service attributes.

The analysis of the major components of the SLA between the International Federation and the service provider of *E-Recruitment* shows that, on the one hand, the sections identified by previous studies are all valid as they are all present in this SLA, albeit, in more detailed forms (e.g., *performance*); on the other hand, the analysis reveals that some lists identified by previous studies are not sufficient. For instance, while the list of contents identified by Bon et al. (2007) is able to accommodate all the sections in this SLA, the rest of the lists are not adequate in that they may fail to identify a couple of other sections, e.g., *maintenance, customer's responsibilities,* and *support*.

In terms of *performance*, previous studies indicate that it specifies the service metrics a provider strives to meet, including *availability*, *response time*, and *disaster recovery*. In reality, it can cover a wide range of metrics, including *availability*, *response time*, *security* (*data center security*, *data center environment*, *data center network*, *monitoring*, *system backup*, *data security*, and *disaster recovery*), *data recovery*, and *acceptable use policy*. In terms of the *customer's obligation*, this describes the responsibility on the customer's part to ensure the performance of the application, e.g., providing necessary computer system, hardware, software, and telecommunication equipment, installing necessary virus protection software, upgrading its system regularly, and others. The *maintenance* section outlines the *maintenance schedule*, *notification*, *test of the release*, *training*, *influence on the existing setup and data*, and *emergency patches*. The *support* section basically outlines the method and lifecycle of incident resolution: *incident notification*, *incident severity classification*, *incident reporting*, *incident response*, *incident resolution*, and *escalation procedure*.

The inclusion of *maintenance* and *support* in an SLA is necessary. One reason that users prefer SaaS applications is that software upgrades will be delegated to the service provider, which not only guarantees the quality of the application, but also saves users from the lengthy and costly upgrade process. As a service-oriented computing model, *support* should be an essential part of the service provided. With very limited control of the application and the underlying infrastructure, the customer largely relies on the service provider to fix any issues arising in the use of the application.

The analysis of the International Federation's SLA regarding the SaaS quality of service attributes yielded different results. For instance, the list of quality of service attributes identified by Qiu et al. (2013) works very well in the sense that the majority of them are present in the SLA, including *responsibilities of parties*, *availability definition*, *calculation method*, *monitoring mechanism*, *violation penalty*, *violation penalty onus*, *remedy calculation*, *termination policy* and others. The few attributes that are absent from the SLA are *change control*, *privacy policy*, and *service definition* (there is a separate document focusing on documenting the functionalities offered). However, some of attributes (e.g., *interoperability*, *modifiability*, *usability*, *testability*, *resiliency*, *composability*, *extensibility*) identified by Khanjani et al. (2014) and Burkon (2013) are not present in the SLA as these are abstract attributes that usually subject to different interpretations. This means that though the SLA can serve as a good source to evaluate the quality of service of the service provider, the SLA alone is not adequate to determine and guarantee all the quality of service.

Trust

Once an SLA is agreed upon and signed by the service provider and the service customer, a benchmark is established against which the service customer can assess the service delivered and claim for remedies where service failure is detected. However, the post-SLA assessment of the service level might not be as easy as expected. One of the primary difficulties is that customers usually do not know how to monitor the service promised. Cloud monitoring is of vital importance for both cloud service providers and cloud service consumers, as, on one hand, it can assist providers in operating and managing cloud service in order to continuously adapt to customers' demands; on the other hand, it provides information and Key Performance Indicators (KPI) so that consumers can see the service they received (Aceto et al. 2013). There are many cloud operational areas requiring cloud monitoring, such as accounting and billing, SLA management, service/resource provisioning, security management, and fault management (Fatema et al. 2014). Yet, thus far, there are insufficient studies on cloud monitoring (Aceto et al. 2013). Analyses of general-purpose monitoring services (e.g., Nagios) and cloud monitoring services (including provider dependent monitoring services, e.g., CloudKick and provider independent monitoring services, e.g., Monitis) show that "current monitoring tools are least sup-

portive of security and privacy assurance management in Clouds" (Fatema et al. 2014, p.2928).
In the case of *E-Recruitment*, cloud monitoring is absent from the cloud services management portfolio. This lack of technical assurance is balanced by several other factors: trust, due diligence in the selection of the service, not putting highly sensitive information in the cloud, the SLA, and others. Among these, trust is one of the most intangible factors and is often associated with a reactive approach in SLA commitment monitoring and potential risks. The risk of misplaced trust in a service provider can be very high if due diligence on the part of the customer is absent, but in this case the risk of trust being misplaced is limited due to the factors listed above.

Partnership
Technically, there is a customer-provider relationship between the service provider and the service user. And as per the SLA, the service provider has the responsibility to deliver the service promised and provide remedies where promises are unmet. But in reality, according to our interviewees, the customer and provider work more like partners in the delivery of the SaaS application.
One common term in SLA is service credit, which usually is expressed as a percentage of the weekly, monthly, or yearly subscription fee and will be credited to the next invoice. The onus is often on the customer to claim the credit within certain time period. From the customer's perspective, they want the service purchased to perform well all the time. But from the provider's perspective, certainly, they will not feel comfortable with committing to one hundred percent performance. As a result, the service provider often uses service credit as an option to compensate any service failure. But for the customer, service credits certainly cannot compensate for whatever the organization might suffer from the service failure. As interviewee4 observed,

> "This is why the relationship side is more important than the contractual document. It is really working on how you make that relationship work, what you get out of them, how you shift the balance of power, what respective benefits both parties get."

Indeed, the willingness to enter a five-year contract in this case requires flexibility, and cooperation from both parties to make this relationship work and to guarantee the quality of the service delivered. For instance, immediately after *E-Recruitment* went live, there were some issues with the service provider's platform. As a result, the International Federation experienced a series of incidents with *E-Recruitment*. To make matters worse, there were some issues with the communication with the support desk because of the way the ticketing system worked. An incident report would go from one level to another, in a long process to finally reach the support technician who could help resolve the issues.

As interviewee7 commented, they were not satisfied with the service at that time. The International Federation discussed the issues with the service provider, which took them seriously. As a result, not only the International Federation got the service credits they deserved but also the communication channel between customer and service provider has been modified so that the International Federation can now contact the support technician directly concerning any technical issues.

Conclusion

In comparison with previous software delivery models (e.g., in-house developed software, and Application Service Provider-oriented application {ASP}), SaaS applications have some unique features which require special consideration with regard to service quality management. As a binding contract, the SLA outlines the service level service providers should deliver and customers should expect. But the fulfillment of the SLA requires proactive monitoring of the service, which the customer might not know how to carry out. More research is needed to develop a customer-friendly service monitoring tool. Moreover, SLA is simply one link in the chain of the quality management process. To obtain high service quality requires due diligence in the selection of service, negotiating a balanced contract, and other procedures. A good understanding of SLA will be gained if it can be situated in the context of the SaaS application adoption, implementation, and use lifecycle.

In addition, our research found that service quality often relies on more than contractual documents. Trust and a partnership relationship also play an important role in guaranteeing and improving the SaaS application. Future research is needed to explore the relationships of different factors in the service quality of SaaS application, e.g., trust, partnership, contract, due diligence in the use of SaaS application and others.

Acknowledgement

We gratefully acknowledge the financial support from the InterPARES Trust project. We also would like to thank the accommodation and assistance of the staff of the International Federation where this case study was conducted, which made this study possible.

References

Aceto, Giuseppe; Botta Alessio; De Donato, Walter; Pescape, Antonio. Cloud monitoring: A Survey. // *Computer Networks The International Journal of Computer and Telecommunications Networking*. 57(2013),9;2093-2115.

Baset, Salman. Cloud SLAs: Present and Future. // *ACM SIGOPS Operating Systems Review* 46 (2012), 2; 57-66

Benlian, Alexander; Koufaris Marios; and Hess, Thomas. Service Quality in Software-As-A-Service: Developing the Saas-Qual Measure and Examining its Role in Usage Continuance. // *Journal of Management Information Systems* 28 (2012), 3; 85-126

Bon, Jan van; Jong, Arjen de; Kolthof, Axel; Pieper, Mike; Rozemeijer, Eric; Tjassing, Ruby; Veen, Annelies van der; Verheijen, Tieneke. IT Service Management: An Introduction. Zaltbommel: Van Haran Publishing, 2007.

Burkon, Lukas. Quality of Service Attributes for Software as a Service. // *Journal of Systems Integration* 4 (2013), 3; 38-47

Fatema, Kaniz; Emeakaroha, C. Vincent; Healy, D. Philip; Morrison, P. John; Lynn, Theo. A Survey of Cloud Monitoring Tools: Taxonomy, Capabilities and Objectives. // *Journal of Parallel and Distributed Computing*, 74 (2014) 10; 2918-2933

Gulia, Preeti; Sood, Sumedha. 2013. Comparative Analysis of Present Day Clouds Using Service Level Agreements. // *International Journal of Computer Applications* 71 (2013), 3

Khanjani, Atieh; Wan Nurhayati Wan Ab Rahman, Abdul Azim Abd Ghani, and Abu Bakar Md Sultan. SaaS Quality of Service Attributes. // *Journal of Applied Sciences* 14 (2014), 24; 3613-3619

Mell, Peter; Grance, Tim. The NIST Definition of Cloud Computing. *National Institute of Standards and Technology* 53 (2009), 6.

O'Brien, Liam; Merson, Paulo; and Bass, Len. Quality Attributes for Service-Oriented Architectures. // *SDSOA '07 Proceedings of the International Workshop on Systems Development in SOA Environments* / IEEE, 2007

Parasuraman, A.; Zeithaml, Valarie A.; and Berry, Leonard L. SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. // *Journal of Retailing*, 64 (1988), 1; 12–40.

Praeg, Claus-Peter; Spath, Dieter. Perspectives of IT-Service Quality Management: A Concept for Life Cycle Based Quality Management of IT-Services. // *Information Technology Governance and Service Management: Frameworks and Adaptations* / Cater-Steel, A. (ed.) Hershey: Information Science Reference, 2009

Qiu, Meng Maggie; Zhou, Ying; and Wang, Chen. Systematic Analysis of Public Cloud Service Level Agreements and Related Business Values. // *2013 IEEE International Conference on Services Computing* (2013); 729-736

Spath, Dieter; Bauer, Wilhelm; Praeg, Claus-Peter. IT Service Quality Management: Assumptions, Frameworks and Effects on Business Performance // *Quality Management for IT Services: Perspectives on Business and Process Performance* / Praeg, Claus-Peter; Spath, Dieter (eds.) Hershey, PA, USA: IGI Global, 2010, 1-21

Sun, Le; Singh, Jaipal; Hussain, Omar. Service Level Agreement (SLA) Assurance For Cloud Services: A Survey From A Transactional Risk Perspective. // *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia* / Khalil, Ismail (ed.). New York: ACM, 2012, 263-266

Thoss, Yvonne; Pohl, Christoph; Hoffmann, Madlain; Spillner, Josef; and Schill, Alexander. User-Friendly Visualization of Cloud Quality. // *2014 IEEE 7th International Conference on Cloud Computing* / (2014), 890-897

Wen, Pang Xiong; Dong, Li. 2013. Quality Model for Evaluating SaaS Service. // *2013 Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)* / IEEE, 2013, 83-87

Zheng, Xianrong; Martin, Patrick; Brohman, Kathryn; and Xu, Li Da. CLOUDQUAL: A Quality Model for Cloud Services. // *IEEE Transactions on Industrial Informatics* 10(2014), (2); 1527-1536

Big Data in Data Warehouses

Tomislav Šubić, Patrizia Poščić, Danijela Jakšić
Department of Informatics, University of Rijeka
Radmile Matejčić 2, Rijeka, Croatia
tsubic@student.uniri.hr, patrizia@inf.uniri.hr, dsubotic@inf.uniri.hr

Summary

The amount of data grows every day. Along with that, new challenges and needs appear. We wonder how to collect and process this enormous amount of data and to get the information we need for business management. Tools and technologies for that task should be scalable, distributed and with high bandwidth so we can quickly get to the information we want. As a solution impose NoSQL databases. But databases without a high quality system for processing and analysis of data are not very helpful. Some solutions that have started to address these problems appeared, but there is still room for development. Some of these solutions are Apache Hadoop and Hive, which are used every day in big companies like Facebook and Google. This paper provides insight into existing technologies that implement such solutions, which have their advantages and disadvantages, and we will try to predict how this story will develop in the future.

Keywords: big data, data warehouse, NoSQL, Hadoop, Hive

Introduction

The information revolution that is now taking place leaves a great impact on all types of businesses and business organizations. With all the tools available it is becoming easier to collect data of the operations of an organization, the user needs etc. and based on the analysis of this data to make better business decisions. The first step was made in 1970 by E. Cobb, when he introduced the relational database model, which remains the standard to this day for the storage and processing of data in information systems. However, 40 years later, customer needs are changing, and in spite of all the adjustments and optimizations of the relational model there are new technologies that are held behind by this model. For better analysis it is necessary to collect as many as possible high quality data from all aspects of the business. A new trend called Big Data appears; this represents the large amounts of data that usually come from sources such as social networks, publicly available information from various agencies, subscription information, news, documents, e-mails etc. (and include structured or unstructured text), as well as digital devices and sensors, location data from smartphones, weather etc. Most companies are not familiar with collecting and

processing large amounts of data from these sources. So much of this data remains lost and unstructured, and thus aren't included into reports and business insights which help in making better business decisions. We can say that most business organisations do not realize the true value of this data. The companies that are trying to collect enormous amounts of data have problems, because they do not fit into the relational model. Unstructured data is the data that has no predefined data model, and the data which does not fit into existing models. Such data is accumulating very fast; these are videos, photos and generally all possible output files of various programs. Furthermore, new technologies such as Web 2.0 applications used by large numbers of users depend on the storage and processing of large amounts of data. Such a large database needs to be distributed, scalable and fast. All this presents great challenges for the relational model and the tools it implements, such as SQL (Stonebraker et al., 2007). Because of these reasons, a big number of organizations such as Facebook, Amazon, Yahoo and Google are using non relational solutions (for specific tasks); popularly known as NoSQL databases. This term first occurred in 1998, to become real competition to the relational model around 2009. Because of the open model, a lot of different companies develop their own non relational solutions, and most of them are available to everyone because they are open source. Of course, apart from raw data storage, there is a more important part, which takes care of the data processing and analysis. Today, we aim to business intelligence systems and their major component is the data warehouse. The technologies that have been developed to solve this problem are sharing some common characteristics associated with the scalability, flexibility and high availability. MapReduce technology together with the so-called Hadoop Distributed File System (HDFS) and the HBase database, as part of the Apache Hadoop project is a new and innovative approach to the processing and analysis of such unstructured data. In this paper we deal with unstructured data, the possibilities of Apache Hadoop technology and the integration of non relational databases into a data warehouse. We will clarify some main concepts, present some existing solutions, and try to show whether Hadoop is a suitable replacement for existing data warehouses. In the second chapter we will speak generally about data warehouses and technologies that are used, in the third chapter we look at non relational databases and give an insight into the different types, further in the fourth chapter we deal with the current Big Data technologies, while in the fifth chapter we discuss data warehouses that implement these technologies. Finally, the sixth chapter provides a description and overview of the Hive data warehouse that uses the described technology.

Data Warehousing

The main purpose of SQL and relational databases is to store and access data from a variety of transactions that occur in everyday business such as orders, invoices, salaries etc., execute queries on this data and generate reports. Further

development of business systems, development of strategic planning and management which occurred because of the globalization of a competitive market, but also the development of hardware and software technology greatly relied on the relational data model and the analytical processing of such data. In recent years we notice the intensive development of Web 2.0 applications. These are applications such as massive online stores, social networks etc., which must process large amounts of data to millions of users simultaneously in real time.
Limitations occur when handling large amounts of unstructured data in traditional data warehouses. Structured data requires a specific scheme to match the relational model. This of course is not the case in unstructured data that does not have a specific scheme. Although SQL supports unstructured text search by keywords, it is limited to just text (Jain et al., 2007). All that led to the development of a completely new model of data warehouses based on the new data models. These models are called NoSQL. The main idea of systems, which are trying to take advantage of the Hadoop technology, is the integration and construction of interfaces with existing data warehouses, or possibly the construction of new types of data warehouses.

Non relational database models

Over the years the term NoSQL turned into something that represents non relational, distributed, horizontally scalable databases. In addition, the most widely used ones are open source. Their main task is to ensure quick retrieval of existing data, as well as storing new data. Since OLAP and other business intelligence systems are critical to many business organisations, the question is how do NoSQL databases fit in such systems? There are several major models that that meet the above criteria and have become the standard (Grolinger et al., 2013):

- column oriented
- document oriented
- key-value database
- graph database.

In this paper, we primarily deal with *column* and *document oriented* databases because they are commonly used in the fields that we discuss in this paper.

Column oriented databases

Unlike the relational model, which stores data in rows (tuples), column oriented databases store data by column, which means that each row has only one data type. Thus, the data from the database is consistently written to the computer storage, without the fragmentation which occurs in implementation by rows. Other benefits include easier retrieval of data; all data fetching usually occurs within one table (no need to merge data from multiple tables), that goes for iterative processing too. Some examples of such databases are Apache Cassandra,

Google BigTable and Hbase. Often is the wrong terminology used to describe NoSQL column oriented non relational databases, namely there is a relational model that distributes data by columns, and is called column store.

Document oriented databases

Databases which simply store data in the form of documents are considered to be the main form of non relational databases. They support data storage in documents of any format. The data itself within the document is structured using XML, JSON or something similar. Today, due to the simple notation, most commonly is used JSON. Popular databases that implement this model are MongoDB, CourchAB and OrientDB.

We will discuss technologies based on non relational databases. Hadoop and Hive both support a variety of NoSql databases.

Big data technologies

Data Vault and Anchor modelling

We cannot talk about data warehouses with a large amount of data without mentioning these two models. Data Vault is a hybrid model that combines the best characteristics of the third normal form and the star schema developed specifically for EDW-a (Enterprise Data Warehousing). Its main concepts are so-called Hub, Link and Satellite entities. All data is stored in the Satellite entities, the keys are stored in the Hubs, while the mutual relation keys are specified in Links (Linstedt, 2011). Anchor modelling is a database modelling technique that is based around the claim that the data warehouse environment is constantly changing and that big changes outside the model should lead to small changes in the model. The aim of using this type of modelling is to achieve a highly decompressed implementation that can deal with the growth of data warehouses. These are highly normalized (6NF) models which are characterized by the ability to store changes in the business environment (Krneta et al., 2014). Still, these are relational models and in this paper we discuss non relational solutions, we will refer to them much.

MapReduce

Developed by Google as a "programming model with the implementation for processing and generating large data sets (Dean and Ghemawat). It is a method which greatly simplifies the distribution of tasks across multiple nodes. Each node processes the data stored on that node (Kumar et al.). It consists of two parts, the first function that is responsible for the so-called mapping of key-value coming from the input data in some other suitable key-value pairs:

$$Map((K1, V1) \Rightarrow (K2, V2))$$

The second, so-called reduce function reduces the value so that the values are first sorted and grouped according to a given key; it gives the final output:

$$Reduce((K2, list(V2)) \Rightarrow list(K3, V3))$$

MapReduce is popular for many reasons:

- automatic parallelization and distribution
- fault tolerant
- easy to use and develop
- programs are written in Java.

Unfortunately, the MapReduce programming framework is Google's proprietary software. However, there are equally good open source alternatives that work on the same principle. The best known and most widely used is Apache Hadoop. This implementation relies on HDFS in processing and storing large amounts of data. Map and reduce platforms such as this play a key role in data warehouses and business intelligence systems, because they enable us to do complex data processing such as aggregation and filtering very effectively.

Comparison of Hadoop and traditional ETL tools

Data analysis built on top of Hadoop in some cases, can get instant access to a large amount of high quality data without the need for long ETL processes (Duda, 2012). Hadoop can implement ETL processes by writing custom programs, which sometimes is not really effective and cannot always replace the traditional ETL tools from the existing data warehouses. ETL tools contain a lot of built-in capabilities suitable for cleaning, coordination and transformation of data. Such tasks need to be manually written in Hadoop. Their disadvantage is they are not particularly scalable and encounter problems when attempting to process several hundred gigabytes of data. It is recommended to use ETL tools for small amounts of data, because it saves data warehouse development time, while with the large amount of data it is better to develop custom programs in Hadoop. There are also hybrid solutions such as IBM InfoSphere DataStage that uses Hadoop for processing unstructured data and then forwards them to ETL tools for further processing (IBM, 2014).

Hadoop

Hadoop is an open source programming framework which supports processing of large amounts of data in a distributed computing environment (Kumar et al.). It is part of the Apache Project and is funded by the Apache Software Foundation. Hadoop includes an interface that allows parallel storage and processing of data. It runs MapReduce programs in a parallel environment of thousands of nodes. It is capable of collecting data from various sources, and can be further

customized to fit the type of data. It is worth noting that Hadoop is a kind of MAD (Magnetic, Agile, Deep) type system (Cohen et al., 2009).
These types of systems are distinguished by their characteristics:

- magnetic – attracts all types of data from various data sources
- agile – able to adapt processing and analysis to possible changes in the data structure
- deep – capable of high scalability in-depth analysis of large amounts of data.

Because of these characteristics as its biggest advantage stands out the ability of in-depth analysis of unstructured data collected from various sources much better and faster than analytics tools based on SQL (Cuzzocrea et al., 2011).

Hadoop Distributed File System

The Hadoop Distributed File System (HDFS) is a scalable distributed file system that provides instant access to application data. Computer clusters with HDFS work on the master and worker nodes principle (Date et al., 1986). The master node is responsible for management of the entire file system, metadata of all files and directories in the file system and the locations of all the blocks of data stored in the worker nodes. Worker nodes store and return information when requested by the main node.

Data Warehouses with non relational databases

One of the definitions of business intelligence gave E. Turban et al., "An umbrella term that encompasses tools, architectures, databases, data warehouses, performance management, methodologies, and so forth, all of which are integrated into a unified software suite" (Turban et al., 2010). During the execution of ETL processes the data is loaded from ERP, CRM and other systems, as well from the database, and if necessary, from different spreadsheets or other formats. After the data is loaded, we can do OLAP analysis, use tools for data mining or just do custom reports (Duda, 2012). There are several models by which data warehouses are built; ROLAP, MOLAP and their hybrid HOLAP. However, current systems and data warehouses do not have all the necessary tools and information necessary to make business decisions in today's dynamic and complex economic environment. Large amounts of data (Big Data) simply cannot be collected and processed by traditional systems and existing OLAP servers that are based on the relational model. Because of their limitations and new needs, solutions that are based on the non relational model are developed.
Map and reduce platforms that we described earlier together with some of the open NoSQL databases that support Hadoop (such as HBase) can be an alternative or supplement for traditional data warehouses. An example of such a solution is Apache Hive. Although is supports an interface similar to SQL state-

ments, queries are translated into map-reduce jobs, which are then performed on the Hadoop platform (Thusoo et al., 2010).

Hadoop as a data warehouse

Hadoop can be considered as the next step in the development of data warehouses, with a particular focus on the ETL phase. In Hadoop there is no referential integrity, which results in better performance. In comparison, the Data Vault model supports full referential integrity through the entire model, which of course has its advantages (Linstedt, 2011). Partitioning supports the separation of files, even to different computer nodes. We load raw data into Hadoop, data without pre-treatment, which means that they do not have to go through special ETL processes. At first glance a similar idea has a one-layer architecture of traditional data warehouses, but it does not support any form of ETL and in practice it is rarely used. The data is copied to the Hadoop platform, and adaptation (extract and load) rules should be written in the source code of Hadoop implementations. We emphasize that Hadoop is not an ETL tool but a platform that supports the ETL processes (Thusoo et al., 2010). This approach has many advantages such as fast loading of data, distributed processing, automated compression (in some implementations) and many others. Also, one of the advantages is freer access to the data storage moral in comparison to relational. We must adhere to certain rules and we have to use some consistent model such as the XML to make it easier to handle data.

Hadoop and data warehouse architecture

The hadoop platform is used together with some existing data warehousing systems. Processes in the Hadoop platform can use the data warehouse as a data source. MapReduce can issue SQL tasks to the data warehouse. We take advantage of the benefits of both technologies. MapReduce task is just another program, a data warehouse acts as a database (Das and Mohapatro, 2014).

Hive

Hadoop gives us a lot of new features but also has some drawbacks. MapReduce programming interface is low-level and tends to become very abstract and requires developers write custom programs that are usually difficult to maintain and reuse. When it comes to processing large amounts of data, the use of Hadoop technologies can reduce the processing time from one day to a few hours. But the use of Hadoop is not easy for end users. For example, it is necessary to write MapReduce programs for simple inquiries such as getting clean totals or averages of some data. Hadoop is missing some interface that works in a similar way as SQL. For these reasons Hive was developed. Hive is an open source solution built on top of the Hadoop platform (Thusoo et al., 2010). The goal was to develop an interface that can interact with unstructured data in Hadoop, and that brings some familiar concepts of tables, rows and columns and part of the

functions and methods that we know from SQL, of course adhering to Hadoops rules with the intention of avoiding noticeable restrictions. It was developed by Facebook where it contained over 700TB of stored data back in 2010 and is still used to provide reports and analytics.

Structure and data mode

Hive structures data in the well-known concepts such as tables, columns, and rows. It supports all major data types; integer, float, doubles and strings, but also more complex types such as folders, lists, and structures. More complex types of support nesting and thus can build more complex types. Also, Hive gives users the ability to upgrade the system to their custom data types (Thusoo et al., 2010). If the data is prepared by another program, or from the relational model, Hive has the ability and tools to incorporate the data to the table, without data transformation, which is a great advantage in case of large data sets.

Hive query language

The language with which are queries built, HiveQL, is very similar to SQL, so users who are familiar with SQL can quickly adapt. Traditional features of SQL as well as various types of joint, carthesian products, grouping, aggregation, creating tables, SELECT command and other useful functions that can be performed on the primitive and complex types make this language similar to SQL. This allows users who are familiar with SQL to a Hive to immediately open a command prompt and start writing queries. Some useful functions for returning metadata tables also present which saves time writing programs or queries. With a single command the data structure can be found out and which are used types of data without further testing various queries or similar methods. There are some limitations with respect to SQL, such as the inability to add new elements to the table or data partition, but if you want to add data, we have to rewrite the entire table or partition. However, this and similar restrictions have not created problems in using (Thusoo et al., 2010). As the data in such systems is retrieved periodically, for example daily or every hour, the data is easily loaded into a new table or data partition. Hive contains plug-ins for analysis using the map-reduce programs which the users write, in the programming language that suits them best. This allows advanced users build much more complex algorithms.

Hive architecture and components

Hive is assembled from a few main components. Megastore is a component that records the system log, and metadata about tables, partitions etc. Driver is a component that takes care of the lifetime of the HiveQL tasks while processing. Query compiler is a component that translates HiveQL inquiries into tasks and subroutines that are suitable for running in Hive. Execution engine is a component that runs the tasks that the compiler gives in the correct order. It cooperates

with the lower layers of the architecture. HiveServer is a component which allows you to interact with other applications. Client components such as CLI, web interface and JDBC/ODBC drivers. HiveQL queries are entered through one of the user interfaces like CLI or web interface. The Driver then sends the query to the compiler to check the syntax and semantic accuracy using the same metadata stored the Metastore. The compiler generates a logical sequence of tasks and then optimizes with a simple optimizer based on certain rules. Finally, generate the optimized sequence of tasks in the form of a map-reduce routine. Execution engine running tasks using Hadoop.

Conclusion

If we want to have a reliable and precise analysis in our data warehouse, or any other system of business intelligence, we need to collect and analyze all available data from the organization itself, but also from its environment, and other data sources. Nowadays, more and more of this data is unstructured and does not fit into the relational model. We are no longer limited to collect certain types of data structures for decision making. In this paper, we explained some of the concepts and technologies that are closely related to the storage, processing and analysis of large amounts of unstructured data. We explained the differences between relational and non relational models. We reviewed the current technologies that use non relational solutions in processing and analyzing large amounts of data. Also, we pointed out some of the flaws and limitations of the relational model. We examined the interface and tools that can come with Hadoop and its integration into the data warehouse. Such a data warehouse can store and process large amounts of data and thereby uses Hadoops MapReduce technology for parallel data processing. One of these is the Hive data warehouse. Although it is still in active development, Facebook uses it daily for processing and analysis of large amounts of sensitive data. We mentioned the Data Vault and Anchor modelling methods that are designed just for the purpose of storing and processing large amounts of data in data warehouses. These models may have come too late to the market, because most of the big companies like Google, Amazon and Facebook have already switched to non relational solutions. Because of the nature of the open source movement and openness of these non relational solutions, they will be the ones to lead the further development of technology in this domain, and that is a big advantage. It is important to note that non relational solutions are not always suitable, and that for certain cases it is still better to use a relational model. We are now at the stage when the non relational databases and technologies that implement them have become a great competition to the relational model and existing, traditional technologies. The development on the integration of Hadoop in existing and new data warehouses is currently active, but in the future we hope to see Hadoop integration with some business intelligence systems, which could eventually lead to some competitive open source solutions.

References

Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., and Welton, C. (2009). MAD skills: new analysis practices for big data. Proc. VLDB Endow. *2*, 1481–1492.

Cuzzocrea, A., Song, I.-Y., and Davis, K.C. (2011). Analytics over large-scale multidimensional data: the big data revolution! In Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP, (ACM), pp. 101–104.

Das, T.K., and Mohapatro, A. (2014). A Study on Big Data Integration with Data Warehouse. Int. J. Comput. Trends Technol. IJCTT–volume *9*.

Date, C.J., Date, C.J., and Date, C.J. (1986). An introduction to database systems (Addison-wesley Reading, Mass.).

Dean, E., and Ghemawat, S. MapRednce: Simplified Data Processing on Large Clusters.

Duda, J. (2012). Business intelligence and NoSQL databases. Inf Syst Manag *1*, 25–37.

Grolinger, K., Higashino, W.A., Tiwari, A., and Capretz, M.A. (2013). Data management in cloud environments: NoSQL and NewSQL data stores. J. Cloud Comput. Adv. Syst. Appl. *2*, 22.

IBM (2014). Bolstering data confidence with comprehensive information integration. (IBM Corporation),.

Jain, A., Doan, A., and Gravano, L. (2007). SQL queries over unstructured text databases. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on, (IEEE), pp. 1255–1257.

Krneta, D., Jovanovic, V., and Marjanovic, Z. (2014). A direct approach to physical Data Vault design. Comput. Sci. Inf. Syst. *11*, 569–599.

Kumar, R., Parashar, B.B., Gupta, S., Sharma, Y., and Gupta, N. Apache Hadoop, NoSQL and NewSQL Solutions of Big Data. Int. J. Adv. Found. Res. Sci. Eng. IJAFRSE *1*, 28–36.

Linstedt, D. (2011). Super Charge Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault (Lexington, KY: CreateSpace Independent Publishing Platform).

Stonebraker, M., Madden, S., Abadi, D.J., Harizopoulos, S., Hachem, N., and Helland, P. (2007). The end of an architectural era:(it's time for a complete rewrite). In Proceedings of the 33rd International Conference on Very Large Data Bases, (VLDB Endowment), pp. 1150–1160.

Thusoo, A., Sarma, J.S., Jain, N., Shao, Z., Chakka, P., Zhang, N., Antony, S., Liu, H., and Murthy, R. (2010). Hive-a petabyte scale data warehouse using hadoop. In Data Engineering (ICDE), 2010 IEEE 26th International Conference on, (IEEE), pp. 996–1005.

Turban, E., Sharda, R., Delen, D., and King, D. (2010). Business Intelligence (Boston: Prentice Hall).

MULTIMEDIA SYSTEMS AND APPLICATIONS

Original scientific paper

A Rationale for Multi-modality in Multimedia Instructional Design

Sanja Kišiček, Tomislava Lauc
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lucica 3, 10 000 Zagreb, Croatia
{smatic, tlauc}@ffzg.hr

Summary

This paper describes a multi-modal approach to multimedia instructional design in a virtual learning environment. The research was conducted within an online course offering multimedia and multi-modal lessons, each of which was comprised of three different types of resources, specifically designed for the acquisition of computer literacy skills. The resources vary from textual resources, pictorial resources accompanied by text and video resources. All the resources are designed by the course lecturers, i.e. the authors of this paper. The purpose of the research was to investigate user perceptual modalities in terms of their preference towards educational multimedia for the acquisition of computer literacy. The methodology of this research is based on data mining techniques through log data which represents user navigation and behavior within the learning management system. The approach fosters self-regulated technology enhanced learning where the end user has control over the choice of learning resources according to personal preferences.
The results of cluster analysis show that there are consistent user preferences in selecting a particular type of resources representing particular modality. The results provide a rationale for a multi-modal approach in designing a learning system based on multimedia instructional design and developed specifically for the acquisition of computer literacy, as a requirement for information literacy of the 21st century.

Keywords: virtual learning environment, multimedia instructional design, perceptual modalities, user preferences, multi-modality

Introduction

Courses supported by learning management systems (LMS) manifest various contexts, activities and resources for learners. In the field of computer literacy, a great amount of learning content is accessible in a variety of multimedia formats, with the resources ranging from text over image to video. In 1955, Jakob Nielsen anticipated that by 2010, video will be the major multimedia format for information presentation on a personal computer (Nielsen, 1955). He also anticipated that today there would be large amounts of hypertext throughout uni-

versities in the form of shared virtual information. Was Nielsen right? The amount of massive open online courses (MOOCs) today is a witness that he was.
Technology Enhanced Learning attracts great interest, promising the possibility of individualized learning structured upon unique circumstances, user preferences and knowledge background (Mulwa et al., 2012). User preferences in a multimedia learning environment can be supported if users are provided with a possibility to choose the way in which they will perceive information, as it is described in this paper. The question is how to facilitate acquiring computer literacy by making it appealing to one user and at the same time to any other? A possible answer is: by accommodating different perceptual modalities in a virtual learning environment and conducting research on user behaviour.

Multimedia Instructional Design

Instructional design is a discipline that is founded on scientifically based theoretical principles of learning and instruction (Smith and Ragan, 2005). Accordingly, multimedia instructional design should be grounded on scientifically based theoretical principles of learning through multimedia. In an earlier educational practice, multimedia distance learning indicated the transmittance of educational content remotely via multiple media, such as television, radio and newspapers. Today, computer technology is the main medium in multimedia distance education. Despite the differences, both behavioural and cognitive approaches are used in multimedia design. Both approaches assume analysis, decomposition and simplifying tasks. Both approaches tend to attract the user, and put users' attention into focus, as well as emphasize the importance of the built-in feedback. Furthermore, both approaches highlight meaningful learning in realistic contexts for the application of knowledge and skills (Atkins, 1993).
Mayer (Mayer, 2001; 2005) developed a Cognitive Theory of Multimedia Learning that is based upon Paivio's Dual Coding Theory (Paivio, 1990) and Sweller's Theory of Cognitive Load (Sweller, 1994). The Cognitive Theory of Multimedia Learning is based on the idea that human beings have two separate, but interrelated memory channels for information processing, the visual and the verbal channel. Image processing occurs mainly in the visual channel, and spoken word processing in the verbal. However, the processing of printed words begins in the visual channel, and then moves to the verbal. According to Sweller, the content load should be dosed, taking into account the cognitive processes and limited cognitive capacity. Each of these channels is limited in capacity, while active learning requires coordination of cognitive processes. In multimedia environments, students construct knowledge by choosing words and images from different materials and by combining the verbal and the visual with one another.
There are five cognitive processes in learning through multimedia (Mayer, 2001): 1. Selection of relevant words from displayed text or narration, 2. Selection of relevant images from illustrations, 3. The organization of selected words

in a coherent verbal presentation, 4. The organization of selected images into a coherent visual representation and 5. Integration of the visual and the verbal and the existing knowledge.
Learning through multimedia is learning from words and pictures (Mayer, 2001). Words can be spoken or written and pictures can be static or dynamic. There are five different elements of multimedia: text, images, sound, animation and video. Two of these are static: text and images, and three are dynamic: sound, animation and video (Boyle, 1997). Video can consist of all of the multimedia elements, and there are different approaches in designing a video.

Video
Video is a powerful tool for conveying information, and can include all elements of multimedia. It requires high-quality software and hardware solutions. The emergence of digital video has enabled users to interact with the displayed content. Interactive video encourages attention and user activity in an educational context. There are several different types of video display among which instructors can choose the appropriate type of video to support e-learning, in accordance with the capabilities and resources that are available (Boyle, 1997).

Video as a text replacement
Instructors can record a video in addition to the explanations that follow a paragraph of text or lecture presentations. This kind of video makes teaching more interactive and allows users to process the content at their own pace.

Narration
Instructors can retell the entire contents of a page or screen. This type of video can be tedious and seemingly uninteresting since interactivity is not emphasized. However, the suitability of this technique depends significantly on the educational content.

Scenario
The scenario is played by actors and recorded to demonstrate certain situations. It is mainly used as a representation of certain activities in various courses.

Simulation
Users are fully involved in the course of such a video, and it is best suited for interactive online learning. This kind of videos suit learners who prefer to learn in a practical way. Creating a simulation is challenging, but the final product has many advantages. The main feature of the simulation is the interactive environment in which the user's actions affect the future course of events (Boyle, 1997).

Demonstration
This kind of video is suitable for detailed guidance on how to do something. Such videos fit well in a teaching process of how to acquire procedural knowledge and skills. Video materials used in the online course from this research are designed as demonstration videos.

Multimodal and adaptive learning

For the learning process to be effective, it is necessary that the system adapts to individual needs. The design of a LMS tends to go beyond the universal and traditional *one-size-fits-all* approach (Marshall, 2011). Adaptation of learning is a process in which users are changing the way of learning indirectly, by navigating through the system. In traditional systems, the teacher adapts the teaching contents to the target users. With adaptive learning systems, personalized access to content can rely on the availability of pre-designed resources, and not necessarily on an open corpus of material (Brusilovsky, 2001), which today is a challenge in the field of hypermedia learning or adaptive hypermedia. Although adaptive learning systems are increasingly being developed in the area of TEL, it is still an area where there is a lack of publicly available, comparable, interoperable and reusable data sets covering formal and informal learning (Sosnovsky and Brusilovsky, 2012). Recommender systems are increasingly being used in education. A recommender system, in the context of e-learning, is an agent trying to recommend specific actions based on data gathered from students' previous actions and achievements as well as data on "similar" students. Through educational hypermedia, recommender systems tend to recommend activities to students depending on their goals, interests, previous knowledge and other aspects. Different individualization strategies in conventionally adaptive learning systems support individuals in the process of knowledge acquisition (Mulwa et al., 2012).
Adaptive hypermedia systems use adaptive presentation and support adaptive navigation. A method of adapting various explanations is that the same information is presented via different multimedia forms. In this way, a user is given control over the selection of resources in a multimodal environment. To encourage self-regulated learning, a user is empowered to choose among the resources with regard to personal preferences. At the same time, the diversity of elements in instructional methods could be present in a multimodal environment. Multimodality refers to the multiplicity of methods available for transmitting information. Multimedia means multiplicity of elements in the transmission of multimedia information. Multimodality and multimedia, in the context of multimedia learning, are often used interchangeably. However, it should be stressed that multimodal learning environments allow instructional elements to be presented in more than one sensory mode (visual, aural, written) (Sankey and Gardiner, 2010). Considering that the average user is multi-modal and owns several different preferences in the adoption of information, we support the realization of a

multi-modal learning environment for learning computer literacy, as described in this paper.

Research and Methodology

The research was conducted at the Department of Information and Communication Sciences at the Faculty of Humanities and Social Sciences in Zagreb University in the winter term of the Academic Year 2012-2013. The research was conducted within Socio-Humanistic Informatics, an online elective course offered to all graduate students at the Faculty. The skills set within the course covers advanced techniques in using MS Office tools in the context of natural language processing and computer literacy. The students attended the course via Moodle. We researched the patterns of user behaviour in interaction with the LMS by investigating the preferences in selecting various multimedia resources: text, pictures accompanied by text and video resources. Every weekly lesson in the course was supported by three resources containing the same content, but in different form (textual resource, pictorial resource and demonstration video). The study included 98 students; 82 females and 16 males. Teaching materials in the form of multimedia resources were structured in the *lesson* activity module on Moodle. Data analysis in this paper is based on log file data.

Course Material

The course content was presented through ten weekly lessons structured as activities. In this research, the content for acquiring the necessary knowledge and skills in computer literacy is presented in three types of resources: text, picture accompanied by text and video.

The multi-modal approach to acquiring computer literacy in this paper is based on a multiple choice of learning resources, different in form but equivalent in meaning. Multimedia in this paper is an integral part of the instructional design of educational resources taking into account that text can be written (including text on the screen) or spoken (including soundtracks), and static images (including graphical representation on the screen) or dynamic display (including video).

The content that is offered through the resources is aimed at mastering the required procedural knowledge and skills.

Learning Resources

Availability of resources is required across learning contexts and models to enable use for varied purposes. Resource-based learning, particularly in digital environments, offers promise for broad applicability across a variety of contexts to provide access to an expanding global library of digital resources (Sankey and Gardiner, 2010). Resource Based Learning (RBL) is an active learning technique based on preparing the students for retrieval and evaluation of information (Kamer, 2011; Hannafin and Hill, 2008). Reusable Learning Objects

(RLOs) are entities usable in various teaching situations and they fit well in instructional design that supports collaborative learning (Sylvain et al., 2011). RLOs are a reflection of the growing need for fast creation of re-usable materials in e-learning systems, and they are made of small pieces of learning content. This method of learning goes beyond the high cognitive saturation, by supporting the user in finding information, and not necessarily memorizing information.

It is important to obtain correspondence between the perceptual salience or "noticeability" of a feature and its thematic relevance. In static display, various graphical devices such as arrows, circles and boxes that rely on visual contrast can be used to direct viewers' attention of what is relevant. The same can be used to set the direction of attention within a dynamic display (Lowe and Schnott, 2008).

Taking direction of attention into consideration, within all of the three types of resources, there are specifically marked areas in the form of; bolded keywords in text, marked key points in images, or key targets in the dynamic display, i.e. video. Textual resources are designed as a series of structured and concise instructions displaying text only. Pictorial resources are designed as a combination of text and pictures. The content is structured so that images are combined with text. Video resources are designed as demonstration videos. A sequence of procedures was recorded and aligned with the accompanying narration. Also, there is an option to display the spoken text, beneath the animation, in a line that can be displayed or hidden.

According to the Principle of Redundancy (Mayer, 2001; Sweller, 2005) it is better to learn through animation accompanied by narration, than by animation accompanied by both narration and on screen text. However, due to the simplicity of linear information search, we included the option to display text in video resources. Students were able to choose to display videos with accompanying text without sound, or to display animated actions with accompanying text and sound, or display animated action without accompanying text, with the included sound.

Clustering of Access Logs

Access logs are a type of log data containing records of information on user access. In other words, every user's activity in the system is recorded in log files. We analysed log files of all the students enrolled in the course in order to learn about the nature of student access towards all three types of resources for each of the ten lessons. The purpose was to determine user behaviour in accessing the resources with the aim of identifying preferences in selecting the resources.

Table 1 shows the descriptive statistics of cumulative access towards a resource.

Table 1. Descriptive statistics of students' access to resources, obtained from log files

	Mean	*Median*	*Mode*	*Min*	*Max*	*Range*
Textual	9,7	6	1	0	48	48
Pictorial	25,7	26	26	0	79	79
Video	9,7	4	1	0	53	53

As a data mining approach we used clustering, a method of unsupervised classification of entities displayed as data points or vector features into groups called clusters. Clustering is used in a variety of professional and scientific fields, and is also one of the commonly used methods of data mining in education. It is used for the purpose of grouping users in the so-called clusters according to similarities in behaviour. In e-learning, clustering can be used to find clusters of students with similar characteristics (Romero et al., 2007.). We used the clustering method in order to group the students with similar behaviour patterns into groups. This method was chosen because the students were not limited or directed towards using one type of resource. Instead, the students were able to access all three resources in every lesson, at any time. Therefore, it was possible to group the students according to their access to the resources throughout 10 lessons, each of which contains three different types of resources.

We used a hierarchical clustering method, where it is not necessary to predefine the number of final clusters. Connection type determines the way in which the distance between two clusters is calculated, and thus distinguishes between the following: the distance between two nearest entities (single linkage), the distance between two furthest entities (complete linkage) and the average distance between all entities (average linkage). We chose the complete linkage algorithm due to clearer results.

The process of clustering involves three main steps: displaying entities, calculation of the similarity matrix and the clustering procedure. The entities are most often represented as a set of features that describe the entity. Vectors may contain discrete or continuous values indicating the degree of relevance of features for a particular entity. In this research, the frequencies of students' access to particular type of resources are indicated as values.

Moreover, it is necessary to determine a function that allows the calculation of the similarity matrix expressing similarity between the data points. In this research, the cosine similarity measure is obtained. It is a measure that is identical to the scalar product of two normalized vectors.

The advantage of the cosine similarity is that the measure is outlier-resistant (Ljubešić, 2009). Furthermore, the cosine similarity is identical to Pearson of Spearman correlation coefficient if data is normalized by the Z value.

Modelling user activity

The strategy for modelling user activity in accessing resources (adapted according to the "strategy to modelling student activity in online discussion forums" by Cobo et al. (2011) was conducted as following:

1. Specifying data
   a) defining the number of users (98 students),
   b) choosing objects for tracking access (30 resources)
   c) defining the type of access (resource view)
2. Constructing the series
   d) defining the type and the number of samples (30 resources)
   e) defining the values from the series' data (the frequency of access)
3. Obtaining the similarity matrix
   f) defining the similarity measure (cosine)
   g) defining the linkage method (complete linkage)
4. Identifying the obtained clusters which are visualized by a dendrogram (see Figures 1 and 2).

Data analysis and results

The aim of the research was to establish the existence of user preferences in the use of resources in order to test the consistency of the students choosing a particular type of resource over time. User access to resources is monitored for a time period of 1 semester, tracked through 10 teaching lessons, each containing 3 types of resources, making a total of 30 features. Time continuity of access to resources is given by the following expression:

$$\boldsymbol{x^{(i)}} = (\boldsymbol{x_1^{(i)}}, \boldsymbol{x_2^{(i)}}, \dots, \boldsymbol{x_n^{(i)}}, \dots, \boldsymbol{x_N^{(i)}}) \quad (1)$$

In the expression, $x^{(i)}$ marks the i^{th} student, $x_n^{(i)}$ marks the value that $x^{(i)}$has for the n^{th} resource, and N is a total number of resources included in the time series (N=30). Moreover, $x_n^{(i)}$ is defined the following way where $r_n^{(i)}$represents the access frequency of the i^{th} student towards the n^{th} resource.

$$\boldsymbol{x_n^{(i)}} = \boldsymbol{r_n^{(i)}} \quad (2)$$

The steps in the process of the agglomerative hierarchical clustering, following complete linkage and by calculating the cosine similarity, are the following:

- Vectors are shown quantitatively, expressed as values that represent the frequency of access to a variety of multimedia resources. Those values are obtained by log file analysis.

- The similarity between entities (a total of 30) is calculated using the cosine measure in order to obtain the similarity matrix showing distance between any two data points from a total of 30 data points.
- Agglomerative hierarchical clustering identified three clusters of access to multimedia resources (see Figure 2):
    - pictorial cluster
    - textual cluster
    - video cluster.

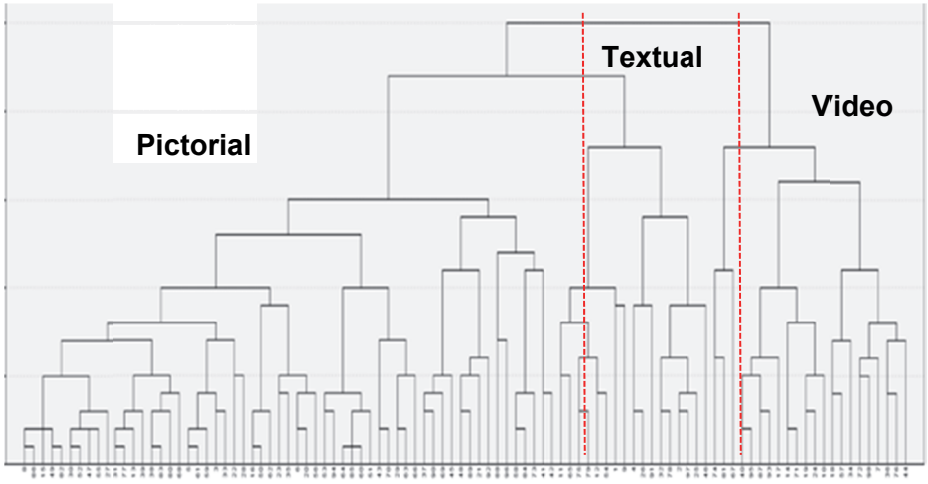


Figure1. Dendrogram shows 3 clusters of students

Cluster analysis shows that the students were consistent in selecting a single type of the resources, due to the fact that three coherent clusters were isolated. According to Figure 1, 60% of students were selecting pictorial resources, 23% selected the video resources and 17% of students selected textual resources. As displayed in Figure 2, the activity of users is grouped into three distinct clusters: text, picture, and video cluster, with a 100% purity of the clusters. Data mining results obtained by cluster analysis confirm consistent user preferences in selecting multimedia resources.

Discussion

Multimedia presentation, multimodality, *adaptivity* and interactivity set different design challenges, so d*ifferent learning theories such as adaptive learning theory, multimedia learning, as well as learning style theories should be taken into consideration.*

Adaptive hypermedia systems use adaptive presentation and support adaptive navigation. Multimodality refers to the multiplicity of methods available for transmitting information. Multimedia includes a combination of media for in-

formation presentation. While adaptive hypermedia systems support content and presentation adaptivity, in self-regulated learning a user is empowered to choose among the learning resources with regard to personal learning preferences, as described in this paper.

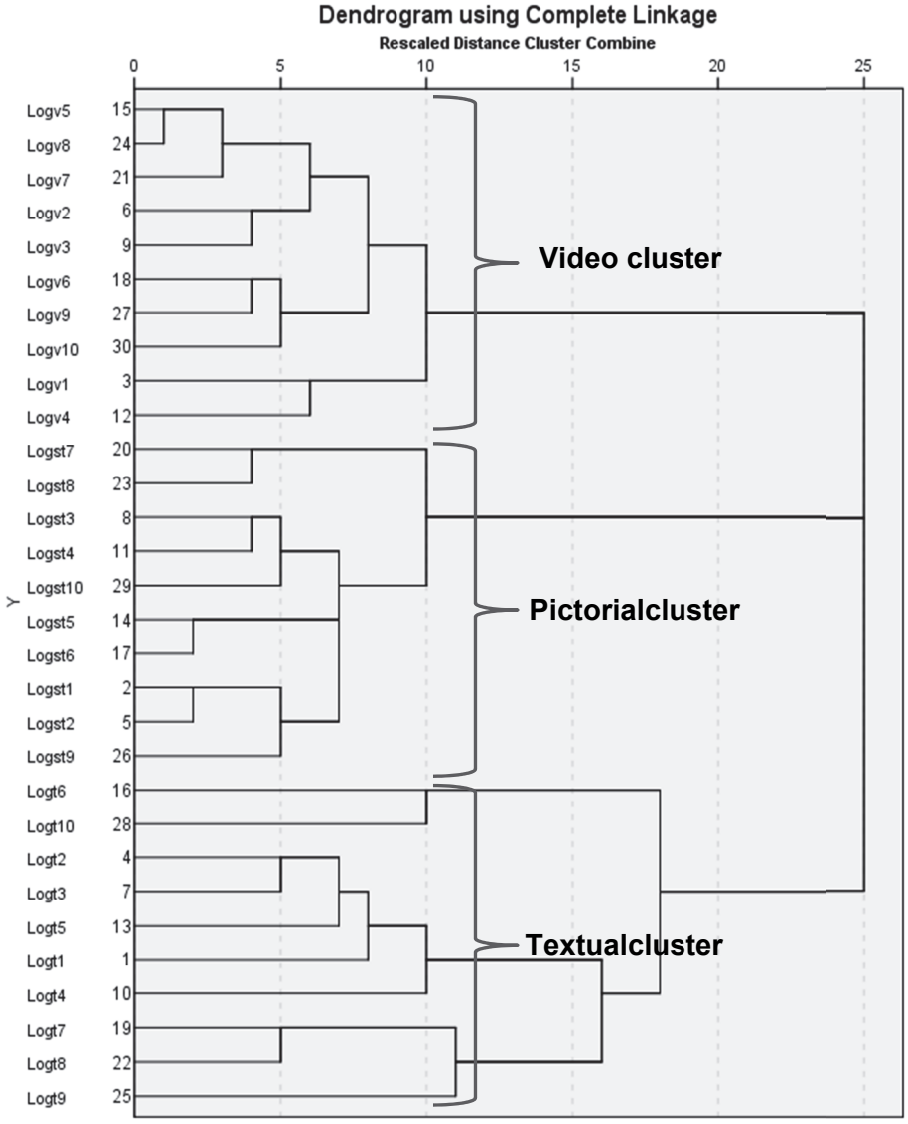


Figure 2. Dendrogram shows 3 coherent clusters of resources, Log=frequency of access; t=textual, st=pictorial, v=video; n=1,2,…,10 – considering 10 lessons

There are four main categories of learning preferences considering sensory modality. As defined by VARK learning styles model they are described as following: Visual, Aural, Visual, Read/Write, Kinesthetic. Visual learners learn best by seeing the material from visual displays. Aural learners learn best when processing the information by listening. Read/Write learners are those who prefer information displayed as words. Kinesthetic learners prefer examples, labs, demonstrations, simulations, videos, etc.

It is important to take into consideration that the average user is multi-modal and owns several different preferences in the adoption of information. In order to support user to choose among the learning resources with regard to personal learning preferences, we support the realization of a multi-modal learning environment. The vast majority of students agreed that the possibility of selecting a preferred type of resource helped them in acquiring the required knowledge and skills.

Furthermore, we confirmed that the students have consistent preferences in selecting multimedia resources. Also, we found matching results by obtaining feedback responses about users learning preferences in our preliminary research, (Kišiček et al., 2012.). In our previous work (Lauc et al., 2012a; Lauc et al., 2012b; Lauc et al., 2014) we conducted a research on students' perceptual modes and their learning activity with respect to multimedia learning resources. The feedback results contributed to the fact that user preferences are consistent through time and that students do have their own preferred mode for gaining computer literacy.

This finding should encourage instructors to offer different multimedia resources to their learners in a virtual learning environment, whether the resources are purposefully made or found on the Internet. When teaching know-how skills, efficiency and practicality are the values that matter, hence the user experience is crucial.

Finally, the main goal of multimodal and multimedia learning is to combine different types of media as well as different types of learning resources n order to improve acquiring of knowledge.

Conclusion and Further Research

Considering multi-modality in the framework of multimedia learning, certain media types can support different learning strategies. Individualization strategies in instruction can affect user experience while learning. In order to foster satisfaction in interaction with the learning material, users should be provided with the possibility to choose the way in which they will perceive information according to their personal preferences. Furthermore, in order to develop their own learning strategies, learners need to have control over their learning process and they need to be empowered in deciding what format of instruction to use. Self-regulation in the learning process can be accomplished through the use of re-usable learning objects, in resource based learning, as described in this paper.

In this research we described the design of a multimedia e-learning course for gaining computer literacy and investigated student behaviour regarding their access to different types of multimedia resources. By clustering on access log data, as well as obtaining feedback about user learning preferences in preliminary research, we confirmed that students have consistent preferences in selecting multimedia resources throughout the course, meaning that user behaviour patterns could be determined by the regularity in choosing a preferred resource over time.

Interactive video material could be added as an additional type of multimedia resource. In the upcoming research, conducting a study in order to compare the learning behaviour and learning outcomes, should also be taken into consideration.

Acknowledgements

We would like to thank our project partners from the University Computing Centre in Zagreb for the collaborative work in creating video tutorials within the Project of developing the e-course.

References

Atkins, M. J. Theories of Learning and Multimedia Applications: An overview. Research Papers in Education, 8(2), 251-271 (1993)

Boyle, T. Design for Multimedia Learning (1st ed.). Prentice Hall Europe (1997)

Brusilovsky, P. Adaptive Hypermedia. User Modelling and User-Adapted Interaction, 11(1-2), 87-110 (2001)

Cobo, G.; García, D.; Santamaría, E.; Morán, J.; Melenchón, J.; Monzo, C. Modeling students' Activity in Online Discussion Forums: a Strategy Based on Time Series and Agglomerative Hierarchical Clustering. In Proceedings of the 4th International Conference on Educational Data Mining (pp. 253-258) (2011)

Hannafin, M. J.,; Hill, J. Resource-based learning. In Handbook of Research on educational communications and technology, pp. 525-536 (2008)

Kammer, J.;Thompson, T. Information-Literate Avatars: Resource-Based Learning in Second life. In Teacjing Information Literacy Online (pp. 109-132). USA: Facet Publishing (2011)

Kišiček, Sanja; Lauc, Tomislava; Golubić, Kruno. Students' Learning Preferences in a Multimedia Online Course. // International journal of education and information technologies. 6 (4), pp: 319-326 (2012)

Lauc, Tomislava; Kišiček, Sanja; Bago, Petra. Multimedia Resources in an Online Course: Access and Usage with Respect to Sensory Modality. // Croatian Journal of Education. 16 (2014), 3; 155-173

Lauc, Tomislava; Kišiček, Sanja; Bago, Petra. Students' Usage and Access to Multimedia Learning Resources in an Online Course with Respect to Individual Learning Styles as Identified by the VARK model // 21st Century Learning for 21st Century Skills (LNCS 7563) / Ravenscroft, Andrew ; Lindstaedt, Stefanie ; Delgado Kloos, Carlos ; Hernández-Leo, Davinia,editor(s). Saarbrücken : Springer, 2012. 548-548

Ljubešić, N. Event detection in parallel information sources (doctoral thesis). The University of Zagreb (2009)

Lowe R., W. Schnot. A Unified View of Learning from Animated and Static Graphics. In Learning with Animation: Research Implications for Design. Ed Lowe R, W. Schnot, pp. 304-357. Cambridge University Press (2008)

Marshall, K. Choose the Right Video Type for Your Learning Goals. LectoraUniversity (2011)
Mayer, R., E. Multimedia Learning, pp. 41-61, Cambridge University Press (2001)
Mayer, Richard E. Cognitive Theory of Multimedia Learning. In The Cambridge handbook of multimedia learning. Cambridge University Press (2005)
Mulwa, C.; Lawless, S.; O'Keeffe, I.; Sharp, M.; Wade, V. (2012). A recommender framework for the evaluation of end user experience in adaptive technology enhance learning. International Journal of Technology Enhanced Learning, 4(1-2), 67-84 (2012)
Mulwa, C., Lawless, S., O'Keeffe, I., Sharp, M., &Wade, V. (2012). A Recommender Framework for the Evaluation of End User Experience in Adaptive Technology Enhance Learning. International Journal of Technology Enhanced Learning, 4(1-2), 67-84 (2012)
Nielsen, Jakob. Multimedia and hypertext: the Internet and beyond, Academic Press ; Morgan Kaufmann (1995)
Paivio, A. Mental Representations: Ad Dual Coding Approach. USA: Oxford University Press (1990)
Romero, C.; Ventura, S., García, E. Data Mining in course management systems: Moodlecase study and tutorial. Computers & Education (2007)
Sankey, M.; Birch, D.; Gardiner, M. Engaging Students Through Multimodal Learning Environments: The journey continues. In Proceedings ASCILITE 2010: 27th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Curriculum, Technology and Transformation for an Unknown Future (pp. 852-863). University of Queensland (2010)
Smith, P. L.; Ragan, T.; J. Instructional Design (3rd ed.). USA: Wiley Publishing (2005)
Sosnovsky, S.; Hsiao, I.-H.; Brusilovsky, P. Adaptation "inthewild": Ontology-Based Personalization of Open-Corpus Learning Material. In21st Century Learning for 21st Century Skills, Lecture Notes in Computer Science (Vol. LNCS 7563, pp. 425-431). Springer (2012)
Sweller, J. Cognitive Load Theory, Learning Difficulty and Instructional Design. Learning and Instruction, 4, 295-312 (1994)
Sweller, J. The Redundancy Principle in Multimeda Learning. The Cambridge Handbook of Multimedia Learning (Edited by R.E. Mayer.). Cambridge University Press (2005)
Sylvain, M. C.; Mofford, K.; Lehr, E.; Riley, J. E. Reusable learning objects: Developing Online Information Literacy Instruction through Collaborative Design. In Teaching Information Literacy Online (pp. 25-43). Facet Publishing (2011)

Original scientific paper

Multimedia functions of MMO games

Winton Afrić
University North
Trg dr. Žarka Dolinara 1, Koprivnica, Croatia
wafric@unin.hr

Vjekoslav Afrić
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
vafric@ffzg.hr

Summary

This paper is a research of functions of multimedia impacts of MMO games stemming from the analytical framework of J. A. DeVito's functions of Mass Communication. In this paper the analytical concepts of DeVito's theory are applied to a qualitative analysis of the most popular, most played and most represented MMO games, taking into account a so called pattern of intensity, to point out new functions and impacts of MMO games. Our qualitative research showed that the functions of MMOs although comparable to the theoretical framework we started from, create a new theoretical model that discovers specific functions with specific impacts exclusive to MMO games which in the end only share their name and heritage with the theoretical frame we started from.

Keywords: Multimedia functions, MMO games, Multimedia impacts,

Introduction

Since the new millennium the multimedia that is internet has become ever more important, encompassing other media and becoming the prime communication means of today. One of the most complex and interesting types of multimedia is that of massive multiplayer online games. The growing amount of players and the growth of the games industry is a clear indicator of the contemporary trend – the audiences are no longer satisfied with active and/or linear multimedia, instead they want interactive non-linear multimedia to satisfy their various needs. The development of internet as a means of communication and a network for sharing information, commercial or otherwise, has slowly shifted us as the audience/ receivers into an interactive role. Instead of being offered different types of media we are accustomed to "taking charge" and creating our own communication experiences. What better way to cater to this experience then to use games. The difference between MMO games and their other offline computer games is in the forming of a new online society within the diegetic frame of the online world of an MMO game. This society forms all kinds of diegetic and ex-

tra diegetic communication/ interaction in which any individual can take part. (W. Afrić, V. Šimović, M. Milković, "The sense of (RPG) Role Playing Games Systems" 2013.) So instead of offering narratives as film, newspapers, or television MMO games offer settings with different rules (structural and processual code) and let players forge their individual experiences interacting with each other and the game world. This paper will therefore focus mostly on MMORPGs since the societal structure and relations are much more complex and elaborate in MMORPGs than in other MMO games. Granted, lots of MMO games have RPG elements which may be important for the games structure so it would be unfair to say we focus only on MMORPGs.

Background research

For the purpose of this analysis we view multimedia as a system of different communication media which function as a whole. In this context any form of communication which involves multiple media (or multiple channels) is multimedia, so interpersonal communication, group communication; public and mass communication are all different types of multimedia. Different types of technology which enable differing multimedia are only extensions of ourselves (McLuhan, 1964) in this process of multimedia communication and add themselves to the total capacity of communication we possess. This technology enables new structural codes which, added onto the old structural code system, form new different types of processual code and behavior and in turn form new communication patterns and new experience. *(J. D. Johansen, S. E. Larsen, "Introduction to Semiotics", 2000.)* We have to point out that when we mention communication we are referring to all types of communication but mainly all different types of audio and visual code, followed thirdly by tactile communication which in MMO games translates roughly to muscle memory, and coordination of movements that enables us to play and fulfill various win conditions. There are rare games today in which communication through the tactile is very important. One such example is "Brothers: A Tale of Two Sons" (Starbreeze studios, 2013). Brothers has a section where the two brothers must climb a cliff and the controls make your hands start to hurt from having to hold and press different buttons, thus immersing you in the experience of "hands hurting from hanging for your life".

As the multimedia of choice has changed for many so did the structure of their multimedia experience and the impacts this multimedia has on its audience. It is safe to say that this led to a change to many of the functions multimedia has today as well as some new functions. We will describe the contemporary functions of MMO multimedia by looking at mass communication functions, choosing a rather extensive list of mass communication functions, proposed by J. A. Devito (1982, 526). DeVito analyzes mass communication impacts through nine factors where he tries to envelop both good and bad influences of media and how they affect audiences. This rather large list of factors is particu-

larly interesting because it offers a thorough analytical framework. It shows numerous ways in which mass media affect our everyday life, and we find it very suitable to analyze how MMOs affect our everyday lives. Not all games have all factors equally represented, nor are all factors equally important, but this extensive list is necessary so that we can analyze many different types of MMOs and their different impacts. We have taken the approach of looking at mass multimedia of contemporary MMOs trough the factors that DeVito proposes and analyze how those factors are applicable to the multimedia of MMO's. We have chosen to do a qualitative research of the most popular MMO games with RPG elements. We have selected games using an intensity pattern firstly taking into account games that are most popular, most recent and most updated but we also looked at games featuring different monetization models, games with strong history (and therefore significant audience), and games exploring different / new play styles and methods. In this way we have analyzed games that cover all different player types and cover the largest player base of over 30 million players playing these games on a daily basis. Regarding player types, the time investment differentiates the casual, the serious and the die-hard elite, and the type of activity differentiates the explorer, the achiever and the power gamer, or the soloist and the group player. Most of these different categories stem from the more traditional difference in player preferences proposed by Blacow (1980) called the "Fourwald Way taxonomy" which differentiates the Wargamer, Powergamer, Storyteller and Roleplayer. The MMOs we used in this analysis are: Guild Wars 2, World of Warcraft, Dragon Prophet, Wildstar, Diablo 3, Path of Exile, Grim Dawn, Heroes of the Storm, League of Legends, Dota 2, Dungeons & Dragons Online, TERA, Lord of the Rings Online, Guild Wars, Torchlight 2, StarCraft 2, Heartstone, Hex, MtG Online. In the rest of the paper we will go through the different factors of the multimedia of MMO's and their features.

The first function of Multimedia: Entertainment

We can all agree that most games exist first and foremost for entertainment purposes. The entertainment of MMO games is different to other more traditional forms of multimedia entertainment in the sense that it is first and foremost interactive, it forms a unique experience and in many cases strives to give players more control over how they spend their time entertaining and how much time they spend. Most MMO games today especially MMORPG games go through great lengths to deliver an experience where each player and group of players feels like a central piece in the story (GW2, WoW, D3 etc.). Their choices have significant impacts on the world around them. Players develop their characters immersed into various roles and have a unique perspective on the world, one that will change if they change their character. This type of story-forming experience is one of the most entertaining aspects of MMO games. The entertainment boils down to two opposing types of play PvE (Player

versus Environment) and PvP (Player versus Player). PvE focuses on player collaboration against more or less pre structured encounters where win conditions range between certain levels of character power needed along with player knowledge and skill. This type of play is very entertaining when players and groups win at these encounters receive various rewards and acknowledgements and advance in the game (Raids, Dungeons, Open World Boss encounters, BAMs etc.). PvP pits players and groups of players against each other and is more about skill and spontaneous player reaction then PvE (Battlegrounds, Duels, Arenas, and Open World PvP etc.). The satisfaction here comes firstly from besting other players and secondly from rewards won inside the game. Ultimately there are different types of players which strive for different types of satisfactions and find different activities entertaining. Game developers try to cater to all of the different player types and make aspects of the game that are entertaining to all of them, as well as specific fragments made for each individual group. Different types of players with different preferences can include different players in accordance of how much time they spend playing the game or what activities they prefer when playing the game. As MMO's develop they try to incorporate new entertaining and unique ways to satisfy each individual player type. For instance "Guild Wars 2" (ArenaNet, NCSoft 2012) introduced Vistas as a means to entertain the explorer. Players need to find secret and hidden routes to difficult-to-reach places and once they get there players get to see an animation of the panoramic view from that place, that beautiful vista. Blizzard are introducing a new way to cater to Storytellers in their (currently in beta development patch for the) "Diablo 3" (Blizzard 2012) where the various ways to play the game are now structured into a unique character journey which becomes a players own character story.

The second function of Multimedia: Education

The multimedia of MMO games educates us in various ways. This type of education although non-formal stretches through various fields and practices. Surely the most important education we receive from RPG in general is the fact that we learn to recognize different types of societal roles and learn to understand differing perspectives of these roles. When we tie roles to attributes such as race, gender, age, intellect, charm etc. we understand different characters as different constellations of attributes, with unique perspectives stemming from those attributes. Therefore playing we learn about differences in a way where we see typically problematic differences in race or gender roles as variance in attributes and perspectives and learn to tolerate differences and acquire a multicultural frame of mind. This type of sensitivity training is also true for experiences which stem from roleplaying different roles and learning what is their perspective and diegetic frame. Here roleplaying educates us about Empathy, Equality, and Openness. MMORPG games are more narrow and specific when looking at these educational aspects due to the fact that most of the diegetic

frame is communicated to us through the multimedia of the game system and is therefore limited by the scope of the game. In MMORPG games we learn about roles through the perspective of our class or character role, we learn about player power dependence (the given matrix, effective matrix and dispositional matrix) *(R. West, L. Turner, "Introducing Communication Theory", 2007)*. We learn how to treat other roles, and other players and what to expect from them. Each MMO is in part a simulation of real world relations and we learn about causal processes and behaviors which correspond to real life situations. The simulation and the education process can be more or less obvious (we can play in a real world environment, or in a real world society structure, we can use real world military strategy or tactics, fly a real life plane etc.) We also learn about the virtual component, in MMORPGs especially about the virtual society. What values, opinions and rules the virtual society judges to be just and valid.

The third function of Multimedia: Conference of Status

Typically this function has to do with recognizing important individuals and those individuals have been given lots of media exposure. In MMO multimedia we don't only recognize important exposed individuals we can learn what it is that makes them important. Becoming exposed has also became interactive and a matter of player preference and desire. The most important individuals here are skilled players as well as the so called streamers and enthuse journalists. All of those have lots of media coverage which comes in various extradiegetic forms, but they are all focused on the game. Examples of these forms are game forums, YouTube channels, web sites, facebook & twitter profiles, reddit posts, twitch streams etc. be it educating players about the game, doing commentary while playing or making shows with discussions about the game world. Many of these have players following their channels and streams and paying them monthly to create game related content and thus playing and creating new multimedia content about a game or games in fact becomes their job. Since any player or group of players can make their own YouTube channel, a website or join Twitch or another service and stream while playing, anyone can take steps to become a celebrity and earn status. In this way the media not only tells us what is important but it also shows us what steps we can take to become important. However it is not only individuals that can be recognized as important, through the celebrity created content we can recognize game related concepts and behaviors as important and less important. Game developers keep a close eye on this type of content as it helps them determine on which things the public is concentrated, why it is concentrated on them, whether that is positive or negative for the scope of the game and this often leads to changes in the game. The question that remains is would the changes occur if the media has not focused on this content or not?

The fourth function of Multimedia: Reinforcement

When we speak of reinforcement we mean the process of strengthening ones beliefs, attitudes, values and opinions. The multimedia of MMO games does this in various ways. Firstly as games develop they will reinforce us to expect and display certain types of behavior within games and to develop certain values and opinions towards various aspects of the game and games in general. Reinforcement in MMO's works firstly for concepts learnt within the game. So players develop behaviors towards different types of players, value different player behaviors differently and they develop a sense of what they want from a game. One of the popular, major values that is constantly reinforced is how much time should an individual spend playing? It is hard to get people that are not interested in playing a game to play it. But getting people that play it to play more is much easier. Also getting players that enjoy in one type of game activity to try another game activity is also done by reinforcement. Another popular topic is how much should games cost, and is it ok for games to charge a monthly fee to be able to play them? How large should this monthly fee be? Is it ok that a game is free but has an Item mall where you can spend real money to buy advantage through Power-ups and Quality of Life changes? General different opinions regarding the answers to these questions can be found all over the internet and are reinforced with both diegetic and extradiegetic means. A popular enthuse journalist "TotalBiscuit, The Cynical Brit" stated that it is relatively easy to determine how much one wants to pay for a video game, be it monthly or otherwise and he offered the following reasoning. Since people play games to satisfy various needs but mostly to have fun it is easy to compare it to costs of going to a movie or the theater. In Croatia an average movie ticket costs 4.2 Euros and it offers 2 hours of entertainment. A digital copy of Diablo 3 for instance costs 39.99 Euros and the first play trough offers over 20 hours. So they seem roughly the same. However given the nature of the game most players invest hundreds of hours over the years exploring different options with different characters and that's not even mentioning free content that is being constantly developed and patched to the game. Playing Diablo 3 thus becomes 10 times cheaper than going to the movies. We should also take Wilbur Schramm's formula where he elaborates the probability of the audience to choose a particular media into account *(W. L. Schramm," Men, messages, and media: A look at human communication", 1973)*. Schramm proposes that: Promise of reward / Effort required = Probability of selection. It is important to note that the Promise of reward takes into account both immediate and delayed rewards, which are in fact satisfactions of various needs. It is easy to deduce that it takes less effort to play a computer game than to go to a movie. Looking at these reasoning's we can understand why it is easy for MMO multimedia to reinforce audiences to spend money on games, game items, game time etc.

The fifth function of Multimedia: Activation

The goal of each MMO game is to get people to play it. Most successful games do this rather well. Since people have limited time to spend on any activity choosing to play always comes with a cost and when we think how many games out there compete for player attention it becomes clear that there is a media war for attention going on. Most games of the same genre are aware of what the other games developers are doing. When is any particular patch day, any expansion etc. But more than that MMO's evolve with time which is done by gathering as much feedback from the player community as possible. The recent arrival of early access games that become playable in alpha and beta states and frequent open betas where players test new game content activate players not only to play but to take part in creating this next bit of content to the game they love and invest so much in. One of the greatest challenges of game development is to maintain the balance of focus - where the game is going and scope – what the player community wants from the game, and all the while keep the player base active and playing. Activating players to do various activities within the game world is also crucial for any successful MMO game specially RPGs. Developers think of new ways to activate players to visit certain places in the game world (gather in one specific area over another), play certain content (map, dungeon, boss encounter), play in a certain way (heal, group, raid, pvp etc.) but also activate players to write about the game, make YouTube videos, Fan art, Discussions, Fan web sites etc.

The sixth function of Multimedia: The Creation of ties of Union

The multimedia of MMO games creates stronger ties of union than any form of mass communication before. What is more important the ties of union created by television and films are artificial but ties of union created in MMO games are real and they compete with different types of belonging from outside of the game-world. While playing in virtual societies we create groups, clans, guilds etc. which become the fundamental social institution of the game world. These ties of union revolve around game activities and primarily diegetic communication but can step out of the diegetic frame of the game. Players fulfilling various roles within these groups achieve together and form unique experiences within the game world which forms real and strong ties of union with common interests, values, beliefs, opinions and attitudes. These ties of union are strengthened trough play time within the group and trough achieving various win conditions together as a group. They strengthen trough common activities, education, reinforcement, the building and conference of status, and ultimately having fun as a group. It is common for long standing guild members to develop meaningful relationships, become close friends outside of the game and do various activities in the real world. Other types of multimedia have been tied with the function opposite of Creating Ties of Union, that of Privatization *(J.A.DeVito, "Communicology" 1982)*. That concept refers to an individual shutting themselves

into their private world conferred by the media and focusing on problems that don't involve any social ties. The multimedia of MMO games doesn't suffer from that kind of problem being oriented towards player interaction and communication. We may however use the term Privatization to describe players which become more interested with the game world and game problems than the real world. Such players spend over 10 hours per day playing a certain game or games. This can naturally be a bad thing but can however lead to players becoming professional gamers and E-Sports develop at an amazing rate. In 2014 professional gamer Chen, Zhihao has earned $1,112,281.00 playing Dota 2 and it is common for professional gamers to earn several hundred thousands of dollars per year.

The seventh function of Multimedia: Persuasion

This function has primarily to do with the change in the audience's behavior. Persuasion goes hand in hand with Reinforcement. It can be seen as reverse or extreme reinforcement where instead of strengthening ones belief, opinions, values and attitudes they end up being changed. Traditional mass media were quick to induce changes of a smaller scale but larger changes like political preferences, religious attitudes or social commitments were very hard to change. MMO's are however another story. As most MMO's form virtual societies and groups they bring new social commitments and they often force choice upon the player of how to prioritize their own social obligations. The rewards of MMO's are frequently more immediate than those of sports, education or most real life commitments so younger people are quick to institute social commitments of games into their everyday life. This can become highly problematic as players can be easily persuaded to play instead of doing a plethora of other things and in turn they can sever their real life ties and substitute them with ties from the game world. Political and religious attitudes are often besides the point and not directly influenced by games but as people invest more and more time into a game they get more and more disconnected with political and religious developments in their community and the world. Thus the importance of their political and religious attitudes will surely lessen. The way MMO's are structured today makes them easy to persuade players to invest more and more of their life into the game, the rewards offered are wide in scope, ranging from immediate character advancement and power, immediate social status that comes with progress and skill, and the promise of becoming a successful streamer or professional gamer. The problem is that most games will occupy more than 50% of one's waking hours in order to achieve this promise and therefore change a person's entire life. Obviously there is high risk of ruining one's own life when choosing this road.

The eighth function of Multimedia: Providing Ethics

Traditional mass media have provided audiences with ethical systems for decades. This collective ethic was formed through various information's made public, and overall media coverage of various events. This kind of coverage forces the audience into a degree of action against something that cannot be tolerated, or for something that the public suddenly approves. It seems that MMO's reflect this behavior but mostly centered on problems that have to do with games. It has little to do with religion, sexuality and race but can have lots to do with similar concepts if they belong to the game world. Factions can express hate towards one another, classes and character types can be labeled inferior and create different concepts of right and wrong. Most games teach us how to treat other players, especially new players. Many games (particularly Mobas or FPS games) have very toxic communities which will insult new or less skilled players to that extent that it may lead to people stop playing. One of the great examples of MMO ethics is the great debate among player communities on the theme of DLC (downloadable content). For years various game developers created DLC as free additions to the game with the goal of keeping the player base active. But as the games switched to different monetization models DLC became a way of monetizing the game and creating a successful business model. New areas such as maps, new items, new encounters, new characters, new cosmetic items or appearance changes etc. all represent various forms of DLC which is popularly monetized. Most DLC are expendable as to persuade players to keep buying them. The problems arise when we question whether or not these DLC are needed to get a good experience of the game and how much they cost. The public verdict as of now, which is present in forums, game chat, YouTube shows by enthuse journalists and such, is that DLC is fine as long as it doesn't give a player a significant advantage in the game making it mandatory. Cosmetic DLC only is generally fine since it gives players no real advantage but it can be frowned upon if it is "day1 DLC". Day1 DLC is DLC which is available for purchase on the day the game is released. This is deemed non ethical because if it is developed already why isn't it in the game we bought? If the game is however free, than it is acceptable that there is day1 DLC of the cosmetic nature.

The ninth function of Multimedia: Narcotization

In traditional mass media this function was about audiences mistaking knowing about problems for doing something about them. Narcotization in MMO's is similar in that it lets game related problems substitute other life problems and drugs players into inactivity in all non game related fields. Character progress, achievements, guild progress, rank and status, team standing, ones place on a ladder, personal score etc can become problems of such importance that they shadow ones job, school and any other obligations. People may still go to their job and attend school but they will use every minute they can to focus on the

problems from the game. They will watch videos from the game, write on forums, and try to stay connected to the game world in any way they can until they get the chance to play and tackle their primary concerns. Often people can be in the game doing nothing for hours because they are waiting for a group activity to start, or trying to find a way to get satisfaction trough any kind of advancement. Games can in this way become highly addictive.

Conclusion

This analysis shows that we have to keep the various functions of MMO games into account when choosing them. These various functions are not separate but they work together in a package. Games will activate players into playing, reinforce certain behaviors, promise rewards and persuade them to go after those rewards. They will be entertaining and fulfilling, create strong ties of union with other players inside the discourse of the game world and create fertile ground to narcotize the players getting them addicted. More importantly we cannot sit idly while our children choose games they will play not knowing the structure of the factors of that game. We need to be involved and understand the factors of various games and what their impacts are. The trend is that most games today have an expert level of maximum rewards which is reserved only for the professional players, the die-hard elite. The promise of those rewards the games tease us with is not realistic for most players and the question remains what to do with the allure they pose, and the potential hazard the road to that promise means? Is it a question of ethics to find a way to separate professional gaming from casual gaming? It is our obligation as parents, teachers, and elders to step in reshape and redefine the road which MMO games of today represent. MMO games today influence our society and shape our youth and it is our obligation to stop sitting idly while this process is going on. We have to explore and establish how we want the games to be tied into our society and what kinds of rewards do we want to attribute games with. We have to define where that thin line between entertainment and work, leisure and obligation is. It is high time to intervene and redefine what it means to be n00b and what it means to be pr0!

References

Afrić, Winton; Šimović, Vladimir; Milković, Marin. „The sense of (RPG) Role Playing Games Systems“. Koprivnica: Medijsko Sveučilište. 2013.
Athique Adrian. „Digital Media and Society: An Introduction“. Polity, 2013.
DeVito, Joseph A. “Communicology”. New York: Harper & Row, 1982.
Fine, Gary A. „Shared Fantasy Role playing games as social worlds“. University of Chicago. 2002.
Johansen, Jorgen D.; Larsen Svend E. „Uvod u semiotiku“ Zagreb: CroatiaLiber, 2000.
McLuhan, Marshall; Fiore, Quentin. „The Medium is the Massage“. Gingko Press; 9th edition, 2001.
McLuhan, Marshall; Gordon, Terrence W. „Understanding Media: The Extensions of Man : Critical Edition“. Gingko Press; Critical Edition, 2003.
Schramm, Wilbur L. “Men, messages, and media: A look at human communication”, New York: Harper & Row, 1973.
West, Richard; Turner, Lynn. “Introducing Communication Theory”, McGraw Hill. 2007.
Williams, Patrick J.; Hendricks, Sean Q.; Winkler, Keith W. „Gaming asa Culture: Essays on Reality, Identity and Experience in Fantasy Games“. McFarland & Company, Inc. 2006.

Cooperation of a museum institution and students in creating virtual exhibitions using the MOVIO tool

Goran Zlodi, Josip Mihaljević, Anica Šimunec, Paula Mikeli, Ivana Rudić
Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
gzlodi@ffzg.hr, jmihalje2@ffzg.hr, anica.simunec@gmail.com,
pmikeli@ffzg.hr, irudic@ffzg.hr

Summary

MOVIO is one of the creative tools from AthenaPlus network. Its purpose is to facilitate the creation of virtual exhibitions. MOVIO is very similar to other content management systems (CMS) but it focuses more on creating interactive multimedia page contents, such as digital photo galleries and exhibitions, games, interactive timelines, and digital storytellers. MOVIO tool provides and also shows how its different options can be easily and efficiently used by users who do not have any specific IT knowledge such as advanced programming or web design techniques, to create and show different types of virtual exhibitions. Virtual exhibition "For Better or Worse… / Wedding Fashion from 1865 till Present" is a virtual extension of a physical exhibition held in the Museum of Arts and Crafts in Zagreb. Virtual exhibition catalogue was developed using the Timeline tool as a page type which proved to be very effective in presenting the exhibition conceptualized chronologically. The Timeline allows easy browsing through virtual objects within the timeframe we have previously determined. The other virtual exhibition, "Herman Bollé – Builder of the Croatian Capital" is also a virtual extension of the real exhibition held in the Museum of Arts and Crafts. Using the MOVIO tool and information from the catalogue of the real exhibition, the virtual exhibition was made making the most MOVIO options. The FFZG team used the following page types Home Page, Page, Google Map, Storyteller and Timeline. Furthermore, Bolle's virtual exhibition has a unique content as compared to other exhibitions. It is an interactive game (among few simple puzzle and memory games) made in the Unity game engine which allows players to try to place many of Bollé's architecture designs and buildings on the appropriate location on the map. Tools like MOVIO show that technology has evolved and has become simplified for end users so that they can easily create virtual exhibitions production of which was complex or costly several years ago.

Keywords: virtual exhibition, multimedia content, interactive game, interactive timelines

Introduction

The aim of the European founded project AthenaPlus is providing new tools to support cultural institutions to narrate their resources. These new tools where created by the Italian enterprise GruppoMeta on MOVIO platform which serves as innovative application for creating digital exhibitions and thematic routes. MOVIO offers different page templates that use different tools for showcasing digital content that can combine multiple media such as text, photo galleries, animations, timelines, games, Google map, and probably more content in the future. Layout of each of these pages is responsive, so it fits screens of all sizes and there is also multilingual support for creating content in many different languages (Minelli et al. 2015). This paper is a result of collaboration that was established between the Museum of Arts and Crafts in Zagreb and the Faculty of Humanities and Social Sciences in Zagreb. Museum of Arts and Crafts participates in the project AthenaPlus, whose goal is to upgrade the terminology and enhance the *Europeana* search engine. At the same time there is great emphasis on re-use of content and the development of tools and applications that support new possibilities of presenting cultural heritage. As part of the project, MOVIO tool for creating online digital exhibitions is in the developing process and several partners are testing the tool. Museum of Arts and Crafts created its own digital exhibition entitled "A Century of the Wristwatch" (http://54.247.69.120/build/movio/movioTraining12) and furthermore, museum's staff (Petra Milovac and Iva Meštrović, curators) held a training on how to use the tool at the Faculty of Humanities and Social Sciences (course Virtual Museum, Goran Zlodi PhD, assistant professor). After the training, students tested the tool as well, creating two digital exhibitions: "For Better or Worse… / Wedding Clothes from 1865 to Nowadays" (http://54.247.69.120/build/movio/movioTraining39/) and "Herman Bollé – The Builder of the Croatian Capital" (http://54.247.69.120/build/movio/movioTraining40/en/). In this paper authors will present their experiences of creating online virtual exhibitions with the MOVIO tool.

MOVIO – online virtual exhibitions

MOVIO is one of the creative tools from AthenaPlus network. Its purpose is to facilitate the creation of virtual exhibitions (athenaplus.eu). MOVIO is very similar to other content management systems (CMS) for creating websites such as WordPress and Joomla!, but it focuses more on creating interactive multimedia page contents, such as digital photo galleries and exhibitions, games, interactive timelines, and digital storytellers. Most of MOVIO's functions are intended for presenting the content of museums and galleries in a variety of ways. For example, Partenon frieze is a part of a plaster cast of ancient sculptures at the Faculty of Humanities and Social Sciences of the University of Zagreb. It can be shown as a simple photo gallery with a slider for images, as a storyteller or a timeline, where each part of the frieze is chronologically explained with images and texts (wiki.athenaplus.eu).

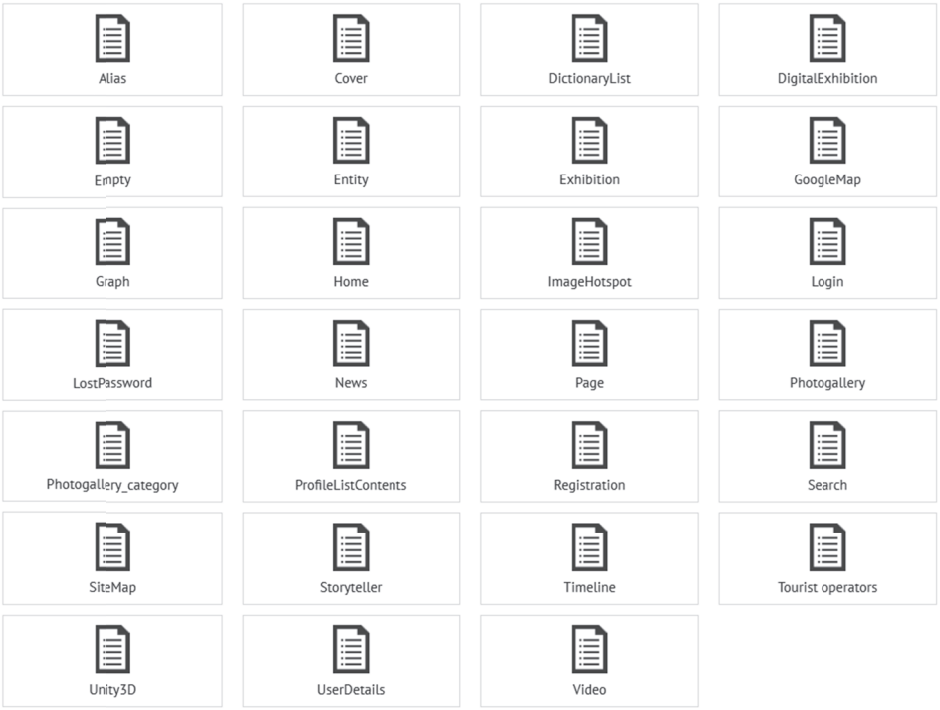


Figure 1. Different types of page contents available with MOVIO

Only users with a username and a password can access the back-end interface. The administrator can create multiple accounts for other people to enter and create site content. However, only the administrator has the right to delete pages, hide them from visitors or change them to private. The back-end is the area for the Ontology and Content Management of MOVIO. It consists of three sections:

A. The section with the list of items on which it is possible to work in the back-end;
B. The section with the index of the structure of pages;
C. The section to create and edit the page content and the ontology structure.

B section is a list of created pages in a MOVIO instance. The Home page from which others originate is always on the top. Underneath it there are three pre-defined folders Metanavigation, Footer and Utility. Other pages and folders are created by the user or administrator. Pages can be rearranged via *drag and drop*: the user simply selects a page and drags it to the desired position. Pages can also be nested one into the other to create better site structure (wiki.athenaplus.eu). The CMS system allows its users to edit page content, therefore each page type can be changed, including its text and media. MOVIO uses

WYSIWYG (What You See Is What You Get) text editor, which is very similar to text editors that are mostly used on web forums. It has basic text formatting buttons such as bold and italic, text alignment buttons, citation button, undo and redo buttons, insert symbol button, hyperlink button, insert table button, anchor button, image button, and a HTML button which allows coding in the editor (wiki.athenaplus.eu).

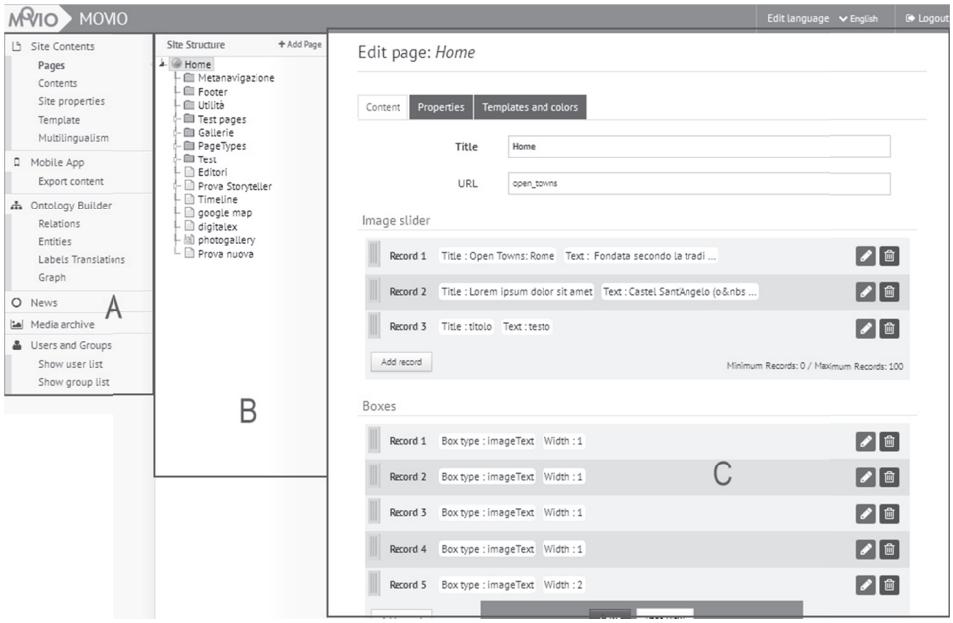


Figure 2. Three sections of MOVIO back-end

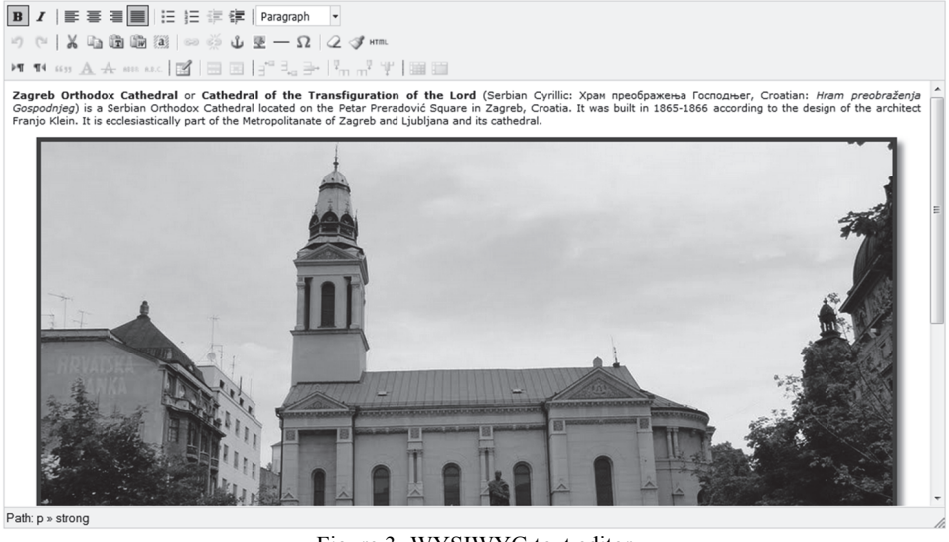


Figure 3. WYSIWYG text editor

One of the very important characteristics of MOVIO is the possibility to add an unlimited number of new languages, and choose the default one. Enabled languages are always shown on the right hand side of the top bar of MOVIO in the back-end. When a page is displayed on the front-end part, language options are switched to the left side of the top bar. When creating content for multiple pages, it is important to have in mind which language is set as the default – one should always create new content in the default language, and MOVIO will transfer the content to other languages automatically. Author then only has to translate the texts to other languages (wiki.athenaplus.eu). The back-end part of MOVIO has one part for the management of multimedia objects in the system called Media Archive. Media Archive stores and organizes multimedia content such as images, audio and video files, PDF and Word documents, as well as Unity game extensions for future use. When the user uploads content to the Media Archive, he can add metadata such as author, category, date, copyright type, and description. There is also an option which allows zooming and downloading of pictures and the insertion of the watermark option. After media is uploaded in the archive, it can be used for creating pages. Finding the required item in the Media Archive is facilitated by the search bar that allows filtering based on file metadata such as author, title, file name, category, and even description. Users can also filter content based on the type of the media (wiki.athenaplus.eu).

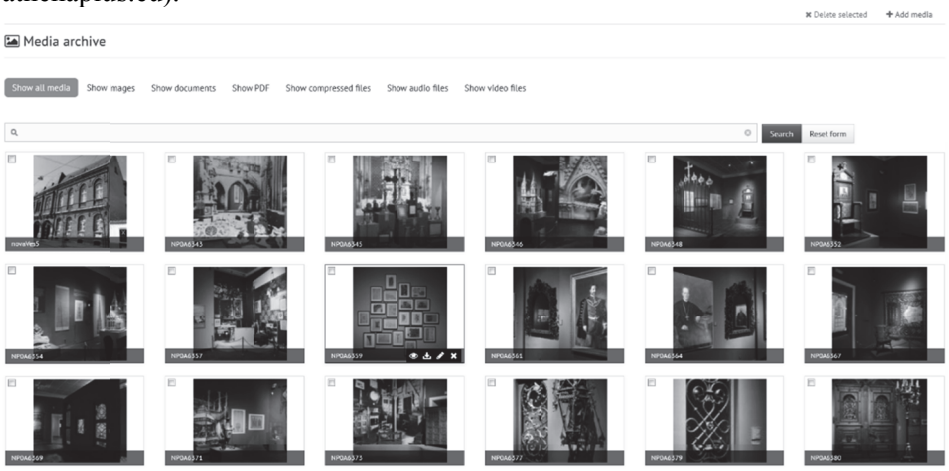


Figure 4. Media archive interface

Speaking of various MOVIO services, the creators strongly feel that the Ontology Builder function/tool is one with the most potential. The Ontology Builder can be used as a solution for displaying the content of a virtual exhibition in a non-linear way. Entities of artists, artwork, and the departments of the University, as well as the relations between them were developed with this tool/ function (Minelli et al. 2015).

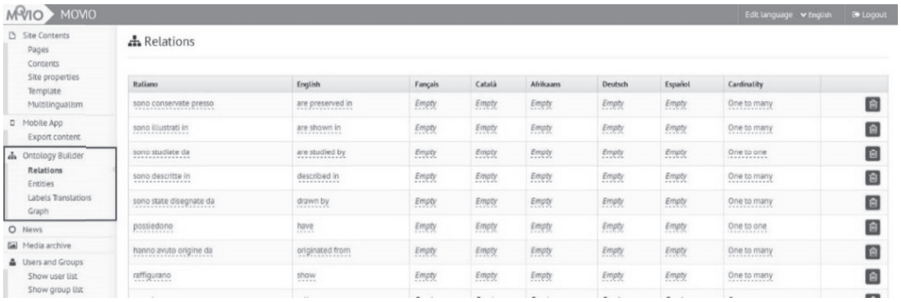


Figure 5. Example of relations in Ontology Builder

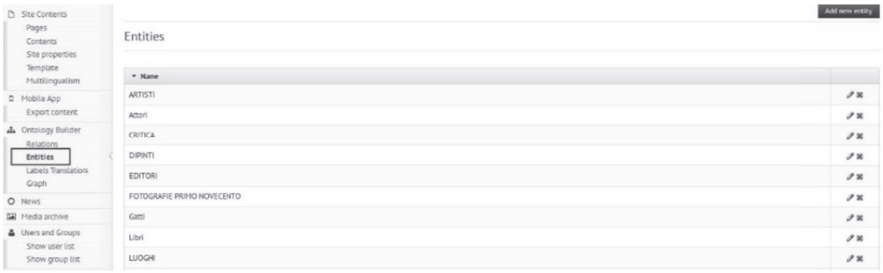


Figure 6. Example of entity in the Ontology Builder

Figure 7. Example of using the created ontology for searching content (wiki.athenaplus.eu)

Case Studies

Virtual exhibition "For Better or Worse… – Wedding Clothing since 1865 until Nowadays"

Virtual exhibition "For Better or Worse… – Wedding Clothes from 1865 to Nowadays'" is a virtual extension of the temporary exhibition held in the Museum of Arts and Crafts in Zagreb from February to May 2015 (author: Andrea Klobučar, senior curator). Virtual exhibition catalogue was developed using the MOVIO Timeline tool which proved to be very effective in presenting the exhibition conceptualized chronologically.

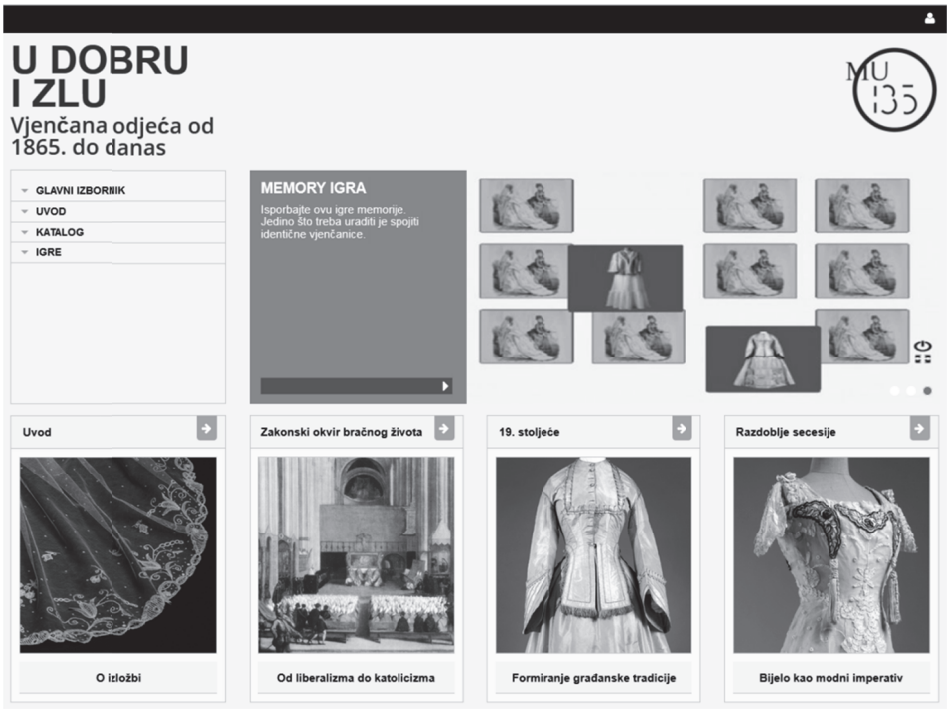


Figure 8. Home page

Timeline allows easy browsing through virtual objects within previously determined timeline. During the development of the timeline, a time period was selected, within which the objects were located automatically and chronologically, depending on the year, month or day assigned to them. For every entered record, MOVIO allows contents: text, images, videos, etc. (wiki.athenaplus.eu). If the content exceeds the provided space for object presentation, in the front-end on the left, a scroll bar automatically appears – this allows the rest of the page to remain static, while the presentation space scrolls up or down. In the front-end, MOVIO tool allows browsing using pointers (left or right), but the user can also

select the exact object he is interested in on the timeline, which is always presented beneath the presentation space. If, however, the user decides to use pointers to navigate left or right, bellow every pointer the heading of the following virtual object is written, which makes browsing easier. MOVIO tool also enables zooming in or zooming out the timeline, which allows the user to enlarge records and see them more clearly or make them appear smaller but more numerous. In the virtual exhibition "For Better or for Worse – Wedding Fashion from 1865 till Present" the main timeline of this exhibition spans over 150 years and it is divided chronologically into shorter periods of a few decades or even less. For example, within those 150 years there are periods such as 19th century wedding clothing, Art Nouveau period, Inter-war period, etc. By clicking on the preferred period another timeline opens in which the objects are again chronologically presented with pictures and a short text. The Timeline allows an excellent overview of a chronologically designed exhibition and the tool is able to do a great part of the work by itself. The user only has to enter the content and determine the timeframe within which the presentation will be made, which makes this tool easy to use. The FFZG team suggested the expansion of the number of records for the Timeline tool. Also, there is a team working at the Museum. However, currently there is a team working in the Museum of Arts and Crafts that reports all current mistakes, bugs and possible improvements to the AthenaPlus network. The programmers from the GruppoMeta are fixing the issues and working on the improvement of the tool.

Figure 9. The Timeline

Virtual exhibition "Herman Bollé – Builder of the Croatian Capital"

Virtual exhibition "Herman Bollé – Builder of the Croatian Capital" is a virtual extension of the real exhibition held in the Museum of Arts and Crafts in Zagreb from April to August 2015 (author: Dragan Damjanović PhD., associate professor). Using the MOVIO tool and information from the catalogue of the MUO temporary exhibition the Faculty of Humanities and Social Sciences team of students created the virtual exhibition. The team tried to make the most of this tool, so they used various different page types such as Home Page, Page, Google Map, Storyteller and Timeline. All page types have common features which are Title and URL, and differ in specific features unique for every page type. In this chapter we will explain the possibilities of these pages and share our experience with each of them.

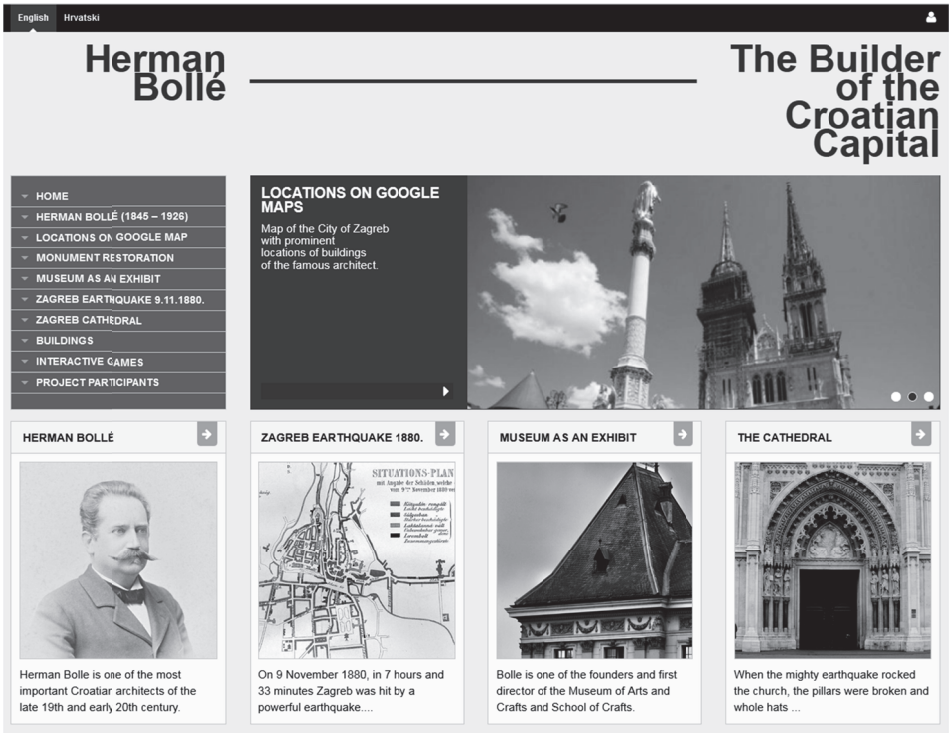


Figure 10. Home page

The Homepage Page Type created in MOVIO allows the creation of a page consisting of a title, one or more Sliders and one or more Boxes. A new Home Page Type is created by selecting Home from the menu Page Types (wiki.athenaplus.eu). For the Slider of the homepage the user can add a title, text, an

image and a reference to one of the pages in the MOVIO exhibition. For the homepage of the exhibition the team created three slides (Timeline, Locations on Google maps and Herman Bollé – Builder of the Croatian Capital), each of them having a reference to a specific page of the exhibition, image, and a short introductory text.

Figure 11. Homepage slider

A Box is an element which is able to host two different types of media combinations: an image with the text or an image with a link. The team created seven boxes, all of them contain an image, a short text and a link. Links of boxes lead to main themes of the exhibition which are: Herman Bollé, Zagreb Earthquake, Zagreb Cathedral, Museum as an Exhibit, Location on Google Maps, Buildings and the interactive Game.

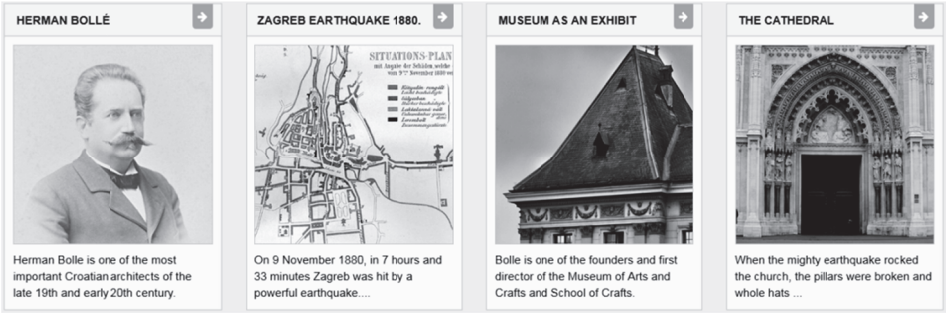


Figure 12. Boxes

Page type: Page

The Page Type Page allows creation of a classic page for a website, a page that can be composed of a title, text, a sequence of images, and a sequence of attachments (wiki.athenaplus.eu). This page type was most useful in creating this virtual exhibition and the most comprehensive part of it. The team used this page type in the process of creating main themes (Zagreb Earthquake, Zagreb Cathedral, Museum as an Exhibit and Buildings) and their subtopics. As the

basic source of information the team selected the catalogue of the real exhibition and relevant secondary sources as well as popular web pages with information about buildings and events. The team used all of the possibilities provided by this page type.

Page type: Timeline

Since the possibilities of the page type Timeline have already been described in this study, this part of the article will focus on the team's experience. The timeline of the exhibition represents the life of Herman Bollé and his work. The team has set up six timelines, which cover the time period from 1845 to 1926. Each item of a specific Timeline has an image, a short text and a date.

Page type: GoogleMap

The GoogleMap Page Type allows the Point of Interest (POI) to be graphically displayed on a map. The tool Google Maps used in MOVIO is a Google service (developers.google.com/maps/). In particular, the interactive geographic maps use the following library: http://hpneo.github.com/gmaps/ (wiki.athenaplus.eu). This page type was very useful for the exhibition – it enabled the team to represent the work of an architect whose buildings spread across Zagreb. The team did not include buildings that were not in Zagreb. The map has 29 locations, each of them has an image and a title. Adding locations on the Google maps with MOVIO is done with Markers, which are composed of POI and texts (address and name of the building).

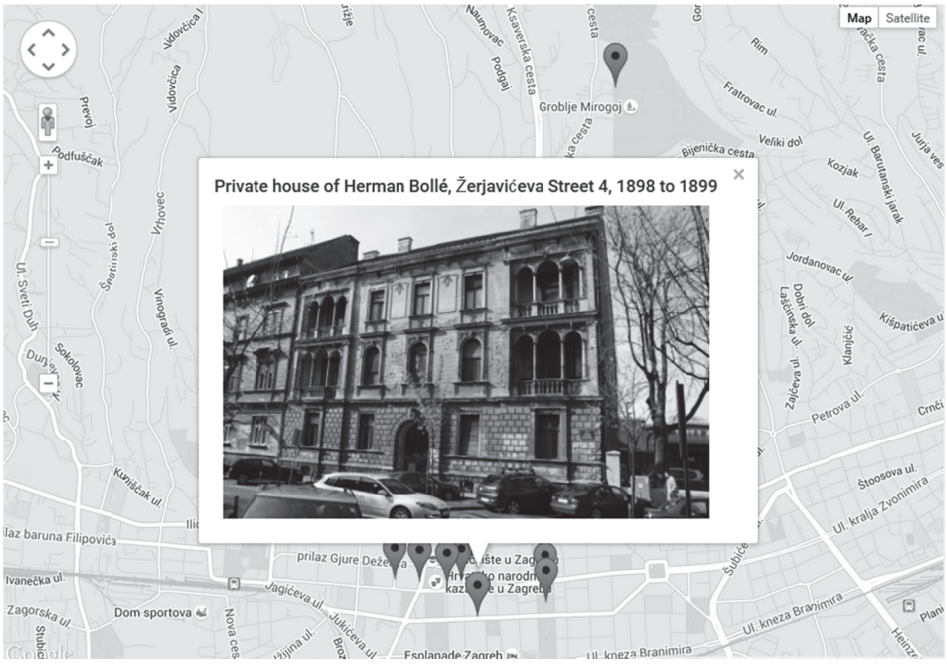


Figure 13. Google maps

Interactive game

In addition to the functions available in MOVIO, in for this exhibition a member of the team has integrated additional content. It is an interactive game (http://54.247.69.120/build/movio/movioTraining40/?pageId=89) which allows players to attempt placing many of Bollé's architecture designs and buildings to appropriate locations on the map. The game was originally planned as an interactive map made with the *ImageHotspot* page template, but the team realized that it would be better to seize the opportunity of the content and create a simple, but interesting and effective, drag and drop game. The Game is made in the Unity game engine because MOVIO supports Unity. The map with the locations was given to students as one of the referential materials for creating site content. It was the perfect material for a drag and drop game although that was not its original purpose. Because all buildings and architecture designs where shaped as round circles with the exact same size, they could easily be cropped as a piece that can be placed on any of the locations.

Figure 14. The map used for the interactive game

There are 28 building that where cropped from the map using a free and open-source raster graphics editor GIMP. Each of the cropped buildings was saved as a small JPEG image that was later imported as an asset into the Unity game engine along with the map with empty round-shaped spaces in which the buildings were to be inserted. Each asset in Unity has its own code which defines the actions that the user will have if he moves one of the objects with a mouse or drags it on a touch screen. If the player guesses the approximate position of the location of the building, it sticks to that location. This game is interesting because it is educational and meant for players who do not know much about certain building locations. Most of the adults in Zagreb recognize the buildings but are not sure where they are positioned. It is also more complex than common drag and drop games because there are 28 locations for which the user should know the exact location. While similar children games are composed of less objects, this game requires more time and certain geographic memory, therefore it is more appropriate for adults. Regardless of this fact, children can also easily play it and learn about some culturally relevant locations on the map such as St. Mark's Church and the Zagreb Cathedral. If the player is not sure where a certain location is, he or she can simply click the Solution button which places all buildings in the correct place. The game is functional but it could be updated in the future by designing better user interface and adding solutions for one location at a time. There is also a problem that Unity has with browsers, since it requires the installation of the Unity web player plug-in to work with a certain browser. MOVIO tool does not support the more widely used Flash plug-in and it does not have options to support HTML5 games which would work equally well on all browsers and platforms without the need for additional plug-in installation. However, with future updates such options will probably become possible. This interactive game can be used as an example of how easy it is to create interesting and simple multimedia content by using only programming tools that are free.

Conclusion

Tools like MOVIO show that technology has evolved and has been simplified for end users so that they can easily create virtual exhibitions that were just few years ago too complicated even for expert programmers. With MOVIO tool archives, museums and libraries can create rich and interactive multimedia content for presentation of their assets. This article also shows that usage of contemporary information and communication technology tools like MOVIO gives new opportunities of cooperation between cultural heritage institutions and university students what could result in benefits for both parts and, what is most important, benefits for end users.

References

ATHENA Plus – Access to Cultural Heritage Networks for Europeana. http://www.athenaplus.eu/ (2015/06/25)

Athena plus wiki. The media archive. 15 April 2014. http://wiki.athenaplus.eu/index.php/The_media_archive (2015/06/25)

Athena plus wiki. Multilingualism in MOVIO. 21 April 2014. http://wiki.athenaplus.eu/index.php/Multilingualism_in_MOVIO (2015/06/25)

Athena plus wiki. The WYSIWYG text editor. 21 April 2014. http://wiki.athenaplus.eu/index.php/The_WYSIWYG_text_editor (2015/06/25)

Athena plus wiki. PageType: Timeline. 21 April 2014. http://wiki.athenaplus.eu/index.php/MOVIO_Pagetype:_Timeline (2015/06/21)

Athena plus wiki. PageType: Google Map. 21 April 2014. http://wiki.athenaplus.eu/index.php/MOVIO_Pagetype:_Google_Map (2015/06/30)

Athena plus wiki. PageType: Page Map. 21 April 2014. http://wiki.athenaplus.eu/index.php/MOVIO_Pagetype:_Homepage (2015/06/30)

Athena plus wiki. PageType: Home Page Map. 21 April 2014. http://wiki.athenaplus.eu/index.php/MOVIO_Pagetype:_Page (2015/06/30)

Athena plus wiki. Introduction to MOVIO back-end. 2 June 2014. http://wiki.athenaplus.eu/index.php?title=Introduction_to_MOVIO_back-end (2015/06/25)

Meštrović, I. (2015). Digital and Virtual Exhibitions – A Glance at the Situation in Croatia. UncommonCulture: Virtual Exhibitions. 6(1): 114-119.

Minelli, Habibi, S., Meštrović, I., Milovac, P., et al.
(2015). Museums' Experiences in Creating Cultural Narrations Using the AthenaPlus Tool Called MOVIO. Uncommon Culture 6(11): 67-87.

MOBILE SOLUTIONS

Original scientific paper

Wireless Network Security recommendations Using the Application for Security Evaluation

Aleksandar Skendžić
Polytechnic "Nikola Tesla" Gospić
Bana Ivana Karlovića 16, Gospić, Croatia
askendzic@velegs-nikolatesla.hr

Božidar Kovačić
Department of Informatics, University of Rijeka
Radmile Matejčić 2, Rijeka, Croatia
bkovacic@inf.uniri.hr

Edvard Tijan
Faculty of Maritime Studies, University of Rijeka
Studentska ul. 2, Rijeka, Croatia
etijan@pfri.hr

Summary

The proposed system of security recommendations of wireless local area network allows applications to achieve higher levels of security. In order to build a security model, it is crucial to pre-evaluate the parameters that affect the security of the wireless network. When evaluating the parameters, expert literature along with practical experience of network administrators has been used. The results of evaluation parameters are included in the constructed security model of the proposed application. The proposed model contributes to a simpler problem solving of wireless network security through the evaluation of safety parameters. In addition, the proposed system gives recommendations regarding security at two levels, together with an appropriate security evaluation. The chosen safety parameters were evaluated using a questionnaire among CARNet system engineers in educational institutions. The results obtained may help to efficiently prevent wireless network security breaches.

Keywords: open source e-bus system, wireless network security, evaluation

Introduction

Configuring security is one of the main problems of wireless networks. It can be hypothesized that the security of wireless networks is lower than security of wire networks [1].

Security is a key element in wireless communication because the communication occurs via an unreliable media (air) [2]. Safety of networks, services and

transactions is essential for the creation of trust in various forms of personal communication. A threat in network environment is defined as a circumstance, condition or event that can harm the network and computing resources in the form of destruction, disclosure, modification of data, denial of service, fraud and abuse [6]. In order to protect the wireless network communication channel, numerous algorithms [8], certificates and protective mechanisms have been defined and used for the protection of wireless local area network (WLAN). They are an integral part of the security policy of institutions or organizations, and are carried out to a certain degree.

In the development of the proposed security model, the protective measures to be employed rely on the use of wireless networks security mechanisms in order to reduce the risk of security breaches. The choice of mechanisms for protection of wireless networks, with regard to the purpose of the local network, can result in optimal security solution that can be applied. If the effectiveness of wireless network security is confirmed by expert evaluation, the risk is reduced, and security is not compromised. If safeguards are not effective, security could be directly compromised. Although the security level cannot reach 100%, it is necessary to attempt all the necessary means of increasing the security level. Consequently, a higher security level requires greater financial investments, which implies a higher cost of planning and setting up the active wireless network equipment. In determining the concept of wireless network security, special attention should be given to the following segments:

- protection of an institution's information system,
- protection of personal data (on networked computers),
- restricted user access (user levels and user rights),
- use of standard encryption algorithms,
- use of compatible active network equipment,
- ease of network access,
- existence and enforcement of security policies [10].

The rest of this paper is structured as follows: Chapter 2 gives a description of the security system; Chapter 3 presents the methodology and tools used to develop the system for wireless network security evaluation; Chapter 4 describes the development and structure of the system; Chapter 5 and 6 offers security parameters evaluation and the interpretation of security evaluation values; finally, we conclude the paper (Chapter 7) and list references.

Security System

At the beginning, it was necessary to restrict the parameters that are an integral part of the overall security system. In the first phase of the study, the parameters that affect the security of the network were analyzed. Expert literature has been used for the purpose of determining and specifying the security parameters. Based on that, a questionnaire was devised and filled in by network adminis-

trators from state educational institutions (CARNet[1] members). These parameters were evaluated by network administrators based of the existing wireless network system in their home institutions.
Defining the level of security implies the creation of security system based on the optimal selection of parameters that influence it. The effective functionality of the system is associated with identifying the actions that interconnect all the elements of the security system. Regarding the security management process, the often applied methodology is known as Plan-Do-Check-Act (PDCA) [5]. The methodology is also known as the Deming cycle. Deming cycle of improvement always starts by analyzing the current situation, followed by deducing the problem.

Choosing a Tool for Building a System for Wireless Network Security Evaluation

The system for security evaluation is implemented as an application for wireless network security evaluation. The application enables users action at two levels:

- Network administrator (expert) level of action,
- Examinee level of action (application user).

Figure 1 provides an overview of procedures which define the level of user action. These procedures are available to the user through authentication procedure of the system. Additionally, users are able to use certain procedures that do not require authorization.

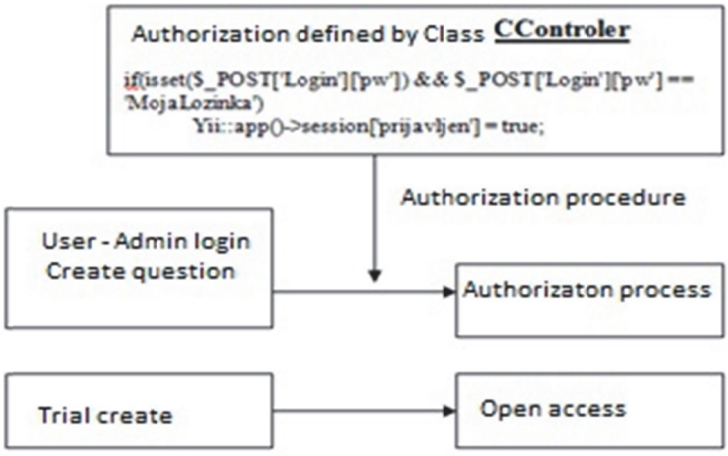


Figure 1. Defined user authorization procedure (source: authors)

The level of action of the network administrator (expert) through the interface provides the ability to create questions and multiple answers. Questions are

[1] Croatian Academic and Research Network (CARNet)

measured by weight grades and higher weight rating implies a greater impact of the given parameter on the overall security of wireless networks. The logging module enables a user to log in and create questions by selecting *Log in*. This module also allows user testing using the created questions. The questions create a structured tree with subordinate and super ordinate relations. By selecting the *Create Questions* option , the user has to fill in the required input fields: the number of the subordinate question, question description and the number of the super ordinate question. The button *Create* creates a question, and has the following required fields: question number and description. A tree structured in this way provides an overview of super ordinate and subordinate relations between questions and provides the option to edit questions and create multiple responses. The button *Create response* gives the opportunity to define all the answers to a question. Weight value is added to the answers based on user evaluation.

The level of action by the respondents relies on the possibility of evaluating wireless network security in the login module of the application. By selecting the *Create Test* option, user fills in a questionnaire which contains previously entered parameters by the administrator. The user selects each parameter and provides an answer from the list of possible answers. In the end, the result is checked through the recommendation module. By comparing the weight value of each parameter, recommendations are given to the user in order to achieve higher security levels. The level of action by the user is achieved by logging into the system via the administrator password. Users wishing to check network security select the option for creating a test.

For the construction of the system, PHP (server side) and Javascript (client side) have been used, driven by MySQL database management system. Development tool YII [9] was used as the framework. The use of scripting languages in the dynamic generation of web sites contains two main elements: (1) server with a programming platform and (2) database with the associated database management system and script language.

Development and Structure of the System for Wireless Network Security Evaluation

The application for wireless network security evaluation is based on the schema shown in Figure 2. Such a model may result from the transformations of the Entity-Relation (ER) data model [3]. The method of constructing an ER model is well known. The model was constructed according to Chen notation and uses key inheritance, while the weak entity type was determined according to the MIRIS[2] notation [4].

[2] Metodologija za Razvoj Informacijskog Sustava (MIRIS) – Methodology for Information System development.

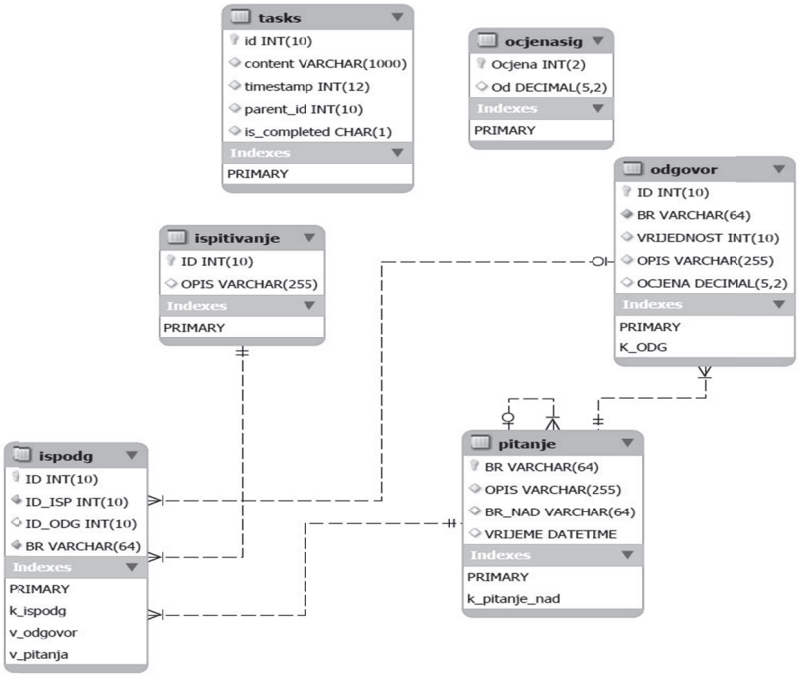


Figure 2. E-R diagram (schema) of the *ispitmreza.sql* database obtained via MySQL Workbench application (source: authors)

Access rules have been defined for every controller, which are executed only if user is logged into the system. For example, *IspitivanjeControler* enables acces to all users without the need for authentication. *IspitivanjeControler* controls the acces to the *Ispitivanje* module.

Evaluating the Selected Wireless Network Security Parameters

Questionnaire was consisted of 19 questions which are oriented on default wireless network security parameters [19]. Questionnaire was distributed through Google docs form in the period of May 2014 – July 2014 to a CARNet system engineers (59 respondents). Every engineer gave an answer and score to the each security parameter. Key questions were based from installed WiFi [18] network in organisation, security policies to evaluation of each security parameter which is shown in Table 1. Collected questionnaire data is used for application development.

The developed system gives recommendations based on the evaluation of security parameters which were defined using expert literature and network administrators (CARNet system engineers). The limitation of the system is the way parameters are evaluated, which relies on the weight value of each parameter. In

cases where two security parameters cannot be evaluated differently, the evaluation system will offer both logical solutions. Table 1 offers an overview of security parameters being evaluated. According to the calculated averages of the questionnaire results in which CARNet system engineers evaluated security parameters, the weight factors have been calculated, as shown in Table 1.

Table 1: Awarded weight factors (ponders) of security parameters based on research results

No.	Parameter	Weight factors (research)	Weight factors (%)	Parameter/ average
1.	Encryption	0.15	15.00%	4.03
2.	Network hardware model (type)	0.08	8.00%	3.74
3.	WIFI coverage	0.18	18.00%	4.15
4.	Firewall	0.21	21.00%	4.24
5.	Number of WIFI users	0.09	9.00%	3.77
6.	Defined security policy	0.13	13.00%	3.93
7.	Services	0.1	10.00%	3.81
8.	VPN	0.02	2.00%	3.41
9.	RADIUS + LDAP directory	0.04	4.00%	3.66
	Total	1	100 %	

As seen in Table 1, the weight factor 0.15 has been assigned to the encryption parameter (WEP[3], WPA[4], WPA2) [7], because it is considered that the encryption directly affects the security of wireless networks, i.e., wireless networks of an open type are extremely vulnerable to security threats and the use of encryption is strongly recommended in order to achieve a satisfactory level of security [12].

The model (type) of network equipment has been assigned a weight value of 0.08. On the network equipment market, various manufacturers offer diverse active network equipment. Active network equipment with greater capabilities can affect the security of the wireless network. Wireless network signal coverage has been assigned a weight value of 0.18. If there is a need for greater availability of wireless network signal, inside or outside the building, network security is decreased because the network covers a larger area that is accessible to more people and is more vulnerable to security breaches [13]. Firewall usage has been assigned a weight value of 0.21. LAN or wireless network firewall is extremely important because it filters the network traffic from the sender to the receiver and vice versa.

[3] Wired Equivalent Privacy (WEP)

[4] Wi-Fi Protected Access (WPA)

The number of wireless network users is proportional to the wireless network signal spatial coverage, and has been assigned a weight value of 0.09. Greater wireless network coverage generally implies more users and lower security level, and vice versa [14]. The definition and the existence of security policies in the organization have been assigned a weight value of 0.13. Encryption, firewall and security policy are key elements in both LAN and wireless network security within the organization. The security policy defines the planning and describes the goals or procedures of security. The network services have been assigned a weight value of 0.1. Educational institutions basically use e-mail services, Web services and file transfer services (FTP[5]).
The minimal weight value of 0.02 refers to the use of virtual private networks (VPN[6]). Such WLAN setup is neither standardized nor mandatory, but contributes to increasing the overall network security. The weight value of 0.04 refers to using the RADIUS protocol and LDAP directory service that are well represented in higher education institutions. Through LDAP + RADIUS, user accounts are assigned certain rights regarding the use of network resources (access to services and applications), which contributes to network (and overall) security. According to the survey, the calculated weights indicate a high correlation (strong association defined within ± 0.70 to ± 0.90) with the results of the study r = 0.78 (Table 1). Two parameters (VPN, RADIUS[7] + LDAP[8]) have low weight values (2% and 4%) and are almost negligible in the final evaluation, because they show low influence on the final result.

The Range of Values of Security Grades

The grades for assessing wireless network security are shown as values between 1 and 5, where 5 represents the highest level of security (Table 2). The range of values of security grades as well as the percentage of each evaluated parameter within the overall wireless network security is given in Table 2.

Table 2: The range of values of security grades

Security grade (level)	Grade (MIN)	Average grade (AVG)	Grade (MAX)		
1	0	20	29		
2	20	40	49		
3	50	60	69		
4	70	80	89	Correlation (r)	
5	90	100	99	Ponders - research	0,779604014

Source: Authors

[5] File Transfer Protocol (FTP)
[6] Virtual Private Network (VPN)
[7] Remote Authentication Dial-In User Service (RADIUS)
[8] Lightweight Directory Access Protocol (LDAP)

System engineers have evaluated the parameters that are used in the proposed model of evaluation. A random sample of examinees was used (N=59), all of whom are employees (system engineers) of the CARNet member institutions. The questionnaire consisted of ten questions. Each examinee needed to evaluate wireless network security parameters individually.

Questionnaire items consisted of statements, and by evaluating each of them, average values for individual statements and parameters which influence wireless network security were obtained. In order to evaluate preparatory tasks, a five-point Liker scale was used (1 = security parameter not important, 2 = important to a small degree, 3 = security parameter is important, 4 = security parameter is very important, 5 = security parameter is crucial). Average values of security parameter grades were used as the basis for establishing the level of wireless network safety in institutions. The main assumption is that if the average grade value of a particular wireless network security parameter is lesser than 2 (<2), it does not represent a satisfactory level within the overall wireless network security, as it directly lessens wireless network security. On the other hand, the average grade value of each parameter which contributes to wireless network security larger than 2, represents a satisfactory security level, but offers a different level of security. Security levels based on each parameter grade and the percentage of parameter grade value in the overall security are given in Table 3.

Table 3: The range of security grade values

Parameter	Grade (INPUT)	Result	Grade (INPUT)	Result	Grade (INPUT)
1.	1	3.00%	2	6.00%	3
2.	1	1.60%	2	3.20%	3
3.	1	3.60%	2	7.20%	3
4.	1	4.20%	2	8.40%	3
5.	1	1.80%	2	3.60%	3
6.	1	2.60%	2	5.20%	3
7.	1	2.00%	2	4.00%	3
8.	1	0.40%	2	0.80%	3
9.	1	0.80%	2	1.60%	3
Parameter No.	Overall percentage	20.00%		40.00%	
See table 1!	Security level	**1**		**2**	

Parameter	Result	Grade (INPUT)	Result	Grade (INPUT)	Result
1.	9.00%	4	12.00%	5	15.00%
2.	4.80%	4	6.40%	5	8.00%
3.	10.80%	4	14.40%	5	18.00%
4.	12.60%	4	16.80%	5	21.00%
5.	5.40%	4	7.20%	5	9.00%
6.	7.80%	4	10.40%	5	13.00%
7.	6.00%	4	8.00%	5	10.00%
8.	1.20%	4	1.60%	5	2.00%
9.	2.40%	4	3.20%	5	4.00%
Parameter No.	60.00%		80.00%		100.00%
See Table 1!	**3**		**4**		**5**

Source: Authors

Network administrators have given the highest grades to parameters relating to encryption and firewall usage (see Table 1). The lowest grade was given to VPN, which can be explained by the fact that VPN is not often found in educational institutions, or is not employed by the current sample of examinees. Authors recommend further research.

Figure 4 illustrates the security level given the value 2 (32% overall), based on the parameter value input. After making a choice, recommendations are given at two levels: level 1 recommendation (named "Preporuka") compares evaluated parameters in case of two choices being made, while level 2 recommendation (named "Preporuka 2") compares evaluated parameters in case of three or more choices being made (Figure 3).

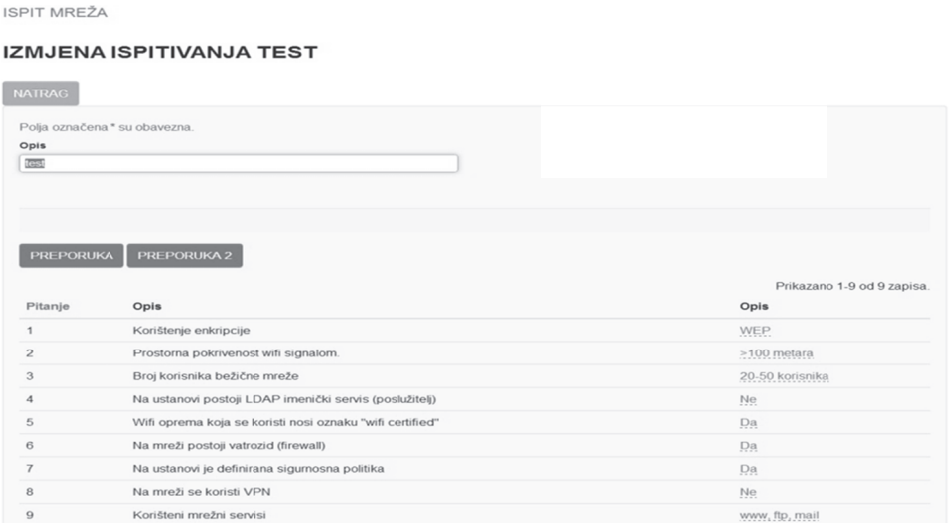


Figure 3: Data input within the *Ispitivanje* module for wireless network security check

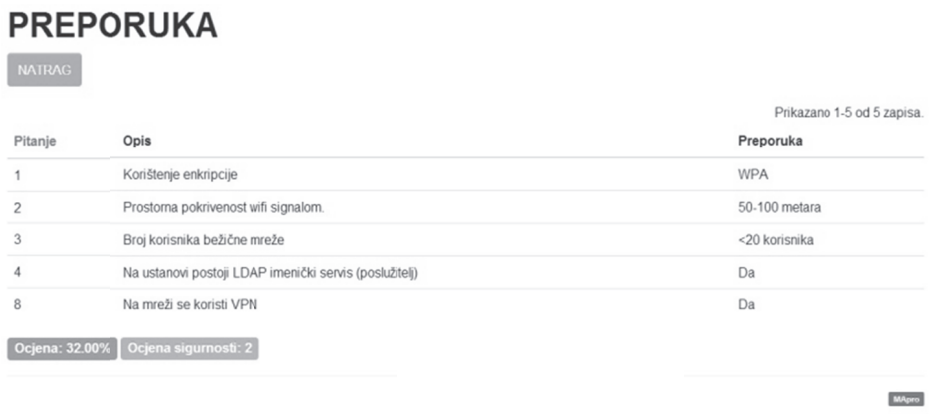
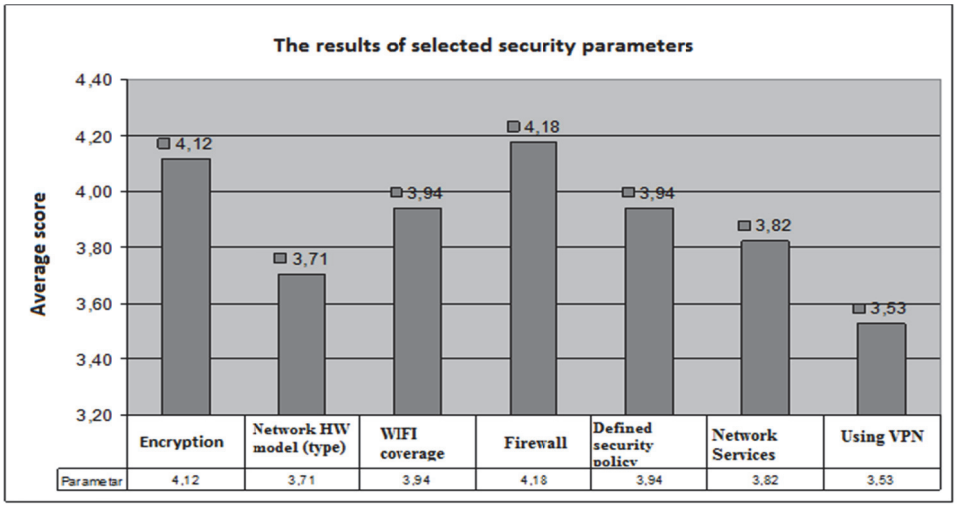


Figure 4: An example of wireless network security recommendation ("Preporuka" - Level 1) with the overall security grade ("Ocjena") (32%) and the corresponding security level (2) ("Ocjena sigurnosti"). At the end of the research and parameter evaluation, the examinees (network administrators) have evaluated the selected security parameters.

Graph 1: Network administrator satisfaction with the selected parameters included in the system model (average values)



Source: Authors

As shown in Graph 1, network administrators have given the highest grades to encryption algorithms (average value 4.12) and firewall usage (average value 4.18) if compared to other selected parameters which influence wireless network security. The lowest grade (3.53) was given to the use of VPN, which can be explained by the fact that VPN is not often employed in educational institutions or is not used by the sample of examinees included in this research [15]. Overall score of network administrator satisfaction with selected parameters was 3.59. Authors recommend further research.

Conclusion

In order to define the specifications for the security model, an analysis of all the key aspects of security settings of active wireless network equipment was undertaken. Using a questionnaire, data was collected from administrators (system engineers) at educational institutions regarding security of the wireless network they administrate, security criteria and the structure and evaluation of these criteria. The collected data was analysed statistically. The collected data was used to set the criteria for assessing threats, and weight factors for each criteria. Data analysis has yielded properties important for the set research goals. Statistical analysis was used as the basis for developing application for security evaluation. In addition, the evaluation of the model was conducted. It was concluded that the proposed application enables recommendations of security measures which enhance the level of security. In the process of security planning, it is very important to choose network equipment by an established manufacturer.

On today's market, different equipment is available, but when making a choice, certain qualities should be taken into consideration, namely, the possibility of maximum adaptability of network equipment. If case a network device got lost, there should be a procedure for reporting it. Moreover, it is important to define a procedure in case of intrusion, i.e., in case of a security breach. In addition to developing the system for evaluating the security of wireless local area networks , this work is also sample research of wireless networks security in educational institutions in Croatia.

This research represents a contribution to the theoretical and practical considering the areas of security of wireless local area networks and provides exceptional importance on recruiting value of each parameter active safety wireless network equipment for the purpose of determining the level of required security protection. Security tests regarding wireless network vulnerability should be conducted periodically, and it is necessary to evaluate security risks [16]. Everything mentioned above should be incorporated into the security policy. In order to enhance the model of evaluation, the authors recommend further research.

References

[1] A. S. Tanenbaum, D. J. Wetherall, Computer networks, 5th ed. SAD: Prentice Hall, 2011.
[2] H. Hamidović, WLAN - Bežične lokalne računalne mreže. Zagreb: Info press, 2009.
[3] C. J. Date, An introduction to database systems, 6th edition. Reading MA: Addison Wesley, 1995.
[4] M. Pavlić, Razvoj informacijskih sustava, Znak, Zagreb, 1996.
[5] M. Piškor, V. Kondić, Đ. Mađerić, "Proces implementacije lean-a u malim organizacijama". Tehnički glasnik, Vol. 5 No. 1, 2011.
[6] G. Gledec, M. Mikuc, M. Kos, Sigurnost u privatnim komunikacijskim mrežama. Zagreb: FER, 2008.
[7] http://www.cert.hr/sites/default/files/CCERT-PUBDOC-2009-06-267.pdf (12.2.2013.)
[8] http://www.cert.hr/sites/default/files/NCERT-PUBDOC-2010-12-001_0.pdf (12.2.2013.)
[9] http://www.yiiframework.com/about/ (06.02.2015)
[10] Radojević, B. Problematika provođenja sigurnosne politike u visokoškolskim ustanovama u RH. Opatija: MIPRO, 2011.
[11] Skendžić, A. Sigurnost infrastrukturnog načina rada bežične mreže standarda IEEE 802.11. Zbornik Veleučilišta u Rijeci, Vol. 2 (2014), No. 1.
[12] Prodanović, R., Simić, D. A Survey of Wireless Security. Journal of Computing and Information Technology - CIT 15, 2007, 3, 237–255.
[13] http://blogs.aerohive.com/blog/the-wireless-lan-training-blog/wifi-back-to-basics-24-ghz-channel-planning. (08.02.2015)
[14] http://wireless-spot.blogspot.com/2009/11/ad-hoc-and-infrastructure-modes.html (02.10.2015)
[15] http://www.cis.hr/www.edicija/LinkedDocuments/CCERT-PUBDOC-2003-02-05.pdf (09.10.2015)
[16] http://www.cis.hr/sigurosni-alati/ispitivanje-sigurnosti-bezicnih-mreza.html 01.10.2015)
[17] https://docs.google.com/forms/d/1G4kpM52yhh5z2U3oon92iEyRy0dQeSi2jdFr6_mzWs/viewform (05.6.2014.)
[18] http://www.wi-fi.org/ (01.10.2015)

An Analysis of Twitter Usage Among Startups in Europe

Sergej Lugović
Polytechnic of Zagreb
Vrbik 8, Zagreb, Croatia
lugovicsergej@gmail.com

Wasim Ahmed
Health Informatics Research Group, Information School,
University of Sheffield
Sheffield, UK
Wahmed1@sheffield.ac.uk

Summary

Startups are becoming part of the development of policies by the European Commission and corporate strategies. The business press widely covers startups, as does the other media. Local governments in Europe are also developing financial and administrative programs to support startups. Most industries now use social media, and in particular Twitter, to reach their customers. Twitter provides startups the opportunity to reach a global audience at a relatively low cost and still support both promotion and branding.
However, there's a lack of evidence-based research, which examines the Twitter usage among startups within the European Union (EU). This paper outlines preliminary findings about Twitter usage among EU startup founders and contributes to an ongoing study that seeks to define the concept of a startup.
Exploring Twitter usage among European startup founders could provide insight into their behavior and expand overall knowledge of how to best define the concept of a startup. The authors retrieved data used in the analysis from one of the most extensive EU databases, F6S.com. The authors collected the Twitter data user handles of EU startups in June and August 2015 and manually verified each Twitter account, which allowed the retrieval of data. The authors then entered the verified data (i.e., verified Twitter accounts) into the Twitter Archiving Google Sheets (TAGS) system that collected data for the purpose this longitudinal study.
The results of this research provide insight into temporal dynamics of Twitter usage, including Twitter account creation, followers, and tweets. There are statistically significant correlations between the Twitter activity of EU startups and the invested amount in that country overall and per capita. This paper offers a methodological contribution in regard to the social media data retrieval

techniques, which can provide an insight into Twitter usage among EU startups founders.

Keywords: Twitter, Technology Entrepreneurship, Startup, Information Behavior, Big Data

Introduction

The Financial Times (McLannahan, 2015) recently published an article about Twitter usage as a source of valuable data for the financial industry. Investors using reports based on Twitter data could gain up to two-and-a-half percent when China Black Monday (Phillips & Farrell, 2015**)** occurred. The main difference between Twitter and other social media data is that user-generated content is publicly available and is more easily accessible (Ahmed, 2015) because of Twitter's free API ecosystem. Twitter usage for the purpose of entrepreneurial activities is not just limited to marketing activities or as a medium of communication but also to support the whole process from opportunity recognition to bringing those opportunities to life (Fischer & Reuber, 2011).

Twitter has the potential to open new opportunities because of its global reach and low cost, and it can support promotion and branding since it enables community interactions. Moreover, by using Twitter, it's possible to observe the competition and industry, more generally, by analyzing streams of tweets. Twitter is also useful for better understanding and engaging with customers (including customer support) (Curran at al. 2012).

Twitter has more than 316 million monthly active users, and there are more than 500 million tweets per day (Twitter.com, n.d.). Most companies acknowledge the importance of social media platforms: 93 percent declare that they employ social media in their marketing strategy, Facebook is top-rated (92%), followed by Twitter (84%) and LinkedIn (71%) (Stelzner, 2011). Twitter is also useful as a tool integrated into product development, marketing communication, and recruitment (Savulescu & Mihalcea, 2013). Small and medium-sized enterprises (SMEs) are, on one hand, drivers of the EU economy employing the majority of the people and creating the most value, and, on the other hand, have limited resources (ec.europa.eu, 2015). According to Wamba & Carter (2013), there's a statistically significant relationship between a firm's innovativeness and the adoption of Twitter.

Another important aspect of using Twitter is its usage in investor relations (IR). IR is a strategic management responsibility that's integrated into finance, communication, marketing, and law compliance to enable two-way communications between company, the financial community, and other stakeholders. It contributes to a company's security to achieve fair valuation (niri.org, 2015). According to Coe (2013), investors match the take-up on social media by companies, and social media has the potential to move financial markets. The same report noted that professional investor use of Twitter usage is increasing, 63 percent of

finance professionals (including brokers and heads of trading desks) believe that valuation of individual stocks can directly link to public sentiment contained in social media channels and 63 percent of 800 publicly traded companies use Twitter for investor-related activities.

Palanissamy (2014) suggested that three out of five financial bloggers use Twitter as their primary news source. Additionally, Twitter usage in investor relationships is also supported; Joyce (2013) noted that Twitter is the most widely used channel for IR, and that 72 percent out of 890 companies surveyed used Twitter to communicate investment materials in 2013. Another study (niri.org, 2013) found that investors viewed financial blogs, along with LinkedIn and Twitter, as the most valuable social media resources. The literature presented above indicates that companies for the purpose of IR use Twitter.

Venture capital funds in the EU are increasing activity (Grilli & Murtinu, 2014). At the same time, "startup" as a phenomenon is mentioned frequently within the media (ft.com, 2015), within political speeches, and European commission programs (ec.europa.eu, 2015). However, there's still no clear definition of what a startup *is*. On one side of the spectrum it's the definition used by the Ewing Marion Kauffman Foundation (Morelix et al., 2015) stating that "Startup businesses here are defined as employer firms less than one year old employing at least one person besides the owner."

On the other side of the spectrum, there's a definition by NESTA (Dee at al., 2015) that defines startups as "[A] young, innovative, growth–oriented business (employees/revenue/customers) in search of a sustainable and scalable business model." This definition expands on Steve Blank's (2013) definition of startups as organizations formed to search for repeatable and scalable business models.

According to NESTA, the emphasis is on defining startups as *young companies*. They differ from overall SMEs and need to develop different policies to support young companies and different ones to support small firms. Therefore, there's no consensus on what a startup is and what it does. Instead of seeking a definition, it's necessary to look at what those companies are doing. Such an inquiry aligns with the Low and MacMillan (1988) suggestion that research into entrepreneurial behavior should consider contextual issues and identify the processes that explain rather than merely describe the entrepreneurial phenomenon.

Cybernetics does not ask, "What is this?" but instead it asks, "What does it do?" It doesn't consider objects; rather, it considers the ways of behaving (Ashby, 1957). So, what companies actually do is the result of their founders' action. This paper examines the correlation of the Twitter usage of startup founders across the EU with the total country investment in startups in order to ascertain the extent to which we can predict how the startups will behave in the future.

Methodology

The authors retrieved the data for the purpose of the research in the following way. First, they retrieved data related to EU startup founders from one of the most widely used databases, F6S.com, which has more than 110,000 companies in the database and 2 million unique visitors a month (f6s.com, 2015). In total, the authors retrieved 50,433 founders from 29 EU countries (including Norway), with the assistance of four freelance workers hired through the Upwork platform (Upwork.com, n.d.). There were more records in total; however, some founders related to more than one company, so the data set had to be de-duped (i.e., the removal of duplicates). Registering founders had the option to submit a company Twitter account and personal Twitter account. For the purpose of this research, the authors focused on the personal Twitter accounts the founders submitted.

After the identifying the Twitter accounts, the authors retrieved data from the accounts, in particular, the date the founders joined Twitter, how many tweets they posted, how many accounts they were following, and how many followers they had, in total 15,192 accounts. We also classified whether an account was private for the purpose of this study, and this was inferred from whether the founder submitted it as a private Twitter account. The authors retrieved data from twenty-nine countries from the EU, including Norway.

In parallel, the analytical company Funderbeam (2015) provided the financial data addressing EU-wide startup investments. This financial data broke down into investment per capital per country and total amount of investment in the startup ecosystem per particular country. Further, we selected Twitter accounts that had more than one hundred tweets and then more than one hundred followers. Those accounts, when filtered, were entered into the retrieval system Twitter Arching Google Sheets (TAGS) (Hawksey, 2013), which the researchers set to retrieve data on daily basis from those accounts. In total, there are 9,696 startup founders that have more than one hundred tweets and more than one hundred followers, and the accounts went into the TAGS system. The data retrieval period was from July 15, 2015 to September 9, 2015.

For the purpose of this preliminary analysis, the following hypotheses were formulated.

H1 Median time spent on Twitter of total country startup founders will have a statistically significant positive correlation with a) total investment in that country and b) investment per capita.

H2 Median number of tweets of total country startup founders will have a statistically significant positive correlation with a) total investment in that country and b) investment per capita.

H3 Median number of followers of total country startup founders will be a statistically significant positive correlation with a) total investment in that country and b) investment per capita.

H4 Number of startup founders with more than one hundred tweets and followers per country will be a statistically significant positive correlation with a) total investment in that country and b) investment per capita.
H5 Percentage of total number of startup founders in the F6S.com database will be a statistically significant positive correlation with a) total investment in that country and b) investment per capita.
The results are presented in following section, including breakdown per country and correlations related to hypothesis.

Results

Results of our analysis are provided below. Table 1 presents overall results, including a breakdown to a country level, including total number of companies in the F6S.com database, the number of founders with Twitter accounts, the percentage of those who submitted their personal Twitter account to F6S.com, the number of startup founders with more than one hundred tweets and followers, the median number of days spent on Twitter, the number of tweets posted, the number of followers, and the number the founders followed. Table 2 presents correlations that align with the hypothesis mentioned earlier in the text.

Table 1. Overview of collected data

	Total no. in F6S.com	No. of Twitter users	%	No. more than 100 tweets & more than 100 followers	Median no. of days on Twitter
Austria	505	130	25.74	70.00	2,193.0
Belgium	1,031	336	32.59	245.00	2,049.0
Bulgaria	2,970	466	15.69	225.00	1,973.0
Croatia	452	135	29.87	74.00	1,737.5
Cyprus	184	40	21.74	20.00	1,864.0
Czech Republic	611	144	23.57	66.00	1,925.0
Denmark	694	183	26.37	115.00	2,004.0
Estonia	1,303	152	11.67	77.00	2,164.0
Finland	1,135	302	26.61	159.00	2,022.0
France	5,000	1,788	35.76	1,136.00	1,922.0
Germany	3,416	872	25.53	518.00	2,107.0
Greece	957	270	28.21	154.00	1,897.0
Hungary	756	127	16.80	64.00	2,020.5
Ireland	1,944	738	37.96	518.00	1,882.0
Italy	4,851	1,347	27.77	781.00	1,790.0
Latvia	321	78	24.30	54.00	2,162.0
Lithuania	468	76	16.24	41.00	1,977.0

Luxembourg	108	30	27.78	22.00	1,769.0
Malta	53	15	28.30	12.00	1,848.5
Netherlands	2,366	737	31.15	493.00	1,852.0
Poland	913	183	20.04	82.00	1,530.5
Portugal	1,582	364	23.01	113.00	2,204.0
Romania	1,192	296	24.83	138.00	2,039.0
Slovakia	291	87	29.90	44.00	1,927.0
Slovenia	533	128	24.02	64.00	1,939.5
Spain	4,917	1,832	37.26	1,318.00	N/A
Sweden	701	190	27.10	114.00	2,043.0
UK	11,033	4,102	37.18	2,954.00	2,042.0

Table 1 continued

	Median no. of tweets	**Median no. following**	**Median no. followers**	**Amount of money per capita 07/2014-07/2015**	**Investment per country**
Austria	453.5	238.0	134.0	2.64	255,566,415
Belgium	482.0	314.5	282.5	4.35	298,527,566
Bulgaria	267.0	150.0	109.0	0.45	31,029,533
Croatia	346.0	229.0	160.0	0.16	3,733,794
Cyprus	210.0	164.0	97.0	46.18	93,175,435
Czech Republic	283.0	128.0	96.0	0.62	18,244,152
Denmark	469.0	260.0	195.0	26.97	494,836,809
Estonia	284.0	188.5	126.0	41.79	131,018,985
Finland	239.0	231.0	124.0	30.43	470,570,350
France	369.0	246.0	180.0	8.26	2,617,622,183
Germany	363.5	229.0	155.0	24.42	5,274,327,168
Greece	366.5	264.5	186.5	1.16	27,057,929
Hungary	321.5	151.0	94.5	0.24	23,577,952
Ireland	534.5	392.0	802.6	37.21	801,443,259
Italy	377.0	228.5	144.0	1.17	492,868,478
Latvia	745.5	241.0	206.0	8.41	34,484,729
Lithuania	270.5	169.5	139.0	2.35	98,813,906
Luxembourg	414.0	303.0	191.0	38.77	65,648,950
Malta	233.5	237.0	109.0	N/A	936,279
Netherlands	417.5	255.0	214.5	19.47	2,254,461,161
Poland	199.0	146.0	109.0	0.82	208,804,181
Portugal	217.0	184.0	113.0	1.96	64,947,960

Romania	178.0	138.0	97.0	0.06	14,857,154
Slovakia	256.0	169.0	123.0	0.62	5,754,056
Slovenia	316.0	164.5	131.0	0.20	14,769,623
Spain	654.5	293.5	237.0	5.00	798,464,239
Sweden	479.5	239.5	197.0	32.86	1,210,577,278
UK	604.0	353.0	289.0	49.49	12,168,393,685

Table 2. Correlations

H1a & H1b		**Days on Twitter (M)**	**H1b – Invest per capita**	**H1a –Invest per country**
Days on Twitter (M)	Pearson Correlation	1	-.152	.098
	Sig. (2-tailed)		.499	.664
	N	22	22	22
H2a & H2b		**H2b** – Invest per capita	**H2a** – Invest per country	No. of tweets (M)
No. of tweets (M)	Pearson Correlation	.057	.356	1
	Sig. (2-tailed)	.776	.063	
	N	27	28	28
H3a & H3b		**H3b** – Invest per capita	**H3a** – Invest per country	Followers (M)
Followers (M)	Pearson Correlation	.328	.200	1
	Sig. (2-tailed)	.095	.308	
	N	27	28	28
H4a & H4b		**H4b** – Invest per capita	**H4a** – Invest per country	No. more than 100 tweets & more than 100 followers
No. more than 100 tweets & more than 100 followers	Pearson Correlation	.162	.872	1
	Sig. (2-tailed)	.419	.000	
	N	27	28	28
H5a & H5b		**H5b** – Invest per capita	**H5a** – Invest per country	% having Twitter acc in F6S
% having Twitter acc in F6S	Pearson Correlation	.036	.410	1
	Sig. (2-tailed)	.860	.030	
	N	27	28	28

Discussion

As the results in Table 2 demonstrate, only two hypotheses are acceptable, **H4a** showing statistically significant correlations between total number of the startup founders who have more than one hundred tweets and more than one hundred followers with the total investment per country and **H5a** showing statistically significant correlations between percentage of the startup founders with personal Twitter profiles presented in the F6S database and total country investment. Rejection of other hypotheses could be interpreted that the average quantity of the tweets, followers, those the founders followed, and period spent on

Twitter per country do not have an impact on the size of the investment startups receive and distribution of investment per capita.

Therefore, it isn't about how much is tweeted, however it could relate to the country level of Twitteracy (Greenhow & Gleason, 2012). Within the startup ecosystem, much more importance was given to the founders' characteristics than to company traction (sales volume and user base) and other investors already invested.

A recent field experiment by Bernstein, Korteweg, & Laws (2015) showed that average investors when looking for information about startups strongly respond to the founding team but not to firm traction or lead investors. Recently published research (Tata et al., 2015) used data from Twitter accounts of startup founders and analyzed the content of their tweets in relation to funding raised (in particular, they analyzed the temporal focus of tweets and how they related to startup performance). They found relations between tweets content and startup performance, including funding raised. The results demonstrate that high past focus, low future focus, high collective self-categorization, and lower level construals can associate with better performance.

Findings from this research, in particular, collective self-categorization, align with our findings, showing that size of the country ecosystem of startup founders is in positive correlation with country investment. The stronger the country's Twitter community is the more the country receives in funding. Lack of significant correlation between number of tweets, followers/following, and days spent on Twitter lead the authors to question relevance of the content created and distributed over Twitter to the investors. This is something that needs further exploration and in more depth.

The results presented in this paper support the direction of this research project, since there's evidence that data about Twitter usage could provide insight into the performance of a startup in terms of financial results. The hypotheses presented in this paper are put forward on a macro level (correlating within EU countries), and a micro level (individual founders' accounts) analysis has to be explored and analyzed in the next phase of the research.

An additional limitation is that the authors did not analyze content of the tweets and did not look into the reasons why some countries perform better than others in term of Twitter. The research had limits because only one source retrieved data for analysis (F6S.com). Also, the authors analyzed only the founders' private Twitter accounts and used only one source of financial data.

A more comprehensive study may use a variety of databases. However, this may not be feasible since the numbers gathered could become difficult to manage. In this paper, we presented preliminary results from the initial data gathered, but the data were only the starting point for a longitudinal study that is currently underway.

The authors believe that the dynamics of the founders' Twitter activities could correlate with startup performance (i.e. growth of the number of followers when

compared with performance). Still, in order be in a position to infer this, data gathered over a longer period of time will have to be used and analyzed. Moreover, future research aims to conduct a more extensive statistical analysis of the data set that's collected. Additional methods of analysis may include sentiment analysis, network analysis, and the correlation of Twitter data with the companies' income. Further research may also explore even further the correlation with Twitter data with another social media channel, for example, Facebook or LinkedIn, or data such as the companies' Wikipedia pages and the number of visitors to the websites of the startups.

Conclusion

The findings demonstrate that it's possible to achieve great insight into the behavior of an object of interest by using indirect observation, in this instance observing Twitter data indirectly. However, to do so, the analysis has to dig deeper into the large amount of unstructured data processed by using advanced techniques. The authors plan to employ more advanced data mining and computational methods of analyzing the data gathered. Moreover, the research continues to collect data on a continuing basis from the 9,696 startup founders (Twitter accounts), which could provide insight into different aspects of startup as a phenomenon.

These findings and contemporary research methods could apply to different disciplines, such as information science, social media, entrepreneurship, and investors' relationship management. In addition, there's an outline of a robust methodology of capturing and analyzing Twitter data related to specific Twitter accounts.

Acknowledgments

The authors would like to thank Kaidi, Villu, and Nicholas from Funderbeam.com for providing the financial data used in this research paper.

References

About Startup Europe (n.d.). [Online] Retrieved from http://ec.europa.eu/digital-agenda/en/about-startup-europe [Last Accessed 19/09/15]

Ahmed, W., (2015). Using Twitter as a data source: An overview of current social media research tools. *LSE Impact of Social Sciences* blog Retrieved from http://blogs.lse.ac.uk/impactof socialsciences/2012/04/19/blog-tweeting-papers-worth-it/ 10 July 2015 (Last Accessed 11/07/2015).

Ashby, W. Ross, *An introduction to cybernetics*. Chapman & Hall Ltd, London, 1957.

Bernstein, S., Korteweg, A. G., & Laws, K. (2014). Attracting early stage investors: Evidence from a randomized field experiment. *Rock Center for Corporate Governance.* Stanford University working paper, (185), 14-17.

Blank, S., (2013). Why the lean start-up changes everything. *Harvard Business Review*, *91*(5), 63-72.

Coe, D. Companies and shareholder engagement—making the most of social media. *Keeping Good Companies*, (2013), *65*(5), 307.

Curran, K., O'Hara, K., & O'Brien, S. (2012). The role of Twitter in the world of business. *Web-Based Multimedia Advancements in Data Communications and Networking Technologies*, p. 169. http://scisweb.ulster.ac.uk/~kevin/twitter.pdf.

Dee, N., Gill, D., Weinberg, C., and McTavish, S. (2015). Startup support programmes: What's the difference? *NESTA*.

Fischer, E., & Reuber, A. R. (2011). Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behavior? *Journal of Business Venturing*, (2011) 26(1), 1-18. http://ntucriticalinquiry.wikispaces.com/file/view/Social+interaction+via+new+social+media.pdf.

F6S.com (n.d.). [Online] Retrieved from: https://www.f6s.com/f6s [Last Accessed 19/09/15].

FT.com (n.d.). [Online] Retrieved from http://www.ft.com/intl/topics/themes/Tech_start-ups [Last Accessed 19/09/15].

Funderbeam.com. (n.d.). [Online] Retrieved from https://www.funderbeam.com/ [Last Accessed 19/09/15].

Greenhow, C., & Gleason, B. (2012). Twitteracy: Tweeting as a new literacy practice. In *The Educational Forum 76*(4) 464-478. Taylor & Francis Group.

Grilli, L., & Murtinu, S. (2014). Government, venture capital and the growth of European high-tech entrepreneurial firms. *Research Policy*, *43*(9), 1523-1543.

Hawksey, M. (2013). Twitter Archiving Google Spreadsheet (TAGS) v5. *JISC CETIS MASHe: The Musing of Martin Hawksey (EdTech Explorer)* https://tags.hawksey.info/ [Last Accessed 19/09/15].

Joyce, S., (2013). Public company use of social media for investor relationship http://q4websystems.com/files/doc_downloads/whitepapers/Q4-Web_2013-Blog-Whitepaper_FINAL.pdf Retrieved 14.09.2014.[Last Accessed 19/09/15].

Low, M. B., & MacMillan, I. C. (1988). Entrepreneurship: Past research and future challenges. *Journal of Management*, *14*(2): 139-161.

Morelix, A., Reedy, E. J., & Russell, J. (2015). The startup activity national trends. Ewing Marion Kauffman Foundation.

McLannahan, B. (2015). In search of market signals amid Twitter noise, *Financial Times 9* http://www.ft.com/intl/cms/s/2/99a7e7ec-4fea-11e5-8642-53585f2cfcd.html#axzz3ly52ANif. Retrieved 15.09.2015.

Niri.org (2013). (http://www.niri.org/media/News-Releases/NIRI-and-Corbin-Perception-Release-Studies-of-Social-Media-Use-in-Corporate-Investor-Communications-.aspx Retrieved 15.09.2015

Palanissamy, A. (2014). Investor-relations-integrating social media into board – Oman perspective. *European Journal of Business and Management 6*(24), 108-113.

Phillips, T, Farrell, S., (2015). FTSE 100 loses more than £60bn after China's 'Black Monday', *The Guardian, 8* http://www.theguardian.com/business/2015/aug/24/black-monday-chinese-stock-market-loses-all-gains-since-start-of-year. Retrieved 14.09.2015.

Savulescu, R. M., & Mihalcea, A. D. (2013). Social networking sites: Guidelines for creating new business opportunities through Facebook, Twitter and LinkedIn. *Management Dynamics in the Knowledge Economy 1*(1), 39-53.

SMES EC.Europe.eu http://ec.europa.eu/growth/smes/ Retrieved 15.09.2015.

Stelzner, M. A. (2011). Social media marketing industry report. How marketers are using social media to grow their businesses. Retrieved from http://www.socialmediaexaminer.com/SocialMediaMarketingIndustryRe- port2012.pdf.

Tata, A., Martinez, D. L., & Brusoni, S. (2015). Looking backward and forward? Understanding the role of temporal focus on startup performance Paper to be presented at DRUID15, Rome, 15-17, http://druid8.sit.aau.dk/druid/acc_papers/xvmf70lrrol2amebcmt1b8pemo94.pdf.

Twitter.com About, (n.d.). *About Twitter, Inc*. [Online] Retrieved from: https://about.twitter.com/company [Last Accessed 29 Mar. 2015].

US National Investor Relationship Institute (2015). [Online] Retrieved from http://niri.org/FunctionalMenu/About.aspx [Last Accessed 19/09/15].

Upwork https://www.upwork.com/ (n.d.). [Online] [Last Accessed 19/09/15]

Wamba, S. F., & Carter, L. (2013). Twitter adoption and use by SMEs: An empirical study. In *System Sciences (HICSS),* (January) *46th Hawaii International Conference,* 2042-2049). IEEE. http://www.computer.org/csdl/proceedings/hicss/2013/4892/00/4892c042.pdf

HEALTH INFORMATICS

Original scientific paper

Development of Health Care e-Services in the European Union

Ana Stanković, Hrvoje Stančić
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences
Ivana Lučića 3, Zagreb
ana.stankov3@gmail.com, hstancic@ffzg.hr

Summary

Development and implementation of e-services has become a common practice of state and public administrations in the developed countries around the world since the usage of information and communication technologies has enabled a faster processing of user applications and more efficient functioning of the service. This paper analyses the level of development of e-services in health care systems of eight European countries, both in terms of safety and reliability from the perspective of end users, i.e. citizens. Furthermore, the concept of trust in e-services based on the availability of information on the Internet on applied technological solutions, storage and data protection and their availability is considered. Implementation of services like e-health records and e-prescriptions requires a well-developed strategy and a very transparent way of their functioning because of the sensitive nature of health data. For this study official websites of the services were searched through, as were other sources on the Internet. The research results indicate a lack of certain key information, especially those related to data storage, their protection and long-term preservation.

Keywords: e-services, e-health record, e-prescription, health, trust

Introduction

Nowadays ICT has entered all aspects of human activity and even more people prefer using the Internet for dealing with their everyday transactions. Therefore, the state and public administrations of countries around the world have recognised the need to make their services available through the Internet. As one of the most commonly used e-services have proven to be those related to health care. These e-services have positively affected the business optimization in health care sector (Menachemi and Collum, 2011) although the benefits are to be visible several years after the implementation due to the high implementation costs and the time needed for the users to start using the e-services (Saluse et al., 2010). The most commonly implemented health care e-services are e-health records and e-prescriptions. Considering the fact that health-related data are very sensitive, many questions about the security and confidentiality of these e-

services need to be answered (Smith and Eloff, 1999). For example, who has access to such data? How are users' data protected from unwanted access? How are users guaranteed that their data are protected from unauthorised access? Answers to these questions are relevant particularly nowadays when medical data are at risk of misuse by outsiders as well as insiders (Anderson, 1996). The aim of this research was to get answers to these and other similar questions and to gain insight to the way in which different state and public administrations solved some of the problems related to the security and confidentiality of health care e-services.

E-services

One can distinguish two main different categories of e-services developed by the government: G2C (Government-to-Citizen) and G2B (Government-to-Business) services. When searching for the definition of e-services one can come across many formulations from different sources, but regardless of the scope and depth of different definitions of e-services, they all point out that e-services:

a) are available through the Internet,
b) provide a particular service (partially or completely), and
c) optimise operations (Piccinelli and Stammers, 2001).

Table 1. Stages of development of an e-service

Stage	Action	Description
0	No information	Information about the service is not available on-line or the service provider does not have a website.
1	Information	Only information about the service is available on-line.
2	One-way interaction	There are forms in electronic format that are available for downloading. Blank forms can be printed out.
3	Two-way interaction	Forms can be filled out and signed after authentication. By completing forms one runs a specific service.
4	Transaction	The complete service is available on-line: completing forms, authentication, payment and delivery of receipts or other forms of complete service.
5	Iteration	Services are repetitive and automatically executed. Users are notified automatically about the execution of the service.

Since not all of the e-services are developed and implemented at the same level, different stage models are used in order to classify them according to their technical specifications and degree of service they provide to the citizens. One of them was presented at the Lisbon Summit in the year 2000. In this model (Table 1), which was used in this research as well, there are 5 (or 6) stages of development (Digitizing Public Services in Europe, 2010).

Health care e-services

Health care e-services that are the most commonly implemented at the moment are *e-health records* and *e-prescriptions*. E-health record is a set of medical information about an individual in the electronic form, usually containing personal and demographic data, patient medical history, laboratory results and other information related to the health of the individual. Although e-health record is considered to be one of the key decision making tools regarding the patient's health, it is necessary to enable the patient to preserve the privacy of certain data sets (Barrows and Clayton, 1996). It is necessary to provide certain restriction measures to a patient which he/she will be able to use in order to prevent doctors from having access to some data sets of their e-health record.

E-prescription refers to prescribing medications using ICT by the authorized health professionals. The implementation of e-prescription service optimises the entire process of purchasing drugs (Cornford et al., 2014), it prevents the inability to read the recipe due to sloppy handwriting and it makes it easier to track which medications are used most often, speeds up the process, and has ecological side because it saves paper.

The study presented in this paper is part of a larger research conducted within the InterPARES Trust project. InterPARES (International Research into the Preservation of Authentic Records in Electronic Systems) Trust "is a multi-national, interdisciplinary research project exploring issues concerning digital records and data entrusted to the Internet. Its goal is to generate theoretical and methodological frameworks to develop local, national and international policies, procedures, regulations, standards and legislation, in order to ensure public trust grounded on evidence of good governance, a strong digital economy, and a persistent digital memory."[1]

An overview of the implemented health care e-services in the chosen EU countries

Research methodology and limitations

The research was carried out between 15 January and 15 July 2014. The sample covered eight EU countries: Belgium, Croatia, Denmark, Estonia, Germany, Lithuania, Sweden, and the United Kingdom. Prior to the selection of these 8 countries an environmental scan was carried out. Then, the countries that have implemented their e-services on a higher level or have shown the intention to do so have been selected for the in-depth research. A questionnaire consisting of 52 questions divided into six categories was conducted. The six categories are:

A. Basic service information,
B. Users,
C. Business optimisation,

[1] InterPARES Trust project, http://interparestrust.org/ (07.01.2015.)

D. Technological solutions,
E. Storage and long-term content availability,
F. System operation transparency.

The data collection was done between 15 February and 15 May 2014. The data were collected by the researchers. They searched for information that was openly available on the Internet. Sources of information were primarily the official websites of e-services or other sites that offered information about their implementation. After the collection of information was finished, the data was analysed quantitatively and qualitatively. Limitations of the study are reflected in the fact that in some cases the information was not available because of the language barrier, because of technical maintenance of the website, or other technical issues. There is a possibility that certain information was available on the Internet but could be reached only after logging in to the system, and therefore was out of reach for this study.

Research results

The data were collected for each of the 52 questions divided into six categories (A-F) for each of the eight surveyed EU countries. Then, the collected data were analysed at the level of each country and each e-service, and finally for all of the researched countries and for both e-services together. Next, the questions within each of the six main categories will be mentioned first, and then the aggregated results of the research will be given and commented.

A. Basic service information

The questions in this category were aimed to gather information on the service URL, service category, category/type of institution authorized for the e-service, start of service development/implementation, level of informatization, connection with any other governmental services, comparison of official and actual development of the e-service, limitations to the service's work schedule, short description of the service etc.

Developing and implementing e-services is a complicated and lengthy process so even though the interest for implementing health care e-services has occurred back in the 1990s most of the researched services were implemented by governments in the period from 2003 to 2009. In most countries e-health records are at the 3rd level of informatization and e-prescriptions are usually between level 3 and 4 since they meet all of the criteria related to the 3rd level and most of those related to the 4th level. E-health records and e-prescriptions are usually interconnected and offer round-the-clock service.

B. Users

This category consisted of the following questions: Is using the service mandatory for a certain category of users? If yes, which category of users is it mandatory for? Are there different groups of users? How many users are there per user

group? What is the percentage of users who use the service electronically? Which age groups are prevalent in using the service? Is the service adapted for users with disabilities? Are there users' satisfaction information?
Answers from this set of questions were found for Estonia and Denmark, both of which point out high percentage of users of e-health record (in Estonia 47% of citizens and 95% of doctors use e-health record and in Denmark almost 100% of doctors use it). When it comes to e-prescriptions, the percentage of users who use the service electronically in all of the researched countries is quite high, 75% and higher. The percentage of satisfied users in all of the researched countries for both services is quite high and it ranges from 80% to 95%. On the other hand, web sites of services in most of the researched countries have not shown the clear intention to adopt the service to the people with special needs. The only exception is Sweden where they have adapted the service for users with disabilities.

C. Business optimisation

The following questions on business optimisation were considered: Are there positive financial indicators for e-service (for the institution responsible for service and for users)? Has there been a decrease in time required to process user applications? How did the service affect the organisation of work processes in the responsible organization in terms of the required number of workers? What are the plans for upgrading and expanding the service in the future?
Business optimisation is visible on several levels. Estonia had the first financial gains from implementing e-health record in 2010 (2 years after implementing the service). In Croatia it is estimated that the implementation of e-prescription service should result with savings of 15 million HRK per year. In the United Kingdom it is assumed that implementation of e-prescriptions will result with total savings of 179 million pounds a year at the level of the entire health care system[2]. In Denmark implementation of e-health record greatly enhanced the efficiency of the health care system with average savings of 50 minutes per day through reduced paperwork and faster communication. In Estonia the use of e-prescriptions contributed to savings of approximately 30 minutes per day. In conclusion, the use of e-prescription is helping to save time on several levels:

a) it ensures doctors the possibility to repeat the prescribed medicine thus reducing the number of patient visits or calls,
b) pharmacists can prepare the medicine in advance thus reducing waiting time,
c) there is less possibility of errors due to illegible handwriting, and
d) there is no risk of losing the prescription since the data is stored in the central database.

[2] Electronic prescription service, NHS England, http://systems.hscic.gov.uk/eps/library/nhsecomms.pdf (20.6.2014.)

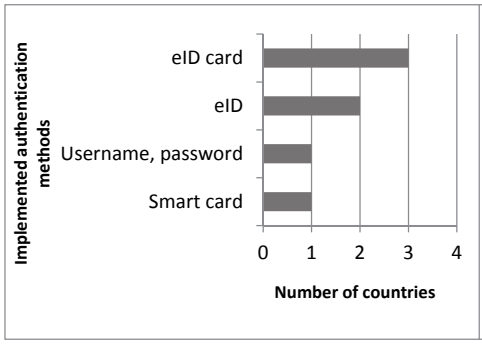
Also, plans for upgrading and expanding the system can be categorized into several groups:

1. technical upgrades (e.g. plans for developing a national eID system, the introduction of smart cards etc.)
2. improving data management, and
3. educating health care professionals.

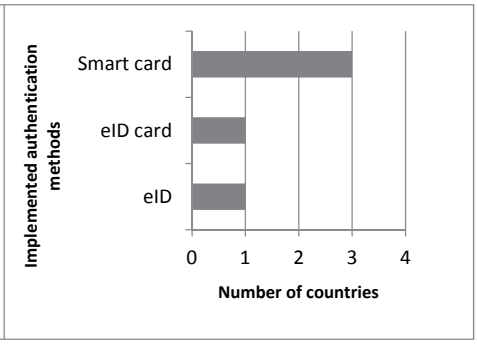
D. Technological solutions

The following questions regarded technological solutions were included in the survey: What type of authentication is used? Is the communication between the server and client station encrypted (SSL, some other protocol)? Does the service use eID? If yes, which one (list all if more than one eID exists)? Does the electronic signature use digital certificates? If yes, which format of electronic signatures is used? In what way does the user fill in and send data? Do the users send attachments with the filled in data? If yes, in what way? Do the users have predetermined types of document formats while sending out data? If yes, which ones? Is the service implemented through open-source or commercial technologies? Which technologies are being used? What type of application is used on the client side? Through which channel(s) is the service available? Is the service hosted within the responsible institution? If the responsible institution is hosting the service, does it have the required certificates? If the service or any of its parts is hosted outside the responsible institution, does it use the Cloud? Is the Cloud/Data centre located within the same country?

All countries that have implemented these e-services offer their users the possibility to access them by using any device with Internet connection and a browser. Most of them use eID cards for authentication when using e-health records (Graph 1) and smart cards when using e-prescriptions (Graph 2). We assumed that doctors usually fill in a web form in order to send data and that they use some of the standardized formats (.doc, .tiff, .pdf, etc.) when sending certain attachments (x-rays, results of different tests, etc.). Information on using digital certificates is available for Denmark and Estonia where they use the XAdES (XML Advanced Electronic Signature) format of electronic signatures. Belgium, Denmark and the United Kingdom provided information on conforming to ISO 27001 and ISO 9001 standards. However, there is no information available weather any of the services use cloud solutions which could be either due to the fact that the cloud solutions still raise a lot of security issues so they are still not implemented within this kind of e-services or the information on the usage is not available.



Graph 1. Type of authentication (e-health records)



Graph 2. Type of authentication (e-prescriptions)

E. Storage and long-term content availability

The questions investigating storage and long-term content availability were: What is the retention period for the data in the system? Is the retention period defined by a law/regulation or some other act? If yes, which one? Are the data deleted after the retention period expires? What is the preferred long-term preservation format(s)? Does the service use a method of materialization of data (conversion from digital to analogue format, e.g. printing, microfilming etc.)? Does the service comply with any of the long-term preservation standards? If yes, which one? Does the service offer use of an electronic archive as an additional service? Are there electronic document safe services? Are the data received through the service stored within the responsible institution's information system? Does the responsible institution possess the required certificates that guarantee security of the stored data? If the data are at least partially stored outside the responsible institution, does it use Data Cloud? Is the Cloud/Data center located within the same country as the responsible institution?

Information on storage and long-term content availability was found for Sweden, the United Kingdom and Croatia. In Sweden data from the patients' health records are kept permanently, while in the United Kingdom the retention period is 30 years. In Croatia the data should be kept for 70 years from their creation or 100 years from a person's birth. In the United Kingdom the retention period for e-prescriptions is 2 years and after the retention period expires the data are deleted in a controlled environment. In most of the surveyed countries, a centralized national database is used for data storing except for Estonia where the national database is not centralized (although it may seem so). Rather they use the X-Road platform for data exchange which allows data storing from different sources (which may use different systems). In all of the researched countries, the data from the e-prescription services are stored in a centralized database.

F. System operation transparency

The system operation transparency was investigated through questions: Is there a defined service use policy? Are there any proclaimed technological measures guaranteeing the users that their data are only used for the defined purpose? Are the employees required to sign a non-disclosure agreement? Can users access and view their data through the service? Can users correct or update any of their data within the service? If yes, can the request for correction be sent electronically? Can users monitor status of their application?

In almost all of the surveyed countries users are able to access and view their data. In Estonia doctors have the possibility to prohibit patients from accessing their e-health records if they estimate it would have a certain negative impact to their health/mental condition. On the other hand, a patient can also prohibit doctors' access to his/hers e-health record. This allows users to control their personal data, but the problem may appear when same rules of prohibition apply in emergencies when a patient may not be in condition to allow a doctor to access his/hers e-health record. In Sweden patients may prohibit doctor's access to certain data, except in emergencies. In Belgium citizens can update data in their e-health records by adding information about any potential complications that have occurred while using certain drugs, information about their health condition etc. On the other hand, in Estonia citizens can update only their demographic data.

It is necessary to provide users with the ability to monitor the state of their e-health record by providing information about who and for what reason accessed their data and/or made changes. In most of the researched countries (Belgium, Denmark, Estonia, Sweden and the United Kingdom) users can access this information and monitor their e-health records more efficiently. Also, certain information should be available to users when using e-prescriptions. Denmark and Estonia recognised the importance to enable their users to monitor which medications were prescribed to them.

Discussion

The study showed that the development of e-services depends on 1) size of the state, 2) informatization of the state, and 3) information literacy of the citizens.

We can determine that countries with a smaller population will have fewer problems with the implementation of e-services at the national level. Countries that appeared to be the most advanced in implementing these e-services were Denmark and Estonia. Both countries have an extremely high percentage of users of these e-services. In Estonia 47% of citizens and 95% of doctors use e-health record and in Denmark almost 100% of doctors use it. When it comes to e-prescriptions, in Denmark almost 88% and in Estonia around 95% of users use the e-service. However, it is important to point out that both countries are relatively small (Denmark with around 5.6 million and Estonia with around 1.3 million citizens). On the other hand, a country like Germany with around 80

million citizens and 16 constituent states will have much more difficulties with the implementation because of the complex state constitution. Also, countries with lower level of informatization or a greater disparity of informatization of rural and urban areas will have more problems introducing e-services. The complete transition from a traditional service to an e-service is possible only in case of complete informatization of health care which still presents a great challenge to certain countries that do not have sufficient financial resources and a clear strategy for the transition to a fully computerized system. Finally, in countries where information literacy is lower it is less likely that citizens will show willingness and interest in using such services. Denmark and Estonia are one of the most developed European countries by the ICT usage and development rankings, which is another reason for a high percentage of users.

Conclusion

There is no doubt that implementing e-services in health care leads to business optimization, but it is also important to point out all of the potential problems and disadvantages when implementing such large-scale and complicated services. When implementing e-services, it is necessary to make an assessment of the current situation and realistically assess the readiness of a country to implement e-health record and e-prescriptions at the national level. A prerequisite when introducing such services is making good foundations by improving the informatization level of the whole country. Also, it is important to be aware that in the first period after implementation not all of the benefits will be visible due to high costs of the development and implementation. Transparency in dealing with confidential data must be a priority so that citizens know what is happening with them at any time. When implementing e-services it is also important to implement technological solutions that will guarantee the security of the stored data. Policies and practices for long-term preservation of authentic and reliable data with preserved integrity and usability during the legally required period of time should be in place. However, it is also important to motivate citizens to use e-services in order to make the development and implementation financially worthwhile. When in doubt, users usually search for the answers online so it is important to make any information regarding the systems' operability available on the Internet. We suggest that e-services should, at minimum, provide information on the 52 questions set out in this research. By providing the basic or more detailed information, e-services will gain trust, users will be motivated to use them, and they will help improve the economy.

Acknowledgements

The research presented here is a part of a much broader research initiative "Comparative Analysis of Implemented Governmental e-Services" which was completed in the context of international multidisciplinary research project InterPARES Trust, http://www.interparestrust.org.

References

Anderson, Ross J., (1996), A Security Privacy Model for Clinical Information Systems. IEEE Symposium on Security and Privacy. Oakland, California, pp. 30-43.

Barrows, C. Randolph and Clayton, D. Paul, (1996), Privacy, Confidentiality and Electronic Medical Records. Journal of the American Medical Informatics Association, Vol. 3, 2.

Cornford, Tony, Hibberd, Ralph; Barber, Nick (2014). The evaluation of the electronic prescription service in primary care. Final report, University of London, 158 pp.

Digitizing Public Services in Europe: Putting ambition into action. 9th Benchmark Measurement (2010), Directorate General for Information Society and Media, European Commission, http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=747, [25.01.2014]

E-government Gateway, (n.d.), Construction related documentation provision. https://www.epaslaugos.lt/portal/citizen/service/177, [01.07.2014]

EHMA, (n.d.), Report on the webinar "Introducing an e-National Patient Summary in Sweden: lessons from practice". http://www.ehma.org/files/Webinar%20Public%20Report;%20Introducing%20an%20e-National%20Patient%20Summary%20in%20Sweden.pdf, [12.07.2014]

ENTRUST, (2007), Understanding Digital Certificates & Secure Sockets Layer. http://www.entrust.net/ssl-resources/pdf/understanding_ssl.pdf , [03.07.2014]

ePractice, (2014), eGovernment Factsheets Lithuania. http://www.epractice.eu/files/eGovernment%20in%20LT%20-%20May%202014%20-%20v.16_0.pdf, [12.06.2014]

ePractice, (2014), eGovernment Factsheets United Kingdom. https://joinup.ec.europa.eu/sites/default/files/98/c1/99/eGov%20in%20UK%20May%202014%20v.16.0.pdf, [12.06.2014]

Ericsson Nikola Tesla d.d, (2005), Informacijski sustav primarne zdravstvene zaštite Republike Hrvatske. http://www.cezih.hr/pzz/dokumenti_pzz/HR_PHCIS_FunctionalSpecification.pdf, [16.07.2014]

Estonian e-Health Foundation, (2010), Overview of Estonian Electronic Health Record (EHR) System. http://www.e-tervis.ee/index.php/en/news-and-arcticles/432-overview-of-estonian-electronic-health-record-ehr-system, [07.05.2014]

Fleming, N. et al., (2011), The financial and nonfinancial costs of implementing electronic health records in primary care practices. PubMed, 30(3).

Gray, T. B., (2011), Electronic Health Records: An International Perspective on "Meaningful Use". http://www.commonwealthfund.org/~/media/Files/Publications/Issue%20Brief/2011/Nov/1565_Gray_electronic_med_records_meaningful_use_intl_brief.pdf, [30.06.2014]

HSCIC, (2013), A new way to get your medicines and appliances. http://systems.hscic.gov.uk/eps/library/summary.pdf, [01.07.2014]

HSCIC, (n.d.), Electronic Prescription Service. http://systems.hscic.gov.uk/eps/library/nhsecomms.pdf, [30.6.2014]

InterPARES Trust. http://interparestrust.org/, [07.01.2015]

J. Devlies, U. W., (2010), eHealth Strategies: Country Brief Belgium. http://ehealth-strategies.eu/database/documents/Belgium_CountryBrief_eHStrategies.pdf, [16.07.2014]

Johansen, I., (2004), e-Prescriptions: ETP-Denmark. http://ehtel.eu/activities/tasks-sources/tf-patient-safety-emedication/ehtel-seminar-eprescribing-an-overview-of-challenges-and-experiences-in-europe-amsterdam-2004/files/dk_ib-johansen_2004.pdf, [23.05.2014]

Krag, B. H., (2012), eHealth in Denmark: eHealth as a part of a coherent Danish health care system. Danish Ministry of Health. http://www.sum.dk/~/media/Filer%20-%20Publikationer_i_pdf/2012/Sundheds-IT/Sundheds_IT_juni_web.ashx, [17.07.2014]

Kushniruk, E. B., (2010), Advances in Electronic Helath Records in Denmark: From National Strategy to Effective Healthcare Sysetm Implementation. http://www.scopemed.org/fulltextpdf.php?mno=2604, [13.4.2014]

Layne, K. and Lee, J., (2001), Developing fully functional E-government: A four stage model. Government Information Quarterly, 18 (2)

M. Püüa, U. V., (2012), Ten years of trusted Estonia. http://www.egov2012.gov.cy/mof/DITS/conference/Europeone.nsf/All/E7916860932FBB22C2257ACB004BAEBE/$file/s2SESSiON2_No3_PUUA_Vallner.pdf, [30.06.2014]

MedTech Europe, (n.d.), Denmark: Electronic patient records. http://www.reforminghealthcare.eu/economist-report/some-roads-ahead-innovative-approaches-in-five-west-european-countries/denmark-electronic-patient-records, [17.07.2014]

Menachemi, N. and Collum, T., (2011), Benefits and drawbacks of electronic health record systems

Murray, S. et al., (2004), Implementation strategies for e-government: A stakeholder analysis Approach

Oh, H., Rizo, C., Enkin, M., Jadad, A., (2005), What Is eHealth: A Systematic Review of Published Definitions, J Med Internet Res,7 (1). http://www.jmir.org/2005/1/e1/, [15.02.2015]

Piccinelli, G., Stammers, E., (2001), From E-Processes to E-Networks: an E-Service-oriented approach. http://researcher.watson.ibm.com/researcher/files/us-bth/piccinelli.pdf, [15.11.2014]

Saluse et al., (2010), Assessing the Economic Impact/Net Benefits of the Estonian Electronic Health Record System. Tallinn

Schmitt, K. and Wofford, D., (2002), Financial analysis projects clear returns from electronic medical records. Healthc Financ Manage. PubMed, 56(1).

Smith, E. and Eloff, J. H. P., (1999), Security in health-care information systems –current trends. International Journal of Medical Informatics, Vol. 54, 1, pp. 39-54

Van Dijk, H. V., (2011), Electronic Prescribing in the United Kingdom and in the Netherlands. http://healthit.ahrq.gov/sites/default/files/docs/citation/europe-e-prescribing-report.pdf, [10.07.2014]

Van Langenhove, P. et al., (2013), eHealth European Interoperability Framework, European Commission

Zakon o arhivskom gradivu i arhivima (Official gazette/Narodne novine, 105/97)

Original scientific paper

Lifelong learning among Croatian physiotherapy professionals

Gilbert Hofmann
PhD student, Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
gilbert@net.hr

Tanja Persoglio
Home for abandoned children "Ivana Brlić Mažuranić"
Rijeka, Croatia
tanjapersoglio@gmail.com

Summary

The aim of the study "Lifelong learning among Croatian physiotherapy professionals" was to examine the attitudes that physiotherapy professionals have towards their professional development. The issue being addressed in the study is their level of motivation for lifelong learning. The basic question to which this study was seeking an answer is outlined in the following hypothesis: "There are statistically significant differences in the dependent variables which relate to lifelong learning with regard to five fundamental independent variables: sex, age, years of service, place of work and qualifications".
1271 respondents participated in the study, i.e. 77% of the defined population of physical therapists. The survey was carried out in 65 institutions at all healthcare levels, in Croatia, in the period from 1 October 2013 to 31 March 2014. The respondents were men and women between 18 and 65 years of age, and with 1 to 40 years of service. These respondents varied in terms of their place of work (private, regional and national institutions) and professional qualifications (ranging from secondary school degrees up to university degrees).

Keywords: physiotherapy, lifelong learning, professional development.

Introduction

Nowadays, in time of enormous economic and social change, faced with a demographic situation that is clearly indicating the ageing of the European population, education and learning require a new approach. The concept of lifelong learning pertains to each learning activity embarked upon during one's life that is aimed at enhancing knowledge, skills and competences in personal, civil, social and professional pursuits. [1, 2] Lifelong learning does not apply exclu-

sively to the working hours – to an individual who is invested in his or her personal and professional development, it applies to the whole working life, as well as the time before and after it.

The concept of lifelong learning is commonly associated with financial goals and one's ability to become more competitive in the job market. The knowledge society contributes to a more active social role of an individual and to his or her social affirmation. Knowledge fosters social inclusion, active personal development and competitiveness on the market. [3] The European Commission's Memorandum on Lifelong Learning defines the new skills that are required for an active participation in the society and knowledge economy. [4] Apart from the traditional skill set, it is essential that one should possess skills in IT, foreign languages, technological culture, entrepreneurship, as well as a set of social skills. Together, they form a new level of the *21st century literacies* that allow for an understanding of natural and social processes, critical thinking, successful communication and participation in a multicultural society, entrepreneurial behaviour and permanent learning.

Physiotherapy and lifelong learning

Physiotherapy is a healthcare profession whose aim is to identify and maximise the quality of life and mobility in the areas of promotion of health, prevention of injury and damage, habilitation and rehabilitation. [5]

Formal education in physiotherapy is based on the Croatian Qualifications Framework (HKO) whose role is to link learning outcomes achieved in all educational institutions, and systematise them at the national and international levels. [6]

Lifelong learning follows after the completion of full-time education and is a statutory requirement for all physical therapists who provide direct healthcare services within their scope of work, as well as for those who work outside the healthcare system as part of the framework for prevention and health promotion. [7] Professional non-formal education available to physiotherapy professionals takes the form of: professional conferences, peer-group meetings that include a lecture, professional advisory sessions, courses and workshops, e-learning (online courses).

Research aim and methods

The subject and the purpose of this study is to examine the attitudes of physiotherapy professionals pertaining tothe lifelong learning they have, more or less, devoted themselves to. The question is to what extent physiotherapy professionals intend to receive education during their working lives. The basic question which this study was seeking to answer is outlined in the following hypothesis: "There are statistically significant differences in the dependent variables which relate to lifelong learning with regard to five fundamental independent variables: sex, age, years of service, employer's organization and qualifications".

According to the available literature and data, no studies on lifelong learning in physiotherapy have been conducted in Croatia as of yet, which is why this paper could contribute both to popularisation of lifelong learning and its recognition as an indispensible part of advancing one's own professional performance and development. The study used the statistical descriptive correlation method. Empirical data were collected by way of a questionnaire that was created specifically for this study, which consisted of two basic parts:

- General information on the respondent, used to conduct the profile analysis of the respondents who participated in the research.
- A questionnaire on the attitudes of the physiotherapy professionals regarding professional development based on Likert scale criteria.

Results

The study included 1271 representative respondents, which is 77% of the defined population when one takes into account the total number of professional employed at institutions where the survey was conducted.The total number of defined population (employees in institutions) is 1650. The survey was carried out in 65 institutions at all healthcare levels, in Croatia, in the period from 1 October 2013 to 31 March 2014. Agroup overview of the results related to the assessment of the attitudes on lifelong learning, based on a Likert scale that was recoded into positive (4 + 5), neutral (3) and negative attitudes (1 + 2), is presented in Table 1.

Table 1. Attitudes towards lifelong learning[1]

Statement		Attitude value based on Likert scale			M	SD
		1 + 2	3	4 + 5		
1	I cover the entire cost of my education (seminars, trainings, lectures ...) myself	22,3	17,7	60	3,65	1,254
2	The employer is interested in my further training and encourages me to transfer/present knowledge to my colleagues	41,2	29,5	29,3	2,74	1,185
3	I regularly read professional literature,and use the internet to keep up with the latest developments in my line of work	12,5	15,3	72,3	3,72	0,935
4	Command of foreign languages can help me in further training or education	4,2	8,9	87	4,15	0,809
5	I think it is necessary for educational institutions to become more involved in the process of lifelong learning	2,4	13,3	84,3	4,05	0,709

A random selection of respondents in the study resulted in a larger number of women (73,6%) than men (26,4%). Sex-disaggregated data are presented in Table 2.

[1] Table legend: (1 = strongly disagree 2 = disagree, 3 = no opinion/neutral, 4 = agree, 5 = strongly agree, M = mean, SD = standard deviation.)

Table 2. Differences in responses by sex

	Statement	Sex	M	SD
1	I cover the entire cost of my education myself	M	3,79	1,21
		F	3,60	1,27
2	The employer is interested in my further training and encourages me to transfer/present knowledge to my colleagues	M	2,72	1,21
		F	2,74	1,18
3	I regularly read professional literature and use the internet to keep up with the latest developments in my line of work	M	3,71	1,01
		F	3,73	0,91
4	Command of foreign languages can help me in further training or education	M	4,21	0,87
		F	4,13	0,79
5	I think it is necessary for educational institutions to become more involved in the process of lifelong learning	M	4,00	0,84
		F	4,07	0,66

With regard to age, the respondents were divided into five groups; 18 – 25 years of age, 26 – 35 years of age, 36 – 45 years of age, 46 – 55 years of age, and 56 – 65 years of age. Information on the respondents and the differences in responses based on age are presented in Tables 3 and 4.

Table 3. Respondents by age

Respondents	Number of respondents	Percentage (%)	Cumulative percentage (%)
18-25	142	11,2	11,2
26-35	369	29,0	40,2
36-45	338	26,6	66,8
46-55	325	25,6	92,4
56-65	97	7,6	100,0
TOTAL	**1271**	**100,0**	

Table 4. Differences in responses by age

	Statement	Group	M	SD
1	I cover the entire cost of my education (seminars, trainings, lectures ...) myself.	18-35	3,89	1,20
		36-55	3,48	1,28
		56-65	3,57	1,19
2	The employer is interested in my further training and encourages me to transfer/present knowledge to my colleagues	18-35	2,86	1,16
		36-55	2,64	1,20
		56-65	2,74	1,22
3	I regularly read professional literature and use the internet to keep up with the latest developments in my line of work	18-35	3,78	0,90
		36-55	3,67	0,96
		56-65	3,73	0,93
4	Command of foreign languages can help me in further training or education	18-35	4,31	0,75
		36-55	4,04	0,83
		56-65	4,04	0,84
5	I think it is necessary for educational institutions to become more involved in the process of lifelong learning	18-35	4,07	0,66
		36-55	4,03	0,75
		56-65	4,07	0,71

According to the level of professional qualifications, the respondents are divided into: 1. university graduates in physiotherapy who have completed the appropriate study in physiotherapy at the University of Applied Health Sciences; 2. bachelors in physical therapy who have completed the appropriate undergraduate professional/academic study in physiotherapy; 3. physiotherapy technicians who have completed a 4-year secondary school programme, and 4. masseurs who have completed the appropriate 3-year secondary school programme. Information on the respondents and the differences in responses are presented in Tables 5 and 6.

Table 5. Respondents by professional qualifications.

Respondents	Number of respondents	Percentage (%)	Cumulative percentage (%)
Graduate physiotherapists	65	5,1	5,1
Bachelors of physiotherapy	753	59,2	64,4
Physiotherapy technicians	437	34,4	98,7
Masseurs	16	1,3	100,0
TOTAL	1271	100,0	

Table 6. Differences in the responses by professional qualifications.

Statement		Group[2]	M	SD
1	I cover the entire cost of my education (seminars, trainings, lectures ...) mselft.	1	3,97	1,311
		2	3,77	1,262
		3	3,41	1,198
		4	3,13	1,088
2	The employer is interested in my further training and encourages me to transfer/present knowledge to my colleagues	1	2,68	1,312
		2	2,70	1,208
		3	2,80	1,129
		4	2,81	1,047
3	I regularly read professional literature and use the internet to keep up with the latest developments in my line of work	1	4,06	0,95
		2	3,76	0,927
		3	3,61	0,929
		4	3,44	0,964
4	Command of foreign languages can help me in further training or education	1	4,40	0,703
		2	4,24	0,765
		3	3,96	0,861
		4	4,00	0,816
5	I think it is necessary for educational institutions to become more involved in the process of lifelong learning	1	4,23	0,844
		2	4,10	0,675
		3	3,94	0,736
		4	4,06	0,574

[2] 1. graduate physical therapist; 2. bachelor of physiotherapy; 3. physiotherapy technician; 4. masseur

With regard to the length of service, the respondents are divided into five groups; 5 years or less, 6 -10 years, 11 – 20 years, 21 -30 years, 31- 40 years, and 40 years of service or more. The data related to the respondents, and the differences in the responses are presented in Tables 7 and 8.

Table 7. Division of respondents by years of service.

Respondents	Number of respondents	Percentage (%)	Cumulative percentage (%)
0-5	259	20,4	20,4
6-10	207	16,3	36,7
11-20	303	23,8	60,5
21-30	292	23,0	83,5
31-40	194	15,3	98,7
40 or more	16	1,3	100,0
TOTAL	**1271**	**100,0**	

Table 8. Differences in the responses by length of service.

Statement		Group	M	SD
1	I cover the entire cost of my education (seminars, trainings, lectures ...) myself.	0-10	3,95	1,171
		11-20	3,58	1,352
		21-30	3,48	1,245
		31 or more	3,31	1,155
2	The employer is interested in my further training and encourages me to transfer/present knowledge to my colleagues	0-10	2,86	1,176
		11-20	2,68	1,187
		21-30	2,65	1,191
		31 or more	2,66	1,177
3	I regularly read professional literature and use the internet to keep up with the latest developments in my line of work	0-10	3,82	0,878
		11-20	3,70	0,96
		21-30	3,64	0,976
		31 or more	3,64	0,949
4	Command of foreign languages can help me in further training or education	0-10	4,35	0,744
		11-20	4,12	0,859
		21-30	4,04	0,775
		31 or more	3,91	0,822
5	I think it is necessary for educational institutions to become more involved in the process of lifelong learning	0-10	4,10	0,641
		11-20	4,00	0,793
		21-30	4,04	0,733
		31 or more	4,01	0,688

The research was conducted at all levels of healthcare. Variables related to the institutions of employment are described in terms of three categories in the following manner: institutions providing primary and secondary healthcare services are included in the category “Founded by regional government”, institutions providing tertiary healthcare services are included in the category “Founded by the central government”, private health institutions providing the physical therapy services are included in the category of “Private institutions”.

Information on the respondents and the differences in responses are presented in Tables 9 and 10.

Table 9 Information on the respondents by the institution of employment.

Respondents			Healthcare institution			Total
			1. Founded by the regional government	2. Founded by the central government	3. Private institution	
Sex	Men	Number of respondents	203	77	55	335
		Percentage by sex	60,6%	23,0%	16,4%	100,0%
		Percentage by category sample	25,8%	24,3%	32,9%	26,4%
	Women	Number of respondents	584	240	112	936
		Percentage by sex	62,4%	25,6%	12,0%	100,0%
		Percentage by category sample	74,2%	75,7%	67,1%	73,6%
Total		Number of respondents	787	317	167	1271
		Total percentage	61,9%	24,9%	13,1%	100,0%

Table 10 Differences in the responses by place of work.

Statement		Group[3]	M	SD
1	I cover the entire cost of my education (seminars, trainings, lectures ...) myself.	1	3,44	1,256
		2	4,06	1,126
		3	3,83	1,259
2	The employer is interested in my further training and encourages me to transfer/present knowledge to my colleagues	1	2,71	1,107
		2	2,61	1,169
		3	3,12	1,468
3	I regularly read professional literature and use the internet to keep up with the latest developments in my line of work	1	3,72	0,922
		2	3,61	0,999
		3	3,93	0,833
4	Command of foreign languages can help me in further training or education	1	4,09	0,810
		2	4,21	0,786
		3	4,29	0,822
5	I think it is necessary that educational institutions become more involved in the process of lifelong learning	1	4,00	0,720
		2	4,08	0,654
		3	4,25	0,726

[3] 1. Founded by the regional government; 2. Founded by the central government; 3. Private institution

Discussion

60% of respondents agree with the statement that they cover the entire cost of their education on their own, whereas 22,3% disagree. Men agree with the above statement more frequently than women. It can be concluded that men have better access to education, but they also cover the costs on their own. With regard to the place of work, those respondents who work at institutions founded by the central government of at private institutions agree more frequently than those who work at an institution founded by the regional government. Employees at institutions founded by the central government work in a less favourable financial conditions, but it is surprising that the employees of private institutions cover the costs on their own more frequently than those working at the institutions founded by the regional government. An entrepreneur is willing to finance the professional training of the employees only if such training will pay off in the future. In comparison with the other three groups of respondents, respondents with 10 or more years of service more frequently agree with the above statement. Highly educated employees agree with the above statement more frequently. They have a greater chance for advancing their career – however, they also cover the costs themselves.

Furthermore, 29,3% of respondents agree with the statement "The employer is interested in my further training and encourages me to convey/present knowledge to my colleagues", whereas 41,2% disagree. Younger employees tend to receive positive feedback from the institution and the employer, as opposed to their older colleagues. Respondents who work at a private institution agree with the statement more often than those who work at an institution founded by the regional or central government. An entrepreneur must have a good reason to allow an employee time and money for training, but at least the entrepreneur will demonstrate a degree of interest and expect appropriate feedback from the employee. On the other hand, an institution founded by the regional or central government may budget for some form of education of their employees, but in real life very few people are interested in information dissemination, further training or knowledge transfer within the institution or beyond.

72,3% of respondents regularly read professional literature, and use the internet to keep up with the latest developments in their line of work in order to advance their professional education, whereas 12,5% do not agree with this statement. Respondents who work at a private institution agree with the statement more often, and at a statistically significant level, than those who work at an institution founded by the regional or central government. Employees of private institutions have greater motivation to stay in touch with their profession and expand their interest and knowledge in this way. There is a difference with respect to the length of service variable. Younger respondents with 10 years of service or less also gather new information and knowledge to a greater extent. Moreo-

ver, staff with a higher level of education claim they read professional literature more and exhibit more interest for the latest developments in their line of work.
87% of respondents agree that a good command of foreign languages can help them in further training or education. Respondents aged 18 to 35 agree more frequently with this statement than those aged 36 to 65. Younger respondents are more aware of the market changes and global developments in terms of future employment opportunities, for example in the EU countries. Naturally, older colleagues are already limited by their age, and therefore less competitive. Those with fewer years of service and those with a higher level of education also opt for additional training through foreign language courses, with the ultimate goal to develop professionally (by reading professional literature, attending lectures, seminars, conferences in a foreign language) and perhaps to further their education beyond Croatian borders.
84,3% percent of respondents agree that it is necessary for educational institutions to become more involved in the process of lifelong learning. Respondents who work at a private institution agree with the statement more frequently than those who work at an institution founded by the regional or central government. Highly educated staff also agree with the said statement more often. The high percentage of positively oriented respondents is indicative of a pressing need for better networking between educational institutions, employers, and the healthcare system in the context of professional lifelong learning and education.

Conclusion

The survival of any society depends on the recognition of the concept of lifelong learning, both in terms of career development and the development of personal competences and resources which can help both individuals and society as a whole to thrive. We are referring to the development of an individual who will live by interacting with the environment, who learns throughout his or her life and career, and who in his or her pursuit should be provided with adequate conditions at all levels. Only through the close cooperation, synergy and transparent actions of all stakeholders in the process of lifelong learning and career development- above all educational institutions, followed by employers, future physical therapists, and the healthcare and social system as a whole – can a clear purpose and direction be formed to serve as a foundation for quality profession and the society of knowledge and excellence. Investing in human capital remains with the individuals who wish to prove to themselves, their peers, and their employers that change is possible.As we see it, the improvement is possible mainly within schooling system; firstly, students should be motivated towards lifelong learning in specific area of physical therapy and secondly, after they become employees should have the obligation to persist in managing a career. Further more, according to their abilities employers should make possible/give opportunity for employees both to work and study/attend various educational programmes. They should encourage working mobility and transparent

approach towards information and scientific achievements within the profession. Lastly, the enormous gap is present between employers and employees, so we urgently need to work on our feedback potentials. Hopefully, this study will make a contribution and steer physiotherapy careers in a new direction.

References

[1] Zelenika, Ratko. Znanje – temelj društva blagostanja: obrazovna i znanstvena industrija. Rijeka : Ekonomski fakultet u Rijeci, 2007
[2] Pastuović, Nikola. Cjeloživotno učenje i promjene u školovanju. Zagreb : Odgojne znanosti, br. 10 (2008)
[3] Agency for Vocational Education and Training and Adult Education. White Paper on Croatian Education. 14. 5. 2009. http://www.asoo.hr. (Viewed 12. 11.2014)
[4] Croatian Association of People's Open Universities. Memorandum on Lifelong Learning: Bruxelles, 30.10.2000, SEC (2000) 1832: Commission staff working paper. http://www.hzpou.hr/?gid=5&aid=26 (Viewed 20. 11.2014)
[5] Official Gazette of the Republic of Croatia. Act on Physiotherapy Profession, (NN 120/2008). 3. 10. 2008. http://narodne-novine.nn.hr/default.aspx (Viewed 5. 11.2014.)
[6] Beljo Lučić, Ružica et al. Hrvatski kvalifikacijski okvir – Uvod u kvalifikacije. Zagreb; Vlada Republike Hrvatske, Ministarstvo znanosti, obrazovanja i športa, 2009.
[7] Croatian Council of Physiotherapy. Rules on registers of the Croatian Council of Physiotherapy. 28. 12. 2012
[8] http://www.hkf.hr/propisiiobrascikomore/tabid/62/default.aspx (Viewed 5. 11. 2014).

The "eSecMed" app: Secure medication through NFC-solutions for a longer, safer & autonomous life of seniors

Gert Breitfuss
evolaris next level GmbH
Hugo-Wolf-Gasse 8-8a, Graz, Austria
gert.breitfuss@evolaris.net

Stefan Raschhofer
evolaris next level GmbH
Hugo-Wolf-Gasse 8-8a, Graz, Austria
stefan.raschhofer@evolaris.net

Elena Stocker
Research Center Pharmaceutical Engineering GmbH
Inffeldgasse 13, Graz, Austria
elena.stocker@rcpe.at

Carolin Kollegger
Infineon Technologies Austria AG
Babenbergerstrasse 10, Graz, Austria
Carolin.Kollegger-EE@infineon.com

Summary

Within the project "eSecMed – Secure Medication through NFC-Solutions for a longer, safer & autonomous Life of Seniors" funded by the Austrian Research Promotion Agency (FFG), a consortium of 3 partners developed an intelligent blister with a NFC chip and integrated booster antenna combined with an mHealth smartphone application. The objective of this system was to enhance health adherence/compliance and authentication possibilities of medications, as well as minimising undesirable side effects and harmful interactions. This paper describes the evaluation of the overall system with the main focus on the development of an application prototype. With a qualitative assessment (usability testing method) the present solution was tested in terms of usefulness, ease of use and technology acceptance.

Keywords: usability, mobile application, medical compliance, NFC, mHealth, technology acceptance

Introduction

"Medicines will not work if you do not take them" — the WHO stated in 2003, addressing the lack of adherence to prescribed long-term therapies and stressing the fact that only half of all chronically ill patients are compliant to treatment recommendations. [1]

This not only has a negative influence on treatment outcomes but also on the number of cases of avoidable hospitalizations, which are estimated to amount to more than $100 billion per year. [2]

Due to the fact that the aging process influences the metabolism and causes certain physical and psychological deficits, elderly persons are more likely to be affected by coexisting chronic diseases. In order to treat these conditions multiple medications are prescribed (polypharmacy) which increase the occurrence of undesired harmful drug interactions. [3]

In 2010, it was estimated that within the age group of 65 plus 44% of men and 57% of women take at least five different medications each week. [4] The probability of adverse drug effects will reach from 13% when two medications are taken to 58% for five and rises further as the number of medications increases. [5]

Unexpected side effects can also occur, when counterfeit drugs with incorrect ingredients or dosages thereof are taken. According to estimates of the Food and Drug Administration (FDA) at least 10% of all medications worldwide are falsified. [6] The proportion of counterfeit drugs even exceeds 50% in some countries. [7]

In addition, the Counterfeiting Superhighway report, published by the European Alliance for Access to Safe Medicines (EAASM) in 2008, indicated that 63% of all compounds which were purchased from more than 100 online pharmacies were counterfeit or at least substandard. [8]

Information-and-Communication-Technology (ICT)-based solutions put themselves forward for countering the described situation. Thus, the applicability of Near Field Communication (NFC)-enabled mobile phones to facilitate data management and exchange as well as patient monitoring with the additional benefits of reducing time and errors and the high availability of the technology has been pointed out by a number of authors. [9,10,11]

Within the project "eSecMed" funded by the Austrian Research Promotion Agency (FFG), a consortium of 3 partners was assigned with the development of an intelligent blister with an NFC chip and integrated booster antenna combined with an mHealth smartphone application. eSecMed aims to improve patient security through ICT solutions from pharmaceutical manufacturers to the intake of pills by seniors with the ultimate goal to enable seniors to live a longer secure & autonomous life at home. Therefore the consortium partners defined the following application-oriented goals:

- To increase the health adherence by application of NFC as a user-friendly technology suitable for seniors, to assist secure medication and to provide additional information for increasing acceptance, usability & compliance.
- To minimize undesirable side effects and harmful interactions of drugs for seniors by development of a NFC supported virtual medicine cabinet as a smartphone app for mobile devices.

To meet these challenging objectives, the industry partner Infineon (an electronic semi-conductor producer), the Research Center for Pharmaceutical Engineering (RCPE) and the competence center for mobile communication and innovation evolaris joined forces in order to maximize the benefit of the complementary expertise for the target group.

Apart from the objectives mentioned above, also the willingness of seniors to use the new technology plays a key role. Especially during the development of the application concept the Quality by Design (QbD) approach was used. The key aspect of QbD is the assumption that the desired quality of a product or service can be achieved by taking targeted measures during the design process. To achieve this, it is necessary to gain knowledge about the customer needs as well as the risks that are associated with essential processes. [12] As a preliminary point the user needs were identified by using the Living Lab[1] approach, where focus groups consisting of chronically ill patients and caregivers were interviewed in the course of several workshops. Also challenges and possible solutions were discussed with experts such as general practitioners and pharmacists. Due to the inclusion of these persons, it was possible to consider the requirements of every actor involved.

This phase was followed by a first testing phase where potential users examined paper prototypes. The feedback of this group in combination with the results from the requirements phase entered the next iteration within the conception phase and provided the foundation for the design of the existing application.

Bearing these facts in mind, the central thesis of the current study is based on the following research question:

- Does the available solution (application and intelligent blister) meet the requirements of seniors in terms of usability and technology acceptance?

The “eSecMed” app

The main goal of the eSecMed project was the development of an easy to use health application, which acts as a virtual medicine cabinet and the creation of an intelligent pillbox with an NFC chip as well as an integrated booster antenna.

[1] Within the evolaris Mobile Living Lab users are being involved in the process of developing mobile communications solutions. Optimization and increase of the acceptance of ICT products and services through user centered design is a main part of the evolaris development process. Since 2008 the evolaris Mobile Living Lab is part of the European Network of Living Labs (ENoLL) and one of three Living Labs in Austria.

By combining these two components, autonomy and safety of elderly patients can be improved. The following features were designed to ensure the aforementioned:

- The 'Medicine Cabinet' is supposed to support patients in taking the right amount of medication at the right time (see third screen on figure 2). For that reason it is possible to incorporate different medicines by simply scanning the blister with the smartphone (NFC enabled). Once this is done, the patient can enter the prescribed dosage, as well as when it has to be taken. This is enhanced by an alert-function, which can be enabled at ones convenience. Above that, every medication in the cabinet offers information on how many pills are left in the blister (including a warning, when the number reaches a certain value), expiration date, last documented intakes, compliance and storage conditions in terms of temperature.
- Documentation of intake: Every time a pill is removed from the blister, a timestamp is created. This timestamp and the number of pills removed at that time are compared to the prescription data in order to determine the patient's compliance. Documentation is conducted automatically every time the user scans the blister. A brief message appears on the screen to inform the user, whether the documentation has been performed successfully or not.
- Check for adverse effects: By utilizing a database, different medicines can be selected to be checked for adverse effects, regardless of whether the former have already been added to the medicine cabinet or not. This is especially crucial when it comes to polypharmacy.
- Check for authenticity of medications: Also by means of a database, the genuineness of medications can be proved (see second screen on figure 2)
- Daily reports are generated which provide information about the patient's compliance. If there are nonconformities with the prescription in terms of dosage or time of intake, they are highlighted in the report to help patients to improve their compliance.
- Medi Buddies: In order to involve caregivers who can support the patient emotionally, the patient is given the possibility to share the daily reports with them.
- It is also possible for the patient to manually enter data, in case the NFC-feature is not performing as designed.

Figure 1 shows the interaction of an NFC-able smartphone with the intelligent blister. When the smart blister and a smartphone, on which the eSecMed application is installed and active, are brought into immediate proximity of each other, the NFC-connection is established and data will be transferred from the blister to the smartphone. The application starts via the start screen (shown in figure 2) from where the user navigates to the required function.

Figure 1: NFC interaction of application and intelligent blister

The application is designed to meet the needs of elderly patients. For that reason during the conception phase we focused on short navigation paths, as little manual input as possible and reasonable font sizes.
The electrical components enable the described smart blister to store a unique identification number and to keep track of temperature as well as resistance values, which allow a detection of how many pills, have been removed so far. By means of NFC technology this data can be retrieved by an accordingly equipped smartphone.

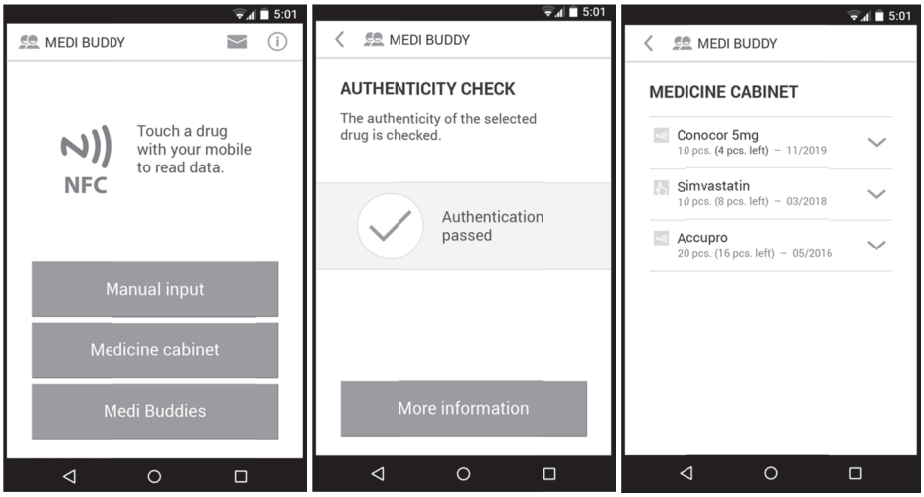


Figure 2: Screen examples

Method

This study is based on the method of usability evaluation. Usability is the ease of use and learnability of a human-made object [13]. The primary notion of usability is that an object designed with a generalized users' psychology and physiology in mind is: [14]

- More efficient to use – takes less time to accomplish a particular task
- Easier to learn – operation can be learned by observing the object
- More satisfying to use

As a general rule for delivering a great experience to mobile users, it is important to think like a mobile user.

- What are customers looking for from a mobile experience?
- What goal do they want to achieve? [15]

The "eSecMed" application was tested with 10 Austrian respondents (6 male, 4 female) using a Nexus 5 (Android 4.1.2) who shared the following sociodemographic characteristics:

- Age: 52 – 69 years
- Medical treatment for a longer period of time within the last 12 months
- Owner of a smartphone with experience in using it
- Not using an application within medical context

The project partner evolaris outlined the usability guidelines for the application and the intelligent blister. The basic concept (paper prototype) was evaluated through the evolaris mobile Living Lab. Based on these results and the underlying research question, an appropriate guideline for this study was compiled. The focus lay on the personal opinion of the respondents. Both positive and negative impressions were taken into account. The interviews were analyzed anonymously and serve as a basis for optimizing the application. This guideline has been divided into three areas.

In the first part (pre-test-questionnaire) experiences with medication handling and first impressions were established. In the second part (tasks-use cases) the respondents had to solve specific tasks relating to the functions of the present solution. Finally (post-test-questionnaire) the respondents were asked questions about the usability of the application and the technology acceptance of the entire system.

Results

The results of the interviews were recorded as follows:

Pre-test: Medication handling in everyday life and first impressions

All respondents have developed experience and strategies in the handling of their medication in everyday life to a certain degree. Strategies reach from using particular kinds of pillboxes, using the alarm function on their smartphone, daily routine (intake at a certain time at a certain place) or support from their

partner. Information about adverse effects are obtained hardly ever, the respondents fully trust their doctors and pharmacists. In some cases the patient information leaflet or the internet is consulted.

- 4 out of 10 users were unfamiliar with the term “NFC technology”, although the concept of contactless payment via NFC was known by everyone.
- Respondents tried to tap the NFC logo on the home screen to ‘activate’ the NFC interface.
- ‘Medicine cabinet’ and ‘manual data input’ were the first options that struck the eye of the respondents.
- Large size of buttons and fonts were appreciated.
- Elements were found neatly arranged.

Task 1: Put this medication into the virtual cabinet

Without preceding instruction the act of ‘scanning’ the pillbox with NFC caused major difficulties among most of the respondents. Some of the subjects handled the smartphone and the blister with extra care as they were afraid of breaking the equipment. Also the time necessary to properly ‘scan’ the blister was misjudged. This lead the respondents to just tap the blister with the smartphone for a split-second, which was insufficient to complete the ‘scanning’-process. Therefore most subjects turned to the option ‘manual data input’. Because of the difficulties using NFC the subjective opinion arose that it would take longer to use it than to manually input the data. At the same time entering complicated names of medications using the keyboard of the smartphone was experienced to be cumbersome by some.

Entering the prescription data could be performed without major problems and was used by 80% of the respondents intuitively. All of the respondents appreciated the amount of information, the clear structure and the information on how many pills are left in the box.

Task 2: Perform an authenticity check

Most respondents (80%) expected this feature to be located inside the medicine cabinet. NFC was not used until the feature could be found in the medicine cabinet. The amount of information provided by the authenticity check increased trust in the application among the respondents especially the information on the expiration date. This feature was mostly seen as a nice-to-have feature as the respondents generally trust their pharmacists and would not expect a medication purchased in a pharmacy to be counterfeit.

Task 3: Push out a pill and document the intake

While one half of the respondents embraced the speed and convenience of the documentation, the other half was insecure whether the documentation was

completed or conducted successfully. The message to inform them was stated to be displayed not long enough to read completely or it was not noticed at all. Manual documentation was experienced to be tedious. One respondent expressed the wish for a text box for personal notes.

Task 4: Perform a check on adverse effects with two different medications
Same as with the authenticity check, many respondents looked for this feature inside the medicine cabinet. Also NFC was not used right away. Once used the feature itself was reported to be helpful. Although using color codes was found useful. For some respondents the codes were not entirely clear, as was the difference between a warning and the urge for caution.

Post-Test: General statement about usability and technology acceptance
The idea of scanning a blister via NFC got a positive rating. All respondents rated the idea of scanning a blister via a smartphone application on a 5 point rating scale (1= very good, 5= not good at all) with 1 (4 respondents) and 2 (4 respondents). In general the usability was rated positive with valuable feedback for improvements in the area of understandability and performance of the scanning function.

Conclusion

Although the target group knew about the existence of NFC technology, the majority was not familiar with alternative domains of NFC besides payment solutions. Also handling this technology is still an unknown territory and causes fears. However confidence increased rapidly after introducing users to the technology and the correct way of handling it. In this respect, the user recognized the usefulness of the solution.
In addition to the already implemented functions of the NFC chip there are plans for enhancements. Through further miniaturization an integrated battery will be available to realize data logging e.g. temperature and humidity sensing across the entire supply chain to monitor transport conditions. Also integration in fully automated primary and/or secondary packaging process is possible. Furthermore electrochemical sensors could be realized for the instant analysis of saliva or blood. It remains to be seen if pharmaceutical companies and semiconductor manufactures are willing to further develop the prototypical solution in future.

References

[1] World Health Organization (WHO). Adherence to long-term therapies: Evidence for Action. Geneva : WHO, 2003

[2] Osterberg, Lars; Blaschke, Terrence. Adherence to medication. // *New England Journal of Medicine.* 353 (2005), 3; 487-497

[3] Sergi, Guiseppe; De Rui, Marina; Sarti, Silvia; Manzato, Enzo. Polypharmacy in the Elderly – Can comprehensive Geriatric Assessment Reduce Inappropriate Medication Use?. // *Drugs and Aging*. 28 (2011), 7; 509-518.

[4] Woodruff, Kathleen. Preventing Polypharmacy in Older Adults. // *American Nurse Today*. 5 (2010), 10

[5] Prybys, Katherine; Melville, Kraig; Hanna, Jeahan; Gee, Andrew; Chyka, Peter. Polypharmacy in the elderly: Clinical challenges in emergency practice: Part 1 overview, etiology and drug interactions. // *Emergency Medicine Reports*, 23(2002); 145-153

[6] Cockburn, Robert; Newton, N. Paul; Agyarko, E. Kyeremateng; Akunyili, Dora; White, J. Nicholas. The global threat of counterfeit drugs: why industry and governments must communicate the dangers. // *PLoS Medicine*. 2 (2005)

[7] Sfetcu, Nicholas. Health & Drugs: Disease, Prescription & Medication. Raleigh : Lulu Press, 2014

[8] European Alliance for Access to Safe Medicines (EAASM). The Counterfeiting Superhighway. Surrey. 1.7.2015. https://lakemedelsverket.se/upload/press/olaga/455_EAASM_counterfeiting_020608.pdf

[9] Benelli, Giuliano; Pozzebon, Alessandro. Near Field Communication and eHealth: Turning a mobile phone into a multipurpose assistant in health scenarios. // *CCIS*. 52 (2010); 356-368

[10] Marcus, Adam; Davidzon, Guido; Law, Denise; Verma, Namrata; Fletcher, Rich; Khan, Aamir; Sarmenta, Luis. Using NFC-enabled Mobile Phones for Public Health in Developing Countries. // *First International Workshop on Near Field Communication*, Hagenberg : IEEE, 2009, 30–35

[11] Pinsker Matthias; Schindler, Karin; Hayn, Dieter; Morak, Jürgen; Kastner, Peter; Riedl, Michaela; Ludvik, Bernhard; Schreier, Günter. Experiences Using Mobile Phones as Patient-Terminal for Telemedical Home Care and Therapy Monitoring of Patients Suffering from Chronic Diseases. // *Lecture Notes in Computer Science*. 5105 (2008); 1305–1312

[12] Juran, M. Joseph. Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services. New York : Free Press, 1992

[13] Dirnbauer, Kurt. Usability – Grundlagen, Beispiele, Trends. Vienna : eye.type design, 2000.

[14] Jordan, W. Patrick; Thomas, Bruce; Weerdmeester, A. Bernard. Usability evaluation in industry. London : Taylor & Francis, 1996

[15] Kim, Y. et al. Protecting Mobile Devices from Adversarial User by Fine-Grained Analysis of User Behavior, Department of Computer Science, Korea Advanced Institute of Science and Technology (KAIST), 2014, Daejeon 305-701, Republic of Korea. http://link.springer.com/chapter/10.1007/978-3-642-41674-3_63#page-1

Long-term Preservation of Longitudinal Statistical Surveys in Psycholinguistic Research

Hrvoje Stančić
Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
hstancic@ffzg.hr

Martina Poljičak Sušec
Croatian Bureau of Statistics
Ilica 3, Zagreb, Croatia
poljicakm@gmail.com

Anabela Lendić
Department of Linguistics,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
alendic@ffzg.hr

Summary

Psycholinguistics deals with different types of evidence and obtained data, including confidential information which needs to be protected from disclosure and other security threats. When it comes to speech-language pathologies, researchers in psycholinguistics are especially interested in aphasia. Aphasia is a loss of language ability as a consequence of brain damage, which may result from head injury or stroke. Research data has to be adequately stored, processed, protected and if possible, preserved for secondary use. Authors are proposing possible application of models and tools used in official statistics and concepts from the archival science that could contribute to solving the so far unresolved issues in the research on aphasia and its records management requirements in the context of long-term preservation, trust, and reuse.

Keywords: psycholinguistics, statistics, archival science, records management, trust in digital records, anonymization, pseudoanonymization

Introduction

Psycholinguistics can be defined as "the study of mental representations and processes involved in language use, including the production, comprehension and storage of spoken and written language" (Warren 2013, 4). The interdisciplinary nature of the field is reflected in the fact that psycholinguistics is influ-

enced by research (and methodologies) in psychology, neurology, and linguistics, as well as those in cognitive science, computer science, and philosophy (Erdeljac 2009; Warren 2013). Since "psycholinguistics tends to blend the theoretical and descriptive insights of linguistics with the experimental methodology and rigor of psychology" (Warren 2013, 6), psycholinguistics has to deal with different combinations of types of evidence and obtained data. When it comes to speech-language pathologies, researchers in psycholinguistics are especially interested in aphasia. According to Warren (2013, 236), aphasia is: "loss of language ability as consequences of brain damage, which may result from head injury or stroke". In the field of psycholinguistics, aphasia research necessarily begins with researchers' request for access to aphasic subjects taking part in clinical therapy[1], and the research proposal needs to include a standard informed consent form approved by an ethic committee. The two most widely used diagnostic instruments for aphasia assessment in English-speaking countries are Boston Diagnostic Aphasia Examination (2001) and Psycholinguistic Assessment of Language Processing in Aphasia (PALPA) (1992). Practical comparisons and normalizations cannot be carried out in aphasia research when it comes to different languages, and data obtained from aphasia research in other languages can only serve as a (more or less "useful") general guideline when it comes to Croatian. There have been no large-scale studies of aphasia in Croatian so far. It is very important to note here that by agreeing to participate in a specific research, the participants also agree to share some of the personal information with the researchers. This paper analyses issues relevant in the context of the development of longitudinal studies on aphasia, of enabling secondary use of research data, and of long-term preservation of the collected data as authentic, reliable, and usable records with their integrity intact.

Psycholinguistics – legal and ethical aspect of protecting sensitive data

Protection of privacy, protection of personal data, and protection of private and family life is a basic human right[2] and is included in the European data protection laws[3]. Medical data are especially sensitive as their disclosure can harm persons by contributing to a social stigma and possible restriction or limitation of persons' rights. The Act on Personal Data Protection[4] regulates the protection of personal data regarding natural persons and the supervision of collecting,

[1] The most prominent clinical institution engaged in aphasia treatment in Croatia is The SUVAG Polyclinic, Section for Therapy of Speech Disorders – Speech Pathology, http://www.suvag.hr/en/, http://www.suvag.hr/en/sluzba-za-govorne-poremecaje-logopedija/.

[2] EU Parliament, The Council and the Commission, 2010.

[3] European Union Agency for Fundamental Rights, 2014.

[4] Legislation Committee of the Croatian Parliament, 2012.

processing, and using personal data in the Republic of Croatia. Concrete actions to prevent data disclosure are taken in the institutions handling data collection and other processes involving usage of sensitive data[5]. The Official Statistics Act[6] regulates official statistical system of the Republic of Croatia[7] and its principles, as well as confidentiality and data protection among other topics.
In aphasia research information such as gender, age, handedness, place of birth, level of education, and knowledge of other languages has to be included alongside other sources of data[8], rather than being separated from them. However, while some of the personal information such as those noted above can be provided by the participant himself, some of the information about the patient, for example his medical diagnosis (type of aphasia), needs to be provided by the clinical institution carrying out the therapy. Over time, the patient's medical record accumulates significant personal information including identification, history of medical diagnosis, the received treatment, medication history, psychological profile, and physicians' subjective assessments of personality and mental state, among others[9].
Abiding to the legal regulations and protection of privacy, researchers from different fields participating in a specific research should be able to access collected data on isolated aspects based on their research agendas.

Data collection – classification in clinical research

Classification in the area of medicine can be defined as a system of categories with criteria sorting diseases and conditions (or disorders, interventions in health, their costs, etc.) in discrete groups. A statistical classification is a classification having a set of discrete categories, which may be assigned to a specific variable registered in a statistical survey or in an administrative file, and used in the production and presentation of statistics[10]. Psycholinguistics should be able to consistently collect, analyse, compare, and interpret different language-related phenomena needed for statistical research processes. The methods of statistical classification could be applied in order to facilitate that process.The World Health Organization (WHO) is the main authority for classification systems in medicine and is responsible for the family of classifications used for statistical research, such as the International Statistical Classification of Diseases and Related Health Problems (ICD), International Classification of Func-

[5] Poljičak & Stančić, 2014.

[6] Official Gazette, No. 12, 2013.

[7] Hrvatski statistički sustav (HSS), http://www.dzs.hr/.

[8] For example, audio and video recordings of the participants performing language-related tasks.

[9] Mercury, 2004.

[10] Hoffmann & Chamie, 1999.

tioning, Disability and Health (ICF), and International Classification of Health Interventions (ICHI). ICD is widely used for mortality and morbidity statistics. Its 10th revision has been used by WHO Member States since 1994. The International Classification for Health Accounts (ICHA) forms a basis for the System of Health Accounts (SHA), an internationally accepted tool developed jointly by OECD, The European Commission, and WHO, and used for describing, summarizing, and analyzing expenditure on health and its financing. The Eurostat's "European shortlist" of 86 causes of death is based on ICD10. The International Shortlist for Hospital Morbidity Tabulation (2005) is likewise based on ICD-10 and is used by Eurostat, WHO, OECD and NOMESCO to collect and present data on hospital discharges[11]. In the course of a statistical survey there can appear a need to include other sorts of classifications, such as Nomenclature of Territorial Units for Statistics (NUTS[12]) to sort and analyze data by territory units. The Neuchâtel Terminology Model (NTM[13]) provides the conceptual framework for the development of a classification database. It defines the key concepts relevant for structuring classification metadata. NTM belongs to the semantic and conceptual sphere of metadata, and it does not include object types and attributes related solely to the technical aspects of a classification database.

Data should be classified according to the sensitivity levels and disclosure risks involved. Sensitivity is a measure of how freely the data can be handled. The sensitivity of a resource falls into two categories: restricted and unrestricted[14]. Data can be otherwise classified as confidential data, internal/private data, and public data[15]. Data (variables) can be classified as identifier, quasi-identifier, sensitive attributes, and non-sensitive attributes[16].

Data processing – preserving information and protecting privacy

Although it might seem contradictory, it is possible to both preserve collected information and protect patients' privacy, while at the same time making the results available to the researchers for further investigations. This can be achieved by preserving the raw data, or original datasets, and creating the processed datasets still retaining all characteristics of the original datasets, except for the pri-

[11] International classification, European Commission http://ec.europa.eu/health/indicators/international_classification/index_en.htm.

[12] Nomenclature of territorial units for statistics (NUTS), Eurostat, http://ec.europa.eu/eurostat/web/nuts/overview.

[13] Hustof, Born, Dunstan, & Mair, 2013.

[14] Sensitivity and Criticality of Data, 2010.

[15] Data Classification and Handling Policy, 2011.

[16] Fung, Wang, Chen, & Yu, 2010.

vacy-related variables. The first and the most basic step in maintaining privacy is to remove variables such as name, social security number, and home address. Removing obvious identifiers from the data is not always adequate to maintain privacy of an individual. Therefore, more rigorous procedures are required to achieve privacy. After removing obvious identifiers, some of the most basic methods for maintaining privacy include limitation of details, top/bottom coding, suppression, rounding and addition of noise[17]. After the data disclosure control methods have been applied, access to confidential data (identifiers and identifying variables) should be possible only if there is a reasonable need and authorization level required to access the variables. Access rights should be administered according to the sensitivity levels and authorization privileges for the data. The second step is the creation and implementation of specialized management system to handle data sensitivity levels and access rights.

Statistical metadata systems play a fundamental role in statistical organizations. A statistical metadata system (SMS[18]) is an important tool for ensuring the goals of the statistical information system are met. SMS should go beyond the function of support for production of official statistics to address other requirements. It should be a tool to facilitate efficient functioning and further development of the whole statistical information system. The possibility to assign access rights to some objects or variables of the database can be implemented on a metadata level to define and manage users and groups of users. Authorization rules can be assigned on specific items of metadata[19] and different levels of access can also be implemented. However, the ethical and legal constraints placed on data collectors processing medical records often limit secondary use of the material, even when anonymized, if prior legal consent has not been obtained[20]. Nevertheless, a system with functionalities similar to those of SMS could be used to manage sensitive health-related data.

In statistical surveys different specialists could be involved in statistical processes. Official statistics is constantly evolving in order to harmonize and standardize the area and even develop official statistics into a standardized industry for production and dissemination of official statistics' data. Usage of the statistical metadata system enables metadata-driven approach to architecture[21] and integration of applications[22]. However, sometimes integration is not possible, or it does not amount to positive outcomes. In these cases a balance be-

[17] Matthews & Harel, 2011.

[18] UNECE, 2009.

[19] Ryals.

[20] The Royal Statistical Society; the UK Data Archive, 2002.

[21] Rivera, Wall, & Glasson, 2013.

[22] Zeila, 2009.

tween integrated and tailored approach to some processes should be met[23]. These areas often include data editing and imputation in statistics. High-Level Group for the Modernisation of Statistical Production and Services (HLG)[24] created in 2010 by the Bureau of the Conference of European Statisticians has a mandate to reflect on and guide strategic developments in the ways in which official statistics are produced. In November 2012 HLG decided that developing Common Statistical Production Architecture (CSPA) was a key priority. Vale states that CSPA "builds on existing standards such as the Generic Statistical Business Process Model (GSBPM) and the Generic Statistical Information Model (GSIM) to create an agreed set of common principles and standards designed to promote greater interoperability within and between statistical organizations"[25]. It was developed to provide the "industry architecture" for official statistics. The Generic Statistical Information Model (GSIM[26]) is a conceptual model that provides a set of standardized, consistently-described information objects, which are the inputs and outputs in the design and production of statistics. Generic Statistical Business Process Model (GSBPM[27]) is a flexible tool to describe and define the set of business processes needed to produce official statistics. The value of the architecture is that it enables collaboration in developing and using services. One of the key enablers for the CSPA is Catalogues. It is envisaged that each statistical organization will have catalogues of processes, information objects, and statistical services. There will be a Global Artefact Catalogue which contains the shareable/reusable artefacts (i.e. processes, information objects and statistical services) from the statistical organizations. The Architecture Sprint proposed eight ways in which CSPA may be used[28].

Data preservation

As seen from the discussion so far, the official statistics has developed a sophisticated ecosystem of the models used by statistical organizations. Those models or parts of them (or even concepts) could be used in collecting, processing, and preserving health-related digital records. Further we will focus on the issues of the data storage and preservation. Data versioning is one of the concepts important in the context of preservation of authenticity of preserved records.

[23] Seljak, 2013.

[24] HLG, http://www1.unece.org/stat/platform/display/hlgbas.

[25] Vale, 2014.

[26] UNECE, 2013.

[27] Vale, Generic Statistical Business Process Modev v4.0, 2009.

[28] UNECE, 2013.

Versioning of statistical surveys in longitudinal studies should be possible using statistical metadata system in an information system. The collected data should be stored in the standardized, globally accepted file formats that enable long-term preservation and preservation of the authenticity and integrity of records, as well as assuring their accessibility. Retention policies for clinical records are defined by the legal framework of each country. Therefore, the minimum period of preservation should be set accordingly. The legal regulations vary considerably, but some types of records could be destroyed after several years, while others should be kept for 70 or more years. Nevertheless, many health care organizations hold records longer than mandated, but over time much of the clinical data become difficult to access[29]. In the context of digital data, long-term preservation means a significant number of media migrations and file type conversions during the preservation period. However, the data should stay authentic, reliable, usable, and its integrity should stay preserved at all times.

Institutions producing statistical survey data can manage data preservation with their own resources – persons with adequate skills to handle data preservation, technological requirements, and data storage facilities. Another approach would be to store the data in a digital, cloud-based archive and consider options available in that case. A Data Management Plan should be produced for all research projects that are creating or capturing data[30].

The choice of metadata standard to be used for data documentation is crucial for ensuring interoperability with other institutions' systems. Prevailing standards for data documentation in social sciences are presented in Table 1 which shows some of the available metadata standards.

Table 1: Standards for depositing social science data

Disciplinary area	Metadata standard	Description
Social Sciences	Data Documentation Initiative (DDI)	A metadata specification for the social and behavioral sciences created by the Data Documentation Initiative. Used to document data through its lifecycle and to enhance dataset interoperability.
	Statistical Data and Metadata Exchange (SDMX)	A self-describing data format that provides both metadata and a method of data transmission. It is primarily used in "the world of official statistics", such as the EU, WHO, UNESCO, World Bank, and US Reserve Banks.

Source: Data Management Resources and Services, University of Idaho Library, http://www.lib.uidaho.edu/services/data/data-management/#standards.

[29] Corn, 2009.

[30] London School of Hygiene and Tropical Medicine, 2014.

Both, SDMX and DDI-L (i.e. DDI-Lifecycle), as presented in Figure 1, are dedicated to repurposing data in social sciences and enabling re-use of results for secondary analysis.

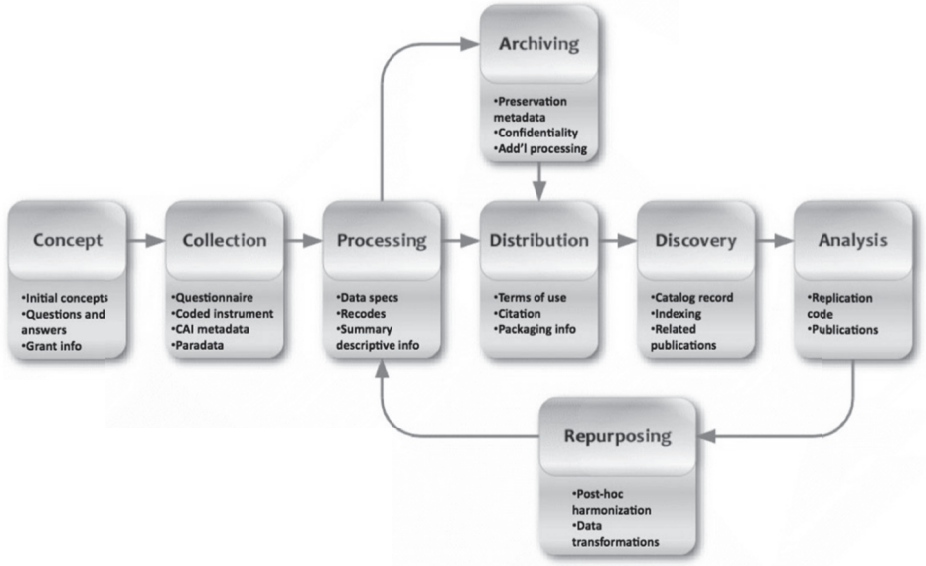


Figure 1. The DDI-Lifecycle Model fostering metadata reuse. DDI Alliance, http://www.ddialliance.org.

The ethical and legal constraints placed on data collectors processing medical records often limit secondary use of the material, even when anonymised, if prior legal consent has not been obtained[31]. The *Processing* phase of the DDI-L Model is where the data about patients is to be properly described. If this phase is done correctly, the anonymization and pseudoanonymization could be achieved. This phase is also the basis for further research made possible through the *Repurposing* phase. Although the two mentioned standards facilitate data and metadata reuse, the key issue is to enable longitudinal studies on aphasia and secondary use of research data. By applying the DDI-L Model to the records management in the aphasia research it could be possible to avoid privacy issues.

[31] The Royal Statistical Society; the UK Data Archive, 2002.

Conclusion

This paper aims to show interrelationships between the needs of psycholinguistic research and the available models and standards in official statistics. Solutions considered range from the ones for data collection, statistical survey processing, and records management. Official statistics' tools can be used to support interdisciplinary research with underlying processes correlating to official statistics' processes in statistical surveys. Furthermore, practices in documenting information objects can be useful as well, as they are based on conceptual frameworks and models – enabling reuse of artefacts. It is the generic nature of standards and models in official statistics and strong metadata orientation, as well as experience in metadata systems' implementation that is seen as valuable to consider for involvement in psycholinguistics in order to have strong methodological ground for data documentation, as well as for enforcement of sensitive data protection and risk management. When it comes to the area of personal data protection in psycholinguistic research, we believe that the records creators can utilize the appropriate methods for statistical disclosure control in planning, realizing, and using the aphasia records management system. A system built in the way in which it is possible to control the metadata, and in turn control the sensitive records, could be than freely used by researchers. By being able to access data and records with no restrictions, the researchers will be able to improve the level of knowledge about aphasia and the patients will be provided with a more successful treatment. Also, if realised as proposed, the records management system will gain trust by patients', their families, and researchers since it would implement all the necessary standards, methods, procedures, metadata schemes and control mechanisms already tested and trusted. Therefore, it can be concluded that the aphasia-related psycholinguistic field of research will become more productive if it incorporates knowledge and solutions from official statistics and modern archival science. We strongly believe that by combining well known and established methods of official statistics with long-term digital records preservation methods and applying them to the field of aphasia research, the reuse of collected patients' data, longitudinal studies on aphasia, secondary use of research data, and protection of patient' privacy rights could be seamlessly achieved.

References

London School of Hygene and Tropical Medicine. (2014.). *Research Data Management Policy.* Retrieved 2015., from http://researchonline.lshtm.ac.uk/612422/1/RDM-Policy-v10.pdf

Corn, M. (2009.). *Archiving the Phenome: Clinical Records Deserve Long-term Preservation.* Retrieved 2015., from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2605592/pdf/1.S1067502708001849.main.pdf

Erdeljac, V. (2009). *Mentalni leksikon: modeli i činjenice.* Zagreb: Ibis grafika.

EU Parliament, The Council and the Commission. (2010). *Charter of Fundamental Rights of the European Union.* Retrieved 2015., from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2010.083.01.0389.01.ENG

European Union Agency for Fundamental Rights. (2014.). *Handbook on European data protection law.* Retrieved 2015., from http://www.echr.coe.int/Documents/Handbook_data_protection_ENG.pdf

Fung, B. C., Wang, K., Chen, R., & Yu, P. S. (2010.). *Privacy Preserving Data Publishing: A Survey of Recent Developments.* Retrieved from https://www.cs.sfu.ca/~wangk/pub/FWCY10csur.pdf

Goodglass, H., & Kaplan, E. (2001). *Boston diagnostic aphasia examination.* Philadelphia, PA: Lippincott, Williams & Wilkins.

Hoffmann, E., & Chamie, M. (1999.). *Standard Statistical Classifications: Basic Principles.* Retrieved from http://unstats.un.org/unsd/class/family/bestprac.pdf

Hustof, A. G., Born, A., Dunstan, T., & Mair, D. (2013.). *NEUCHATEL TERMINOLOGY MODEL: CLASSIFICATION DATABASE OBJECT TYPES AND THEIR ATTRIBUTES. REVISION 2013.* Retrieved 2015., from http://www.google.hr/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0CBwQFjAA&url=http%3A%2F%2Fwww.dst.dk%2Fext%2F595717909%2F0%2Fukraine%2FENG_Statistical-Metadata-Working-Paper-Part-7--pdf&ei=DO6PVf7NK4nWU5j7toAC&usg=AFQjCNHcwJrueqMojwloAB0iRl3IGk1h

Kay, J., Lesser, R., & Max, C. (1992). *Psycholinguistic Assessment of Language Processing in Aphasia (PALPA).* Hove, England: Lawrence Erlbaum Associates.

Legislation Committee of the Croatian Parliament. (2012.). *The Act on Personal Data Protection.* Retrieved 2015., from http://narodne-novine.nn.hr/clanci/sluzbeni/2012_09_106_2300.html

Matthews, G. J., & Harel, O. (2011.). *Data Confidentiality: A review of methods for statistical disclosure limitation and methods for assessing privacy.* Retrieved from https://projecteuclid.org/download/pdfview_1/euclid.ssu/1296828958

Mercury, R. (2004.). *The HIPAA-potamus inn Health Care Data Security, Communication of the ACM, vol . 47, no.7.* Retrieved 2015.

Michigan Technological University Information Technology Services and Security. (2011.). *Data Classification and Handling Policy.* Retrieved 2015., from https://security.mtu.edu/policies-procedures/DataClassificationAndHandlingPolicy.pdf

Official Gazette No. 12. (2013.). Zakon o službenoj statistici. *Narodne novine*.

Poljičak, M., & Stančić, H. (2014.). *Statistical Microdata: Production of Safe Datasets, Transparent Presentation of Contents and Advanced Services for Users through Metadata Authorisation System.* Retrieved 2015., from https://bib.irb.hr/datoteka/714198.Poljicak_ Stancic_ Statistical_Microdata.pdf

Rivera, A., Wall, S., & Glasson, M. (2013.). *Metadata Driven Business Process in the Australian Bureau of Statistics.* Retrieved 2015., from http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP21.pdf

Ryals, M. (n.d.). *SAS Metadata, Authorization and Management Services - Working Together for You.* Retrieved 2015., from http://www.lexjansen.com/nesug/nesug03/et/s1059.pdf

Seljak, R. (2013). Integrated statistical systems and their flexibility - How to find the balance. *NTTS - Conferences on New Techniques and Technologies for Statistics*, (pp. 269-278). Brussels.

Seljak, R. (n.d.). *Integrated statistical systems and their flexibility - How to find the balance.* Retrieved 2015., from http://www.stat.si/dokument/5287

The Royal Statistical Society; the UK Data Archive. (2002). *Preserving and Sharing Statistical Material.* Retrieved 2015, from http://www.rss.org.uk/Images/PDF/publications/rss-reports-preserving-sharing-statistical-data-2002.pdf

The Royal Statistical Society; the UK Data Archive. (2002.). *Preserving and Sharing Statistical Material.* Retrieved 2015., from rss-reports-preserving-sharing-statistical-data-2002

UNECE. (2009.). *Common Metadata Framework Part A: Statistical Metadata in a Corporate Context.* Retrieved 2015., from http://www1.unece.org/stat/platform/display/metis/The+Common+Metadata+Framework

UNECE. (2013.). *Fostering Interoperability in Official Statistics: Common Statistical Production Architecture.* Retrieved 2015.

UNECE. (2013.). *Generic Statistical Information Model.* Retrieved 2015., from http://www1.unece.org/stat/platform/display/gsim/GSIM+Specification

UNECE. (2013.). *Generic Statistical Information Model (GSIM): Specification.* Retrieved 2015., from http://www1.unece.org/stat/platform/display/gsim/GSIM+Specification

University of South Florida. (2010, 6 11). *Sensitivity and Criticallity of Data. Securiy Standard.* Retrieved 2015, from http://www.usf.edu/it/documents/issp001-3-3-10.pdf

University of South Florida. (n.d.). *Sensitivity and Criticallity of Data, Securiy Standard.* Retrieved 2015., from http://www.usf.edu/it/documents/issp001-3-3-10.pdf

Vale, S. (2009). *Generic Statistical Business Process Modev v4.0.* Retrieved 7 4, 2015

Vale, S. (2009.). *Generic Statistical Business Process Modev v4.0.* Retrieved 2015.

Vale, S. (2014). The Common Statistical Production Architecture: An Important New Tool for Process Standardisation. *European Conference on Quality in Official Statistics (Q2014).* Vienna.

Vale, S. (n.d.). *The Common Statistical Production Architecture: An Important New Tool for Process Standardisation.*

Warren, P. (2013). *Introducing Psycholinguistics.* New York: Cambridge University Press.

Zeila, K. (2009.). *Metadata Driven Integrated Statistical Data Processing and Management System.* Retrieved 2015.

TECHNOLOGICAL SOLUTIONS FOR SOCIETY AND EDUCATION

Digital repositories and scientific communication challenge

Radovan Vrana
Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
rvrana@ffzg.hr

Summary

Digital technology has a great influence on scientific communication and scientific research in general. As a result, science has become more highly collaborative, network-based, and data-intensive seeking for different kinds of support for research as well for other activities that follow research such as rapid publication of research results and dissemination and preservation of research data. Having these issues in mind, the traditional system of scientific communication has also begun search for a new publishing and communication platform. At the beginning of the 1990s digital repositories appeared as a possible solution offering several key advantages over the existing publishing and preservation solutions. At the moment digital repositories of scientific information are closest to the position of new mainstream publishing and communication platform in science as they have become more mature and developed networked information systems; as their number in the world is on the rise and as they have become more and more accepted by the scientific community. While first digital repositories were isolated systems, today they distributed federated and highly integrated applications and services. In combination with the open access initiative, digital repositories of scientific information started to offer many benefits for different groups of users. Furthermore, digital repositories could help smaller countries to become more visible in the global scientific community by offering online open access digital peer reviewed scientific content. Digital repositories should solve problems inherited from the existing paper based system of scientific communication: access to scientific information for different categories of users, storage space, slow publishing, quality and quantity of scientific knowledge, organization of scientific knowledge in transparent way, interoperability of information systems and many other issues scientific publishers, universities and their libraries haven't solved so far. This paper offers an overview of the important characteristics that make digital repositories so important in modern science.

Keywords: Scientific communication, E-science, Scientific publishing, Digital repositories

Introduction

Digital technology has a great influence on scientific communication and scientific research in general. Under the influence of digital technology the daily practice of scientists has changed: digital technology has transformed the way in which scientists conduct literature reviews; access research libraries; collect, store and mine research data; publish written research outcomes; communicate with editors and publishers; apply for grants; exchange preprints and reprints; and maintain informal networks with their peers (Genoni, Merick and Wilson 2005). Consequently, science has become more highly collaborative, network-based, and data-intensive. To conduct and publish their research more effectively, scientists need new and improved platforms for publishing which will support their efforts and facilitate better communication with their colleagues. The system of scientific communication that is still in place is now over 300 years old and it has become clear that it should be changed to meet the information needs of modern scientists. The widespread availability of scientific digital content on the internet is already creating opportunities for new forms of research and scholarship that are qualitatively different from the traditional way of using academic publications and research data (Arms and Larsen 2007). Use of digital technology and changes in the ways in which scientists conduct research, collect and analyze data, and communicate their findings to colleagues and scholarly communities (Kafel 2012) has sped up the development of one new information system – the digital repository with aim for the digital repository to become publishing and communication platform. At the moment digital repositories of scientific information are closest to the position of a new mainstream publishing and communication platform in science as their number in the world is on the rise and as they become more widely accepted by scientists and their universities. This paper will discuss the characteristics of digital repositories of scientific information in the prism of change in the scientific communication paradigm.

Scientific communication paradigm change

Like many other areas of human endeavor, scientific communication has been facing challenges introduced partially but significantly by the rapid development of the internet in the 1990s. Before the 1990s, scientific communication was based on printed journals (and other printed publications), there was a substantial lag in refereeing articles and publishing journals, and the distribution system for scientific papers was clumsy and slow, resulting in difficulty for researchers trying to keep up with new developments in their field (Cullen and Chawner 2010). Additionally, scholars and researchers in many countries suffered from a lack of access to the current literature in their field because of high journal subscription prices as well as per article or conference paper prices (Shearer 2003). By being in some cases unable to pay for access to scientific journals, universities lost control over their published scientific output. During

the 1990s scientific communication was undergoing a revolution comparable to the one occasioned by the invention of printing: new ways of creating, reviewing and presenting scientific work were introduced as well as a new information system – thorough and systematic in its distribution, potentially global in scale, almost instantaneous in speed and interactive (Harnad 1990). This new information system is called a digital repository i.e. a digital archive of the scientific output of the members of a scientific community. Digital repositories were designed with the idea of solving the problem of the organization of large quantities of scientific knowledge which exceed the ability of individual scientists to identify and utilize all relevant information needed for research (Lawrence 2001). Digital repositories were also created to become a new publishing platform for academic intuitions enabling them to regain complete control over their scientific output and to make this output more visible.

Another significant initiative that appeared in the 1990s and that helped digital repositories to become more accepted globally by scientists and visible was open access. "The open access literature is digital, online, free of charge, and free of most copyright and licensing restrictions" (Sawant 2013: 16). Open access "means that a reader of a scientific publication can read it over the internet, print it out and even further distribute it for non-commercial purposes without any payments or restrictions" (Björk 2004). The European Commission points out that definitions of open access "(…) describe "access" in the context of open access as including not only basic elements such as the right to read, download and print, but also the right to copy, distribute, search, link, crawl, and mine" (Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 2013: 2). "Open access to a repository adds easy availability of content that may come from various sources to the advantages of no cost or unlimited access to information (Calderón y Enar Ruiz 2013: 194).

Open access offers many benefits for different groups of users. Researchers can benefit from open access in the following ways: their articles are freely accessible online; the authors hold copyright for their work and grant anyone the right to reproduce and disseminate the article, provided that it is correctly cited; a copy of the full text of each open access article is permanently archived in an digital repository separate from the journal; authors are assured that their work is disseminated to the widest possible audience, the widespread availability of articles will enhance literature searching; the results of publicly funded research will be accessible to all interested readers and not just those with access to a library with a subscription; a country's economy will not influence its scientists' ability to access articles (Chiao and Schmidt 2004). Scientists and institutions they work in will attract more funding as their scientific outputs will be more visible and accessible in open access digital repositories. As a result, digital repositories hold the potential to increase the use and application of research so the people who decide on funding the development of digital repositories will

decide more easily which repository project to support (Zuccala, Thelwall, Oppenheim and Dhiensa, 2006).
The general public can also benefit from open access: greater access to research outputs without the often prohibitive costs associated with traditional publishing; no requirement for passwords or other forms of authentication; greater access to academic research findings for those researchers, academic and professional, who work outside academia; ensuring that those who actually provide the money for publicly-funded research (i.e. tax-payers) have access to the outputs they have funded (Suber 2007). Open access has revolutionized the scientific communication paradigm by opening a new chapter in the development of the global system of scientific communication that will enable scientists and readers to publish and access scientific knowledge more easily than ever.
While open access is accepted more widely than ever before (judging from the constantly increasing number of open access digital repositories registered in the Registry of Open Access Repositories (http://roar.eprints.org/) and in The Directory of Open Access Repositories – OpenDOAR (http://www.opendoar.org/), it is still not universally accepted and because of that and because of unwillingness of major world publishers to move their businesses completely to the digital environment for publishing adapted for more cheaper access to the scientific output, the paradigm shift in scientific communication and scientific publishing hasn't occurred yet.

Digital repositories: responding to the communication challenge of modern science

Open access and digital repositories are closely connected as open access literature can usually be delivered in two ways (Zainab 2010): open access digital repositories and open access journals. Digital repositories are more interesting as they aggregate scientific journals (and other types of publications) and offer access to the content of a variety of publications in digital format.
Digital repositories are digital archives of the intellectual product created by the faculty, research staff, and students of an institution and are accessible to end users both within and outside the institution, with few if any barriers to access (Johnson 2002). A digital repository can be owned and managed by an individual, a small group, an institution or commercial organization, a consortium of organizations or a government entity (Branin 2003). That makes it applicable in many different situations where reliable scientific information resources are needed. Digital repositories offer scientists a way to manage and disseminate the digital materials created by the individual academic institution and by the scientific community members. Essentially they represent an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as their organization and access or distribution (Lynch 2003). Digital repositories are archival, stewardship and dissemination systems for content that have a fairly heavy policy component in

terms of who can deposit, what metadata is required for deposit, acceptable formats and the implications of format choices for institutional preservation guarantees (McLean and Lynch 2003). Digital repositories are also information systems and service models designed to collect, organize, store, share, and preserve an institution's digital information or knowledge assets worthy of such investment (Branin 2003).

Digital repositories aim to replace the traditional systems for publishing and preservation of scientific information. Because of that, their functions can be compared to those of a traditional printed journal to see whether this idea is possible. Prosser (2005) chose traditional functions of journals: registration – the author wishes to ensure that she is acknowledged as the person who carried out a specific piece of research and made a specific discovery; certification – through the process of peer review it is determined that the author's claims are reasonable; awareness – the research is communicated to the author's peer group; archiving – the research is retained for posterity. Then he compared journals with digital repositories which showed that digital repositories mirror at least three functions of traditional journals: registration – by depositing in the repository the researcher would make claim to their discovery; awareness – by constructing the repository to open access initiative standards the institution would ensure that the researcher's work would be found by search engines and available to their peers. New alerting services could be developed that would inform readers of new papers deposited in any repository that matched their research interests (in the same way that journal table of contents can be received); archiving – the institution would be responsible for maintaining the long term archive of all the work produced by members of that institution. This would place the focus of archiving back onto the library community where it has rested for centuries, rather than on the publisher community where it has migrated following the transfer from print to online. In many cases the research library will be best placed to maintain over many decades an archive of its own research. This comparison proves that digital repositories are ready for taking over role of main publishing medium in science.

In addition to the enumerated three functions of the traditional journal, Prosser suggested that there are many benefits to institutional digital repositories (Prosser 2005: 1) For the individual: they provide a central archive of the researcher's work; by being free and open they increase the dissemination and impact of the individual's research; they act as a full CV for the researcher 2.) For the institution: they increase the institution's visibility and prestige by bringing together the full range and extent of that institution's research interests; they act as an advertisement for the institution to funding sources, potential new researchers and students, etc. 3.) For society: they provide access to the world's research; they ensure long-term preservation of institutes' academic output; they can accommodate increased volumes of research output (no page limits, can accept large data-sets, 'null-results', etc.). Visibility of scientific output of a

country is very important as it helps scientists and institutions they work in to attract funds for new research. By visibility we mean how avidly published work is received by the academic or scientific community (Miguel, Chinchilla-Rodríguez and de Moya-Anegón 2011: 1130).

Digital repositories are of extreme importance in e-science for care of published scientific works and research data. E-science is most commonly referring to scientific activities supported by high bandwidth computer-mediated telecommunications networks, and particularly to encompass the variety of such digital information-processing applications that are expected to be enabled by the grid i.e. the general purpose network technology which will serve to facilitate new, computationally intensive forms of scientific inquiry (David 2003). E-science offers better support to collaborative science i.e. the ability to have broader interactions through the sharing of data, experimental approaches and both intermediate and final results in systems that will maintain a history of the data, processes, outcomes and conversations among scientists (Wright, Sumner, Moore and Koch 2007). For e-science to be successful, new communication and publishing infrastructure is needed. This new infrastructure must facilitate broader interactions through the sharing of data, experimental approaches and both intermediate and final results in systems that will maintain a history of the data, processes, outcomes and conversations among scientists (Wright, Sumner, Moore and Koch 2007). Almost all these are already possible in digital repositories or will be possible in the near future.

While rather well developed, digital repositories are far from being perfect and without problems. Some of these include: low awareness of their existence among scientists and in scientific institutions; scientists being unmotivated to deposit the results of their work in digital repositories; long term financing; lack of technical support; lack of long term development policies; lack of clear policies or guidelines for the identification of material to be stored; copyright issues that prevent different publications being stored or made openly accessible; issues around the mandatory peer review of digital content as well as to its authenticity and long term preservation bot of published works and research data. Calderón (2013, 199) analyzed digital repositories at the top 50 European universities and suggested aspects of their development and operation which put them in the leading position. These aspects are: the language of the repository, the diversity of the content, the size of the institution or its funding. In another study Koulouris, Kyriaki-Manessi, Giannakopoulos and Zervos (2013: 770) emphasized another problem related to the success of digital repositories in academic community: faculty members' attitude towards use of digital repositories. They were not familiar with open access benefits, were suspicious of open access policies and were hesitant because publishing rights and publishing policies are not clear. As a result, scientists are reluctant to publish results of their work in digital repositories, but, instead they still use journals as the primary channel for publishing and communication. Obviously, every university should

promote digital repositories to its members and have very clear written policies about publishing and/or archiving scientific works in digital repositories. Furthermore, scientists should be given incentives to publish and store results of their research and other forms of work in digital repositories. So far, there are not very many cases in which scientists' academic advancement is related to their active use of digital repositories. It is also expected that digital repositories will act as data sharing platforms (e.g. for the re-use of research data and learning objects) (Poynder 2006: 15). The main point in accepting digital repositories in the daily scientific routine is receiving the political, organizational, financial, technical and human resources support from the university (or other scientific institution) in which the digital repository exists. Unfortunately, this is not the case in many universities, so digital repositories are developed under the care of university libraries which are often understaffed, unequipped and not properly funded to maintain digital repositories in the long run. The global economic crisis might advance the development of digital repositories as universities have tighter budgets for scientific journal subscriptions so part of these funds can be re-directed to the development and management of digital repositories. In such cases digital repositories might prove to be a very feasible solution to the publishing, dissemination and preservation problems in scientific communities of different sizes (Business Models, Costs and ROI, 2009).

Open access journals and digital repositories in Croatia

Open access is present in Croatia in form of open access journals, books and open access digital repositories. Journals in Croatia benefited from the open access as it introduced "a great advantage and enhanced transparency of the whole research and publishing process and helped in promotion science in Croatia globally" (Stojanovski 2015: 20). The portal "Hrčak" ("Hamster" in English)(http://hrcak.srce.hr/) currently offers access to 388 journals with 127197 full text articles which is the largest number of available full text articles in Croatia in one place. These articles and other open access resources are becoming more extensively used in research and teaching in higher education institutions in Croatia as a replacement for fee based information resources (Vrana 2014). The current list of the Directory of Open Access Repositories at http://www.opendoar.org/countrylist.php?cContinent=Europe includes only seven digital repositories from Croatia. While this number is very small in comparison to some other European countries, these digital repositories have an important role as promoters of open access in the Croatian academic community and serve as role models for future digital repositories in Croatia. Academic libraries which are most closely connected to digital repositories still do not have the adequate support in form of financial resources, hardware, software, written policies, education and training and political support at their universities (Vrana 2010; Vrana 2013). "Libraries at the Croatian universities are facing many challenges and are making efforts to improve the conditions in areas such as de-

velopment of information infrastructure, education and training of library staff, securing financial resources, etc. which are all important for the development and management of digital repositories thus creating the bridge towards e-science" (Vrana 2013: 150). As a result of these sometimes unfavorable conditions, the number of digital repositories in Croatia hasn't increased in years. Nevertheless, the existing digital repositories are important for scientific communities in scientifically peripheral counties as they help in making scientific output of such countries more visible (Stojanovski, Petrak and Macan 2009; Markulin and Šember 2014).

Conclusion

The system of scientific communication has been on the verge of change for some time. In spite of numerous announcements of paradigm change, the existing communication paradigm hasn't changed very much. It has been improved by digital technology to facilitate easier and faster access to published scientific knowledge but it didn't change scientific publishing radically. It has become clear that this change will not happen instantly; instead it will happen over time in several different phases in which different parts of the scientific communication system will be improved and in which these improvements will be carefully implemented. Small countries with limited visibility in the scientific community could benefit greatly from the latest changes in scientific communication oriented towards the development and use of digital repositories of scientific information. Digital repositories of scientific information might be included in one of the most important phases of change, because they could improve directly publishing and offer use of open access materials and ensure their long term preservation. The proof of their advancement can be found in the history of their development. In the last 20 years, digital repositories have come a long way from isolated systems towards distributed, federated and highly integrated applications and services (Jones 2007: 155). Today, digital repositories still have the same goals to fulfill i.e. to "(…) centralize, preserve, and make accessible an institution's intellectual capital, at the same time they will form part of a global system of distributed, interoperable repositories that provides the foundation for a new disaggregated model of scholarly publishing" (Johnson 2002). Development of digital repositories is still burdened by problems such as long term funding, lack of infrastructure and support from the academic community (Vrana, 2014) which slows down their development. In spite of that, their number is constantly growing. Generally speaking, digital repositories already offer the outputs of the institution or community to the world; they maximize the visibility of outputs; they showcase and sell the institution to interested constituencies; they collect and curate digital outputs; they manage and measure research and teaching activities; they provide and promote a workspace for work-in-progress, and for collaborative or large-scale projects; they facilitate and further the development and sharing of digital teaching materials and aids; and

they support and sustain student endeavors (Swan 2008; Kelly and Eells, 2015). These are all reasons why digital repositories could be accepted even more widely by many universities around the world and especially in smaller countries where they might become crucial for the development of the national scientific communication infrastructure and improve the visibility of these countries in the global scientific community.

References

Arms, W.; Larsen R. The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship. 2007. http://www.sis.pitt.edu/~repwkshop/SISNSFReport2.pdf (15.12.2014.)

Björk, B. C. Open access to scientific publications – an analysis of the barriers to change. // Information research. 9(2004), 2 http://www.informationr.net/ir/9-2/paper170.html (15.12. 2014.)

Branin, J. Institutional repositories. 2003. https://kb.osu.edu/dspace/bitstream/1811/441/1/inst_ repos.pdf (15.12.2014.)

Business Models, Costs and ROI (2009). http://www.openoasis.org/index.php?option=com_ content&view=article&id=166&Itemid=357 (9.7.2015.)

Calderón, Aurora Ruiz. The Participation and Web Visibility of University Digital Repositories in the European Context. // Comunicar. 20(2013), 40; 193-200.

Chiao, P.J.; Schmidt, C. Open Access gains attention in scholar communication. // Molecular Cancer. 23(2004), 3; 1-2

David, P. A.; Spence, M. Towards institutional infrastructures for e-Science: the scope of the challenge. Oxford : The University of Oxford, 2003

Directory of Open Access Repositories at http://www.opendoar.org (9.7.2015.)

Genoni, P.; Merrick, H.; Willson, M. The use of the internet to activate latent ties in scholarly communities. // First Monday. 10(2005), 12 http://firstmonday.org/issues/issue10_12/gen oni/index.html (15.12.2014.)

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf (15.12.2014.)

Hrčak portal. http://hrcak.srce.hr/ (9.7.2015.)

Johnson, R. K. Institutional repositories: partnering with faculty to enhance scholarly communication. // D-Lib Magazine. 8(2002), 11 http://www.dlib.org/dlib/november02/johnson/ (15.12. 2014.)

Jones, R.. Giving birth to next generation repositories. // International Journal of Information Management. 27(2007); 154-158

Kelly, J.; Eells, L. Global scholarship The role of subject repositories in advancing research from the developing world. // College & Research Libraries News. 76(2015), 5; 265-268

Koulouris, A.; Kyriaki-Manessi, D.; Giannakopouloa, G. Institutional repository policies: best practices for encouraging self-archiving. // Procedia – Social and Behavioral Sciences. 73(2013); 769–776

Lawrence, S. Free online availability substantially increases a paper's impact. // Nature. (2001), 411; 521

Lynch, C. Institutional repositories: essential infrastructure for scholarship in the digital age. // ARL Bimonthly Report. (2003), 226, http://www.arl.org/resources/pubs/br/br226/br226ir. shtml (15.12.2014.)

Markulin, H.; Šember, M. University of Zagreb Medical School Repository: promoting institutional visibility. // Croatian medical journal, 55(2014); 89-92.

McLean, N.; Lynch, C. Interoperability between Information and Learning Environments: Bridging the Gaps. 2003 http://www.imsglobal.org/DLims_white_paper_publicdraft_1.pdf (15.12.2014.)

Miguel, S.; Chinchilla-Rodríguez, Z.; de Moya-Anegón, F. Open Access and Scopus: A New Approach to Scientific Visibility From the Standpoint of Access. // Journal of the American society for information science and technology, 62(2011), 6; 1130–1145.

Poynder, R.. Clear blue water. 2006 http://www.openscholarship.org/upload/docs/application/pdf/2009-01/poynder_-_irs.pdf (15.12.2014.)

Prosser, D. C. The Next Information Revolution – How Open Access will Transform Scholarly Communications. // International Yearbook of Library and Information Management 2004-2005: Scholarly Publishing in an Electronic Era. London : Facet Publishing, 2005, 99-117

Sawant, S. Open access resources useful in LIS education. // Library Hi Tech News. 30(2013), 7; 16-20

Shearer, K. Institutional Repositories: Towards the Identification of Critical Success Factors. // Canadian journal of information and library science. 27(2003), 3; 89–108

Stojanovski, J. Do Croatian open access journals support ethical research? Content analysis of instructions to authors. // Biochemia Medica, 25(2015), 1; 12-21.

Stojanovski, J.; Petrak, J.; Macan, B. The Croatian national open access journal platform. (2009) https://bib.irb.hr/datoteka/430563.stojanovski_et_al.pdf (9.7.2015.)

Suber, P. Open Access in 2007. // The Journal of Electronic Publishing. 11(2008), 1 http://dx.doi.org/10.3998/3336451.0011.110 (15.12.2014.)

Swan, A. The business of digital repositories in A DRIVER's Guide to European Repositories. Amsterdam: Amsterdan university press, 2008, 15-47

Vrana, R. Access to digital information resources as a support to academic achievement. // Central European Conference on Information and Intelligent Systems – CECIIS 2014. Varaždin : Faculty of Organization and Informatics, University of Zagreb, 2014, 144-151.

Vrana, R. Digital repositories at the Croatian universities: providing support for the enhancement of research and teaching. // MIPRO 2010 33rd International convention on information and communication technology, electronics, and microelectronics, Rijeka : Croatian society for information and communication technology, electronics and microelecronics, 2010, 141-146.

Vrana, Radovan. Digitalni repozitoriji i stvaranje informacijske infrastrukture // Četvrti festival hrvatskih digitalizacijskih projekata / Petrušić, Renata ; Marasović, Nela ; Hraste, Matilda ; Zajec, Jasenka (ur.). Zagreb : Nacionalna i sveučilišna knjižnica u Zagrebu, 2014. 41-42

Vrana, R. Digital repositories of scientific information at the Croatian universities: Developing the bridge towards e-science. // Information Technology Interfaces (ITI), Proceedings of the ITI 2013 35th International Conference. Zagreb : University Computing Centre ; University of Zagreb, 2013. 145-150

Zainab, A. N. Open Access repositories and journals for visibility: Implications for Malaysian libraries. // Malaysian Journal of Library & Information Science. 15(2010), 3; 97-119

Zuccala, A; Thelwall, M.; Oppenheim, C. Dhiensa, R. Digital Repository Management Practices, User Needs and Potential Users: An Integrated Analysis. http://cybermetrics.wlv.ac.uk/DigitalRepositories/FinalReport.pdf, 2006. (9.7.2015.)

Wright, M.; Sumner, T.; Moore, R.; Koch T. Connecting digital libraries to eScience: the future of scientic scholarship. // International Journal on Digital Libraries. 7(2007), 1; 1-4

Welcome to the jungle: science communication in the mediatized society

Tea Čonč, Denis Kos
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10 000 Zagreb
{tconc, dkos}@ffzg.hr

Summary

This paper discusses the influence of the media on science communication, on both the institutional and personal level with regard to the changing public nature of science. Authors analyze three heavily intertwined parts of the modern system of science: regulation of science, communication of science and the processes of its production. Looking from the institutional perspective the following tendencies of science mediatization are observed: institutional and individual self-promotion (i. e. PR), increase of scientific (and/or political) impact, self-regulatory behaviour and simplistic mediation of scientific content. At the micro-level, or in the arena where science actually is produced three areas of particular media influence are highlighted: achieving and improving scientific status, increase of informal scientific communication and changes in scholarly publishing patterns. Authors conclude by emphasizing the need of embedding the media and information literacy content into all levels of science education and production.

Keywords: science, science communication, mediatization theory, media

Introduction

In this paper we explore the influence of the media on science communication. Today, science, scientists and scientific topics are much more exposed to the media than they were ever before. At the same time the overall social influence of the media is increasing. Stig Hjarvard, a media theoretician, observes that the media „mould the way people communicate, act and sustain relationships with each other" [2]. This author points to a necessity of examining how science as an older activity is changing under the influence of the rather recent phenomenon (i.e. the media). This particular influence is the focus of study of a new media concept called mediatization, which has emerged in the last few years and which aims to research how different existing social systems change their inner rules and the ways they function to adopt (more or less consciously) a rationale based on the media logic. Although there have already appeared critical debates

on the subject and the method of mediatization theory[1], we attempt a review of articles on the topic of mediatization of science[2]. To achieve this, we combine insights from the field of information and communication sciences that can shed valuable light on this topic of interest.

There are two main approaches in studying the processes of mediatization, and this paper adopts both of them, each with regard to the theoretical background it originates from. First is the institutionalist approach where mediatization is seen as a process in which non-media social actors have to adapt to media's rules, aims, production logics [2] but also its constraints – an approach that originates from the discourse of the traditional mass media and the "broadcast era" [4]. The second approach looks at media from a socio-technical standpoint which sees mediatization as a process in which the changing information and communication technologies drive the changing communicative construction of culture and society [2] and which is usually connected to the uprise of the new media. In the end, we look how these two approaches affect the scientist himself who is charged with the responsibility of science production.

Underlying these approaches is a concrete definition of the media as a transformational pervasive instrument where the conflicts in different social systems are being reworked in a media space and time according to the principles of media logic. Such a definition constitutes the media as a growing colonizing force; a form through which phenomena are presented and which ultimately wants to become the (most) relevant point of the legitimation and construction of the social order, consequently reshaping other social subsystems according to its logic. This outlook has been called the "centrality of the media" [1] and will be considered as a point of reference in this paper.

Still, we look at science and scientific communication, as one of the older and most prominent of social systems, in context of a society that is becoming ever more globalized, technological, economized but also, as the theory argues – mediatized. Not putting aside the common problems of mediatization theory we will attempt to engage the problem of a mediatized science by taking into account the role of the non-media factors such as cultural, political and social variables considering at the same time relevant aspects at the macro-, meso- and the micro- level, as proposed by Deacon and Stanyer [2] . Moreover, our main focus will not be the media and the mediatic turn themselves, so much as the processes that embody the contemporary practice of science. The key question

[1] For the insight in the most recent debate see [2], [3].

[2] For a concise introduction into the mediatization theory see the short overviews in Hepp, Andreas. Cultures of mediatization. Cambrigde: Polity Press, 2013, and Hjarvard, Stig. The mediatization of culture and society. Park Square: Routlege, 2013. Politics were the first research field of this uprising theory, because of the easily perceived connections it has with the media. Strömback proposes four phases of mediatization of politics which imply a long-term process but which do not necessarily have to coincide with any specific time periods. [15]

to be answered is how can science become mediatized and is it possible to pinpoint the exact processes affected? Science is a complex endeavour. It is impossible to examine all of its aspects since its functions, aims and purposes are much more varied than those of the media system. To find answers we look at science through a model consisting of three heavily intertwined moments or parts of the modern system of science which can encompass the institutionalist and the socio-technical approaches, namely: the regulation of science, communication of science and the processes of its production.

Regulation of science and its changing public nature

We regard the processes of regulation in science as the overarching structure that organizes and legitimizes scientific work in the society. It refers to the institutional and managerial arrangements that are formed through legal regulations and executive academic and governmental instances. From the viewpoint of communication theory, regulation of science is regarded as a conversation between relevant stakeholders. Due to the historically achieved autonomy of science this conversation has the quality of a debate. The way science is and can be done today is a result of a long history of an ongoing negotiation between the scientists, the general public and the state, and nowadays also the corporate sector. Since the 17th century science fought for a primacy in the value of the knowledge it produces and has enjoyed a privileged status. However, due to the quality of its relations with the aforementioned stakeholders this status has been challenged. This problem, dubbed in literature the changing public nature of science (e.g. [10]), is becoming complicated even further by the role that the media play in it.

To understand the problem of the public nature of science we have to look closer at the context in which science exists today. According to Nowotny [10] we can distinguish two main ways the public nature of science is changing. These two influences can be regarded as somewhat opposite to each other. On one hand, there is the proprietization of science – the produced data and achievements that are increasingly being regarded as something that is private ownership of the scientist or the institution that produced them – that "support[s] the transformation of public science into privately owned knowledge domains" [10]. On the other hand, there is a growing demand for access and public oversight of new achievements that are guided by a will "to preserve a sense of control over one's own life in a bewildering world of scientific-technological complexity, intertwined with the relentlessly ongoing process of globalization" [10]. Still, the inner workings of science, the rigour of its method and approach, remain the same. Furthermore, regardless of the current demands, either towards democratization or privatization, its task to systematically discover and present gained understandings about the universe did not change.

At the first look one notices that, in the scope of the regulation of science, the problem of the public nature of science arises in the field of scientific communication, since the produced knowledge (that one can subsequently claim ownership of) has to be communicated and validated. At the second look, as Nowotny has also mentioned and still in the field of communication, the problem also exists in the shaken authority of science to communicate the Truth on Nature's behalf since the general public feels left out i.e. kept on the border of true understanding of contemporary scientific endeavours. Such a definition of the problem identifies two audiences [10]: the "insider" scientific and the "outsider" general public that interact with the newly produced tentative contributions to the ever-changing body of scientific knowledge.

These insights lead us to conclude that because science is increasingly adopting an economic logic [14], it has removed itself from the field of perceived public interest. Under the influence of proprietization and by adoption of economic criteria public science is changing its research priorities and presents itself as obliged to get public approval. On the contrary, as Marcinkowski and Kohring [9] argue: "given the new schemes of public funding, the public communication of science primarily serves the purpose of enabling academic institutions to promote themselves in a competition that has been forced upon them by the political domain" [9]. Science is only in appearance becoming increasingly accountable to the public. If we consider that proprietization needs a climate of deregulation, politics is seen as withdrawing "from its original responsibility of making binding decisions" [9] that configure a public accountability of science. Then, science itself becomes less focused on the need to be accountable to the public and more bound by the need to achieve efficiency. Following the insights gained from mediatization theory it is the media who have taken upon them to hold science accountable and make its inner struggles visible to the public, while they are also providing space and time to individual scientists and institutions for their much needed promotion.

Communicating science to the public

In this chapter we first deal with the "outsider" public of scientific communication. In the reviewed literature, communication to this public is looked through the lens of the field of public relations (PR) mediated mostly through the traditional mass media. To start off with some clarifications on what kind of PR we are at this point referring to, Marcinkowski and Kohring point out that we are not dealing with the question whether the public can play any significant role in deciding what counts as scientific knowledge, since the truthfulness of insights does not change if they are reviewed by the public. The mechanism for declaring something is scientific knowledge still predominantly is the peer-review process coupled with formal scientific criteria that communicated contributions need to integrate. Science is not a genuine public business [9] and its main audience is not the general public [14]. Even so, the pervasive impact of the media

seen mainly from the *institutionalist* perspective reveals four notable tendencies of science mediatization:

- institutional and individual self-promotion,
- increase of scientific (and/or political) impact,
- self-regulatory behaviour and
- simplistic mediation of scientific content.

All of the various forms of public communication are increasingly entering into the media domain. Scientific institutions have reacted to this fact, just like many other institutions. They have adopted the non-scientific game of self-representation in order to publicly legitimize themselves and acquire socio-political relevance. As a result, science related media communications have increased, both within institutions and among individual scientists. [12] At the institutional level we can recognize that new public information (or PR) departments are being established in order to develop appropriate media strategies for public presentation of the institution and create and manage opportunities for scientists to appear in the media "as they preselect and produce stories anticipating the journalistic criteria"[11]. We have to note that the aim of these efforts is not to publicly present the latest scientific achievements, institution's ongoing projects or its future plans etc., rather their aim is to make sure that the institution "looks good" – that it has the possibility to manage and control its public image. Institutions can also "look good" by taking stands in public discussions and political debates or give opinions on emerging social crises (be those political, health or related to other issues). According to Peters, [12] these opinions will later be adopted and repeated by politicians who have the executive power to enable that the scientific expertise has the chance to become a relevant influence in policymaking. This inclusion might be a double-edged sword if politicians use this expertise for justification – to manage the appearance of their proposals, while the content and the estimation of value come in second. This relationship opens up a space for various other kinds of political pressures on science which recurrently stimulate a scientific institution to plan a so-called "crisis management strategy" to defend itself from journalists' attacks when, for example, financial or revision reports, or some incorrect or misleading information is published. To reiterate, science is distraught between the need to promote itself, defend its image, manage its influence and actually be publicly accountable, true to the principles of scientificity and efficient in economic terms without having the chance to rely on any (legally based) mechanisms of control. A further aspect of science mediatization is revealed in scientific institutions attempt to develop their own "media and communication strategy" that can ensure popularity and increase their influence. Peters observes that this media orientation originates from an understanding that the interaction with media is a part of the definition of a modern scientist's role, especially regarding his leadership capabilities. At the same time, institutions are trying to monitor and influence its scien-

tists' media related activities, mostly by subtle means, to make sure they conform to the adopted media strategy. [11]

Mediatization of science spreads the criteria of media logic even between individual scientists. They observe the media interactions of their colleagues and evaluate how they established their competence as scientists in the media, the content of their statements, their self-presentation, the reputation of the news organization they appear at and the level of priority they give to a scientific way of speaking. [11] We can conclude that scientists estimate their colleagues' media interactions in order to learn from their experiences and regulate their own behaviour with regard to "what works" based on the goal of their media appearance.

Finally, the most familiar aspect of the mediatization of science probably refers to media presentations of scientific topics and the expression of journalistic views on science in the media. From the scientific point of view reporting on scientific topics is often too simplistic and exclusive in a combination with the dominant sensationalist approach (for the Croatian example see Šuljok and Vuković Brajdić [16]). Research topics are chosen according to their current potential to entice attention [9] resulting in behaviour typical for the media [17].

Although traditional mass media still present the most commonly used way in which scientific institutions communicate their messages, the process of their recognition of the new media has already started. An example is given by Lövgren and Pallas where a university used blog posts by its vice-chancellor for public promotion that favour market and competition-oriented claims with a dominance of reputation, legitimacy and status based communication. [7] A similar example we find in the work of Lüthje who emphasised that most of the scientific blogs are in fact official PR blogs of scientific institutions. [8]

Producing science in an embedded communication environment

We have talked about how practices of scientific institutions change under the influence of mediatization. In this last chapter we focus on the individual scientist and move away from communication of science mediated through traditional mass media towards the entangled structure of democratized communication and science production through the new media. The new media establish an embedded communication environment where science production is immersed, soaked in a network created by a plethora of technological gadgets and applications which are not just tools, but also communicational devices that organize, facilitate and connect scientists and/or their contributions.[3] Through the interactive new media like social networks, forums, blogging and microblogging services, collaborative content creation platforms such as wikis, etc. the border between internal and external communication of science is getting blurred.

[3] Herein we refer to research data, presentations, and textual scientific contributions.

Since it is mediated through these network-based media science communication gains an aura of accessibility. Web 2.0 based technologies offer a decentralized structure where individual scientists or science enthusiasts (as members of the general public) can create a voice of their own and publish content about science. This process, called disintermediation, undermines centralized mediation, a trait found in the mass media system. Still, although there is no intermediary, the question remains whether Web 2.0 based media construct a logic of their own, and give way to a different, socio-technical mediatization of science.

At this micro-level, we noticed three aspects of the mediatization of science that have direct impact on its production:

- achieving and improving scientific status,
- increase of informal scientific communication, and
- changes in scholarly publishing patterns.

With the help of the new media, scientific institutions or individual scientists can completely avoid journalist mediation and thus "maintain control over the process and content of communication" [11]. Activities carried out online, may directly improve not only public but also the scientific status of the individual scientist. New research is needed that would correlate the scientific impact of individual scientists with the creation of their media presence on the web. On the other side, mediatization might not reveal itself only as an increase but also as a decrease in scientific impact. In this decentralized media space characterized by disintermediation questionable content can easily gain influence. To combat these influences in science the scientific community is using Web 2.0 based media as an approachable way to reveal pseudoscientific work, for example *DC's improbable science* blog [5]. Still, the truth is only in the eye of the beholder – ultimately, he will determine whether these accusations are grounded, or new media are used to undermine particular scientists or research agendas. It is in these cases that mediatization of science presents its influence to reconstruct science as an autonomous social system into a contested area for resources and primacy of outlooks.

The intensification of scientist's online activities leads to an overall increase in informal scientific communication. Public informal communication of science is also getting more and more popular, for example between scientists and the interested public or only among citizens. [8] Examples of informal communication between scientists, experts and the public can best be found on (micro)blogging platforms. Coupled with the fragmentation of communication, important topics[4] are being discussed e.g. on Twitter. This kind of communication has accepted the boundaries posed by the media form. Sentences are short, one

[4] One such case can be found under #acrlrevisions where a number of academics and practitioners discuss the new proposal to revise the American College and Research Libraries Association's Information literacy competency standards for higher education. These debates were later summarized on blogs and articles published in online journals in the field.

has to use a lot of abbreviations to save space and thoughts are maximally compressed. Still, an estimation of the true influence of this kind of communication eludes us. Furthermore, these kinds of informal discussions appear to be used to establish what positions exist and what information is available. Before these particular opinions, interpretations and local knowledge gained from (e.g. professional) experience can sway the dominant position they could serve to establish dialogue in an extended peer community during public debate (about legal propositions, state-wide practical guidelines, strategy formation etc.). Interestingly enough, research shows that Web 2.0 tools are not widely used among scientists [6] and that science still resists the processes of mediatization. [13] It has also been shown that younger scientists are keener to use these tools. [6] [8] This leads us to question how will the dominant scientific communication and production processes look like when young scientists, some of them digital natives, become the leading generation. In this highly mediatized information jungle, besides the scientific skills and knowledge, scientists have to achieve a very high level of media and information literacy. Finally, the mediatized environment is undoubtedly changing patterns of scholarly publishing. Online social networks for scientists enable self-archiving of scientific contributions. Scientists can act on their own due to the possibilities offered by the media and communicate their contributions on their own terms backed by licensing management options like the Creative Commons licences. With the wide-spread of open access initiatives and the increasing number of platforms and tools more and more scientists are becoming aware of the benefit of publishing in open access. Decentralized new media stimulate the openness of science and create a space of influence to change the dominant publishing patterns in science. The appeal of such practices resides in the possibilities the technology offers, since it resonates with the principle of transparency of scientific work – science is given the chance by the new media to mediate its processes publicly from allowing open corroboration of results by posting online the raw research data, publishing pre- and post-print versions of scientific contributions, to the transparent peer-review process and the final act of publishing in open access. Multinational publishing corporations are facing important decisions whether to accept open access initiatives as relevant stakeholders. Rejecting any adjustments to their publishing policies and keeping their original positions could make them become obsolete, since open platforms like arXiv.org enable that scientists themselves maintain the control over the publishing process as a whole.

It is important that (future) scientists have all the tools, skills, knowledge and information needed for orientation in this new scientific (and social) environment. With respect to these changes a scientist's education at all levels has to integrate an emancipatory quality so he/she is able to distinguish the useful from invasive influences. The field of information literacy already debates these needs in context of Research 2.0. In a paper by Koltay, Špiranec and Karvalics [6] information literacy practices of academic libraries are put on the crossroads

as support amidst the changes in the research process brought about by Web 2.0 on one hand and the information needs of individual researchers on the other. From the standpoint of the complex of science, media and information literacies, all involved in Research 2.0, we return to what we have emphasised in the beginning – to the individual researcher and his/her responsibility to uphold the principles of scientificity who understands how these principles are changing under the influence of mediatization and as an engaged and critical information literate preventively ensures that "good" publicly accountable science is done in the first place.

Conclusion and further research recommendations

In this paper we identified two pressure directions that act upon today's science. Above, from the institutional level there are political and economic pressures which make science distraught between the needs to publicly promote itself, defend its image and manage its influence and accountability. While struggling with these pressures in the mediatized environment (presented mainly through traditional, centralized mass media) the influence of the media logic is a potential threat to the autonomy and quality of scientific research. Further research could show whether criteria of scientific efficiency have succumbed to media logic and explore the policy makers' awareness of this influence. On the other hand, there is a growing demand for access and public oversight, but according to the dominant opinion among the theoreticians, public opinion is not indispensable for science credibility. Not to be misunderstood, we do not want to deny that science has to answer to society, and especially so if it is publicly funded, but we argue that media related PR activities might not be the adequate mechanism of science accountability. It should be regulated with mechanisms that are independent from the media influence before any kind of media representation takes place. Qualitative research of media content could elicit media involvement in science regulation. Also, with respect to the institutional communication of science, the foci of future research need to be the practices of institutions and individual scientists that show how they construct media and communication strategies in order to identify and promote best practices. The second part of the article looks at science mediatization processes from below adopting the socio-technical view that reveals how decentralised new media democratize communication and access to scientific information. These processes change the way scientists do science by strengthening informal science communication and by changing publishing practices. Still, a crucial area for further developments in the field needs to be the investigation of the impact of digital scientific waste that overflows the web amplifying the need to develop eco-friendly models of science production. Finally, we believe that to be prepared to deal with these changes it is of great importance that scientists have all the media and information literacy skills and knowledge needed to work in this new scientific (and social) environment.

References

[1] Adolf, Marian. Clarifying mediatization. // *Empedocles*. 3 (2011), 2; 153–175.

[2] Deacon, David; Stanyer, James. Mediatization: key concept or conceptual bandwagon? // *Media, culture & society*. 36(2014); 1032–1044.

[3] Hepp, Andreas; Hjarvard, Stig; Lundby, Knut. Mediatization: theorizing the interplay between media, culture and society. // *Media, culture & society*. 37(2015); 314–324.

[4] Hoskins, Andrew. Flashbulb memories, psychology and media studies. // *Memory studies*. 2(2009), 2; 147–150.

[5] DC's improbable science: truth, falsehood and evidence: [blog]. http://www.dcscience.net (Access date: 2015-16-06)

[6] Koltay, Tibor; Špiranec, Sonja; Karvalics, Laszlo Z. The shift of information literacy towards research 2.0. // *The journal of academic librarianship*. 41(2015); 87–93.

[7] Lövgren, Daniel; Pallas, Josef. Blogging for reputation. // *The governance of modern universities, 21-23 August 2013, Reykjavik, Iceland* / Engwall, Lars (ed.), 2013.

[8] Lüthje, Corinna. Mediatisierte wissenschaftsinterne Kommunikation: Stand der Forschung und theoretische Rahmung. // *kommunikation @ gesellschaft*. 15(2014).

[9] Marcinkowski, Frank; Kohring, Matthias. The changing rationale of science communication: a challenge to scientific autonomy. // *Journal of science communication*. 13(2014), 3.

[10] Nowotny, Helga. The changing nature of public science // *The public nature of science under assault*. / Novotny, Helga [et al.] (ed.). Berlin ; New York : Springer, 2005, 1-27.

[11] Peters, Hans Peter. Gap between science and media revisited. // *Proceedings of the National Academy of Sciences*. 110(2013); 14102–14109.

[12] Peters, Hans Peter; Heinrichs, Harald; Jung, Arlena; Kallfass, Monika; Petersen, Imme. Medialization of science as a prerequisite of its legitimization and political relevance. // *Communicating Science in Social Contexts* / Cheng, Donghong [et al.] (ed.). Dodrecht; London : Springer Netherlands, 2008, 71–92.

[13] Rhoten, Diana; Powell, Walter.W. The frontiers of intellectual property. // *Annual review of law and social science*. 3(2007), 1; 345–373.

[14] Rödder, Simone; Schäfer, Mike. S. Repercussion and resistance: an empirical study on the interrelation between science and mass media. // *Communications*. 35(2010) 3; 249–267.

[15] Strömback, Jesper. Four phases of mediatization. // *The international journal of press/politics*. 13(2008), 3; 228–246.

[16] Šuljok, Adrijana; Vuković Brajdić, Marija. How the Croatian daily press presents science news. // *Science & technology studies* 26(2013), 1; 92–112.

[17] Weingart, Peter. Science and the media. // *Research policy*. 27(1998), 8; 869–879.

Analysis of studies learning outcomes according to Bloom's taxonomy

Vedran Juričić, Monika Pejičić, Gabrijela Vidaković
Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences
I. Lučića 3, Zagreb, Croatia
vjuricic@ffzg.hr, mpejicic@ffzg.hr, gvidakovic@ffzg.hr

Summary

The paper analyses learning outcomes of undergraduate and graduate studies of Faculty of humanities and social sciences, University of Zagreb, and correlates them with the categories of Bloom's taxonomy. By observing the representation of categories in different study programs, the paper analyses the similarities and differences between undergraduate and graduate studies, as well as between single and double major studies.

Keywords: learning outcomes, analysis, correlation, similarity, levels of study

Introduction

A Learning Outcome is a statement of what a student is expected to know, understand and/or be able to demonstrate after completion of a process of learning [5]. Outcomes may refer to one level of study, one subject or one module, however, they are cumulative and interconnected so that they together contribute to the overall learning outcomes of the study program. Outcomes state the competences of the learning program that can also be used to correlate it with its level in the European Qualifications Framework.

Defining what a student will know or understand is a complex process, especially when it is required to be expressed in a short sentence. Teachers often rely on a certain model that helps them to design, assess, and evaluate student learning [7].

American educational psychologist Benjamin Bloom led a group of educators who studied the thinking process and correlated it with educational goals and objectives [4]. This classification is known as Bloom's taxonomy, a model of ordering and classifying thinking. Bloom identified three domains of educational activities: cognitive, affective and psychomotor.

The cognitive domain involves knowledge and the development of intellectual skills [6]. There are six categories that can be thought of as degrees or difficulties: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. Those categories are successive, the previous one must be mastered before

the next one can take place. For example, student cannot understand something if he does not first remember it [2].
The affective domain involves changes in feelings, attitudes and values that shape students' thinking and behaviour [9]. It includes the manner in which the students deal with problems emotionally, for example: feelings, appreciation and motivation [1]. There are five categories in affective domain, ordered by the level of complexity: Receiving, Responding, Valuing, Organization and Characterising.
The psychomotor domain includes physical movement, coordination and use of the motor-skill areas [10]. These skills are measured in terms of speed, precision, distance, procedures or techniques in execution. There are seven psychomotor categories, ordered by the level of complexity: Perception, Set, Guided response, Mechanism, Complex overt response, Adaptation and Origination.
Each category (in all three domains) has corresponding action verbs that teachers can use to form learning outcomes. This paper analyses verbs for learning outcomes of studies on Faculty of humanities and social sciences, and analyses the relation between undergraduate and graduate studies, as well as single and double major studies.

Data acquisition and initial processing

The first step of research was the acquisition of study programs learning outcomes, which was made during the faculty's accreditation process. Learning outcomes were collected for each study program separately using a web application, and then inserted into the database. From total of 138 study programs, there were 16 programs that had no associated learning outcomes and were not taken into further analysis.
Table 1 shows the number of single and double major studies in each level of study. It can be seen that double major studies (form little less than 75% (90) and that graduate studies form almost 66% (80) of all studies. The only integral study (undergraduate and graduate) is History and Geography that has no double major variant. This paper observes various differences between levels of study programs, and because the number of integral studies is inadequate for quality analysis, History and Geography was also omitted from analysis. In other words, the research was based on total number of 121 study programs.

Table 1. The structure of study programs

	Undergraduate	Graduate	Integral	Total
Single major	10	21	1	32
Double major	31	59	0	90
Total	41	80	1	122

The second step was the acquisition of taxonomy action verbs, which were collected from different internet sources, manuals and literature. They were also loaded into the database and associated with one of the categories in their learning domains: cognitive, affective or psychomotor. Each category was assigned a numerical value that was in later analysis used for calculating the value of learning outcomes. In the remainder of the paper, those categories are referred as levels.

The third step was determination of relationships between outcomes and action verbs. The majority of learning outcomes were associated with only one action verb, in which case they were assigned a numerical value of their action verb. For example, outcome "Analyse the basic concept…" has one action verb, *analyse*, (cognitive domain, 4th level) and thus its assigned value was 4.

There were two exceptional cases:

- learning outcome contained more than one action verb, e.g. *define and describe concepts*…
- learning outcome contained an action verb that existed in more than one level or learning domain, e.g. *explain* exists in cognitive (Comprehension) and affective domain (Valuing and Organization)

These problems were solved equally: learning outcome was split for each existing verb or level. For example, *define and describe concepts* was treated as two outcomes, *define concepts* and *describe concepts*.

Table 2 shows the statistics of processed learning outcomes. From total of 151 action verbs in cognitive domain, 77.48% were used in at least one learning outcome. Also, there are 3592 learning outcomes that were classified in cognitive domain because of their relations with cognitive action verbs. The cognitive domain consists of 7 levels, thus the average level can take value from interval [1, 7]. The calculated value of 3.41 shows that the average level of all study programs is approximately in the middle between Application and Analysis level. Similar applies to the affective domain, where the average level is between Valuing and Organization. The weakest level is found in the psychomotor domain, which was expected because of the field of the observed studies.

Table 2. Action verbs usage and average levels

	Total verbs	Used verbs (%)	Outcomes	Average level
Cognitive	151	77.48	3592	3.41 / 7
Affective	129	51.94	2336	3.33 / 5
Psychomotor	69	42.03	825	2.68 / 7

The analysis of common outcomes

A relationship between outcomes and studies was analysed in a way to discover which outcomes are common to all or majority of study programs, i.e. which outcomes are not subject specific. Those outcomes should be generic, which

means that they are not related to any discipline or field, but to the level of study.

Analysis resulted in a very small number of common outcomes. As it is shown in Table 3, the most frequent outcome appears in less than 19% of study programs. It is included only in undergraduate studies, which is understandable because it also refers to the selection of further education and future graduate study. The other four most frequent outcomes appear only in teaching-oriented studies.

Table 3. Common learning outcomes

Learning outcome	Percentage
Assess own interests and competencies and select appropriate areas for further education	18.85
Prepare and transmit course materials in accordance with the fundamental articulation models	9.84
State, explain and apply basic psychological factors of successful learning and teaching	8.20
Apply the skills of lifelong learning and language training in certain field, depending on the needs of the job	5.74
Explain didactic theories and models and apply them in teaching	5.74

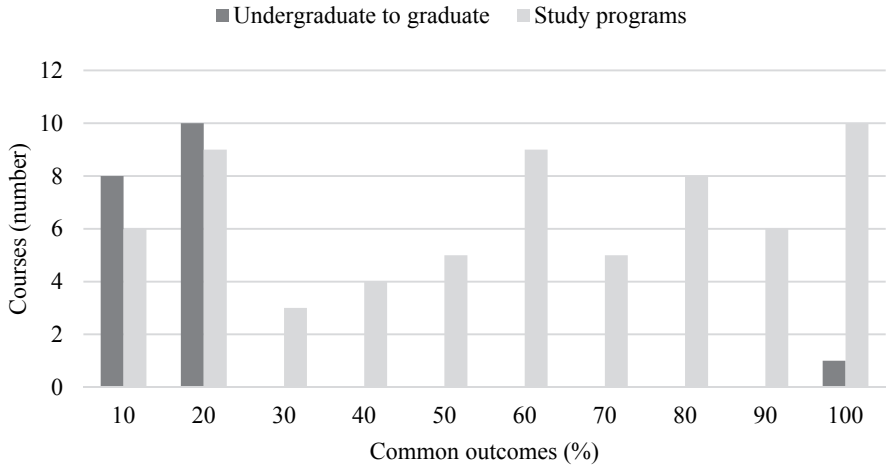


Figure 1: Common outcomes analysis

Data were further analysed for common outcomes between undergraduate and graduate level of each study program (Figure 1, darker fill). Most study programs overlap from 0% to 20% of their outcomes, which corresponds to one or two common outcomes. In this interval there is a total of 18 (out of 41) study programs. For example, *Undergraduate study of Information sciences* and

Graduate programme of library science overlap in one learning outcome: Implement programs of material and data protection. It is interesting to observe that one study program on graduate level completely overlaps with its undergraduate level.
Figure 1 also shows the overlapping of learning outcomes between study programs of their corresponding study (e.g. Graduate study of Informatics, Teaching-oriented and Research-oriented study). There are 10 study programs with completely overlapping outcomes, 6 study programs with 90% overlap, etc. There exists a certain regularity: higher similarity level (above 50%) is noticed only between single and double major variants of the same study program. Other combinations (e.g. between two single major studies) have a lower level of similarity.

Differences between undergraduate and graduate studies

Undergraduate level of study should be formed in a way that it provides general knowledge in the field of study, that is, students are supposed to gain principles and concepts relating to the field [8]. Graduate level of study provides highly specialized knowledge and students gain specialized skills that enable problem solving, integration with other fields and developing new knowledge [3].
Figure 2 shows the relationship between learning outcome levels in each domain and two observed levels of study. Horizontal axis represents learning outcome levels, and vertical axis shows the percentage of each outcome level in undergraduate or graduate studies.

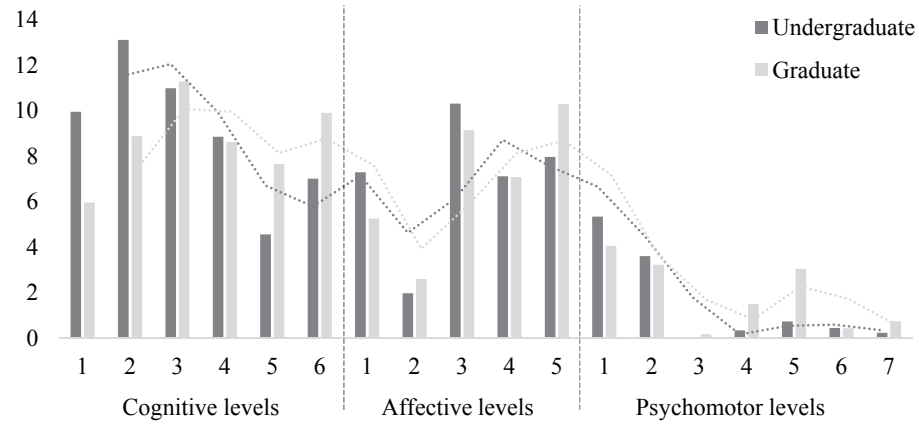


Figure 2: Relationship between outcome levels and study levels

The analysis showed that graduate studies are more represented in higher levels of cognitive domain, i.e. 10% of their learning outcomes is associated to 6th cognitive level, as opposed to 7% in the undergraduate studies. Synthesis level

(5^{th}) refers to integrating and combining ideas and knowledge into new ones, and Evaluation level (6^{th}) refers to assessment and critique of values and ideas. Both levels are emphasised in graduate studies, which is in accordance with differences between undergraduate and graduate studies (explained in the previous paragraph).

In the affective domain, there exists a very small difference between levels of study. The average level of undergraduate studies is 3.19, while graduate studies have average level of 3.42. Graduate studies are noticeably more represented only in the 5^{th} level of the affective domain.

As it is shown in Table 2, psychomotor domain is the least represented one, with only 12% of learning outcomes, versus 53% in cognitive and 35% in affective domain. Thus, each increase in action verb usage has strong influence on study level representation in that domain. It is noticeable that graduate studies are more represented from 3^{rd} to 7^{th} level of the psychomotor domain.

Difference in distributions between outcome and study levels is statistically significant; calculated Chi-Square value was 61.95, which is above the critical value of 26.3 with 16 degrees of freedom. The null hypothesis (there is no difference between distributions) is rejected and the alternative hypothesis (there exists a difference between distributions) is accepted.

Differences between levels of the same study

Data was also analysed to determine differences between undergraduate and graduate level of the same study (e.g. from *Undergraduate study of Information sciences* to *Graduate study of Informatics – Teaching-oriented study*). Values of learning outcomes were summed together for each study level, so study levels could take a value from the interval [0, 15] ([0, 5] cognitive value + [0, 4] affective value + [0, 6] psychomotor value. The increment was calculated so that the final value of the undergraduate study was subtracted from the final value of graduate study. For example, final value for the *Undergraduate study of Information sciences* was 7.15, and for the *Graduate study of Informatics – Teaching-oriented study* was 9.99 (increment is 2.83).

The importance of increment is that it is a measure that shows how much a graduate level of study is above an undergraduate level. In the example above, an increment of 2.83 means that a graduate study is almost 3 levels above an undergraduate study. Those levels can be located within the same domain, but a difference may be caused by any combination of levels.

Maximum determined positive value was 6.69 (almost 7 levels) and maximum negative value was 1.77, which means that undergraduate study is almost 2 levels above graduate study.

Figure 3 shows the number of study programs in each increment interval. It can be seen that there are seven study programs with almost no change in outcome levels, for which the increment takes value between 0 and 0.5 (dashed vertical line). Full vertical line represents an average increment value, which is 1.74.

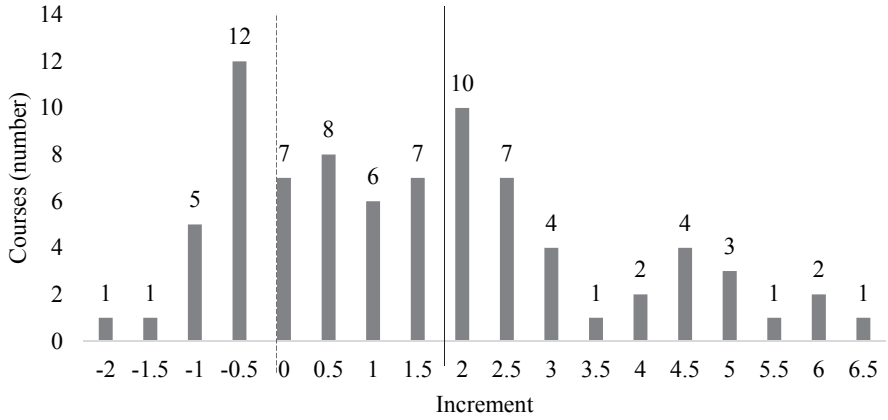


Figure 3: Increments from undergraduate to graduate studies

There are 19 (23%) courses located left of the dashed vertical line, which means they have negative increment and their graduate level is below their undergraduate level. Detailed analysis showed that the maximum decrement is -1.2 for cognitive, and -1.4 for affective and psychomotor domain.
Still, the majority of studies have a positive increment; 70% of them have an increment greater than 0.5. The greatest increment was 6.7 which is quite surprising, as it shows a big leap from undergraduate to graduate level. It is composed of the following increments: 2.27 in the cognitive, 1.00 in the effective and 3.42 in the psychomotor domain.

Differences between single and double major studies

Double major studies are generally combined with each other to provide students broader view and interconnections between two fields. It is possible that single major students receive better knowledge or competencies than the double major students, as latter gain only half ECTS credits in that field.
Figure 4 shows the differences in study outcomes between single and double major studies. Horizontal axis shows the levels of cognitive, affective and psychomotor domain, and vertical axis shows the number of single or double major studies in each level.
It can be seen that single major studies are more represented in higher levels of cognitive, affective and psychomotor domain. The average cognitive level of single major studies is 3.64, versus 3.32 for double major studies. In other words, single and double major studies differ in only 0.3 cognitive level, which is only a small difference.

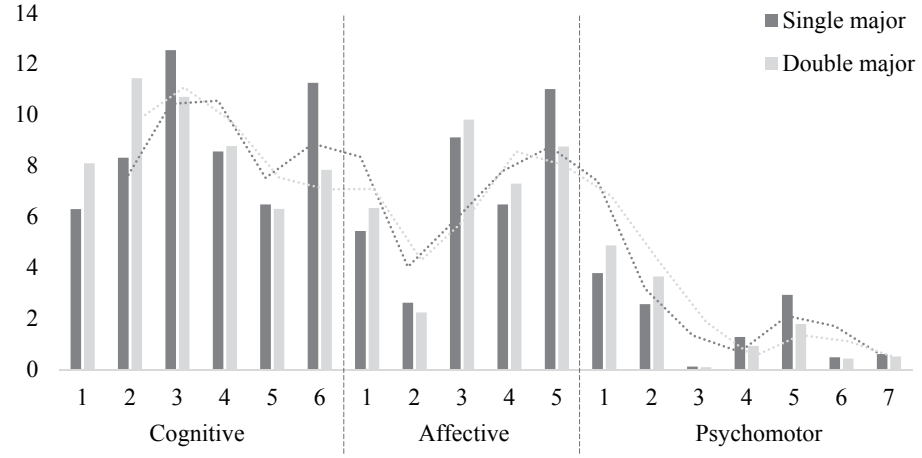


Figure 4: Relationship between single and double major studies

Although a difference between averages of single and double major studies is visually small, a difference in distributions between them is statistically significant; calculated Chi-Square value was 168.73, which is above critical value of 26.3 with 16 degrees of freedom. The null hypothesis (there is no difference between distributions) is rejected and the alternative hypothesis (there exists a difference between distributions) is accepted.

Conclusion

The research in this paper is based on more than 6000 learning outcomes and 121 study programs that are offered on Faculty of humanities and social sciences. The paper analyses a relationship between learning outcomes and the levels of cognitive, affective and psychomotor domain, in order to determine their correlation with the undergraduate and graduate studies, and with single and double major studies.

Analysis showed that the differences between undergraduate and graduate studies are statistically significant. Graduate studies are more represented in higher levels of cognitive, affective and psychomotor domain.

Also, there exists a statistically significant difference between single major and double major studies, where single major studies in average have higher levels of learning outcomes.

The future step in research will be directed towards the analysis of learning outcomes of individual courses, especially on the relationship between them and the learning outcomes of their studies.

References

[1] Bloom's Taxonomy: The Affective Domain. http://www.nwlink.com/~donclark/hrd/Bloom/affective_domain.html (20.06.2015.)
[2] Churches, A. Bloom's Taxonomy Blooms Digitally. http://teachnology.pbworks.com/f/Bloom%5C's+Taxonomy+Blooms+Digitally.pdf (12.06.2015.)
[3] Descriptors defining levels in the European Qualifications Framework (EQF). https://ec.europa.eu/ploteus/content/descriptors-page (23.06.2015.)
[4] Forehand, M. Bloom's Taxonomy: From Emerging Perspectives on Learning, Teaching and Technology. http://www4.edumoodle.at/gwk/pluginfile.php/109/mod_resource/content/5/forehand_bloomschetaxonomie02.pdf (11.06.2015.)
[5] Kennedy, D. Hyland, A. Ryan, N. Writing and Using Learning Outcomes: a Practical Guide. http://www.procesbolonski.uw.edu.pl/dane/learning-outcomes.pdf (29.09.2015)
[6] Learning Domains or Bloom's Taxonomy.
[7] http://users.manchester.edu/Student/GJTribbett/Webpage/Bloom's%20Taxonomies.pdf (15.06.2015.)
[8] Levels of study. http://www.makeyourmark.edu.au/study-options/choosing-what-to-study/levels-of-study/ (24.06.2015.)
[9] Lord, T. Baviskar, S. Moving Students From Information Recitation to Information Understanding: Exploiting Bloom's Taxonomy in Creating Science Questions. Journal of College Science Teaching, 2007.
[10] Neuman Allen, K. Friedman, B.D. Affective learning: A taxonomy for teaching social work values. Journal of Social Work Values and Ethics 7.2, 2010.
[11] Simpson, E. Educational objectives in the psychomotor domain. Behavioral Objectives in Curriculum Development: Selected Readings and Bibliography, 1971.
[12] Vodič za korisnike ECTS-a. http://www.unizd.hr/Portals/0/bolonjski_proces/pdf/Vodic_za_korisnike_ECTS-a. pdf (10.06.2015)

Original scientific paper

Learning with smartphones: a Hong Kong experience

Zvjezdana Dukić
The University of Hong Kong
Hong Kong
dana.dukic@gmail.com

Summary

The study provides an overview of smartphone use for academic learning in higher education. The research was conducted on undergraduate and graduate students enrolled in Library and Information Science (LIS) study program at the University of Hong Kong. The research method is survey and an online questionnaire was used to collect data. The study findings show that LIS students commonly use smartphones for communication, socializing, entertainment, casual reading and other daily needs. They use smartphones for academic learning as well. Most frequent learning activities with smartphones are browsing and reading relevant materials on websites, watching videos, searching for information, using productivity tools and communicating with classmates. Major barriers to using smartphone for academic learning are the smartphone's small screen, websites not formatted for smartphones and long loading time.

Keywords: smartphones, higher education, mobile learning, library and information science

Introduction

Latest developments in mobile and wireless technologies opened up new opportunities to mobile devices owners for knowledge acquisition. Mobile technologies enable learners to access relevant information without time and location restrictions and to communicate easily with others. Although laptops are still the most often used mobile device for learning and study (Dahlstrom, Dziuban and Walker, 2014), smartphones and tablets are getting extremely popular among university students. Smartphones have recently become cheaper and more common among students in higher education, therefore, more students use smartphones for academic purposes (Dahlstrom *et al.*, 2014).

This study aims to provide an insight into smartphone use for academic learning by Library and Information Science (LIS) students from Hong Kong. An inquiry into students' smartphones use for their daily needs is included too, aspiring to provide some background information on smartphone use.

Literature review

There is an extensive literature on smartphone use for learning but only a limited number of studies look specifically at the use of smartphones by students in higher education. A study by Dahlstrom *et al.* (2014) conducted on a sample of 75,000 university students in USA reports a significant growth in students' smartphone use for academic learning over the period from 2012 to 2014. Students' learning activities with smartphones are mostly related to accessing a course management system for teaching materials, news, grades etc. Dresselhaus and Shrode (2012) looked at undergraduate and graduate students at one university in USA and found that 43% of study participants use smartphones for academic purposes daily or weekly. Bradley & Holley (2011) examined mobile phone use for academic learning by undergraduate students at London Metropolitan University and found that students use mobile phones to communicate with classmates, search for information, access learning materials, generate contents, record presentations and take notes. Bomhold (2013) found that more than two thirds of undergraduate students at one university in United States used smartphones for academic purposes, such as retrieving information with search engines and accessing online reference sources. On the other hand very few of these respondents used smartphones for searching library catalogs or subscription databases. Kim, Ilon and Altmann (2013) found that most of education and engineering university students at one Korean university widely used smartphones for learning purposes. Students use smartphones to search the Internet, check dictionary, view documents, translate texts, e-mail and text to classmates and teachers, take notes, manage schedules, access media and social networking websites.

A few studies on smartphone use for academic purposes involve higher education students from Hong Kong (Kukulska-Hulme, Pettit, Bradley, Herrington, Kennedy and Walker, 2011; Cheung, 2014; Dukic, Chiu and Lo, 2015). Kukulska-Hulme *et al.* (2011) explored the use of mobile devices for learning purposes from the learners' perspective on students in master and doctoral programs from Australia, Hong Kong, Portugal, Sweden and UK. The study shows that students in all five countries use mobile phones for learning-related activities like searching for information, reading e-books, listening to education materials, communicating with classmates, group work and collaboration, note taking and scheduling. According to Cheung (2014) undergraduate students in marketing and public relations at the Hong Kong Polytechnic University use smartphones to perform the following learning activities: sending emails to classmates, reading notes, searching Google Scholar, posting comments and uploading contents to the course website. Dukic *et al.* (2015) conducted a research on LIS postgraduate students from the University of Hong Kong and University of Tsukuba (Japan) and found that students rarely use smartphones for reading academic resources. They prefer to use smartphones for learning related activities like discussing assignments, accessing resources and news

from the course learning platform, checking course emails, etc. Use of smartphone for knowledge sharing and peer-based collaborative learning activities is also recognized in some other studies on mobile learning (Lippincott, 2010; Rambe & Bere, 2013; Liu et al., 2013).
From this brief literature review it can be concluded that students in higher education widely use mobile phones for academic learning. This is the only study that takes LIS students as subjects, so it will shed more light on this particular student profile.

Research questions

The main goals of this study is to investigate the scope of LIS students' use of smartphone for academic learning purposes, to detect what specific learning activities students perform with their smartphones and to identify possible barriers to smartphone use for learning. The major research questions are listed below:

1. To what extent LIS students use smartphones for academic leaning?
2. What typical learning related activities do LIS students perform with smartphones?
3. What are possible barriers to LIS students' smartphones use for learning?

Research method

The research method applied in this study is survey and an online questionnaire is used to collect data. The questionnaire included seventeen questions, fifteen closed-ended questions and two open-ended questions, all arranged in three parts. In the first part we collected data on respondents' demographic characteristics, on some general features of their smartphones and on circumstances of smartphone use. The second part examined participants' smartphone use for everyday needs. The third part of the questionnaire gathered data on participants' activities with smartphone related to learning and study needs. This part also looked at possible barriers to smartphone use for learning. Most of questions in the second and third part applied a Likert scale.
The study was conducted on LIS students enrolled in bachelor and master program at the University of Hong Kong. The questionnaire was posted to the online polling platform *SurveyMonkey* and the link to the questionnaire was sent to students by email. The participation in the research was voluntary and anonymous. Ethical Clearance for this research was obtained from the Human Research Ethics Committee for Non-Clinical Faculties of the University of Hong Kong. Data were collected during the spring and autumn academic terms in 2014. Ninety-three valid responses were received.

Findings and discussion

Major demographic characteristics of study participants can be seen in Table 1. The data show that there are more female students than male students. Only a few of study participants are in the age group ranging from 41 to 50. The percentage of master students is slightly higher than bachelor students. Comparing the age and study level our data show that 43% of students in the age group up to 30 are master students while in the age group from 31 to 40 master students make 77%. It can be noticed that 33% in the sample are mature bachelor students.

Table 1. Gender, age and study level

Gender		Age		Study level	
Male	35%	20 - 30	54%	Bachelor degree	42%
Female	65%	31 - 40	43%	Master degree	58%
		41 - 50	3%		

When study participants were asked how often they accessed the Internet with their smartphones, it turned out that even when a computer with Internet access was easily available most of them used smartphones. Our data show that 51% of study participants use the Internet from smartphone 'very often', 20% do it 'often', 22% 'sometimes' and only 7% 'rarely'. Of course, smartphones can't replace computers in performing many complex tasks but it is obvious that smartphones are widely used to supplement computer use for completing some simpler activities.

Smartphone use for everyday needs

Daily smartphone activities are considered to be a good indicator of study participants' general acceptance of smartphones. Therefore an insight into LIS students' smartphone use for their everyday needs helps us better understand their smartphone use for learning purposes.

Table 2. Smartphone activities for everyday needs

Types of activities	Daily/weekly
Communicate with friends & family (e.g. email, SMS, chat)	95%
Use search engines (e.g. Google, Yahoo, Baidu)	89%
Social activities with social media (e.g. Facebook, Twitter)	83%
Games, music, movies, TV series, etc.	82%
Using productivity tools (e.g. calendar, notes, to-do lists)	75%
Finding locations, like streets, restaurants etc.	59%
Accessing reference sources (e.g. encyclopedia, dictionary)	52%
Hobbies, sports, fitness, travel	46%
Casual reading: books, comics, magazines, newspapers, etc.	75%
Academic reading: articles, e-books, blogs, websites	40%

The study shows (Table 2) that LIS students extensively use their smartphones for various daily needs. Most of them regularly use smartphones for talking, sending email, texting or chatting. Further, they often communicate with friends and family through social media and also use search engines. Other highly popular activities involve entertainment, casual reading and using productivity tools. Data in Table 2 show that 75% of LIS students read casual literature such as books, comics, magazines etc. This finding supports the findings from the study by Dukic et al.(2015) that reading novels and comics with smartphones is very popular among Hong Kong students. Academic reading with smartphone is less popular then casual reading, but still up to 40% of LIS students read academic literature on daily or weekly basis. These findings go in line with the findings in studies by Kukulska-Hulme *et al*. (2011), Bomhold (2013) and Cheung (2014).

Smartphone use for learning purposes

Study findings on LIS students' smartphone activities related to learning and study are organized under four major categories:

- communication and sharing
- browsing, reading, viewing and listening
- searching and accessing information and
- using productivity tools and recording

The research results (Table 3) indicate that more than half of study participants use their smartphones to discuss various study related issues with classmates or teachers. It is interesting that they do it more often by email or texting (SMS/MMS) than by talking on the phone. Less often LIS students share their learning experience on the official course management platform such as Moodle or on some other social networking websites.

Table 3. Smartphone use to communicate and share (talking, texting, posting) for learning

Types of activities	Daily/weekly
Talking to classmates to discuss course materials, assignment etc.	58%
Using email, SMS, MMS or chat apps for study related issues with classmates/teachers	62%
Posting to class forums on the learning management platform (e.g. Moodle)	26%
Posting or commenting study related items to social networking sites (e.g. Facebook, Twitter)	33%

Findings about using smartphones for study related communication and sharing with classmates emphasize the role of smartphone technology in facilitating collaborative learning. A qualitative research conducted by Dukic *et al.* (2015) also found that smartphones are very effective for students' collaboration and sharing for study purposes as well.

From data in Table 4 it can be seen that a high percentage of LIS students browse daily or weekly through various websites on their smartphones for learning purposes. Browsing or reading posts on social networking sites is even more popular. Data indicate that 85% of study participants report using smartphones for learning related activities through social networking sites. These findings are not surprising because research on mobile learning already recognized the close connection between smartphone and social media (Gikas and Grant, 2013). On the other hand, it surprises that only 24% of students access learning management platform with smartphones on daily or weekly basis. This result differs significantly from finding in the study by Dahlstrom *et al*. (2013) showing that over 80% of students use their smartphones to access course manage systems for their learning related needs.

Table 4. Smartphone use for browsing, reading, viewing and listening for learning purposes

Types of activities	Daily/weekly
Browsing through websites, blogs, wikis, micro-blogs etc.	68%
Browsing or reading posts on social networking sites (e.g. Facebook)	85%
Reading articles from academic journals & magazines	22%
Reading e-books	23%
Viewing a video clip (from YouTube, TED talks or similar)	55%
Accessing learning management platform for information or resources (e.g. Moodle)	24%

Using smartphones to watch videos for learning purposes is also popular among LIS students. More than half of respondents watch YouTube, TED talks or some other learning related videos daily or weekly. Less popular is reading academic journals and e-books. In our study only 22 % read academic journals and 23 % read e-books on their smartphones. Students participating in the research by Dukic *et al.* (2015) argue that they prefer to read academic works at their desk at home where they can fully concentrate on the task. Smartphones are usually used *on-the-go*, in the environment that might be distractive and thus not suitable for serious academic work. It is worth mentioning that our data on reading academic journals and e-books differ from findings in study by Kukulska-Hulme *et al*. (2011) where 43 % of Hong Kong students report reading e-books and 35 % reading academic journals with smartphones once a week or more often. An explanation for these disparate findings might be that study participants from Hong Kong in the study by Kukulska-Hulme *et al*. (2011) were science students and that disciplinary profile influences smartphone use for learning.

Table 5. Using smartphone for searching and accessing information for learning

Types of activities	Daily/Weekly
Accessing and searching library catalog	28%
Accessing and searching e-databases	20%
Accessing reference sources (e.g. encyclopedia, dictionary)	44%
Searching with search engines (e.g. Google, Yahoo etc.)	79%

When it comes to information search for learning and study purposes our data show that the most popular searching tools in smartphone use are search engines (Table 5). No less than 79 % of study participants use search engines such as Google, Chrome or Safari to find information for learning purposes. It is interesting that LIS students search library catalog or electronic database with their smartphones less often. Some other studies on mobile learning also report low percentage of searching library resources with smartphones (Dresselhaus and Shrode, 2012; Bomhold, 2013) but they mostly involve students from various disciplines. Since our study focus on LIS students (future librarians), it would be expected that they would be familiar with all advantages of library resources and that they would use these sources in higher proportion. It is worth mentioning here that academic libraries in Hong Kong provide collections rich with electronic resources with user-friendly retrieval interface, which also includes mobile access.

Table 6. Using smartphone productivity tools and recording capabilities for learning

Types of activities	Daily/Weekly
Planning personal schedule (e.g. Google calendar, organizers)	55%
Making notes with note taking tools (e.g. Evernote)	45%
Creating documents (e.g. text, presentation, spreadsheets)	21%
Taking photos to record learning materials (e.g. book pages, slides)	45%
Audio recording presentations, seminars, interview, etc.	28%

More than half of LIS students (Table 6) use smartphones for planning their personal schedules and organizing daily and weekly activities, and a little less than half use note-taking tools. It can also be seen that only 21% create documents with their smartphones weekly or more often. This result does not surprise if we keep in mind how difficult is to type and edit on a small smartphone screen. Smartphone functionality of taking photos is often performed by 45% of respondents while audio recording is performed only by 28% of respondents. Video recording with smartphone is very rarely used for learning and study purposes so it is dropped from Table 6.

According to study findings a major barrier to smartphone use for learning is the small size of smartphone screen (Table 7). Small screen also makes it difficult to read academic papers and also to write and edit longer texts. Difficulties in interacting with a small screen are also reported in some other studies (Ku-

kulska-Hulme *et al*., 2011). Another barrier for many respondents is that webpages are not always formatted for smartphone. Slow load time also discourages our study participants from using smartphones for learning purposes. These barriers can to a certain extent explain LIS students' behavior. For instance, the low percentage of participants reading academic papers and e-books on their smartphones can be easily explained by difficult reading from small screen.

Table 7. Barriers to smartphone use for learning

Types of barriers	Medium/high barrier
Screen size is too small	87%
Reading is difficult	82%
Typing is difficult	72%
Web page is not formatted for smartphone	86%
Load time is slow	72%

Conclusion

The results of this study demonstrate that LIS students from Hong Kong extensively use smartphone applications for their day-to-day needs. They use smartphones for communication and socializing, for finding information, entertainment, leisure, and for managing their daily and weekly activities.

LIS students also use their smartphones for learning and study purposes. They often use smartphones for browsing and reading study related materials found on the Internet and watch video clips from social media websites. However, smartphones are used less frequently for reading academic resources such as academic journals and e-books. Therefore, it can be concluded that LIS students are less inclined to read lengthy and more complex materials with their smartphones.

It further transpires that smartphones are commonly used for communication with classmates and discussing study relates issues, such as course group activities and assignments. Therefore, smartphones can be considered as facilitators of collaborative learning.

When it comes to retrieving and accessing information for academic purposes LIS students still prefer search engines to library resources. It is surprising that even future librarians underuse library resources although they are aware that subscription databases are more reliable resources for academic literature than search engines.

Major barriers to smartphone use for academic learning reported by study participants are small screen which makes it difficult to read and type, absence of smartphone friendly webpages and too long loading time.

References

Ally, M.; Needham, G. M-libraries 5: From devices to people. London: Facet Pub, 2015.

Bradley, C.; Holley, D. Empirical research into students' mobile phones and their use for learning. // *International Journal of Mobile and Blended Learning*. 3(2011), 4; pp. 38-53

Bomhold, C.R. Educational use of smart phone technology: A survey of mobile phone application use by undergraduate university students. // *Program: electronic library and information systems*. 47(2013), 4; pp. 424 – 436

Cheung, R. Predicting user intentions for mobile learning in a project-based environment. // *International Journal of Electronic Commerce Studies.* 4(2014) 2; pp. 263-280

Dahlstrom, E.; Dziuban, C.; Walker, J. ECAR Study of undergraduate students and information technology 2013. Louisville, CO.: EDUCAUSE Center for Applied Research, 2013

Dahlstrom, E.; Dziuban, C.; Walker, J. ECAR Study of undergraduate students and information technology 2014. Louisville, CO.: EDUCAUSE Center for Applied Research, 2014

Dresselhaus, A.; Shrode, F. Mobile technologies & academics: do students use mobile technologies in their academic lives and are librarians ready to meet this challenge? // *Information Technology & Libraries*. 31 (2012), 2; pp.82-101

Dukic, Z.; Chiu, D. K.W.; Lo, P. "Learning with smartphones: higher education students' experiences and practices" in Needham, G.; Ally, M. (eds.). *M-Libraries 5: From devices to people*. London: Facet Publishing, 2015

Gikas, J.; Grant, M. Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. // *Internet & Higher Education*. 19(2013), pp. 18-26

Kim, J.; Ilon, L.; Altmann, J. Adapting smartphones as learning technology in a Korean University. // *Journal of Integrated Design and Process Science*. 17(2013) 1; pp. 5-16

Kukulska-Hulme, A.; Pettit, J.; Bradley, L.; Carvalho, A. A.; Herrington, A.; Kennedy, D. M.; Walker, A. Mature students using mobile devices in life and learning. // *International Journal of Mobile and Blended Learning*, 3 (2011), 1; pp. 18–52

Lippincott, J. K. A mobile future for academic libraries. // *Reference Services Review, 38(2010),* 2; pp. 205 - 213

Liu, M.; Geurtz, R.; Karam, A.; Navarrete, C.; Scordino, R. "Research on mobile learning in adult education" in W. K. S. Marshall (ed.). *On the move: mobile learning for development*. Charlotte, NC: Information Age Publishing, Inc., 2013.

Rambe, P.; Bere, A. "Using social embeddedness to explore ubiquitous learning in mobile environments at a South African University of Technology" in Ivala*, E. (ed.) Proceedings of the 8th International Conference on e-Learning: ICEL 2013*. Cape Town, South Africa: Academic Conferences and Publishing International Limited, 2013

Original scientific paper

Developing reading skills and motivation through mobile phones

Jasminka Pernjek
High school Krapina / Postgraduate study in glottodidactics, Faculty of Humanities and Social Sciences, University of Zagreb
Šetalište hrvatskog narodnog preporoda 6, Krapina, Croatia
jpernjek@gmail.com

Monika Habjanec
High school Krapina / Polytechnic Croatian Zagorje Krapina
Šetalište hrvatskog narodnog preporoda 6, Krapina, Croatia
monika.habjanec@gmail.com

Summary

Continuous development of information and communication technologies has opened the door to a wide range of opportunities for digital learning. Recent didactic concepts, such as student-centered teaching stress the need for learner autonomy, which includes non-formal learning aided by modern technologies such as cell phones. Mobile phones have become an essential part of the teenager's life. If incorporated in the learning process, they could have a positive effect on student's motivation, which in turn would lead to improved achievements. The aim of this study was to determine the attitudes of Croatian students of English and German towards the use of modern technologies in teaching foreign languages. The study evaluates the students' attitudes towards reading comprehension in a foreign language when using traditional methods (multiple choice questions and paper-pencil C-Test method) as opposed to solving tasks with the use of mobile technology. The results of the study show that, although students find online testing more interesting and less stressful, they still prefer to be tested in the classical way.

Keywords: m-learning, Computer-Assisted Language Learning, Learner Perceptions, Learning Experiences

Introduction

Continuous development of information and communication technologies has opened the door to a wide range of opportunities for digital learning. Recent didactic concepts, such as student-centered teaching, involving a focus on both the interests and the needs of students, and learner autonomy, involving a focus on students' responsibility for their own learning, have created space for the

incorporation of non-formal and informal learning and the use of modern technologies in the classroom discourse.

Non-formal and informal education, which is defined as any relatively structured learning outside of the educational system, contributes significantly to the enrichment of the learning experience, as recognized by the European Commission in the White Paper from November 1995 (224). This document clearly draws attention to the importance of the competences acquired outside of the formal education institutions, with a view to, among other things, facilitating the recognition of rich experiential knowledge and practical skills and incorporating that knowledge and skills in the concept of lifelong learning. With the purpose of fostering non-formal learning, education institutions have assumed new responsibilities, e.g. to enable the acquisition of knowledge in real life situations through projects, out-of-school activities aimed at the exploration of the immediate environment, involvement of experts and practitioners into teaching, introduction of students to the principles of learning autonomy and self-regulated learning, involving independent research and collection of information as well as a critical review and presentation of the results. A study on non-formal learning conducted by the German Youth Institute (*Deutches Jugendinstitut*, DJI) has revealed a significant difference between formal and informal learning with respect to the acquired content (Lipinski, 2000). The relevance to one's life and experience makes informal learning much more appealing than learning for the sake of learning, which is far too often the case in formal education. Informal learning is not a process regulated by a teacher, but rather a situational process regulated by the student. Therefore, information communications technologies, which provide access to any content at any given time, have become very important in this context.

A good example of an opportunity for informal learning is the use of computers and mobile phones. Mobile phones have become an essential part of the teenager's life. If incorporated in the learning process, they could have a positive effect on student's motivation, which in turn would lead to improved achievements. The role of an instant feedback should also not be neglected. The m-learning opportunities have been discussed for a long time. However, whereas ten years ago the central question was how to make certain contents available on mobile phones, the discussion today is aimed at the integration and orchestration of mobile technologies into the models supporting ubiquitous learning.

Theoretical Background

M-learning is a term derived from e-learning. It refers to any learning with the help of portable media, such as Smart Devices, by using Apps for learning on mobile phones, taking place at any time and any location. O'Malley et al. (2003) define e-learning as any type of learning that occurs when the student is not at a fixed, predetermined location or any learning involving the utilisation of learning opportunities provided by mobile technology. M-learning is a type of u-

learning which Sakamura et al. (*in* Roreger and Schmidt, 2012) defines as a learning style treating learning as a process that takes place anywhere and anytime with the help of ubiquitous computing, allowing adaptability to different contexts so the size and the content of teaching learning unit can be adapted to the learner. Petrova and Li (2009) define mLearning as *an ubiquitous learning activity supported by the appropriate mobile technology and pedagogical approach.*The difference between m-learning and u-learning presents itself in the fact that u-learning provides the learning material matching the situation which the learner is in.

An advantage of m-learning is the fact we can learn at any given time and at any place. However, due to technical restrictions it requires a good didactic preparation of material intended for m-learning which involves breaking down the learning content into small chunks (nuggets) which makes them ideal for, e.g. learning vocabulary.

Roreger and Schmidt (2012) put an accent on information exchange and group interviews as the basic mechanisms of active learning process, which are mainly intrinsically motivated and spontaneous. Creating groups on social networks associated with a given content element has a significant potential for successful distance learning. Since mobile devices are today used as a natural means of communication and entertainment, and their users have adopted the concept of constant presence, availability and access of the information on the Internet, they replace the traditional knowledge acquisition and memorization of information so it can be invoked and used in the future with the paradigm of the acquisition of skills to access the information constantly available on the Internet. Modern teaching methods also show a positive attitude towards mobile learning. The Situated Learning Theory (SLT - Collins, Brown & Newman, 1989; Burston, 2011) states that the best type of learning is the one that occurs unintentionally and takes place during authentic activities, context and culture. While it is still perhaps too early to assess the effectiveness of utilization of mobile phones for learning, the didactic concept of student-centered learning undoubtedly constitutes a solid foundation for the development of a mobile learning program.

Mobile learning is a didactic reaction to a change in the learning habits and in the media competences of children and youth, to whom mobile phone is a ubiquitous communication device that offers access to a variety of media as well as a variety of didactic and scientific content. Although m-Learning brings a number of challenges to the educational environment, integrating mobile phones into school teaching and developing new forms of learning could be a motivating factor which would in turn have positive effects on other types of learning. Additionally, mobile learning is a contribution towards the growing significance of informal, ubiquitous learning taking place outside of school. The stages of learning that are moderated by the teacher can be interchanged with the stages

of mobile-supported activities (e.g. the use of dictionaries, searching databases available on the Internet, making short video films, etc.).

Previous research

According to Heift and Chapelle (2012), the idea of using the possibilities offered by technology in language teaching has been around since the 1960s, initially focusing on enhancing the learning process with the use of computer-assisted instruction, while today's research focuses mainly on determining what, why, how and to what extent do new technologies lead to successful learning outcomes. Over the years, computer-assisted language learning (CALL) has developed on the principles of communicative language teaching. It is believed that with the use of CALL students develop learning strategies and improve their intercultural competence by using dictionaries, reading newspapers, watching contents on YouTube, and engaging in online forums. A rapid development of technology has made learning possible at any time and in any place (u-learning). The use of appropriate learning strategies and the development of effective learning activities have a crucial role in the context of CALL. A study conducted by Jung (2014) provided an insight into the views and expectations of English language learners (ELL) with respect to u-learning, showing a significant effect that u-learning has on the learner's motivation.

The development of educational technology offers possibilities for autonomous learning as a supplement to conventional classroom learning by enabling students to use many sources of information for independent learning and helping students take responsibility for learning by encouraging them to look for appropriate material and plan their own progress. The use of online communication technology allows a higher degree of independence in learning a foreign language because students can maximize their opportunities for practice.

Mobile phones are increasingly used in learning vocabulary, grammar and developing speaking and listening skills, as shown by a number of studies. The results indicate that mobile phones could provide an alternative source for learning vocabulary (Chen & Chung, 2008; Hu, 2013; Lu, 2008; Stockwell, 2010). In one study (Lu, 2008) students were required to learn two sets of words by the use of mobile phones or in a traditional way, on paper. The students who studied using text messages showed a better understanding of vocabulary than those who were presented with the vocabulary on paper. Liaw (2006) shows that e-learning is not restricted only to learning vocabulary, but can also be applied in other learning contexts, e.g. to enhance intercultural competence by reading about your culture and sharing the new findings with speakers of another culture.

Adapting to the use of digital interfaces also means that the modes of learning and the perception of classroom material is changing, with smaller tasks, such as mini-essays and grammar quizzes becoming more suitable for a better learning experience with the use of mobile phones. Krashen (1989) suggested that

the acquisition is enhanced if learning occurs in understandable and easily manageable sequences. With this in mind, Wong and Smith (2013) started research in 2009 where students were given small modular parts of classroom material for reading and learning grammar, which turned out to be a good approach in promoting m-learning.

Mobile hardware is evolving rapidly, enabling high-speed data processing and with the size and the resolution of the screen and the storage capacity increasing. Mobile phones are packed with a rich assortment of features, such as photo and video cameras, QR code reader, voice recorder, MP3 / MP4 players, internet access and the respective numerous services and social networks. As such, they are, in terms of technical characteristics, becoming comparable to personal computers. Smart phones can be easily connected to wireless networks which provide users with much greater flexibility. Some studies have shown that m-learning has become popular among students in Korea (Park, 2012) and the United States, where the number of students opting for m-learning is on the rise, and that the majority of students in Japan prefer to download teaching materials on their mobile phones (Wang and Smith, 2013). M-learning still has certain drawbacks in comparison to e-learning, either due to technical restrictions such as small screens or the unwillingness of teachers to apply the new technologies in teaching (Franklin, 2011; Wang and Higgins, 2006.). Despite the unfavorable economic conditions in the country, the use of smart phones has become a rule rather than an exception in Croatia. Nonetheless, a study on the use of mobile devices for educational purposes in high school population has still not been conducted in Croatia. It is often the case that mobile phone use is even prohibited in the classroom.

Research questions

The aim of this study was to determine the attitudes of Croatian students of English and German towards the use of modern technologies in teaching foreign languages. The study evaluates the students' attitudes towards reading comprehension in a foreign language when using traditional methods (multiple choice questions and paper-pencil C-Test method) as opposed to solving tasks with the use of mobile technology. It also examines the level of satisfaction with ubiquitous learning and mobile learning. The question being raised is whether the characteristics of ubiquitous learning, such as the ubiquity, interactivity and autonomy, will affect the satisfaction in learning and motivation as well as the learning results.

Research Question1: Are students ready to use their mobile devices to connect, learn and create new learning experiences?

Research question 2: Do gender and type of school affect the readiness to use mobile technologies for learning purposes?

It was hypothesised that learning motivation could be improved by using online materials. User satisfaction with m-learning was presumed to be a desirable

outcome of the study. The increased motivation due to m-learning could be an effective method for promoting reading comprehension.

Research methodology

Target population and data collection

The participants of this study were students of English and German language learning the language as their first foreign language in a high school near Zagreb. The subjects comprised 40 high school students aged 15 and 16 (n=40, m=23, f=17).

Table 1: Demographic Characteristics

Demographic Categories		Frequency	Percentage
Age	15	26	60.5
	16	17	39.5
Gender	Male	24	55.8
	Female	19	44.2
Foreign language	English	21	48.8
	German	22	51.2
Type of school	Grammar school	21	48.8
	Vocational school	22	51.2

Data were collected in 4 steps. Four instruments were used to collect data: a pre-study questionnaire, two reading comprehension tests and a post-study questionnaire. All participants were required to complete the pre-study questionnaire, which gathered information on the students' views on reading and on their use of the ICT.

After finishing the online test, all students were required to complete the post-study questionnaire, which examined the students' perception of the use of mobile devices in education and their willingness to adopt new technologies.

The study examined the subjects' perception of the use of technology for educational purposes using two questionnaires and two reading comprehension tests in English and in German. The first questionnaire collected information on the perception of reading, the reading habits, and the ways of using technology in leisure time and in the school context. The students were also required to give an estimate of the extent to which technology is used for education purposes in their school.

The subjects were asked to indicate the extent to which they agreed with the items on a Likert scale, with the options ranging from "strongly disagree" to "strongly agree", and from "never" to "always". After that, two reading comprehension tests were conducted. The instrument used were standardized tests of the basic A2 level of the state graduation exam, which is also the level of knowledge required by the curriculum from students in the first year of the secondary education.

The first of the two tests was carried out as a conventional paper-pencil test during the same class period in which the students filled out the questionnaire A, while the second test was to be filled out using one's mobile phone within 24-hours from accessing the test. The access was enabled via an application which was developed for the purpose of the study and made available for download on the school website. It was protected by a password and available only to the subjects of the study. The students had to answer all the questions before they could receive feedback. They were encouraged to study the feedback, re-read the material, identify errors themselves and then change any answers which they thought were incorrect. After the second test, the students filled out the questionnaire 2 in the next class period. The questionnaire examined the students' perception of completing the test in the conventional way as opposed to the online test as well as the motivation for the use of mobile technology in learning and how useful did they find the received feedback for the process of learning.

The study was more of an exploratory than confirmatory nature, since the subjects used their mobile phones to complete a test for school purposes for the first time. The chosen type of interaction for this study was the one in which the student's performance is followed by a computer-generated feedback. Such feedback came from instructional materials containing explicit exercises aimed at providing learners with practice in particular areas of reading, grammar and vocabulary. The learners were provided with immediate feedback on the correctness of their responses to the questions.

Results

The study investigated the students' perception of the use of mobile learning in education. The SPSS program was used in the data analysis. The analysis of the collected data showed that online reading and solving a test using mobile phones is considered to be a positive experience. The enjoyment of using a mobile device had a positive effect on the students' attitudes and improved learning motivation can improve performance.

The analysis of data gathered from subjects, shown in Figure 1 showed that only a small part of the subjects state that reading is their favorite activity. In relation to that, most of them read a maximum of 60 minutes a day and one book a month.

Unlike reading, the Internet is a favorite activity. Students spend 1 or 2 hours online every day, without significant differences between working days and weekends, except in the category of more than four hours a day, as evident in Figure 2.

After the online test, students' attitudes towards the type of test were expressed by the second questionnaire. Although most participants considered the online test more interesting and less stressful, as evident in Figure 3, they would still prefer to write the next test in the classical way.

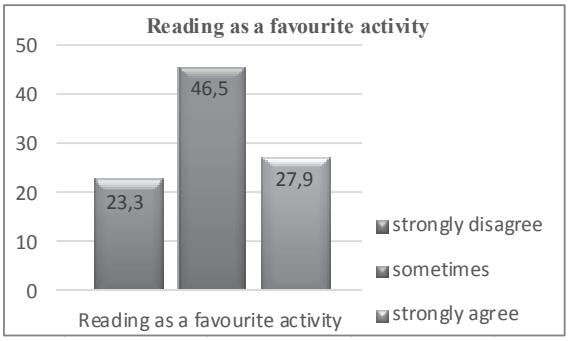


Figure 1: Reading as a favorite activity

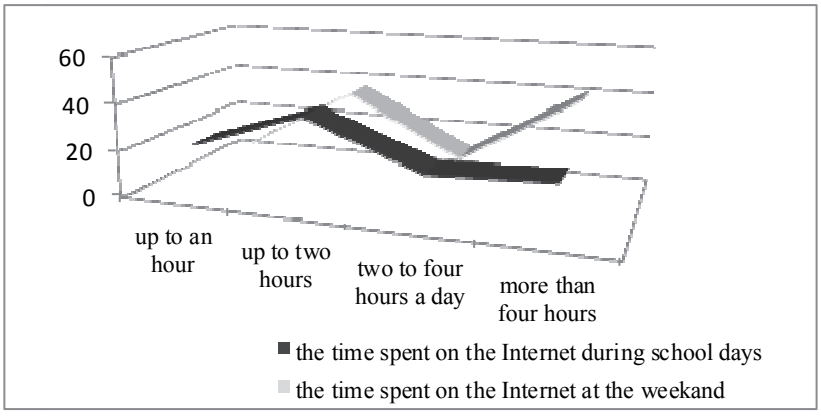


Figure 2: Time spent on the Internet

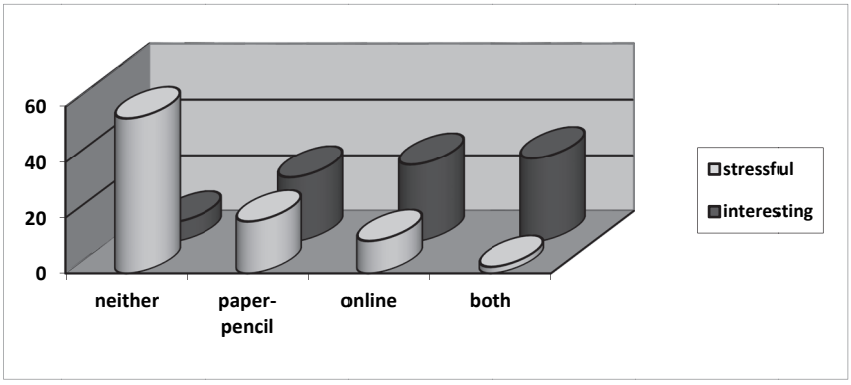


Figure 3: Attitudes towards the type of test

We received a positive response to the first research question. The students were motivated to solve the online test so it is somewhat surprising that in the future they would prefer to write tests in paper-pencil method.
The response to the second research question shows a significant difference between students of different gender and between students attending a different type of school. Female students mainly opt for the grammar school, while male students opt for the vocational school. A significant difference has also been identified between students of different gender and between students attending a different type of school with respect to reading, which is a far more popular activity among grammar school students, i.e. female students. However, no statistically significant correlation has been found between these two variables (gender and the type of school) and the age at which students first used computers and the attitudes towards the use of computers and the Internet for school and learning. A statistically significant difference has been found between the type of school and the attitude towards future testing: although grammar school students found the online test more interesting and less stressful, they would still prefer to take the next test in the paper-pencil method, unlike vocational school students who would prefer to be tested online.

Table 2: Correlations between Constructs

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
(1) Gender	1						
(2) Type of the school	-,682**	1					
(3) Reading as a favorite activity	,468**	-,453**	1				
(4) The age of the first Internet use	,142	-,143	,373*	1			
(5) Attitudes about Internet for school	,047	-,120	,090	-,265	1		
(6) Online test was more interesting	-,091	,041	,065	,095	-,103	1	
(7) I want to take the next test online	-,283	,351*	,214	-,002	-,172	,395	1

**:p<0.05; **:p<0.01*

Conclusion

The findings of this study suggest that mobile technologies have the potential to provide new learning experiences. Students can engage in learning activities outside the classroom, at any time and any location. Although many teachers already use technology in their classes, the results have shown that the ICT is not used often, even in well-equipped schools (in only 40 out of 1800 class periods held in a school year). Mobile learning requires a certain level of technological knowledge, which could be the reason for a lack of readiness of teachers to adopt this method of learning and teaching. Although the use of mobile devices is ubiquitous, the students have not shown readiness for mobile learning.

The reason for this could be the fact that, in comparison to PC's, mobile phone screen size is much smaller.

Limitations and future work

The number of subjects was small (N=43), so the findings cannot be generalized. For further research the student sample should be gathered from different schools and possibly compared with the fourth class to see whether it the age affects the attitudes about m-learning. It would also be interesting to examine whether students' readiness for mobile learning increases if, instead in testing, mobile devices are used, in group work and other collaborative activities which provide opportunities for student interaction so the communication features of smartphones could be used in the best possible way.

References

Chen, C.-M., Hsu, S.-H. Personalized Intelligent Mobile Learning System for Supporting Effective English Learning. //*Educational Technology & Society*, 11(2008), 3; 153-180

Dohmen, G. Das informelle Lernen. Die internationale Erschließung einer bisher vernachlässigten Grundform menschlichen Lernens für das lebenslange Lernen aller. Bonn: Bundesministerium für Bildung und Forschung, 2001

Europäische Kommision: Weißbuch: Lehren und Lernen. Auf dem Weg zur kognitiven Gesellschaft, Brüssel, 1995

Franklin, T. MOBILE LEARNING: AT THE TIPPING POINT. // The Turkish Online Journal of Educational Technology, Vol.10 (2011), 4: 261-275

Heift, T. and Chapelle, C.A. Language learning through technology. // Gass, S.M. and Mackey, A. (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp.555 – 569). New York: Routledge, 2012

Hu, Z. Vocabulary Learning Assisted by Mobile Phones: Perceptions of Chinese Adult Learners. In *Journal of Cambridge Studies*, Vol. 8 (2013), No. 1; 139 - 154

Jung, H.-J. Ubiquitous Learning: Determinants impacting Learnerns' Satisfaction and Performance with Smartphones. // *Language Learning and Technology*, 18(2014), 3; 97-119

Kim, D., Rueckert, D., Kim, D.-J., Seo, D. Students' perceptions and experiences of mobile learning. // *Language Learning and Technology*, Vol. 17 (2013), 3; 52-73.

Krashen, S. We acquire vocabulary and spelling by reading. Additional evidence for the input hypothesis. // *Modern Language Journal*, 73 (1989); 440-464

Liaw, M.- L. E-Learning and the Development of interculturale competence. // *Language Learning & Technology*, Vol. 10 (2006), 3; 49-63.

Lipinski, J. Lernen und Interesse. München Deutsches Jugendinstitut (DJI): *Informeles Lernen in der Freizeit. Erste Ergebnisse des Projekts "Lebenswelten als Lernwelten"*, 2000

Lu, M. Effectiveness of vocabulary learning via mobile phone. // *Journal of Computer Assisted Learning*, 24 (2008), 6; 515-525

O'Malley, C, Vavoula, G., Glew, J., Taylor, J., Sharples, M., Lefrere, P. (2003). Guidelines for learning / teaching / tutoring in a mobile environment. Mobilearn project deliverable. 2015-04-11. https://hal.archives-ouvertes.fr/hal-00696244/document

Park, S. Y., Nam, M.-W, Cha, S.-B. University students' behavioral intention to use mobile learning: Evaluating the technology acceptance model. // *British Journal of Educational Technology*, Vol. 43 (2012), 4; 592-605.

Petrova, K., Li, C. Focus and Setting in Mobile Learning Research: A Review of the Literature. // *Communications of the IBIMA, Vol.10* (2009); 219-226

Roreger, H., Schmidt, T.C. Mobiles Lernen im Zeitalter sozialer Medien. In Schulz, M. and Neusius, A. (Eds.), *Grundlagen der Weiterbildung*. Hergensweiler: ZIEL-Verlag, 2012

Specht, M., Ebner, M. Mobiles und ubituitëres Lernen. Technologien und didaktische Aspekte. In Ebner, M. and Schön, S. (Eds.), *Lehrbuch für Lernen un Lehren mit Technologien.* Berlin, 2011

Stockwell, G. Using mobile phones for vocabulary activities: Examining the effect of the platform. // *Language Learning & Technology*, Vol. 14 (2010), 2; 95 – 110

Wang, S., Higgins, M. Limitations of Mobile Phone Learning. // *The JALT CALL Journal*, Vol.2, (2006) No.1; 3-14

Wang, S., Smith, S. Reading and grammar learning through mobile phones. // *Language Learning and Technology*, 17 (2013), 3; 117-134

Original scientific paper

Why are Croatian Higher Education Institutions Present on Social Networks?

Kruno Golubić
University Computing Centre, University of Zagreb
Josipa Marohnića 5, Zagreb, Croatia
kruno.golubic@srce.hr

Mihaela Banek Zorica
Faculty of Humanities and Social Sciences, University of Zagreb
I. Lučića 3, Zagreb, Croatia
mbanek@ffzg.hr

Summary

Reasons why Croatian higher education institutions are present on social networks and their desired goals have not been researched yet. Previous research studies were mainly focused on the mere presence of institutions and the content they publish, but none of them looked into faculty's opinions (or opinions of persons who are in charge of the official presence of institutions on social networks) about the institutions' presence on social networks and if the outcomes of that presence were as expected. In this paper we will present results of a survey conducted in 2015. The goal of the survey was to find out what do persons that are in charge of the official presence of Croatian higher education institutions on social networks think about it and did it achieve the desired goals. When an institution is deciding whether it should establish presence on any social network or not, experience from other institutions is a valuable input. Therefore, in this paper, we will present some of the key findings on motivation and goal fulfillment of those institutions that have already been engaged on social networks.

Keywords: Croatian higher education, social networks, social network presence, marketing goals, benefits of presence on social networks

Introduction

With the constant prolongation of time that Internet users spend[1] on social networks daily, it seems that social networks can no longer be ignored if an institution wants to adequately present itself on the Internet. For quite some time, institutions' web sites have not been the only place for information dissemination.

[1] Bennett, Shea. 28% of Time Spent Online is Social Networking. January 27th 2015. http://www.adweek.com/socialtimes/time-spent-online/613474 (5th July 2015)

A number of research studies conducted during the last few years have shown that Croatian higher education institutions have increasingly started to accept social networks. There has also been a lot of research done about Croatian libraries and their presence on social networks[2]. Studies dealing with Croatian higher education institutions and their presence on social networks were focused either on the mere form of institutions' presence[3] on social networks, or the use of social networks in communication[4]. The research was mainly based on publicly available data published by institutions.

These research studies gathered valuable data about the institutions' presence, published content and communication, but none of the published papers focused on the reasons why institutions use social networks and whether they achieved their goals. When an institution is deciding about establishing its presence on social networks, experience from other institutions is a valuable input. In this paper we will present some of the key findings on motivation and goal fulfillment of those institutions that have already been engaged on social networks.

Research

This paper is a part of an ongoing research project started a few years ago with a goal to map the usage of social networks among Croatian higher education institutions[5]. The goal of this paper is to find out the reasons why higher education institutions are present on social networks and to provide an insight into institutions' views and opinions about social networks as well as the benefits that their usage brings them. Along with the institutions' motivation and goals, we also wanted to see if institutions are using any kind of marketing activities on social networks and are those activities achieving the desired results.

[2] Umreženi : društvene mreže i knjižnice u Hrvatskoj : zbornik radova / Šalamon-Cindori, B. (ed.). Zagreb : Nacionalna i sveučilišna knjižnica u Zagrebu, 2015. http://nsk.hr/umrezeni-download.php

[3] Golubić, Kruno; Lasić-Lazić, Jadranka. Analysis of On-line Survey about Need for Presence of Higher Education Institutions on Social Networks: a Step Towards Creation of Communication Strategy. // *CIT: Journal of computing and information technology*. 20 (2012) , 3; pp.189-194

[4] Delić, Alen; Grd, Petra; Gregurec, Iva. Analysis of Communication of Croatian Faculties through Facebook - Part I. // *In Proceedings of the ITI 2013 35th International Conference on Information Technology Interfaces* / Luzar-Stiffler, Vesna ; Jarec, Iva (ed.). Zagreb: Srce - University Computing Centre, University of Zagreb, 2013, pp. 43-48.

[5] Golubić, Kruno. The official presence of Croatian higher education institutions on social networks. // *In Proceedings of the 3rd International Conference The Future of Information Sciences (INFuture)* / Billeness, C ; Hemera, A. ; Mateljan V. ; Banek Zorica, M ; Stančić, H. ; Seljan, S. (ed.). Zagreb: Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2011, pp. 263-275.

Since this is a research about online activities, use of an online survey tool seemed like a logical choice. The survey was created using SurveyGizmo[6] online survey tool. The survey consisted out of 23 questions in total. They were spread across 9 pages.

The survey had five main sections:

- General data
- Presence on social networks
- Marketing activities and fulfilment of goals
- Communication over social networks
- Opinions and views about social networks

All of the questions were marked as required. Single and multiple choice questions were used across the survey. Open text fields were used only when additional information was needed from the participants. In the middle of the survey, branching was used to display additional questions to those institutions that use social networking for marketing activities. At the very end, there were three questions that consisted out of 25 statements and opinions. Participants had to select the degree of agreement or disagreement with those statements. For those last three questions, Likert scale[7] was used. The Likert scale allowed us to get a better insight into institutions' opinions and views on specific topics. Due to a large amount of gathered data, especially from the last three questions, not all results will be presented in this paper.

After the internal testing of the survey's form and logic, the next step was sending out invitations. The survey was sent out to those institutions already present on social networks and previously identified by another research[8]. The invitation with a link to the online survey was either sent via direct messaging feature on Facebook, or via e-mail. In total, 53 institutions received the invitation. The survey was active for one month, from 22nd April until 22nd May 2015.

Findings

General data

Out of the 53 invited institutions, 30 of them are publicly funded (54%) and 23 are privately funded (46%) institutions. During the aforementioned one month period, the survey was opened 33 times. Out of those 33 times, in 9 cases the

[6] More information about SurveyGizmo can be found at their web site at http://www.surveygizmo.com/.

[7] Brace, Ian. Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research. London, Philadelphia, Kogan Page, 2008, p. 73.

[8] Golubić, Kruno; Banek Zorica, Mihaela. Presence and Activity of Croatian Higher Education Institutions on Social Networking Sites. // Proceedings of the 25th Central European Conference on Information and Intelligent Systems / Hunjak, T.; Lovrenčić, S.; Tomičić, I. (ed.). Varaždin: Faculty of Organization and Informatics, University of Zagreb, 2014, pp. 102-107.

surveys were partially filled out and in 24 cases, the surveys were completed, which means that the survey's response rate was 45%. No subsequent reminder was sent to those participants who started filling out the survey, but did not finish it. All incomplete surveys were discarded (N=9). None of the completed surveys were duplicates.
We consider this to be a good response rate when compared[9] with some standards[10]. Only the data from 24 completed surveys was used for this research. Both publicly and privately funded institutions were represented with 12 responses each.
At the very beginning of the survey, participants were asked if they were in any way responsible for the institution's presence on social networks. Only those participants that answered "Yes" were allowed to fill out the survey, while the rest were disqualified. There were no disqualified participants, probably due to the direct targeting of the participants.

Content selection
Most of the surveys (N=22) were filed out by one of the institution's employees. None of the respondents were students. One respondent stated that he/she is an employee of an external agency in charge for the presence of the institution on social networks. The remaining respond was marked as "Other".
Creating content and curation for social networks, as well as communicating via social networks, can be a rather demanding tasks. We wanted to find out how many people are involved in those tasks (Table 1). As it can be seen from the results, one third of the institutions has only one person in charge of the content and the large majority of institutions think that this is not a task for only one person. This shows that institutions care about their presence on social networks and they want to be adequately presented.
Almost half of the institutions (Table 2) do not have clear rules on the selection process of content that is being published. This means that people in charge of publishing hold great responsibility regarding the selection of content. As it can be seen from the data, employees and students are contributing to the process of content gathering. We believe that this is an example of good practice. Higher education institutions have very diverse fields of interest within themselves. Inputs from multiple sides allow institutions to cover more fields of interest and consequentially, this brings them better reputation in the community.

[9] Stoop, Ineke A.L. The Hunt for the Last Respondent. Hague: Social and Cultural Planning Office, 2005, page 23

[10] Ficker Jr., Ronald Sampling methods for web and e-mail surveys. // *The SAGE Handbook of Online Research Methods* / Fielding N., Lee R., Blank G. (Ed.). London: SAGE Publications Ltd, 2008, pp. 195-216.

Table 1. Number of people in charge for the presence on social networks

Number of people in charge	Number of institutions
1	8
2	6
3 to 5	10
6 or more	0

Table 2. Rules and content guidelines

Existence of rules	Number of institutions
There are clear rules that define types of content to be published on social networks	13
Content is generated based on information published on institution's web page	21
Employees provide content for publishing	18
Students provide content for publishing	14
I don't know	0

Presence on social networks

According to the survey, all institutions have been present on social networks for at least two years. The longest running ones have been present for seven years (Table 3). If we look at the number of institutions that started using social networks in a particular year, we can see that 2012 was the year when the largest number of individual institution established its presence on social networks. We believe that these higher education institutions are conservative in their essence, and therefore they do not belong to early adopters of technologies and trends. They rather wait for maturity of trends and concepts before adopting them into their everyday routine.

Table 3. Presence of institution on social networks

Year	Number of institutions
2013	3
2012	7
2011	3
2010	3
2009	3
2008	1
I don't know	4

The presence on specific social networks is not the focus of this paper, but we believe that it is important to make a note which social networks were identified as the most popular ones among institutions. Participants were given a multiple choice question, a list of social networks identified by a previous research[11].

[11] Golubić, Kruno; Banek Zorica, Mihaela. Presence and Activity of Croatian Higher Education Institutions on Social Networking Sites. // Proceedings of the 25th Central European Conference on Information and Intelligent Systems / Hunjak, T.; Lovrenčić, S.; Tomičić, I. (ed.). Varaždin: Faculty of Organization and Informatics, University of Zagreb, 2014, pp. 102-107.

The fact that not one institution selected the option "Some other social network" shows that the list of social networks was appropriate for this research. The popularity of social networks is as expected. (Table 4).

Table 4. Presence on social networks

Social network	Number of institutions
Facebook	24
Twitter	16
YouTube	19
LinkedIn	12
Google+	7
Tumblr	0
Instagram	4
Flickr	1
Vimeo	6
I'm not sure	0
Some other social networks	0

Marketing activities and fulfilment of goals

According to the survey, 13 institutions have advertised on social networks. Privately funded institutions (N=9) are majority of institutions that have advertised on social networks. In the multiple choice question about advertising goals (Table 5), all institutions responded that attraction of students was one of their goals. From the other selected goals, it can be seen that institutions are eager about informing the general public about their work and accomplishments.

Table 5. Advertising goals

Advertising goal	Number of institutions
Attracting students to institution	13
Find associates for projects	3
Present work results of institution	9
Increase recognisability of institution	11
Increase visibility in local community	7
Other goals	2
I don't know	0

According to the answers, institutions are satisfied with their advertising activities (Table 6). Not one institution answered that no goals were achieved, and the majority of institutions were satisfied with the achievement of advertising goals. This shows us that advertising over social networks brings benefits to institutions.

Table 6. Fulfilment of advertising goals

Fulfilment of advertising goals	Number of institutions
All goals were accomplished	3
Larger portion of goals was accomplished	9
Smaller portion of goals was accomplished	1

Communication over social networks

Quick and easy communication is one of the key components and features of social networks. At this point, we were not interested in the communication with the general public, but rather in the communication with two groups of people that are tightly connected with institutions (Table 7). Those two groups are students and employees (including associates). A large majority of institutions (N=9) have stated that they do not use social networks for direct (i.e. private) communication with those two groups. Since this was a multiple choice question, we can see that some institutions use social networks to communicate directly with both groups.

Table 7. Use of social networks for communication

Private communication with certain group	Number of institutions
Communication with students	2
Communication with employees and associates	1
Communication with students, employees and associates	1
There is no private communication over social networks	9

All institutions (N=3) that communicate with students provide only general information, e.g. working hours of the registrar or the accounting service. None of them stated that they provide more specific information, such as exam schedule and results. In communication with employees, all institutions (N=2) provide only general information as well. In addition to those information, one of the institutions stated that they provide individual exam results to staff members via social networks.

According to the results, formal communication via social networks is the weakest link in the institutions' presence on social networks. We believe that the lack of formal communication lies in the fact that it is hard to prove someone's identity while communicating via social networks. This limits the communication mainly on providing general information, and in the rare cases when there is trust in someone's identity, more sensitive information and data is communicated.

Opinions and views on social networks

We believe that the support given by the institutions' management shows that the benefit of presence on social networks has been recognized and is considered to be a valuable asset. To find out more about the benefits of that presence for institutions, participants were asked to rate their views on several possible benefits. To rate their opinions and views, we have used the five point Likert scale (Table 8). Ratings on the scale were as following:

- Strongly agree (1)
- Agree (2)
- Neither agree nor disagree (3)

- Disagree (4)
- Strongly disagree (5)

Table 8. Opinions and views on social networks

Statement	Rating on Likert scale and number of responses				
	1	2	3	4	5
My institution has benefits from presence on social networks	16	6	2	0	0
My institution has official communication strategy	4	9	6	4	1
Management of my institution supports presence on social networks	13	8	3	0	0
Due to presence on social networks my institutions is more successful at attracting new students	5	14	5	0	0
Due to presence on social networks my institution is more successful at finding new associates for projects	2	5	14	3	0
Due to social networks it is easier for my institution to present result of its work	8	11	5	0	0
Due to presence on social networks my institution is more recognisable and visible in local community	6	12	5	1	0

Due to the fact that the survey was conducted among institutions that use social networks, it is no surprises that most opinions about social networks are positive ones. Only the statement about finding new associates for projects can be considered as neutral one.

It can be seen that institutions have benefited from presence on social networks in many different ways:

- Students are more easily attracted to the institution
- Institution is more visible in the community
- Presentation of work results is easier

Based on the answers, we can see that even if the institutions do not use social networks for advertising, they can gain the same benefits as those that do. All those benefits can be considered as building blocks, used to build up the institutions' on-line reputation. We believe that all benefits are tightly interlaced, e.g. institutions that present their results on-line are more visible and therefore they attract students more easily.

Conclusion

Institutions' managements are supporting presence on social networks and considerable human resources are dedicated to this task. Content published on social networks is gathered from different sources. Important role in content gathering belongs to the staff members and students that provide content for publishing. It is surprising that some institutions do not have an official communi-

cation strategy. Such strategy can be very useful to prevent unwanted situations, e.g. in situations when several persons are in charge of communication, a clear set of rules would make task handling much easier. We believe that a communication strategy would make a job of content handling much easier for persons in charge of it.

Almost all institutions are taking care of their own presence on social networks and only in rare cases this job is done by specialised agencies. Since institutions have been present on social networks for more than couple of years, and none of them have joined social networks in the last year, we consider social networks mature enough to be considered as an official communication channel. Institutions that are not already present on social networks should take steps towards social networks in the near future, if they also want to benefit from them, for example, attract new students, achieve better visibility of the institution and profit from it.

Presence on social networks does not mean that marketing activities are the imperative, but they are a welcoming addition for achieving the desired goals. Due to a large number of users, social networks make an excellent platform for marketing activities. Many institutions have already recognised this and they are satisfied with the results of their ongoing marketing activities.

There are many benefits for institutions that use social networks. Those benefits are not limited only to institutions that use social networks for marketing activities. Social networks make the presentation of results easier for institutions. They are also more visible in community and better at attracting new students.

Lack of direct communication over social networks can be explained by the fact that, due to the nature of social networks, it is not always possible to prove someone's identity. This limits communication mostly to the exchange of general information, while private date is omitted and not being transferred. We believe this will not change in the near future.

Social networks cannot be considered as a passing trend. As time passes by, they are becoming more and more important as a medium for information dissemination. This means that institutions need to keep up with the latest trends in this ever-changing landscape of Internet. Not all trends should be accepted, but after careful consideration, focus should be put on the most important ones.

References

Bennett, Shea. 28% of Time Spent Online is Social Networking. January 27th 2015. http://www.adweek.com/socialtimes/time-spent-online/613474 (5th July 2015)

Brace, Ian. Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research. London, Philadelphia: Kogan Page, 2008.

Delić, Alen; Grd, Petra; Gregurec, Iva. Analysis of Communication of Croatian Faculties through Facebook - Part I. // *In Proceedings of the ITI 2013 35th International Conference on Information Technology Interfaces* / Luzar-Stiffler, Vesna ; Jarec, Iva (ed.). Zagreb: Srce - University Computing Centre, University of Zagreb, 2013, pp. 43-48.

Ficker Jr., Ronald Sampling methods for web and e-mail surveys. // *The SAGE Handbook of Online Research Methods* / Fielding N., Lee R., Blank G. (Ed.). London: SAGE Publications Ltd, 2008, pp. 195-216.

Golubić, Kruno. The official presence of Croatian higher education institutions on social networks. // *In Proceedings of the 3rd International Conference The Future of Information Sciences (INFuture)* / Billeness, C ; Hemera, A. ; Mateljan V. ; Banek Zorica, M ; Stančić, H. ; Seljan, S. (ed.). Zagreb: Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2011, pp. 263-275.

Golubić, Kruno; Banek Zorica, Mihaela. Presence and Activity of Croatian Higher Education Institutions on Social Networking Sites. // *Proceedings of the 25th Central European Conference on Information and Intelligent Systems* / Hunjak, T.; Lovrenčić, S.; Tomičić, I. (ed.). Varaždin: Faculty of Organization and Informatics, University of Zagreb, 2014, pp. 102-107.

Golubić, Kruno; Lasić-Lazić, Jadranka. Analysis of On-line Survey about Need for Presence of Higher Education Institutions on Social Networks: a Step Towards Creation of Communication Strategy. // *CIT: Journal of computing and information technology*. 20 (2012), 3; pp.189-194

Stoop, Ineke A.L. The Hunt for the Last Respondent. Hague: Social and Cultural Planning Office, 2005.

Umreženi : društvene mreže i knjižnice u Hrvatskoj : zbornik radova / Šalamon-Cindori, B. (ed.). Zagreb : Nacionalna i sveučilišna knjižnica u Zagrebu, 2015. http://nsk.hr/umrezeni-download.php (8th October 2015)

Original scientific paper

Croatian ESP vocational high school pupils and university students' habits of using online sources

Brankica Bošnjak Terzić
Technical School Ruđer Bošković
Getaldićeva 4, Zagreb, Croatia
Department of Technical Foreign Languages, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb
Ivana Lučića 5, Zagreb, Croatia
brankica.bosnjakterzic@inet.hr

Ana Gabrijela Blažević
Department of Romance Studies,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
agblazev@ffzg.hr; anablazevic@gmail.com

Summary

Information literacy as a set of abilities needed to recognize, locate, evaluate and effectively use needed information is extremely important in education, especially in English for Technical Purposes classes due to their specific content and the rapid technological changes connected to it. For this reason, this research has been conducted trying to determine Croatian English for Specific Purposes (ESP) vocational high school pupils and university students' habits of using online sources in ESP classes and the frequency of using online sources in translation. Furthermore, the aim was to get insight into their IL background and expectations about being educated on IL by ESP teachers. The results of the study presented in this paper show that Croatian ESP vocational school pupils and university students have quite high perception of their abilities regarding their IL but their usage of online sources for ESP needs has proven to be low and not specific enough for a meaningful learning.

Keywords: information literacy, Croatian ESP pupils and students, online sources, translation

Introduction

Information literacy (IL) as a set of abilities needed to recognize, locate, evaluate and effectively use needed information is a prerequisite for the use of modern information services (Lasić-Lazić & László, 2014). There are numerous studies showing that the Internet is used increasingly for educational purposes (Lubens, 1998, 1999; Jones & Madden, 2002) and it is becoming the main

information source (Browne, Freeman & Williamson, 2000; Paris, 2002), but often without any formal training in IL (Fidel et. al., 1999; Grimer & Boening, 2001; Jones & Madden, 2002; Metzger, Flanagin & Zwarun, 2003, Frazer, 2010).

Studies about vocational high school pupils and university students' IL suggest they prefer online sources over printed ones, but also that they have difficulties evaluating found information and they have poor searching skills (for a literature review, see Julien & Barker, 2009). Although most of them started to use the Internet early and are heavy and capable users of it (Jones & Madden, 2002), their processing of information is superficial and aims at the quantity of the data and not the quality and critical explanation which are bases for meaningful, deep and true learning (Lubans, 1999; Metzger, Flanagin & Zwarun, 2003). That represents a gap regarding teachers' expectations of the used sources (Grimer & Boening, 2001; Lasić-Lazić, Špiranec & Banek Zorica; 2012) and has important implications on students' academic work, learning and professional development since the Internet contains information of very diverse quality (Špiranec, 2004) and even biased, manipulated or inaccurate information (Alexander & Tate, 1999; Flanagin & Metzger, 2000; Metzger, Flanagin & Zwarun, 2003).

Due to the all above mentioned, many researchers consider IL an educational imperative essential for lifelong learning (Lau, 2006; Špiranec & Banek Zorica, 2008). That is why many educators are incorporating it into their curricula, that is, in a real context and with concrete students' needs, which helps and facilitates the learning of the information skills (Eisenberg, 2008). For example, Seljan (2011) in her research of evaluating online language resources and tools for Croatian language states that Croatian students are very interested in the use of translation technology, especially in translating domain specific texts (like ESP), but a further research also shows that translations performed by Google Translate from English to Croatian are of lower grade than the ones in opposite direction (Seljan, Brkić & Kucis, 2011), which is a fact students should be aware of.

Since Technical English for Specific Purposes (ESP) classes deal with very specific content prone to rapid change and development, internet sources are of crucial importance for a quality education. This research was conducted among Croatian ESP vocational high school pupils and university students with the aim of getting some insight into their IL as a very important reference point for improving ESP teaching.

Purpose of the Study

The aim of this study was to determine Croatian pupils and students' ESP habits and beliefs about using online sources for educational purposes, frequency of used online sources when translating ESP texts and their background on IL

education. Furthermore, the aim was to get insight into their expectations about being educated on IL by ESP teachers.

Hypothesis

In regards to the aim of the research, the study hypothesizes the following:

a) Croatian pupils and students are heavy internet users, both for general and educational purposes.
b) Most participants didn't have any previous IL training, but still are aware of the importance of being educated about IL.
c) The main criteria for choosing an online source is the ranking on the search engine and the web page design.
d) One of the most frequent sources used in translation among ESP pupils and students is Google translate.
e) University students will use more specialized online sources than vocational high school pupils in translation process.

Methodology

Participants

The research included 63 participants (N=63), 30 pupils (all male) and 33 students (N=33; 22 males and 11 females). Pupils (age 17) attended the third grade at Technical School Ruđer Bošković studying ESP for the first time. Two classes which participated in the research were Technicians for Mechatronics. Approval from the school's principal was gained and all parents signed the approval consent for the research.

The students (age 20-21) were second year students at the Faculty of Mechanical Engineering and Naval Architecture. Technical English is taught at the second and the third year at the Faculty. Out of 33 students; 22 were studying ESP for the first time in their education, 7 studied for 2 years and 3 for three years.

For both groups the average number of learning English was 10.51 years (SD=1.6, Mo=11). Conducted t-test confirmed that there was no statistical difference between both groups regarding the duration of learning English.

Instrument

The instrument consisted of two parts. The first part included ESP text participants had to translate. Each group got different text according to their language level while the second part was the questionnaire, same for the both groups. The text for pupils was intended for the intermediate level of language knowledge, covering the basics of *Pneumatics*. The text for students was intended for the upper-intermediate level of language knowledge, covering the basics of the *Gasification Process*. Participants were sent texts that had to be translated from English to Croatian using only online sources. Teachers didn't suggest which online sources they could use.

The second part included a questionnaire in Croatian, which consisted of 19 questions of different types: multiple choices, Likert scale questions, short answers, and yes/no questions. The questionnaire was divided in 3 parts. The first part of the questionnaire included information about students and pupils' education and their beliefs in regards to using online sources. The second part captured insight into students and pupils' IL. The last part covered the frequency of the internet and online sources use for ESP classes and classes in general.

Results[1] and discussion

General and educational internet use

Out of 63 participants, 62 (98%) participants answered that they had internet access. About the time they spend on the computer, out of 63 participants in the study, 24 (38%) use computer between 1 and 3 hours, 20 (32%) use computers for more than 3 hours, 12 (19%) use it for an hour and only 7 (11%) students use computer 30 minutes a day.

The answers to the questions how often they use the Internet for general educational purposes and for ESP classes are shown in the Table 1.

Table 1. Usage of the Internet for educational purposes

Usage of the Internet for...	...general educational purposes	...ESP
Never	0	7
Rarely	7	28
Sometimes	25	3
Often/very often	31	13

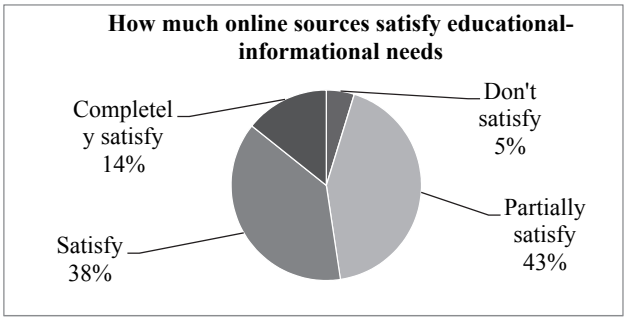
It's interesting to note that despite almost half of the participants (N=31, 49%) stated they used the Internet often and very often for educational purposes, and slightly less than half (N=25, 40%) used it sometimes (altogether 89%), the use of the Internet for ESP classes dropped dramatically and more than half of the participants (N=35, 56%) used it rarely or never.

This supports the first hypothesis, but since the characteristics of Technical LSP (Language for Specific Purposes) impose the Internet as an essential tool, the gap between the use of the Internet for general educational purposes and ESP also shows that the possible benefits of the Internet usage in ESP classes are not properly exploited yet, which leads to the conclusion that there is still much to do in that area.

[1] A t-test was conducted for every item to see if there was any statistically significant difference between pupils and students. Since the significant difference was established only in one item, this chapter will present the results of both groups together, only indicating their difference when it was found statistically significant.

Participants were also asked to grade their satisfaction with online sources in regards to their educational-informational needs. Responses are shown in Chart 1.

Chart 1. Satisfaction with online sources for educational-informational needs



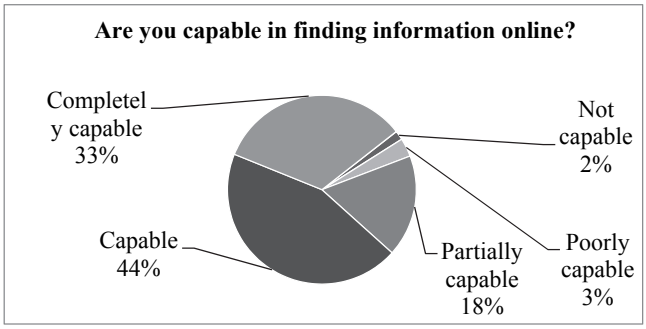
Since only 3 (5%) of the participants said that online sources didn't fulfil their educational-informational needs, the results indicate one more time the well-known potential of the Internet for educational purposes, which is corroborated with the fact that 33 participants (52%) said sources satisfied or completely satisfied their needs. Still, it is important not to disregard 27 participants (43%), who stated that online sources only partially satisfied their educational-informational needs. This could mean two things: either they also value traditional sources or they are not finding on the Internet everything they need, or probably, the combination of both. Knowing the importance of IL for getting the best of the Web without being overwhelmed by the amount of available information, the next aim of the questionnaire was to check participants' background and basic knowledge on IL.

Information literacy: background and awareness

When questioned about their previous formal education regarding informational literacy, only 8 (13%) of the participants said they had some instructions on IL (N=6 in high school, N=2 in elementary school), which confirmed the first part of the second hypothesis of the research. Despite that, when participants were asked to assess their capability to find information on the Internet on the scale from 1 (I am not capable) to 5 (I am completely capable), they assessed their capability quite high, as it's shown in the Chart 2.

Participants rated their capability with the average score 4.05 (Mo=4, SD =.888). While 28 participants (44%) assessed to be capable (4), 21 (33%) assessed themselves as extremely capable (5) and 11 participants (18%) as partially capable (3). The lowest assessment (2) was given by two participants (3%) and only one participant (2%) assessed himself as not capable (1).

Chart 2. Capability in finding information online



Accordingly, to the question 'Is it difficult to find needed information on the Internet*?'* 50 students (79%) answered that it was easy while 11 (18%) claimed that *sometimes* it was difficult. Only 2 students (3%) found it difficult. The given explanations were that difficulty depended on the content they were looking for.

Although a very high percentage (77%) of the participants found that they were capable or completely capable to find information on the Internet, still 65% (N=41) of all participants thought that education in IL was necessary, while 17.5% (N=11) thought that education was not necessary and the same number that they weren't sure, which supports the second part of the second hypothesis.

When asked whether their language teacher should include IL training as part of his/her ESP classes several arguments were given. 29 (46%) participants agreed that ESP teachers should give some IL education and they supported the argument with the following: a language teacher knows what information is correct and where to find it; that kind of training would help them translate specialized text easier and would improve their learning. A total of 15 (23.8%) participants said that a language teacher shouldn't do that since that wasn't his/her role, while 19 (30.1%) participants weren't sure about the role of an ESP language teacher regarding IL.

Regarding the search for information, 28 (45%) of the participants said they looked for the needed information until they found it, 21 (33%) claimed they only checked the results that appeared on the first page of the search engine and 7 (11%) that they looked at the results on the first three pages. Surprisingly, there were 7 (11%) participants who stated they only looked at the first result of the search engine.

To the question whether they checked the information in more sources, altogether 24 (38%) participants stated that they checked the information in more sources, while 21 (33%) said that they never did that. Interestingly, 18 (29%) participants stated that they *sometimes* checked information supporting their arguments that the same depended on the quality of explanation (N= 13; 20%) and the quality of translation (N=4; 6%).

A few more questions of the questionnaire had the aim of getting a little insight into some other elementary aspects of participants' IL. As the second hypothesis was that participants didn't have any IL training, the expectation was that the main criteria for choosing an online source was the ranking of the page in the web searcher and the web page design, while checking the author and the date were not so popular. As is shown in Chart 3, the hypothesis was partially supported.

Chart 3. Criteria in choosing best online source

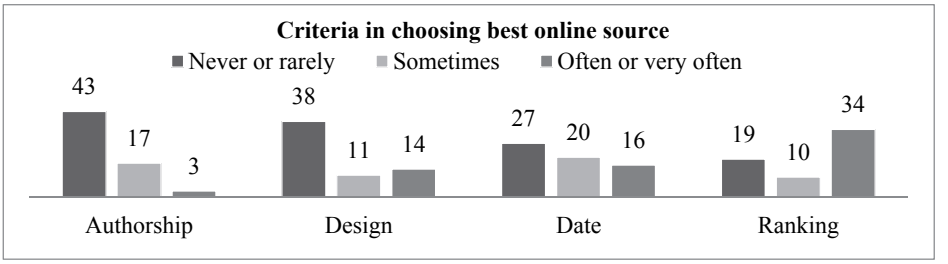


Chart 3 shows that the main criteria of the participants is the ranking of an online source in the search engine (M=3.19, SD=1.24), while authorship is rarely checked (M=2.05, SD=.99). About the date of the online information (M=2.70, SD=1.21), that criteria seems to be more relevant for the participants and it's checked more often, which can be related to the field of their studies because the rapid change of technology makes some technologies quickly obsolete. Web design (M=2.33, SD=1.16) doesn't seem to be relevant for the participants, which can be related to the type of sources they mainly use, as it will show the last part of the research.

Insight into the use of online sources for ESP purposes

Questionnaire answers

The third part of the questionnaire analyses the use and the frequency of online sources for the needs of ESP classes. Participants had to circle on the Likert scale from 1 (never) to 5 (very often) the frequency of online sources use for ESP classes. Even though the participants have quite high perception of their abilities regarding their IL, Table 2 shows not very high frequency of online sources use for ESP needs. Apparently, their need for online sources has proven to be low and not specific enough, especially when we take into consideration the specific sources that are needed for ESP classes.

Regarding ESP texts, it is particularly interesting to note that 89% (N=56) of the participants said they *never* used terminological dictionaries, while other valuable sources like mono and bilingual dictionaries, corporate web pages, web pages of professional association and manuals were used *rarely*. The only two types of online sources they apparently use *sometimes* are videos and wiki

sources (mainly Wikipedia), while they mostly rely on online translation services.

Table 2. Average score of the online sources usage for ESP classes

0-1,49; never	1,5-2,49; rarely	2,5-3,49; sometimes	3,5-4,49; often
• thesaurus (M=1.06; SD=.400) • terminological dictionaries (M=1.16; SD=.549) • encyclopaedias (M=1.16; SD=.578) • blog (M=1.48; SD=.936)	• presentations (M=1.51; SD=.840) • monolingual dictionaries (M=1.66; SD=1.173) • corporate web pages (M=1.68; SD=1.129) • books and magazines (M=1.69; SD=1.034) • individual web pages (M=1.71; SD=.974) • forums (M=1.78; 1.054) • web pages of professional associations (M=1.87 SD=1.129) • portals (M=1.89; SD=1.220) • bilingual dictionaries (M=2.00; SD=1.462) • manuals (M=2.02; SD=1.284) • pictures (M=214; SD=1.378; p=.010**)[2] • spellcheckers (M=2.29; SD=1.349) • social networks (M=2.33; SD=1.513)	• video (M=2.52; SD=1.533) • Wikipedia and other wiki sources (M=3.21; SD=1.297)	• online translation services (M=3.76; SD=1.422)

This general low usage of online sources in ESP classes could compromise students' meaningful learning outside the classroom, so ESP teachers should work on this aspect of IL students' training. Even though, translation services could be good and fast help that can speed some translation processes or resolve some language doubts, their usage in LSP and general language classes should be seen only as an aid, since they still demand a great human effort to process, analyse and correct the received output. Furthermore, when dealing with such specific terminology as the one LSP deals with, online translation services could also be misleading. For this reason we consider that monolingual and bilingual dictionaries should be better and more used sources for ESP needs. It is unusual to note that terminological dictionaries as well as encyclopaedias and thesauruses are almost never used by Croatian ESP vocational high school pupils and students, even though they should be of crucial importance as complementary sources to the ones used in ESP classes since they are not exclusively related to the language, but to the theoretical parts of their study.

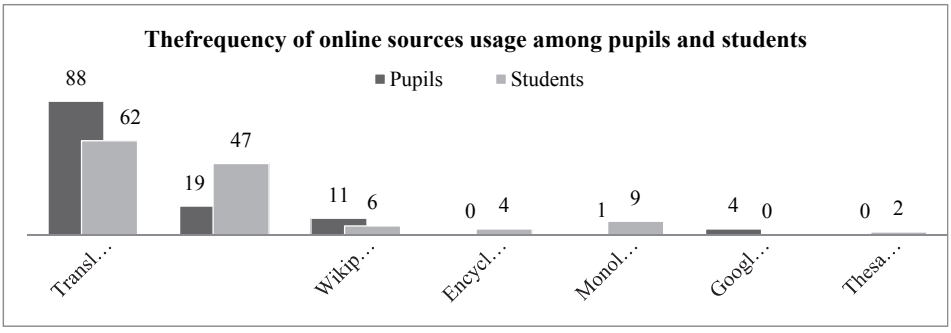
[2] This was the only result in the entire research in which statistically significant difference was found between pupils and students (pupils M = 2.43, SD = 1.547, students M = 1.88, SD = 1.166).

It is also interesting to notice the low position of pictures, manuals and presentations, which could be quite useful in learning ESP. They can be very useful for example when learning about technical parts like parts of a boat, car, engine and so on. In a future and more detailed research, it would be useful to separate the use of online sources for concrete language needs and for other educational purposes, to gain more accurate insight into their habits as information consumers and to see how often they use those sources for other classes. Such results could reveal whether those sources are generally of low usage or students have special problems in using them or finding them in a second language.

List of used sources while translating

In order to see which specific online sources were used in translating ESP texts, the participants had to name the sources they used in translating the given ESP text. The Chart 4 shows which online sources were used by pupils and students.

Chart 4. Used online sources among pupils and students



Pupils: Pupils numbered 124 uses of online sources when translating given ESP text, out of which the most frequent used online source were online translation services (88 times: Google translate 75, Etranslator 13) then bilingual dictionaries (19 times: Eudict 13, Glosbe 3, Crodict 2, English-Croatian dictionary 1). Wikipedia was selected 11 times, Google.hr 4 times and monolingual dictionary only one time (The Free Dictionary)[3].

Students: Students searched for online help 130 times, out of which the most frequent used online sources were translation services (62: Google translate 61, Etranslator 1). Bilingual dictionaries were used 47 times (Glosbe 23, Eudict 13, Crodict 6, English-Croatian dictionary 4 and E-dictionary 1). Monolingual

[3] Since The Free Dictionary (thefreedictionary.com) is an online dictionary, thesaurus and encyclopaedia, an analysis of the links showed that it was used as a dictionary/thesaurus. Because the thesaurus part is at the very end of the page, after numerous and repeated definitions of the word looked-up, we believe it was mainly used as a monolingual dictionary.

dictionaries were selected 9 times (The Free dictionary 6, Cambridge dictionary 2, Vocabulary.com 1). Wikipedia was chosen 6 times, Investopedia 4 and Thesaurus (thesaurus.com) 2 times.

Comparison: The results support the fourth hypothesis, that is, *Google translate* has been confirmed as the most frequent used online source. The fifth hypothesis was partially confirmed. Students use more bilingual and monolingual dictionaries and other sources which are unknown to pupils (online encyclopaedias other than Wikipedia and thesaurus), but still the usage is so poor and most of it is relied on online translation services which are not the best solution for such a specific translation which ESP texts require.

This research opens the door for new researches which could investigate the validity of ESP translation to and from Croatian while using online sources and it gives an idea of needed IL education among students and pupils that can be given only by professional educationalists.

Conclusion

The conducted research has shown frequent habits of using the Internet among pupils and students for general and educational purposes. The study confirmed that most of the participants didn't have any previous IL training. Even though they self-evaluated themselves as highly capable when searching for information, they are still acknowledging the fact that IL education is desirable since online sources are not likely to be useful unless the students can locate them and know how to use them to enhance their language learning experience. This is especially important among Technical ESP students and pupils since their needs are specific in regards to the specific language they are learning and thus the choice of internet sources should be also specific in order to satisfy those needs. The research results show that Croatian ESP pupils and students present low frequency of online sources usage and are mainly oriented to Google translate.

Such results emphasize the fact that education in IL is necessary and extremely important which again heightens the role of language teachers educating students on specific language oriented sources. By implementing IL education on all levels of education, students will become critical users of online technology and will be able to choose the best online information, explanation and translation while scrolling the Internet, which is a prerequisite for a deep and meaningful learning.

References

Alexander, Janet E.; Tate, Marsha A. Web wisdom : How to evaluate and create information quality on the Web. Hillsdale, NJ : Lawrence Erlbaum, 2009.

Browne, M. Niel, Freeman, Kari E.; Williamson, Carrie L. The importance of critical thinking for student use of the Internet. // College Student Journal, 34(2000), 3, 391–398.

Eisenberg, Michael B. Information Literacy : Essential Skills for the Information Age. // DESIDOC Journal of Library & Information Technology. 28(2008), 2, 39-47.

Fidel, Raya et. al. A Visit to the Information Mall : Web Searching Behavior of High School Students. // Journal of the American Society for Information Science. 50(1999), 1, 24-37.

Flanagin, Andrew J.; Metzger, Miriam J. Perceptions of Internet information credibility. // Journalism and Mass Communication Quarterly. 77(2000), 515–540.

Frazer, Megan. Students Leave High School Without Necessary Information Literacy Skills. Massachusetts School Library Association, 2010. http://maschoolibraries.org/content/view/129/105/ (29.5.2015).

Grimer, Deborah J.; Boening, Carl H. Worries with the Web: A Look at Student Use of Web Resources. // College & Research Libraries 2001, 11-[22].

Julien, Heidi; Barker, Susan. How high-school students find and evaluate scientific information: A basis for information literacy skills development. // Library & Information Science Research. 31(2009), 1, 12-17.

Jones, Steve; Madden, Mary A. The Internet goes to college : how students are living in the future with today's technology. Washington, D.C. : Pew Internet & American Life Project, [2002].http://www.pewinternet.org/files/old-media/Files/Reports/2002/PIP_College_Report.pdf.pdf (18.5.2015).

Lasić-Lazić, Jadranka; Špiranec, Sonja; Banek Zorica, Mihaela. Izgubljeni u novim informacijskim okruženjima-pronađeni u informacijskom opismenjavanju. // Medijska istraživanja. 18(2012), 1, 125-143.

Lasić-Lazić, Jadranka.; László, Marija. Mjerila informacijske (ne)pismenosti. // Informacijska tehnologija u obrazovanju : znanstvena monografija / Lasić-Lazić, J. (ed.). Zagreb : Zavod za informacijske studije, 2014, 223-234.

Lau, Jesus. Guidelines on information literacy for lifelong learning : final draft, reviewed July 30, 2006. http://www.ifla.org/files/assets/information-literacy/publications/ifla-guidelines-en.pdf (28.5.2015).

Lubans, John. How First-Year University Students Use and Regard Internet Resources. (1998) http://www.lubans.org/docs/1styear/firstyear.html (28.5.2015).

Lubans, John. Key Findings on Internet Use among Students. (1999) http://www.lubans.org/docs/key/key.html (28.5.2015).

Metzger, Miriam J.; Flanagin, Andrew J.; Zwarun, Lara. College student Web use, perceptions of information credibility, and verification behavior. // Computers & Education. 41(2003), 271-290.

Paris, Paul G. Critical Thinking and the Use of the Internet as a Resource. // International Education Journal. 4(2002), 1, 30-41.

Seljan, Sanja; Brkić, Marija; Kucis, Vlasta. Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs // INFuture2011: The Future of Information Sciences – Information Sciences and e-Society. Zagreb : Department of Information Sciences, 2011, 331-344.

Seljan, Sanja. Translation Technology as Challenge in Education and Business. // Informatologia. 44(2011), 4, 279-286.

Špiranec, Sonja. Virtualna učionica NSK ili kako su knjižnice zakucale na vrata e-učenja. // Edupoint. 25(2004), 8. http://eprints.rclis.org/5757/1/Edupoint_25_Spiranec.pdf (29.5.2015)

Špiranec, Sonja; Banek Zorica, Mihaela. Informacijska pismenost: teorijski okvir i polazišta. Zagreb : Zavod za informacijske studije Odsjeka za informacijske znanosti Filozofskog fakulteta Sveučilišta u Zagrebu, 2008.

Scientific Conferences as a Medium of Academic Communication – a Case of INFuture

Nikolaj Lazić[*], Ivan Dunđer[**]
[*]Department of Phonetics
[**]Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb Croatia
nlazic@ffzg.hr, idundjer@ffzg.hr

Marko Odak
Faculty of Philosophy, University of Mostar
Matice hrvatske b.b., Mostar, Bosnia and Herzegovina
odakmarko@yahoo.com

Summary

Scientific conferences provide opportunities for both formal and informal communication of academic insights among conference participants and make possible the sharing of those insights with the wider academic community as well as the wider public through the publication of conference papers in both print and digital form, and owing to all this they may be considered a significant medium of academic communication. Every system of communication consists of three basic elements: source of information, information user(s) and mediums or channels used for disseminating information. We analysed the system of academic communication using this basic structure, where we approached scientific conferences as a medium used for the dissemination of information. However, it is important to emphasise that in this basic scheme researchers represent the source of information and information users at the same time. Our analysis of academic communication includes both the dissemination and assimilation of scientific information taking place at conferences.

Standard forms of activity and interaction at scientific conferences, ranging from oral communication during the giving of presentations, through written communication involved in the publication of papers as part of conference proceedings to one's very presence at a conference, should carefully be evaluated in terms of their effectiveness as mediums of academic communication.

We will present the results of our preliminary research through demonstrating correlations between the number of authors and number of papers presented at previous INFuture conferences, classifying authors by country, co-authorship, etc., and thus provide an introduction into a more detailed research into the communication space created as a result of scientific conferences and the role of that space in the communication within a particular academic community.

A more extensive research project will include scientometric and bibiliometric analyses, interviews and comparative analyses that should provide us with insights into the significance of conferences as a medium of academic communication and their impact on the academic community.

Keywords: academic communication, evaluation of scientific work, conferences as a medium of academic communication, INFuture

Introduction

Science, as a complex human activity, is based on the innovative work of scientists and researchers and a subsequent dissemination of insights resulting from their work. If we approach science as a process based on research activity and the accompanying communication of research results, then scientific conferences and conference publications may be regarded as a multi-directional communication channel.

Every system of communication consists of three basic elements: source of information, information user(s) and mediums or channels used for disseminating information. We analysed the system of academic communication using this basic structure, where we approached scientific conferences as a medium used for the dissemination of information. However, it is important to emphasise that in this basic scheme researchers represent the source of information and information users at the same time. Our analysis of academic communication includes both the dissemination and assimilation of scientific information taking place at conferences.

Researchers are constantly searching for new information relevant for their work and keep publicising the results of their research and analyses, first in the form of various oral or written reports and subsequently as published academic papers. These two roles of researchers are so heavily intertwined that it is difficult to analyse them separately. It is particularly hard to distinguish between dissemination and assimilation of information when it comes to informal types of communicating knowledge. Thus when researchers give presentations at a conference, they both provide information and seek for new information in the form of comments of their colleagues and the acceptance or rejection of the ideas that they put forward.

Personal encounters and conversations on themes and issues being discussed at a conference are also very significant, regardless of possibilities provided by means of online communication.

Methodology

As the main source of data on conference papers, authors and their institutions, authors' countries of origin, themes and keywords we used digital conference proceedings, which we first analysed separately and then proceeded to make comparisons.

We wanted to use the results of the comparative analysis of the four years of INFuture conferences to extract data that we would subsequently use as the basis for further research on the value of academic communication taking place at conferences and the resulting connections realised as various types of personal and institutional cooperation. Since we hold that a more detailed research on the communicational value of conferences is necessary, this preliminary research represents a basis for the formulation of hypotheses on the value of academic communication occurring at conferences and of the communication space that they create.

The obtained data have been structured and interpreted as part of the Results and Interpretation chapter. They were analysed according to the following parameters: *participating countries, number of authors, participating institutions* and *cooperation between authors*.

Results and interpretation

We analysed the extracted data and will present them by the parameters stated above.

Participating countries

The sole number of countries represented at the conference since its beginning is highly significant since it provides insight into the size of the communication space created by the conference (Table 1). The conference's second year brought about the expansion of that space and the inclusion of a larger number of participating countries.

Table 1: Number of participating countries

	2007	2009	2011	2013
Number of participating countries	11	13	11	6

Table 2: Number of conference participants by country

2007	2009	2011	2013
Croatia (98)	Croatia (145)	Croatia (64)	Croatia (46)
Slovenia (5)	UK (5)	Serbia (6)	USA (2)
UK (2)	Slovenia (5)	Finland (5)	Canada (2)
Italy (2)	China (3)	Germany (3)	Germany (1)
Czech Republic (2)	Germany (3)	Turkey (2)	Australia (1)
Bosnia and Herzegovina (2)	India (3)	Slovenia (2)	Netherlands (1)
	USA (2)	Norway (2)	
Poland (2)	Iran (2)	Portugal (2)	
Nederlands (1)	Italy (2)	China (1)	
Australia (1)	Bosnia and Herzegovina (2)	Ireland (1)	
Ireland (1)	Netherlands (1)	Switzerland (1)	
Lithuania (1)	Serbia (1)		
	Belgium (1)		

Data on the conference's third and fourth year show that the number of participating countries decreased, which may be attributed to social and economic conditions in some countries. This decline also brings us to the issue of the sufficiency of financial resources provided by academic institutions or ministries for conferences as a medium of academic communication.
Table 2 shows that majority of authors are from Croatia.
A questionnaire survey that would include the majority of conference participants since the launching of the conference should be carried out as part of a comprehensive research in order to gain insight into these circumstances.

Number of authors
The analysed data show an increase in the number of authors between the first and second conference year (Table 3). However, a marked decline in this number is observable in the third and fourth conference year.

Table 3: Number of authors by conference year

	2007	2009	2011	2013
Number of authors	117	175	89	53

Reasons for such decline surely have to do with the economic crisis that befell countries in the region and the related insufficiency of public resources provided for education and science.

Participating institutions
We considered this category important for several reasons. First we wanted to gain insight into the type of institutions that authors came from and subsequently analyse if there were any institutions that stood out or were in any way 'unexpected' (as foundations under 'others' in Table 4). The analysis showed that the majority of participating institutions were educational and scientific institutions. Apart from them, authors at the first conference also came from libraries, archives, museums and government institutions and agencies.

Table 4: Participating institutions grouped by type of institution

	2007	2009	2011	2013
Libraries	8 (6.8%)	11 (6.3%)	8 (9.0%)	1 (1.9%)
Archives	2 (1.7%)	1 (0.6%)	1 (1.1%)	0 (0.0%)
Museums	0 (0.0%)	1 (0.6%)	2 (2.2%)	0 (0.0%)
Faculty, university, institute	82 (70.1%)	136 (77.7%)	75 (84.3%)	41 (77.4%)
Industry	4 (3.4%)	6 (3.4%)	1 (1.1%)	5 (9.4%)
Government agencies	11 (9.4%)	10 (5.7%)	1 (1.1%)	3 (5.7%)
Other	10 (8.5%) foundation (2) elem. school (2) sec. school (1) unknown (5)	10 (5.7%) freelance (3) foundation (2) elem. school (2) association (1) unknown (2)	1 (1.1%) sec. school (1)	3 (5.7%) elem. school (3)

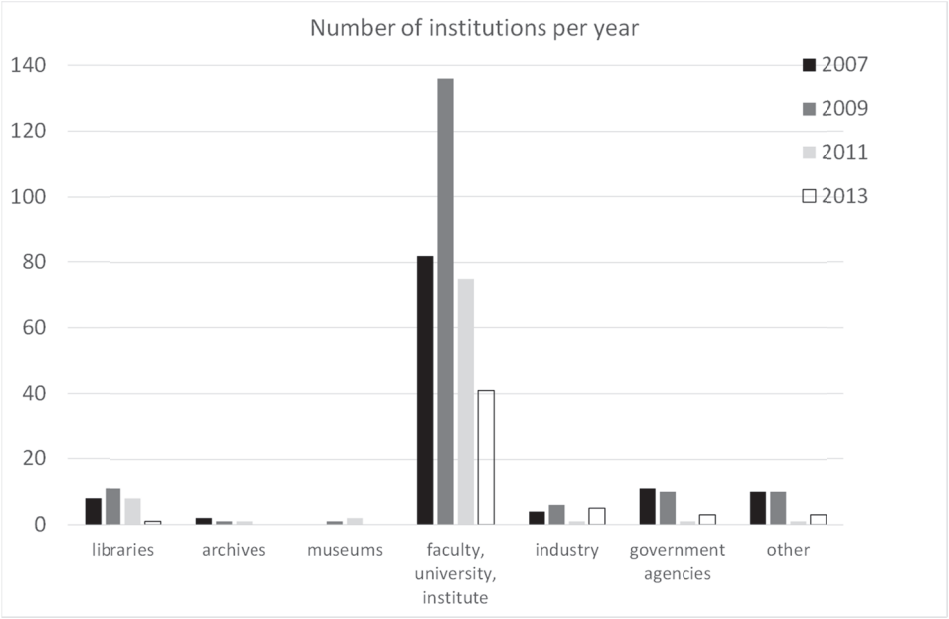


Figure 1. Participating institutions grouped by type of institution

Data for the conference's second year show that an already large number of conference participants coming from educational institutions in the conference's first year increased even further (Figure 1). Such tendencies are not surprising if we take into consideration that the programme of the conference addressed themes and issues related to changes in the curricula of many LIS-oriented study programmes. In our more comprehensive research we will use these indicators to examine the hypothesis that conference's significance in the realm of academic communication is significantly affected by institutions that authors of conference presentations come from.

Cooperation between authors

The concept of cooperation implies the joining of efforts and activities and this aspect should be analysed both at the level of personal cooperation between individual researchers and between institutions. Since we hold that the establishment of cooperation is one of the tasks of a scientific conference, a more comprehensive research should include interviews and questionnaires that would provide us with insight into the extent to which the existing connections and cooperation initiatives came as a result of the conference. Co-authorship is a form of intellectual cooperation, so we approached it as an indicator of cooperation.

Table 5: Number of papers grouped by authorship

	2007	2009	2011	2013
Number of accepted papers	56	77	43	25
Single author	20 (35.7%)	21 (27.3%)	18 (41.9%)	8 (32.0%)
Two authors	16 (28.6%)	19 (24.7%)	11 (25.6%)	10 (40.0%)
Three authors	17 (30.4%)	33 (42.9%)	11 (25.6%)	4 (16.0%)
More than three	3 (5.4%)	4 (5.2%)	3 (7.0%)	3 (12.0%)
More than two	36 (64.3%)	56 (72.7%)	25 (75.8%)	17 (68.0%)

Our analysis demonstrates the relationship between single-author papers and papers by several authors which remained constant throughout all four years of the conference (Table 5), with a percentage of single-author papers ranging from 27% to 41% and a percentage of papers by several authors shifting between 59% and 73%. The predominance of co-authored papers indicates a strong tendency towards the joining of efforts between individual researchers and institutions and further fostering of connections established as a result of the conference as a dynamic context for communication.

Academic cooperation is a subject that has to be looked at separately and we should analyse in greater detail the extent to which various cooperation initiatives were prompted by the conference. Researchers and specialists in various fields use conferences to exchange their academic experiences and present their institutions and countries of origin. Thus specific instances of academic cooperation may be categorised as instances of personal cooperation between authors sharing similar research interests or as instances of institutional cooperation. A thorough analysis of a cooperation network spawned by INFuture conferences that would be carried out as part of a more comprehensive research opens up many possibilities but also necessitates carefully constructed in-depth interviews that would ensure the reliability and objectivity of data interpretation.

Conclusion

Scientific conferences as a medium of academic communication are affected by a wide range of circumstances. Some of them are related to specific characteristics associated with a given scientific field (Lazić, et al., 2013), area or discipline (Nederhof, 2006), while others are related to academic policies and official strategies (Mali, 2010; Abramo, et al., 2013; Bornmann, 2012). Rapid changes in the field of LIS (Lasić-Lazić, et al., 2013; Jokić, et al., 2012; Andreis et al., 2008) and LIS-related scientific disciplines, along with growing requirements for the dissemination of information, sharing of ideas, innovative solutions and more frequent and productive professional encounters and exchanges are making scientific conferences a significant context for and medium of academic communication. In our research we sought to reinforce the INFuture conference's significance as a medium of communication within LIS-oriented academic community and formulate hypotheses for our further research.

Why should the publication of papers as part of conference proceedings be included among the criteria for academic advancement? Apart from citation databases today there is also the Thomson Reuters Conference Proceedings Citation Index and its Social Science & Humanities collection, serving as a multidisciplinary bibliographic and citation database for this type of publications. However, since this is a commercial product that is not readily available to small academic communities, such publications cannot easily be subjected to scientometric analyses, so their evaluation has so far been relatively subjective, i.e. based on individual assessment. Information and insights in the field of library and information sciences quickly become outdated (Lasić-Lazić et al., 2013), so conferences have great importance as a medium of publicising these information, insights, data and ideas and having them either accepted or rejected by other partakers in the academic field. Furthermore, publishing papers as part of conference proceedings is currently growing in importance owing to long periods that researchers have to wait for their papers to be published in scientific journals. For example, the publication of papers in some LIS-oriented scientific journals sometimes takes as long as two years. Many data and insights become outdated during this period which makes conferences, as a context prompting and enabling a dynamic interaction between researchers in a given academic field, an even more significant medium of academic communication. The publication of a bibliography of conference papers would greatly contribute to the dissemination of knowledge and productive academic communication, therefore it should always accompany the publication of conference proceedings (regardless of the fact that some online services (e.g. Google Scholar) are already "compiling" one such huge bibliography).

Our research provided us with data demonstrating that personal cooperation between researchers and institutional cooperation feature as an important communicational element. A closer look at a varying number of authors and papers may lead us to the analysis of various economic and social factors influencing the development of the conference from year to year. The number of conference participants coming from LIS-oriented educational institutions indicates the significance of INFuture conferences for the development of the field of library and information sciences, especially librarianship.

Since the structure of communication associated with scientific conferences includes both formal and informal activities ranging from the first stages of research that a conference paper will be based on, to the actual presentation, the paper's publication as part of conference proceedings and the integration of its content into the corpus of the given academic discipline or field, it is clear that a more comprehensive research that would encompass all these aspects is necessary, and the results of our preliminary analyses provide a useful introduction into such research.

References

Andreis, Mladen; Jokić, Maja. An impact of Croatian journals measured by citation analysis from SCI-expanded database in time span 1975–2001. // *Scientometrics*, 75 (2008), 2, 263-288.

Bornmann, Lutz. The Hawthorne effect in journal peer review // *Scientomerics*, 91 (2012), 3, 857-862.

Jokić, Maja; Zauder, Krešimir; Letina, Srebrenka. Karakteristike hrvatske nacionalne i međunarodne znanstvene produkcije u društveno-humanističkim znanostima i umjetničkom području za razdoblje 1991-2005. Zagreb : Institut za društvena istraživanja u Zagrebu, 2012.

Lazić, Nikolaj; Lasić-Lazić, Jadranka. Analysis of Scientific Activities in the Field of Information Science: the Case of Croatia. // *Medijska istraživanja*. 19 (2013), 1; 5-18.

Lasić-Lazić, Jadranka; Pečarić, Đilda; Lazić, Nikolaj. The Network Map of the Fields of Interest in Information Sciences in Croatia // Recent Advances in Information Science. Proceedings of the 4th European Conference of Computer Science (ECCS '13) / Maurice Margenstern, Kleanthis Psarris, Danimir Mandic (ed.). Paris : WSEAS Press, 2013. 251-257.

Mali, Franc. Problemi javnih politika prema međunarodnoj produktivnosti i vidljivosti društvenih znanosti u zemljama Srednje i Istočne Evrope. // *Sociologija i prostor*, 48 (2010), 39, 415-430.

Nederhof, Anton J. Bibliometric monitoring of research performance in the social sciences and the humanities: A review. // *Scientometrics*, 66 (2006), 1, 81-100.

WORKSHOPS

How to maintain Authenticity and Integrity of Electronic Information without Utilizing Electronic Certificates

Hans Almgren, Mats Stengård
Enigio Time AB
Jakob Westinsgatan 3, Stockholm, Sweden
hans.almgren@enigio.com, mats.stengard@enigio.com

Introduction

How to manage integrity and authenticity are key areas for all kinds of long term preservation of electronic information. Today, many e-archive solutions are built to support the Records Continuum Model, i.e. a record is never definite, it can continuously be provided with more information, such as metadata.
This shift provides further demands on securing integrity and authenticity within the e-archiving systems. One way to address that is by using solutions for qualified electronic time stamps in order to securely track every event of a record.

Workshop Outline

Qualified electronic timestamps are provided either by using technology based on electronic certificates or by using one-way hash functions. The time span of electronic certificates is limited and therefore not suitable for long term applications, thus this workshop will focus on solutions using one-way hash functions.
Advanced one-way hash functions are perfect basis for proving the existence of electronic information. But the problem is how to secure that the proof of existence hasn't been tampered with. The provider of the qualified trusted timestamp has to be able to prove that the proof of existence (i.e. the hash code) is "written in stone".
This workshop will enlighten participants about different technologies for providing a non-manipulated publication channel for the proof of existence and thereafter elaborate on how this can be used also for streaming media (such as: continuous log file tracking, video and voice recording).

Part 1: How to utilize Block Chain Technology for providing proof of existence

- How does it work?
- Bitcoin and alternatives
- Pros and cons

Part 2: How to secure authenticity for streaming media by utilizing timeloop

- How to secure authenticity of streaming media?
- How does it work?
- Use cases

Anyone who has concerns on how to secure long-term authenticity and integrity of data is welcome. This workshop will explain some possibilities and solutions in this area of research.

Initial maturity model for information governance arrangements in organizations

Diogo Proença, Ricardo Vieira, José Borbinha
INESC-ID, Instituto Superior Tecnico, Universidade de Lisboa
Rua Alves Redol nº 9, Lisbon, Portugal
diogo.proenca@tecnico.ulisboa.pt, rjcv@tecnico.ulisboa.pt, jlb@tecnico.ulisboa.pt

Clive Billenness
University of Portsmouth
2 Winston Churchill Avenue, Portsmouth, Hampshire, UK
clive@fp7experts.eu

Introduction

The workshop is based on the research work undertaken in the EC Project E-ARK[1] [620998] as published in public Project Deliverable D7.2, submitted to the EC on 1 August 2015.

Workshop outline

The purpose of the workshop is to introduce delegates to the background and principles of the Maturity Model which measures the effectiveness of the Information Governance Arrangements within an organisation - which are essential to support the archiving of electronic data. This will be followed by a detailed inspection of the evaluation model and an examination of the questions.
Delegates will be invited to take a copy of the model back to their own organisation and complete it to assess their own current maturity level in a number of areas and so develop an action plan to address any shortcomings identified.
Delegates will also be invited to submit their responses anonymously to a central database, thereby contributing to a facility for organisations to benchmark themselves again comparable bodies and measure their progress.
The workshop is intended to be highly interactive and is intended for all involved in the archiving of electronic data or preparing for such activities.

[1] E-ARK (European Archival Records and Knowledge Preservation) Project, http://www.eark-project.com/

INNOVATIVE TECHNOLOGIES

PIQL Preservation Services – A Holistic Approach to Digital Long-term Preservation

Alexander Rych
In-Vision Digital Imaging Optics GmbH
Industriestrasse 9, Guntramsdorf, Austria
rych@piql.at

Summary

Digital storage media (hard disks, magnetic tape) exhibit limited lifetime. Repetitive data migration to overcome rapid obsolescence of hardware and software bears accelerated risk of data loss, data corruption or even manipulation and adds significant repetitive costs for hardware and software investments.
So Piql Preservation Services ("Piql") has been developed as a turnkey solution designed for secure, migration-free long-term preservation of digital data. It sets an open standard for long- term preservation for the future, consisting of equipment and processes needed for writing and retrieving digital data.

Keywords: Piql, long-term preservation, archive, digital data on analogue media, film, digital frames, hybrid preservation

PIQL – Solution

Piql stores any kind of data in its digital as well as analog form securely on a film reel for 500 years. Using photosensitive film polyester base, a very stable material that is known for its immutability over hundreds of years, secure and cost-effective long- term preservation can be provided. The film reel itself is stored in a packaging capable of protecting the optical storage medium. These components have undergone extensive testing to ensure longevity of up to 500 years. In addition to its durability, film is a true WORM (write once-read many) medium. It therefore is resistant to editing or manipulation.
Being able to store any form of data onto the film makes Piql a superior solution for long-term preservation. Paper documents, *images, video or audio sequences* – all of those file formats and documents can be preserved in its native file structure. In order to restore the encoded digital data, only a film scanner, a digital camera or any appropriate optical reading device will be needed in the future. Every film reel includes besides meta-data all necessary information to enable users in the future to read and decode the digital information.

Key features:

- **Migration-Free**. By storing digital data in archival file formats on the unique piqlFilm, repetitive migrations are no longer needed. Users avoid the risk of migration-related corruption and data loss and thus save long-term costs.
- **Searchable**. Operated within a standard IT-environment, the user's data is indexed and fully searchable – today and in the future.
- **Flexible**. Depending on the user's needs, data can either be stored as a human- readable text or image or in a digital format, making data readable and understandable for 500 years.
- **Unalterable**. Data is preserved on a secure true WORM medium, making it impossible to manipulate or delete valuable data.
- **Permanent**. Photosensitive film is known to be robust with proven long-term qualities. PiqlFilm is tested to last for 500 years. Although your time perspective may be shorter, you can be assured that your data will remain intact.

Figure 1: piqlBox and piqlFilm

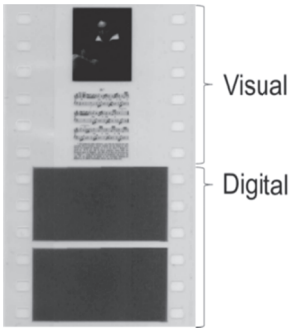


Fig. 2: digital and visual representation of data

The workflow consists in data boxing and integrity check, data recording, data scanning to guarantee integrity and accessibility in the future – storage and data unboxing (process that fully restores the scanned data).

Piql Preservation Services manage to combine some of the most important components to ensure a secure and cost-effective long-term archiving system.

- There is **no limit** of what can be stored onto the piqlFilm. Digital data such as HD audio-visual data, documents or databases can be archived with Piql.
- It is a service that uses a whole new approach to long-term preservation by **removing the need for constant data migration**, reducing hardware and software maintenance.
- The innovation is to use a permanent media, to adopt standards for formats and interfaces and be **independent of proprietary software and specific hardware**.

Piql Preservation Services convert digital files into a physical copy that can be stored as the ultimate digital insurance. Data owners can feel confident about the long-term accessibility of their valuable data – today and in 500 years.

INFuture2015 reviewers

All papers were reviewed by at least two reviewers. INFuture relies on the double-blind peer review process in which the identity of both reviewers and authors as well as their institutions are respectfully concealed from both parties.

INFuture2015 gratefully acknowledges its reviewers:

Reviewer name	Insitutional affiliation
Stefano Allegrezza	Department of History and Preservation of Cultural Heritage, University of Udine, Udine, Italy
Melita Ambrožič	National and University Library, Ljubljana, Slovenia
Karen Anderson	Mid Sweden University, Sweden
Petra Bago	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Regis Bertholon	Haute Ecole Arc, University of Applied Sciences, Western Switzerland
Jenny Bunn	Department of Information Studies, University College London, UK
Tanja Buzina	National University Library, Zagreb, Croatia
André Constantino da Silva	Federal Institute of São Paulo, Brazil
Rodrigo Costas	Center for Science and Technology Studies (CWTS). Leiden University, The Netherlands
Khaled Dehdouh	Universite de Lyon, France
Margarita Fernández Monroy	Department of Economics and Business, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
Antonio García Domínguez	Department of Computer Science and Engineering, University of Cadiz, UCA, Cádiz, Andalusia, Spain
Francisco José García Peñalvo	Departamento de Informática y Automática, Facultad de Ciencias, University of Salamanca, Salamanca, Spain
Katarina Grolinger	Department of Electrical and Computer Engineering, Faculty of Engineering, The University of Western Ontario, London, Ontario, Canada
Sam Habibi Minelli	Gruppo Meta, Italy
Ivana Hebrang Grgić	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Birger Hjørland	University of Copenhagen, Denmark
Wolfgang Hofkirchner	Institute of Design and Technology Assessment, Vienna University of Technology, Austria
Tomislav Ivanjko	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Violeta Janulevičienė	Institute of Philosophy and Humanities, Mykolas Romeris University, Vilnius, Lithuania
Peter Jurgec	Department of Linguistics, University of Toronto, Canada

Reviewer name	Insitutional affiliation
Vedran Juričić	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Declan Kennedy	Department of Education, University College Cork, Ireland
Ksenija Klasnić	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Kristina Kocijan	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Stefanos Kollias	Division of Computer Science, Ethniko Metsovio Polytechnico, Department of Electrical & Computer Engineering, Athens, Greece
Vlasta Kučiš	Department of Translation Studies, University of Maribor, Maribor, Slovenia
Tomislava Lauc	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Nikolaj Lazić	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Hong-Nan Lin	Department of Animation and Game Design, Toko University, Chiayi, Taiwan
Vjera Lopina	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Henrik Lorentzen	Det Danske Sprog- og Litteraturselskab (DSL), Copenhagen, Denmark
Bojan Macan	The Ruđer Bošković Institute Library, Zagreb, Croatia
Segomotso Masegonyana Keakopa	Department of Library and Information Studies, University of Botswana, Botswana
Nicola Matteucci	Universita Politecnica delle Marche, Ancona, Italy
Alexandre Matos	Sistemas do Futuro, Ltd., Porto, Portugal
Clarence Maybee	Purdue University, Syracuse, New York Area, USA
Sue McKemmish	Centre for Organisational and Social Informatics, Monash University, Australia
Carlo Meghini	Networked Multimedia Information Systems Laboratory, Institute of the National Research Council of Italy, Italy
Nives Mikelić Preradović	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Miroslav Milinović	University Computing Centre, University of Zagreb, Croatia
Krishna Mohan Surapaneni	Saveetha Medical College & Hospital, Faculty of Medicine, Saveetha University, Chennai, India
Sandro Nielsen	Center for Leksikografi, Department of Business Communication, Aarhus University, Aarhus, Denmark
Gillian Oliver	School of Information Management, Victoria University of Wellington, New Zealand
Oleksandr Pastukhov	Department of Information Policy and Governance, Faculty of Media and Knowledge Sciences, University of Malta, Malta
Krešimir Pavlina	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

Reviewer name	Insitutional affiliation
Gilda Pečarić	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Arlene Peukert	Institut für Museumsforschung, Staatliche Museen zu Berlin, Berlin-Dahlem, Germany
Saulius Ragaišis	Software Engineering Department, Faculty of Mathematics and Informatics, Vilnius University, Lithuania
Arian Rajh	Agency for Medicinal Products and Medical Devices, Zagreb, Croatia
Günter Schreier	Austrian Institute of Technology, Austria
Hossain Shahriar	Computer Science department, Kennesaw State University, Kennesaw, Georgia, USA
Elizabeth Shepherd	University College London, Department of Information Studies, London, UK
Vjeran Strahonja	Faculty of Organization and Informatics, Varaždin, Croatia
Vladimir Šimović	University North, Croatia
Maarten van Bommel	Faculty of Humanities, Capaciteitsgroep Conservering en Restauratie, Amsterdam, The Nederlands
Alexander Voiskounsky	Psychology Department, Moscow Lomonosov State University, Moscow, Russia
Radovan Vrana	Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Magdalena Wójcik	Institute of Information and Library Science, Jagiellonian University, Kraków, Poland
Krešimir Zauder	freelance expert, Croatia
Vatroslav Zovko	Faculty of Teacher Education, University of Zagreb, Croatia

The INFuture2015 conference is sponsored by

Financial Agency – FINA, Croatia
http://www.fina.hr

b4b, Croatia
http://www.b4b.hr/

PIQL, Norway
http://www.piql.com/

In-Vision Digital Imaging Optics GmbH, Austria
https://www.in-vision.at/

Zagreb Tourist Board
http://www.zagreb-touristinfo.hr/