Review

A Domain-Specific Records Management and Information Governance Solution Designed to Support the Implementation of the General International Standard Archival Description

Arian Rajh
Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
rarian_sp@yahoo.com

Krešimir Meze
Omega Software Ltd.
Oreškovićeva ulica 25, Zagreb, Croatia
kresimir.meze@omega-software.hr

Summary

The authors investigate the creation of a records management (RM)-and-archival system based on the General International Standard Archival Description standard (ISAD(G)) and domain specific language (DSL) with RM and archival functions already built-in. This should enable harmonised archival descriptions and better communication between RM subject-experts and developer experts. This should also shift the focus onto solving the creators' needs related to their instances of RM processes and the characteristics of their archival material. The developed RM-and-archival system solution aims to streamline RM business processes and support them through the implementation of specific automation procedures, information technologies like barcode/RFID and additional controls. This RM-and-archival system supports the process of creating a records schedule, a procedure for registering archival material, the retrieval and usage process, storage management, ISAD(G) compliant archival description, disposal processes, inventory checks and the production of XML records with information about archival material. The authors conclude by announcing future work on aligning RM and archival DSL with the MoReq2010 set of services.

Key words: Archival Description, Domain specific language, ISAD (G), Records Management and Information Governance

Introduction

Public and private organisations have to manage large amounts of records that they receive and create while running their business processes. Record management (RM) practice in public organisations usually differs from RM practice in private organisations. Public institutions receive a large number of applications from citizens, private organisations and other public organisations destined for long-term preservation. Public institutions usually have highly demanding workflows and case management procedures, large quantities of additional support documentation and a large number of different types of documents. Processes that circulate documents between working colleagues are not just support processes in the public sector; they have high significance for successful case management in public institutions.[1] In addition, public institutions are under the auspices of archival authorities and they have to manage their records in a way that meets the regulatory environment, mandatory and voluntary practices and archival instructions.[2] The design of records management practice and the selection of RM or DMS/ERMS (Document Management System/Electronic Records Management System) tools have become very important for public institutions.

Successful RM practice should cover all records-related tasks from receiving them to archiving them according to previously developed rules. RM and archiving in the creators' organisations is a domain that comprises registration, storage and arrangement, description, disposal and cooperation with the archival authority. These processes have their own logic and rules. The functions, concepts, terms and business roles in these processes are carrying a specific RM and archival meaning. In order to build a tool that supports RM and archival business processes, it was presumed that it was necessary to build use cases of the processes and a model of the tool using a domain specific language (DSL). RM and archival DSL were built using terms and concepts gathered from professional archival experience and further developed on specialised platform for DSL and system development (the Rhetos platform). This platform generated code, database schemes, a user interface and web services for RM-and-archival system in development. Using Rhetos boosted the development of the system, decreased the misuse of functions and created possibilities for the permanent updating of the newly-constructed DSL. The final result of this development was the RM-and-archive system for the public sector (at first as a prototype and later as a final product Centrix).

[1] Brumec, Dobrović, Tomičić: The Model of the DMS in the Public Sector. Journal of Information and Organizational Sciences, Vol.30, No.1, 2006, page 32

[2] ISO 15489-1:2001, p. 4-5.

Organisational and process-related transformation

Documents and archival records are used and reused in the business processes of public institutions as a basic input of these processes, their result or their supporting material. Since records managers often lack certain descriptors for case- and business-oriented retrievals at the moment of registering or describing archival units, RM and archival system was built as a modular system that can communicate with main business registries via an intermediary. This is consistent with the MoReq2011 concept that endorses the usage of main business applications and RM functionalities.

The RM-and-archival system was developed upon a precise professional DSL and it was made as completely process-oriented software. The main processes used as the backbone during the development were receiving and registering technical units, filing or placement of technical units in storage, archival description, distribution and usage of material, preservation/disposal and collaboration with the archival authority. The RM and archival processes were streamlined during the development of the tool, and their tasks were reengineered in order to support them with information technology. These streamlined processes can be used as an instrument of organisational transformation in institutions with insufficient RM practice.

Defining the retention schedule, adding an archival description and the process of registering the technical units

RM and archival material consist of various technical units, i.e. conventional units like binders or unconventional units like optical disks etc. Technical units are organised into RM/archival units according to the records schedule. The records schedule lists groups and items of records, their descriptions and information on the retention period. It is linked with information on the level of archival description for a particular RM/archival unit, types of retention period start date[3] and permanent preservation/disposal procedures. The retention schedule is a hierarchical classification scheme that governs RM/archival units with the technical units that belongs to them. If a retention schedule is linked to a case management tool, through a case classification number, with accurate metadata on the retention period start date for particular case, and if the technical units are linked to this case, the records schedule should trigger the automation of the retention procedure. The retention schedule could also be linked with the DMS/EDRM and govern archival storage procedures.

[3] After finishing a case, at the end of the year in which the unit has been created etc.

Table 1: A records schedule extended with metadata for the automation of the retention procedure

Ordinal number	Title	Description	Description level	RP	RP start date	Disposal procedure
n.n.	group	description of the group	series	-	-	-
n.n.n.	item	description of the item	subseries	x years	A	destruction
n.n.n.	item	description of the item	subseries	x years	B	preservation

There is some additional metadata related to this kind of grouping of records. Conditions of access could be added to each item of the schedule. When this is combined with the user management service and ascribed to the workflows of the DMS or business registry/applications, it could help the system to govern the access function and related interfaces.[4] Creators could also use this master data to create an information catalogue etc. Records schedules should be customised by the creators' records manager. After defining the records schedule, it is possible to perform the process of archival description and to define the list of RM/archival units. The list of RM/archival unit contains information on existing items in the records schedule. Items in the list represent real RM/archival units and items in the records schedule represent the set of potential units for one creator (and rules for their preservation). Besides providing detailed information for RM/archival units, archival description fills the gap between usually obsolete content-based retention schedules and the contemporary RM/archival units' hierarchies needed for the function-(or process)-based management of records. The functional organisation of records represents the principle of provenance and links together archival holdings, functions and creators. Archival description metadata is gathered by the areas of description compatible with ISAD(G). By adding an ISAD(G)-based archival description, an RM-and-archival system gains an archival search tool. The process of archival description incorporated into the solution follows the multilevel description rule: metadata depends on the level of description, all descriptions are linked together, and the metadata is not repeated through different levels. Although it is intended that information on more general description units (fond, series) will be entered first and then specific units are described, records managers can always add additional information to hierarchically super-ordinated units. Areas of description are: identity statement area, context area, content and structure area, conditions of access and use, allied materials area, notes area and description control area.[5] The integra-

[4] Reference model for an open archival information system (OAIS), Recommended practice CCSDS 650.0-M-2, Magenta book, June 2012, p. 4-16 (part 4.1.1.7), http://public.ccsds.org/publications/archive/650x0m2.pdf, accessed in March 2013

[5] The identity statement area in a Centrix RM-and-archival system comprises identifiers, superordinated units, level of description, unit title, additional title and dates. The context area com-

tion of the ISAD(G) attuned search tool into this system improves the creators' archival management functionality, it bridges the domains of the records manager and archivist and provides a tool for the real archival organisation of emerging material. This system links a records entrance point and archival point (links between technical units and cases from the case management tool). It also facilitates collaboration between creator and archival authority using XML schemes for export of data on archival holdings to the national archival system.
The process of the registration of technical units in the prototyped and developed tool is designed to be flexible enough to enable an optimal workload for records managers. They can choose to register all the technical units if necessary (e.g. if tasks are executed simultaneously and technical units linked to the same case are used by different employees). In the cases of simpler workflows, they can choose to register larger technical units – it depends on the creator's actual RM practice. Registration of technical units starts when an office employee submits technical unit(s) into the archive. The records manager receives it, links it with the item of records schedule and enters metadata. Metadata categories are related to XML schemes of the ARHiNET – national archival information system. The record manager enters the title of the technical unit, year of creation, type and quantity of technical unit, medium/media, preservation-related metadata and a note. Records manager selects an item of the records schedule the technical unit is related to. The system fills in the RP, RP start date, RP end date, submission to archival authority date, availability rules, identifiers and date of entry.[6] The records manager links the technical unit with the item of the existing RM/archival units list. System then gathers and adds quantitative metadata from technical units to registration/archival unit's level and this can be used in an XML export for ARHiNET. The final task of the registration process is to create a label for the technical unit. The label contains the required data but it can also carry a barcode and RFID tag if warehouse management is supported.

prises the name of the creator and its administrative/biographical history, the archival history of the unit, and the source of acquisition/transfer. In the content and structure area is metadata for: scope and content; appraisal, destruction and scheduling information; foreseen additions to the unit; and system of arrangement. The conditions and access of use area lists information on conditions of access, conditions of reproduction, language and scripts, physical characteristic and technical requirements, existing search aids/registration office aids. In the allied materials area, the record manager should provide users with information on the existence of originals (if the unit consists of copies), the existence of copies (if the unit has been reproduced), related units of description, and bibliographies/publications. The notes and description control area consist of a note, a record manager/archivist note, rules and conventions used for creating the description of the unit, and the date of the description/revision of the description. In addition, archival description functionality also provides users with a list of technical units that belong to a particular RM/archival unit.

[6] This basic set of metadata could be extended, which depends on archival material types and the creator's RM requirements. Describing technical documentation, cartographic material, photographs, microfilm, AV material or electronic records requires specific metadata.

Filing process

The filing process should have the basic characteristic and functionalities of a warehouse management system so it can be supported by imager or radio frequency technology device. Records managers should define the hierarchical structure of the creator's storage. When technical units are placed in some organisational unit's storage or in the archive (with the status: archived), records manager pairs location information with the technical unit. Filling process could be enhanced by using barcode or RFID terminals for reading the barcode of the position and pairing it with the barcodes or tags of technical units, and by using sensors and RFID readers placed beside the entrance of the archive port, for controlling what enters or leaves the archive.[7]

The distribution and usage of archival material

When a user retrieves material and sends an order, the record manager finds technical units in the archival storage and takes them back to the shelf after usage. Records management/archive system should track the orders for archival material and their deadlines. System can be customised to synchronise users' orders with taking archival material from the storage (with an RFID port reader and sensors that activate it) and with filing forms for confirmation of taking over the materials by users (with barcode readers). The distribution and usage function is supported by extended retrieval possibilities. This should be done because the typical user is likely to use the case-related descriptor before the title of the technical unit that he requires. However, a typical record manager can't have entered case-related descriptors because they were usually not known. This is why retrieval possibilities are extended by enabling communication between the RM system, case management system and business registry by the use of an *intermediary* agent.

Permanent preservation/disposal process

The records schedule and archival description areas for the particular RM/archival unit contain temporal and event-related metadata required for calculating the retention period end dates (according to purpose-designed algorithms). Some technical units are intended for long term preservation in the creator's organisation and archival authority, and some are destined to be destroyed after the retention period ends. The disposal process starts with listing the technical units that should be destroyed. The system tracks down temporal metadata (date of decision, metadata on the start of the retention period, retention period), calculates the retention period end date and notifies the record manager. After that, the record manager asks the archival authority for approval

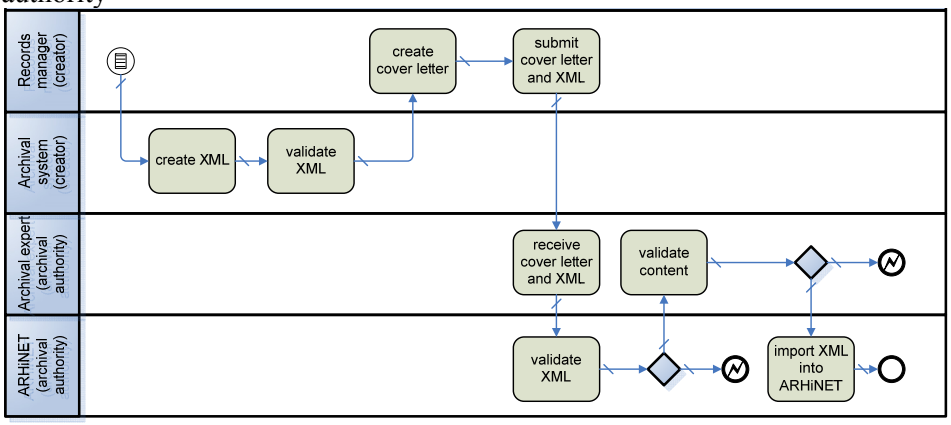
[7] A good example of applying RFID/barcode technologies is an RM project realised in the Croatian Agency for Medicinal Products and Medical Devices in 2013 (the project partners were Omega Software and Selmet).

for the destruction and updates information on what has been disposed of. For archival units that need to be preserved and transferred to the archival authority, the system prepares notifications for creator's records manager and a list for the archival authority (with a temporal trigger for the period set up by the archival legislation).

Collaboration with the archival authority

Legislation requires creators to inform their archival authorities about their RM/archival holdings once a year. This could be done in Croatia directly by using ARHiNET (national archival system) or by submitting an XML document with metadata on RM/archival holdings. This document should provide the archival authority with information on RM/archival units under the creator's care. The business process starts after the event trigger and includes the automatic creation and filing of an XML file,[8] sending a cover letter and the XML file by the creator, content-related and technical validation of the XML file by the archival authority, and the upload of metadata to the ARHiNET.

Business process model 1: The submission of data from the creator to archival authority



A streamlined business process of submitting data for creator's archival material is shown in the model below. Validation in the creator's organisation could be performed by RM/archival systems or XML editors. There are two possible improvements to the model. The first one is related to preventing the creation of invalid XML. The archival system in the creator's organisation could enable the validation of XML at the moment of creation by executing encoding rules. The second improvement is related to building a gateway for transferring XML from the creators' systems to ARHiNET directly.

[8] An XML file is created according to instruction on the ARHiNET web site: http://arhinet.arhiv.hr/_Pages/PreuzmiXMLShemu.aspx, downloaded in April 2013.

The development of the RM-and-archive system for the public sector

Record management is essential to all organizations. The ISO 15489 standard emphasizes that records represent evidence of a business transaction or decision and that they can be used in the pursuance of legal obligations. Despite significant differences in business processes, this applies to both public and private organisations. To fulfil evidential requirements, it is important to assure that every record in an organization is kept as authentic, reliable, integral and usable. It is obvious that an RM database for evidence and ERMS for archival storage can be very helpful in record management practice. In addition to these expectations, a state of the art system must satisfy even more challenges. Legal regulations and professional standards in specific business domains must be met. They usually require a high level of business knowledge from the system architect and developers. Also, each organization or company has its own specific record management needs so the system must be flexible and customizable in order to adapt to them. Due to rapid technological changes, ERMS must now be able to adjust to new trends in a reasonable time frame and within a reasonable cost. Standalone ERMS systems have existed on the market for some time, but customers in the first place focus on solutions that meet their core business requirements and prefer record management functionalities built in line with business applications. The main question is how to produce such a system efficiently?

The key concepts of the development based on the use of RM and archival DSL

The methods, tools and programing languages that are currently widely used in software development are still not immune to errors in the design phase and in performance. This makes IT solutions based on them difficult to develop and expensive to maintain. This is why engineers at *Omega Software* decided to take an approach based on the innovative use of domain-specific languages (DSL) for prototyping and building a RM and archive system. Domain specific language is a kind of language based on terms, keywords and rules originating from some problem-specific domain. DSLs standardize and simplify information exchange between actors in specific domain. Probably the best known example of a DSL term is the H_2O formula. This is a term written in chemical DSL. DSLs are not reserved for purely scientific domains though. They are widely used in applied professional and business areas. Experts in finances use keywords like *invoice* or *payment*. In the governmental domain, we can encounter *case* or *file* and in record management we use *record*, *aggregation* or *class*. These are all well-defined professional concepts with a clear and strict meaning and associated business rules, so they can also be used as concepts in programming languages. Languages based on concepts taken from specific domains are called domain specific programming languages. This means that programmers have abstract concepts for describing the models they want to build.

These concepts are already integrated into business domains so they could be used for information exchange between the programmer and future software users. Hence domain specific languages mainly consist of understandable functions and data types they could significantly simplify communication between development engineers and users. A cleverly designed programming language delivers one additional benefit: the main focus of the programmer is now shifted from *how to do something* to *what to do precisely*. Beside definition, the DSL concept brings certain functionality and business behaviour from the domain it pertains to, so strong expertise in business domain is no longer required for a software architect and developer. Concepts themselves carry behaviour compliant with business rules.

There are two prerequisites for the development of systems based on the above stated paradigm. Firstly, a domain specific programming language and its concepts should be designed. Programmers will later use these concepts to describe the actual business data model with entities, metadata and business rules in DSL scripts. DSL scripts are plain text files that are easy to read and edit in any text editor. Secondly, the DSL execution platform should be implemented. The execution platform is a software component that reads DSL scripts written by the programmer and generates server application, database and web services used for communication with other systems and applications. In particular, the application can be a web or mobile user interface for interaction with human users or another line of the business system. Communication between the application and server side web services happens through the REST (Representation State Transfer) or SOAP (Simple Object Access Protocol) interface. REST and SOAP are industrial standards for message exchange between IT systems and applications.[9] By separating the business logic to a server application, the consistency of data and the execution of business rules are obtained. Any number of client applications can then use services that the server application provides, and none of them need to implement any additional business logic. All data validation procedures, security and authorization rules are also executed in the server application. *Omega Software* has implemented its own execution platform called *Rhetos*. The *Rhetos* platform was constructed using several basic DSL principles. The main purpose of starting the project was the overall desire to improve the quality of the application design process, the solutions themselves and the achievement of long-term support (extension, adjustment to specific businesses, convergence with the general and specific norms of the domain etc.). Applications built on the *Rhetos* platform are not constrained to one particular domain or one particular domain specific language. *Rhetos* can be used for the creation and execution of domain specific languages for any domain and subdomain. This platform is extremely extensible, so programmers can add new concepts

[9] For additional information about these industrial standards, please see: http://www.w3.org/TR/soap/ and http://www.w3.org/TR/ws-arch/ (accessed June 2013).

tied to any business domain. The whole internal infrastructure (e.g. database management system) can be changed in *Rhetos* from the default to any other required by the customer.
Rhetos platform is published as an open source platform. This enables users to fully grasp the content of the platform and to participate in design and changes. The open source *Rhetos* platform brings several meaningful benefits to implementers and users. Most significant is reducing the price of development and facilitating community support. Furthermore, it increases security and user awareness of the organization and implementation of corrections and upgrades, without the need to consult the original implementer of the software solution.
The next step in development will be standardization with MoReq2010, the latest industrial standard for DMS/ERMS. This was written by a pan-European community interested in Information Governance across Europe. MoReq2010 aims to provide a comprehensive but easily understood set of requirements for an RM system that is intended to be adaptable and applicable to divergent information and business activities, industry sectors and types of organization. MoReq2010 is the third version of the MoReq specification and it enables the certification of business applications with standardized RM functionalities, in addition to the certification of standalone ERMS systems. In the future, it is planned to implement the MoReq2010 services and concepts as a particular DSL that will be used for extending other additional business DSLs with harmonized RM functionalities. This will also bring benefits to independent software vendors, especially SMS companies. Although the MoReq2010 standard is written to be simple, implementing the MoReq2010 requirements is a complex and expensive task to do. The MoReq-based DSL will be a useful extension for providing business solutions with standardized record management functionalities.

Conclusion

The new approach used for the development of an RM-and-archive system included the creation of model and system by using RM/archival DSL for most domain-specific processes and by using the ISAD(G) international standard for the implementation of archival descriptions. During this development process, it was shown that using DSLs enables the creation of systems that already contain domain-related professional logic. The usage of ISAD(G) enables the creation of description of archival material according to immanent archival instructions. Finally, Moreq2010 compliant records management domain specific languages could be the basis for the creation of records management and archival systems that would fit a large variety of organisations.

References

ARHiNET Croatia, http://arhinet.arhiv.hr (April 2013)
Brumec, Dobrović, Tomičić: The Model of the DMS in the Public Sector. Journal of Information and Organizational Sciences, Vol.30, No.1, 2006
ISAD(G): General International Standard Archival Description, 2nd ed.
ISO 15489-1:2001 Information and documentation - Records management - Part 1: General
ISO 14589-2:2001 Information and documentation - Records management - Part 2: Guidelines
MoReq2010 Modular Requirements for Records Systems, http://moreq2010.eu/ (April 2013)
Reference model for an open archival information system (OAIS), Recommended practice CCSDS 650.0-M-2, Magenta book, June 2012, http://public.ccsds.org/publications/archive/650x0m2.pdf (March 2013)
World Wide Web Consortium (W3C), http://www.w3.org (June 2013)