DigitalPreservationEurope: A Way Forward in the Long Term Curation of Digital Materials

Chiara Cirinnà
Fondazione Rinascimento Digitale
Via Bufalini 6, Florence, Italy
cirinna@rinascimento-digitale.it

Maurizio Lunghi
Fondazione Rinascimento Digitale
Via Bufalini 6, Florence, Italy
lunghi@rinascimento-digitale.it

Summary

We are shifting from an industrial economy to an economy in which knowledge is an input of increasing significance and this has been triggered by rapid progress in the evolving Age of Information, where repositories of digital information and the tools to mine, analyse and re-purpose them represent a society's intellectual capital. Effective and affordable digital preservation strategies and systems will transform archives into valuable assets.
The European funded coordination action DigitalPreservationEurope (DPE) has reached its conclusion in March 2009 and several results in the field of digital preservation have been accomplished.
The primary objective of the coordination action was to concentrate and share efforts towards the common purpose of assuring effective preservation of digital materials. To this end, the three main projects on digital preservation, CASPAR, Planets, DPE (co-funded under the European Commission Information Society Technology (IST) Sixth Framework Programme) have worked together to create the window on their synergistic activities
Before a repository is created, PLATTER, the Planning Tool for Trusted Electronic Repositories, provides a basis for a digital repository to plan the development of its goals, objectives and performance targets over the course of its lifetime in a manner which will contribute to the repository establishing trusted status amongst its stakeholders.
After the repository has been created, the digital curation can be considered as a risk-management activity. Based on practical research and developed jointly by the DCC (Digital Curation Centre) and DPE, the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) provides a methodology for self-assessment through a metric to enable an auditor to establish the organisa-

tional context and goals of a repository and then to assess how it is achieving these in terms of risk.
Digital preservation solutions are a little less distant then before.

Key words: digital preservation, trusted digital repositories, risk-assessment

Introduction

Digital preservation is a set of activities required to make sure digital objects can be located, rendered, used and understood in the future. This can include managing the object names and locations, updating the storage media, documenting the content and tracking hardware and software changes to make sure objects can still be opened and understood.
But what does 'long-term' mean in the context of digital preservation? DPE agrees with the CCSDS[1], that states “a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future”.
That assumed, what do we need to preserve? Various aspects of the digital objects may be needed to be preserved. The lowest level of preservation requirements includes preservation of the bit stream, this does not however ensure understandability, readability or usefulness of the digital object. The biggest risk in terms of understandability is that the meaning (and even the names) associated with values in a dataset, although known to the data producers, is not available to the users; without this the data is essentially useless. Another aspect is that, even for users within the same sub-discipline, terminology drifts and meaning is lost; users in different (sub)disciplines will require even more help with the semantics of the data.
A more complex approach may strive to preserve also the meaning so that it remains readable and understandable. Such an approach requires the preservation of additional information (representation information, technical metadata etc.)
Even more ambitious preservation approaches try to preserve understandable content in such a way that the provenance and source of the digital object also remains clear. Thus the users can have trust that the object is authentic, accurate, and complete.

Why should we care about digital preservation?

Digital objects are much more 'fragile' than traditional analogue documents such as books or other hard copy mediums. Digital objects are fragile because they

[1] Consultative Committee for Space Data Systems

require various layers of technological mediation before they can be heard, seen or understood by people. Digital objects are also much more venerable to physical damage. One scratch on CD-ROM containing 100 e-books can make the content inaccessible, whereas to damage 100 hard copy books by one scratching move is - fortunately - impossible. A flash memory stick can drop into glass of water or get magnetised, portable hard drive or laptop can slip from your hands and get irreparably damaged in a second.
Digital objects require pro-active intervention to remain accessible. While you can put a book on a shelf and return to it in upwards of 100 years and still open it and see the content as it was intended by the author/publisher, the same approach of benign neglect to a digital object is almost a guarantee that it will be inaccessible in the future.
Alternatively the software or file format can become obsolete for a number of reasons. For example software upgrades may not support legacy files; the format take up is low and the industry does not produce compatible software; software which supports the format may be bought by a competitor and withdrawn form the market place. Without the intervention of digital preservation techniques the information contained will no longer be accessible.
The following paragraphs describe some of the main results achieved by the DPE[2] project.

A successful coalition

In order to reach bigger results in terms of disseminating knowledge, information and practice among a wide community, the DPE project with the two main projects on digital preservation issues, CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) and Planets (Preservation and Long-term Access through NETworked Services) have signed up an agreement to enable synergy and concerted action. The first results of this joint cooperation was the development and delivery of a collaborative web platform shared by the projects to serve as a common entry point to digital preservation and curation projects[3], provided with common services, a calendar of events, information board, resources. Moreover they have collaborated on the development of training and educational events and programmes in Europe and supported the dissemination of publications and the mechanisms to ensure their visibility (e.g. by automatic means such as OAI-PMH).
Several common events have been organised, like the joint conference in Lisbon in 2007 and in Nice in 2008.

[2] http://www.digitalpreservationeurope.eu

[3] http://www.wepreserve.eu

PLATTER, a tool for achieving trust in repository planning

DPE has recognized that a critical step for creation of a repository is to early plan the development of its goals, objectives and performance targets. PLATTER, the Planning Tool for Trusted Electronic Repositories, is not in itself an audit or certification tool but is rather designed to complement existing audit and certification tools by providing a framework which will allow new repositories to incorporate the goal of achieving trust into their planning from an early stage. A repository planned using PLATTER will find itself in a strong position when it subsequently comes to apply one of the existing auditing tools.

When we deal with a “trusted repository”, we can state that a repository is “Trusted” if it can demonstrate its capacity to fulfil its specified functions, and if those specified functions satisfy an agreed set of minimal criteria which all Trusted Repositories are assumed to require.

Among all the different available approaches, a suitable compromise would be to allow repositories to identify their own goals within a broadly accepted framework of basic requirements relevant to all trusted repositories. Precisely such a framework is represented by the Ten Core Principles of Trust Repository Design which have been developed by the Center for Research Libraries (CRL), the Digital Curation Centre (DCC), DigitalPresevationEurope (DPE), and the German Network of Expertise in Digital long-term preservation (nestor) in the field of audit and certification at a meeting hosted by CRL in Chicago in January 2007. The principles state that a repository:

1. Commits to continuing maintenance of digital objects for identified community/communities.
2. Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfil its commitment.
3. Acquires and maintains requisite contractual and legal rights and fulfils responsibilities.
4. Has an effective and efficient policy framework.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as the relevant production, access support, and usage process contexts before preservation.
8. Fulfils requisite dissemination requirements.
9. Has a strategic program for preservation planning and action.
10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.

What remains open, and what PLATTER is designed to address, is how these principles can be incorporated into the design and planning of a repository so that it is “trust-ready” from the start.

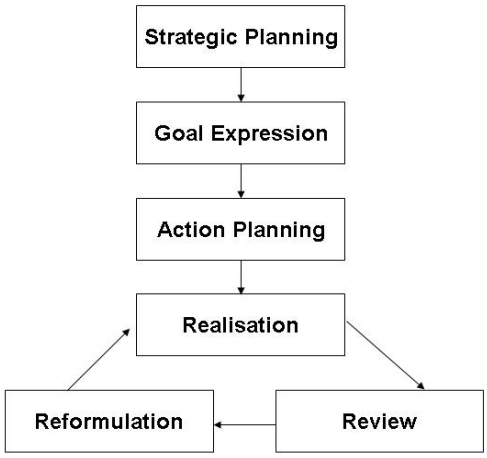
Repository classification

Since no "one-size-fits-all" approach can hope to apply to all types of repository, it is vital for a repository planner to be able to classify their repository in order to be able to compare its policies and practices with other similar repositories. The first stage of the PLATTER analysis is a taxonomic classification which will enable a repository to be compared with other similar repositories. Many possible schemes for such a classification could be developed, and in PLATTER we have chosen to characterise a repository along a number of independent axes grouped into four major descriptive classes:

- Purpose and Function: the purpose of this group of taxonomic axes is to determine the general functional type of the repository. The requirements of a national library may be quite different from those of an institutional or subject-based repository, a scientific data repository, or a national archive.
- Scale: in this group we consider the various factors which together define the overall scale of the repository, whether expressed in human, technical, or financial terms.
- Operation: this group of axes is primarily concerned with how material enters into the repository, the kind of material stored, and the extent to which that material may be accessed by end users.
- Implementation: this group of axes deals with the choices made in the implementation of the repository system.

The PLATTER planning cycle describes a semi-formalised set of steps intended to facilitate the processes of definition and expression of organisational objectives, and implementation and evaluation of the measures intended to meet them.

Table 1-The PLATTER Planning Cycle



The process is a cyclical one, and individual sections conform in many respects to parts of the DRAMBORA risk-analysis process. The following sections seek to describe in some more detail each stage of the PLATTER cycle, outlining their implicit parts and how they interrelate.

Strategic planning. Strategic planning is an invaluable means for maintaining a sufficiently broad and forward-facing organisational perspective, even when individuals are focusing on much more immediate and specific aspects of business activity. One, or more of three fundamental questions are posed during the process of strategic planning:

1. What do we do?
2. Who do we do it for?
3. How can we excel?

Responses to these three questions organisations will encapsulate the repository's mandate (or reference a non-self imposed, e.g., legislative mandate), detail the identities and broad expectations of primary stakeholders and describe in general, but tangible terms, the circumstances and performance levels that will represent success.

Definition of goals. Defined objectives must take into account the expectations and requirements of each major stakeholder. From the perspective of digital repositories, this may include management, funders, information creators, owners and depositors, and end users interested in accessing preserved content. Each must be related, either explicitly or implicitly with more fundamental strategic objectives, most immediately with the organisation's mission statement.

Undertake planning. The planning stage is bridge-building; between the determination of what must be achieved, and the tangible realisation of such achievements. Related experiences in comparable environments can be considered and where appropriate absorbed into the planning cycle. Action planning is only really feasible by adopting a global perspective of organisational constraints and influences. Many will have been formalised in the initial strategic planning stages, most notably those focused on establishing current situational awareness, and perceptions of any emerging contextual influences. Legislation, policy originating from parent organisations, stakeholder expectations and resource availability will all contribute to the success or otherwise of planned actions and must be given adequate consideration.

Deliver, review and reformulate implementation. In an iterative cycle, that may extend beyond the planning and development of the repository into full production phases, these three interrelated activities are fundamental to the ongoing improvement and developing maturity of the repository. An agile approach to all three will benefit the organisation and the pursuit of its objectives; no period of implementation should become too prolonged prior to the initial phases of review and reformulation.

The PLATTER process is centred around a group of Strategic Objective Plans (SOPs) through which a repository specifies its current objectives, targets, or key performance indicators in those areas which have been identified as central to the process of establishing trust.
In the future, PLATTER can and should be used as the basis for an electronic tool in which repositories will be able to compare their targets with those adopted by other similar (suitably anonymised) repositories. The intention is that the SOPs should be living documents which evolve with the repository, and PLATTER therefore defines a planning cycle through which the SOPs can develop symbiotically with the repository organisation.
The PLATTER tool is concerned exclusively with management of the objectives and targets of repository. It is not itself a tool for establishing trust and is not intended to compete with other initiatives in that area.
PLATTER is designed to complement DRAMBORA and a repository planned using PLATTER will be strongly placed to use DRAMBORA as a self-assessment tool.

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)

Most current digital repositories and most databases and collections used to help curate scientific data do not have specific mandates for long term preservation, nor do they have necessary long-term budgets. Instead they are mandated to support access and re-use in the near-term future. Long term preservation may be one of their aims, or at least hopes and wishes, but it is not (yet) a responsibility.
The DRAMBORA toolkit aims to complement other repository and certification work by addressing the full range of repositories, whether they aim for long term preservation or not. The toolkit is intended to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognise their strengths.
Within this toolkit the authentic and understandable digital object is positioned at the centre of a risk-based approach to audit; digital curation is characterised as a process of transforming controllable and uncontrollable uncertainties into a framework of manageable risks, classified according to a repository's activities, assets and regulatory context. The audit tool will encourage repository staff to identify and classify the risks posed at every stage of their activities, to assess the probability of their occurring, to appreciate their potential impact if they should arise.
Throughout a series of interactive stages, auditors are expected to develop a comprehensive image of their organisational objectives, the regulatory context within which they operate and the activities that must consequently be undertaken. Any risk management exercise will include these stages:

- Identifying the context where risks have to be managed

- Identifying risks
- Assessing and evaluating risks
- Defining measures to address and manage risks

The self-audit progresses through six stages:

Stage 1: Identify organisational context. In Stage 1, auditors document the mandate and derive both the goals and objectives of the repository. The ultimate purpose of this stage is to define the scope of the repository work, verifying internal awareness of the organisational framework, and at the same time ensuring that appropriate supporting documentation exists. Within this stage auditors must describe the overall purpose of the repository, in order to determine the characteristics that will undergo risk analysis and subsequent assessment. In particular auditors must identify the repository’s mandate which will be described in an organisational mission statement, then they will identify, within the mandate, each organisational goal and objective relevant to the repository.

Stage 2: Document policy and regulatory framework. This Stage gives auditors the opportunity to provide or refer that the repository:

- operates appropriately with respect to relevant regulatory frameworks;
- has an efficient and effective policy framework;
- is aware of the societal, ethical, juridical, and governance frameworks;
- is aware of the legal, contractual and regulatory requirements to which the repository is subject.

At this stage auditors need to determine what to look for, to collect information from documentary sources, and to compile a list of documents regulating the work of the repository.

Stage 3: Identify activities, assets and their owners. The purpose of Stage 3 is to develop a conceptual model of what the repository does and how it does it, by examining its activities and work processes, key assets and technology, and the staff involved. This Stage requires auditors to split the broad-level mission and goals of the repository into more specific activities or work processes that the repository carries out in order to achieve its aims.

Stage 4: Identify risks. The aim of this stage is to derive from organisational activities and assets a comprehensive selection of pertinent risks faced by the repository. Some risks can be derived from examining the mandate and objectives, regulatory environment and the model of the repository’s work (activities, assets, staffing, technology solutions).The principal outcome is the definition of an organisational ‘worry radius’, detailing the parameters within which risk management must be undertaken.

Stage 5: Assess risks. The aim of this stage is to characterise the risks and risk relationships derived within the previous stage, and to assess the severity of each. Each risk must be enriched with a number of additional attributes; among the most significant are values describing the probability and potential impact of

each, which cumulatively offer a quantitative insight into the overall riskiness of the repository's business activities.
Stage 6: Manage risks. A fundamental imperative with respect to this work is that risks must be managed appropriately. Once a risk has been assessed, a business decision must be made to determine how the risk is to be approached. This should consider the risk's potential impact, its frequency, its owners and its stakeholders. Risk mitigation strategies and tasks should be assigned, with accompanying deadlines for achieving predefined targets. There are several strategies that an organisation can pursue to deal with the negative impact of identified risks. In this Stage, auditors are asked to:

- choose a risk management strategy;
- describe the risk mitigation measure;
- assign responsibility for the risk mitigation activities;
- set target dates and/or results for the risk mitigation activities.

A principal outcome from the successful completion of this stage is a risk register with risk management features included. The risk management exercise cannot and should not stop with the creation of the risk register. Ongoing review and monitoring is essential to ensure that the risk management plan remains relevant. It is therefore necessary to repeat the risk management cycle regularity and review the target outcomes when their deadlines are reached.
Risk management is the final Stage and the end-result of this self-audit. The previous five Stages have created a comprehensive body of information that ultimately informs the risk treatment and management process.
This toolkit was developed as a collaboration between the Joint Information Systems Committee and Core eScience funded Digital Curation Centre (DCC) in the United Kingdom and the European Commission co-funded initiative DigitalPreservationEurope (DPE). These two initiatives will continue to work together to test and refine the toolkit, to manage the online tool, which is available at http://www.repositoryaudit.eu, and to foster its widest possible take up within Europe and broader international contexts.

Conclusions

DPE has recognised that some of the benefits of digital preservation can be identified as the following:

- Legal. National legal frameworks often require organisations to provide adequate records of business processes, communications and many other types of data for many years after their creation.
- Accountability & protection from litigation. Recent legal cases have shown the importance of being able to search and recover archived emails quickly and in a legally admissible manner.
- Protecting the long term view. Access to digital data is critical to ensure business continuity and to support decision making with a long term

view. For research in particular preserving data may be crucial for identifying long-term trends.

- Protecting investment. The valuable intellectual assets of organisations are increasingly in digital form. This data represents both intellectual property and a considerable investment of time, effort and money. It would therefore be foolish not to protect and preserve these assets adequately.
- Reuse. Repositories of digital information and the tools to mine, analyse and re-purpose them represent a society's intellectual capital. Effective and affordable digital preservation solutions are essential to transfer digital data into valuable assets for business.

When PLATTER has been used for planning of objectives, a repository will be in a very strong position to carry out an effective DRAMBORA analysis because all its current objectives will be thoroughly documented. The combination of PLATTER and DRAMBORA therefore represents a powerful tool in the development of Trust.

References

CCSDS (Consultative Committee for Space Data Systems). Reference Model for an Open Archival Information System (OAIS). Blue Book, Issue 1. Washington, DC (US), CCSDS Secretariat, January 2002. Technical report. CCSDS 650.0-B-1. Recommendation for Space Data System Standards. http://public.ccsds.org/publications/archive/650x0b1.pdf.

ERPANET. Workshop on audit and certification in digital preservation. 2004. http://www.erpanet.org/events/antwerpen/index.php.

JISC. Managing risk: a model business preservation strategy for corporate digital assets. 2005.

McHugh, Andrew; Ross, Seamus; Ruusaleep, Raivo; Hofman, Hans. The digital repository audit method based on risk assessment (DRAMBORA), 2007. ISBN: 978-1-906242-00-8. http://www.repositoryaudit.eu.

National Council on Archives. Your data at risk. Why you should be worried about preserving electronic records. 2005.

nestor Working Group on Trusted Repositories Certification. The catalogue of criteria for trusted digital repositories. Version 1. nestor Studies n. 8. 2006.

RLG/NARA Task Force. An audit checklist for the certification of trusted repositories. 2005. http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf.

RLG/OCLC Task Force. Trusted digital repositories: attributes and responsibilities. 2002. http://www.rlg.org/legacy/longterm/repositories.pdf.

Ross, Seamus; McHugh, Andrew. Audit and certification of digital repositories: creating a mandate for the digital curation centre (DCC). // *RLG DigiNews*. Issue index: October 2005. http://www.rlg.org/en/page.php?Page_ID=20793#article1.

Ross, Seamus; McHugh, Andrew. The role of evidence in establishing trust in repositories. *D-Lib Magazine*. July/August, vol.2, nos 7/8. http://www.dlib.org/dlib/july06/ross/07ross.html.

World Bank. Assessment of organisational capacity to manage records: a top level checklist. 2004.