Arch Sci (2015) 15:83–95
DOI 10.1007/s10502-013-9211-z

ORIGINAL PAPER

Digital preservation of knowledge in the public sector: a pre-ingest tool

Tarvo Kärberg

Published online: 13 September 2013
© Springer Science+Business Media Dordrecht 2013

Abstract This paper describes the need for coordinating pre-ingest activities in digital preservation of archival records. As a result of the wide use of electronic records management systems (ERMS) in agencies, the focus is on several issues relating to the interaction of the agency's ERMS and public repositories. This paper indicates the importance of using digital recordkeeping metadata to meet more precisely and at the same time semi-automatically the criteria set by memory institutions. The paper provides an overview of one prospective solution and describes the Estonian National Archives' universal archiving module (UAM). A case study reports the use of the UAM in preserving the digital records of the Estonian Minister for Population and Ethnic Affairs. In this project, the preparation and transfer of archival records was divided into ten phases, starting from the description of the archival creator and ending with controlled transfer. The case study raises questions about how much recordkeeping metadata can be used in archival description and how the interaction of the agency's ERMS and ingest by the archives could be more automated. The main issues (e.g. classification, metadata elements variations, mapping, and computer files conversions) encountered during that project are discussed. Findings show that the Open Archival Information System functional model's *ingest* part should be reconceptualised to take into account preparatory work. Adding detailed metadata about the structure, context and relationships in the right place at the right time could get one step closer to digital codified knowledge archiving by creating synergies with various other digital repositories.

Keywords Digital preservation · Ingest · Pre-ingest · Universal archiving module · Estonia

T. Kärberg (✉)
National Archives of Estonia, J. Liivi 4, 50409 Tartu, Estonia
e-mail: karberg@ut.ee

Introduction

There has been great progress in raising awareness of dealing with long-term preservation difficulties in the public sector. Various organizations and governments have started to build or use some digital preservation system components. The most widely used framework for digital archives is most probably by now the Open Archival Information System (OAIS) as many repositories or preservation tools claim OAIS (CCCDS 2012) influence or compliance: e.g. DSpace, OCLC Digital Archive, METS, and LOCKSS (Ball 2006). As the OAIS model is well known in memory institutions, details relating to it are not addressed here, but it is important to note that the model does not encompass pre-ingest functions (Beedham et al. 2005). The pre-ingest component is a set of processing steps and systems that are orchestrated to facilitate data submission, harmonisation, and enrichment (TMB 2011); according to the experiences of the UK Data Archive (2011), the inclusion of this function within the preservation model has considerable benefits. For example, one of the activities that may be reasonable to perform before ingest is normalisation. If it is necessary to normalise metadata or content, this should be a separate pre-ingest process (Waugh 2007).

There have been several research projects which have investigated the possibilities of improving the ingest process by looking for ways to somehow coordinate the pre-ingest phase more effectively. For example, the European Union Seventh Framework Programme funded PReservation Organizations using Tools in AGent Environments (PROTAGE) set out to investigate the value of software agents to find the suitable submission information package (SIP) format for an agency, determining and applying the correct software to create SIPs. But it was technically too complex, and the prototype did not fully succeed in providing all the intended functionality (NAE 2010).

Recently, a trend has been noted among public memory institutions to look more into the pre-ingest part of the digital preservation process ("pre-ingest" as keyword is appearing more often in the digital preservation literature). For example, Nancy Deromedi says that the "next generation archives" requires developing new work procedures and policies for digital acquisitions (Deromedi 2011). The preservation policy of the UK Data Archive states that a successful application of the pre-ingest process also helps ensure that data collections are submitted at a standard which requires a lower level of processing at the ingest stage and potentially greater levels of usability through the provision of adequate documentation (UK Data Archive 2011). One of the key observations for the future of digital preservation is a slow shift from addressing questions that help to fix problems in maintaining digital information over time to ensuring that the problem will not appear in its full complexity in the first place, reducing the need for specific ex-post fixing (Strodl et al. 2011).

One stage where pre-ingest processes can be carried out is when information is in electronic records management systems (ERMS). As there is a huge amount of metadata already created during the course of the records' lifespan, it is important not to lose the metadata that has value in the archiving process. To meet the criteria set by the memory institutions (Pullonen et al. 2009; NAE 2008), the ingest part of

the archive must be standardised, but on the other hand, it requires flexibility in order to deal with the different records management systems that are used in agencies.

According to a survey carried out in 2011 by the Estonian Information System's Authority and the Ministry of Economic Affairs and Communications (RIA and MKM 2011), there are 11 large-scale ERM systems available in the Estonian public sector. The national records management metadata standard of Estonia (based on EVS-ISO 15489-1 (2004) and EVS-ISO 23081-1 (2006) standards) was issued in 2006. Despite the fact that a national records management metadata standard for Estonia has been available for more than 5 years, it is not widely used. There are many reasons (e.g. the ERMS was established before the standard came out and implementing the standard at a later stage requires too many extra resources; agency-specific metadata are not included, etc.) for explaining the current situation, but the consequence is that the usage of different metadata sets results in poor interoperability between agencies and memory institutions' repositories. As ERMS use variable (proprietary, custom) metadata sets and do not necessarily use standards, it is very important to pay special attention to the arrangement and description of records during the preparation and transfer from various ERMS to the digital repositories.

The pre-ingest phase is a crucial point for digital preservation because it influences all subsequent activities. It is also one of the best times to enrich the information and raise it to a new level, where it is already possible to speak of digital codified knowledge as it could become more than just raw data or pure information when the details about the structure, context, and relationships are added. So it is crucial to have good-quality records' metadata prepared at the earliest possible stage. Memory institutions should help agencies in meeting the requirements specified by regulations. This article covers the practical experience of the Estonian National Archives (NAE) in proving that concept. The paper begins by describing the specially designed pre-ingest tool, the universal archiving module (UAM), followed by a case study of its use by an Estonian government department.

The pre-ingest tool and general usage

Estonian National Archives has created a free, multilingual, universal archiving module (UAM) that allows agencies to prepare records for archiving. The analysis phase was completed in 2007, and the main development and testing process took place in 2008–2009. The creation of the UAM required significant effort as it required the analysis of usage of ERM systems in the public sector and the design of accurate rules and approaches for the pre-ingest phase.

The UAM is intended to be installed on the agency's archivist's computer and can be flexibly configured to be used with specific records management systems.

As shown in Fig. 1, some large agencies simultaneously use several systems for managing records (e.g. ERMS X1, ERMS Xn). These systems may also have different metadata sets (M_{X1} and M_{Xn}) and contain a variety of computer file formats (F_{X1} and F_{Xn}) that do not correspond with the metadata set and file formats

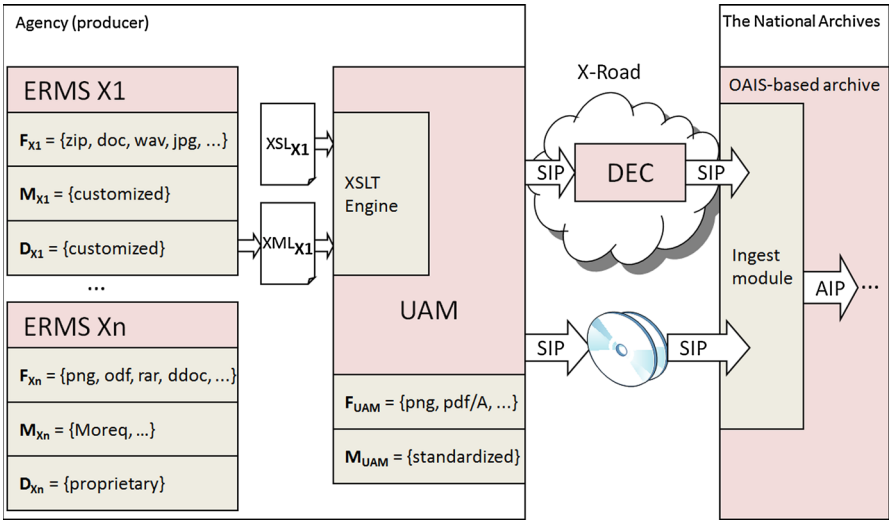

Fig. 1 Schematic outline of pre-ingest activities

specified in the UAM. Therefore, the main criterion for an ERMS is the ability to provide records and metadata in extensible mark-up language (XML) format. It is not necessary to dispose of records in a specific format as it is possible to use the built-in functionalities of ERMS systems (see D_{X1} and D_{Xn}, Fig. 1). The UAM can translate exported XML files with a built-in XSLT (EXtensible Stylesheet Language Transformation) engine into a suitable format. The XSLT is a core technology for XML processing designed to effectively transform XML documents. It is possible to change the metadata (e.g. M_{X1} and M_{Xn}) structure that is used in ERMS by mapping the ERMS output to the UAM's input scheme. Below you can see one simple example of how it is possible to transform an item named *<DocName>* to a UAM item named *<Title>* using XSLT.

```
<o:Title>
  <xsl:value-of select=''i:item[@name='DocName']''/>
<o:Title>
```

It is assumed that ERMS vendors will do the initial mapping between the ERMS export and the UAM's import schemes. The mapping does not require any changes to ERMS source code, but it does need in-depth understanding of the elements of the ERMS metadata set. If the agency does not use a standard configuration of ERMS, additional mapping configuration will be needed.

The UAM uses the format identification, validation, and characterisation tool JHOVE (JSTOR/Harvard Object Validation Environment) and enriches computer file level metadata automatically with additional information (NAE 2013). If the UAM finds computer files while importing that are not in an archival format, it tries to migrate them into the format defined in the match table. For example Microsoft WORD DOC files are converted into PDF/A. The migration tools that are used are

OpenOffice converter with the addition of .NET Bitmap for images. As the migration functionality may not know how to deal with all different file formats that are imported into the UAM, it is reasonable to convert those files into accepted formats beforehand or initially create files in widely used open source formats.

Transfer to the NAE or to some other institution can be done manually (e.g. saving SIPs on DVDs) or using the secured Internet layer X-Road (Estonian Information System's Authority 2012b) and document exchange centre (DEC) (Estonian Information System's Authority 2012a). The packages are delivered to NAE for ingest, where additional checks are made before the final transfer to the digital repository.

It is possible to configure the output by using a built-in XSLT engine for changing the UAM's export elements. One can modify the structure or semantics of the exported XML files very flexibly by setting different XSL transformation files.

The tool allows the structure of records to be arranged (e.g. merge or add new sub-series to the classification scheme, and add additional descriptions) by the originator if the collection is semi- or completely unorganized. It is also possible to check whether the data that is prepared for archiving meet the rules set by the archives institution or not.

The UAM provides built-in functionality to check the validity of descriptions and classification schemes. If there are any issues, a list of warnings and errors will be displayed. Items on the list can be accessed directly, so this provides a fast and convenient way to correct issues. It is possible to add or change some metadata values. Computer files that are not in an archival format can be replaced by migrated files. To ensure authenticity, original and converted files are automatically marked and they are both preserved. The archives institution cannot guarantee the usability of the files that are not in archival format over time, but they can provide for the availability of those files over time. This means that bit streams and descriptive metadata are preserved, but not the knowledge of how to find the right software for correct opening and representation. So-called original files are kept mainly for two reasons. Firstly, the original format may be preferred by researchers. The second reason is to provide additional insurance as file formats become obsolete over time and new migrations are likely to be needed. As every migration carries a risk of causing some loss in information quality, it is more reasonable to migrate from the original file.

Case study

On May 21, 2009, the Office of the Minister for Population and Ethnic Affairs (Rahvastikuministri büroo) was closed. The tasks and responsibilities of that office were divided among other ministries of Estonia. The Government Office (Riigikantselei) whose mission is to support the Government of the Republic and the Prime Minister in policy drafting and implementation was responsible for organising the process of transferring the ministerial office records to the NAE. Appraisal was carried out in April 2010 and identified records of archival value from the period 1997–2009. Digital records were mainly in electronic records management systems,

so it was decided to use the UAM. The Government Office had some knowledge about the UAM already as personnel there had participated in the development and configuration of the import interface of the UAM to increase flexibility.

The mapping between the ERMS and the UAM's metadata model was carried out collaboratively between the NAE and the Government Office. As the ERMS metadata set is proprietary, it is not possible to show the exact mappings of metadata in this paper. However, the mapping process consisted of two parts. Firstly, deep analysis of the ERMS elements was carried out, which included defining common semantics, structure, and granularity of elements. After this logical mapping to the UAM, the second stage was a technical mapping using XSL transformation capabilities of the UAM import module.

The project was planned as a 10-step workflow (Fig. 2) inspired by earlier paper-centric processes. This workflow is designed to be flexible and universal, so it can be used for paper, digital, or hybrid material archiving. Ten main steps are the maximum that need to be carried out—it is necessary to go through all steps only for the first transfer. For subsequent transfers, it is possible to use the previous project as a base and start from step G skipping the previous time-consuming steps.

As the ERMS did not contain descriptions from the archives authority, it was necessary to add them manually into the UAM (step A, Fig. 2). Descriptions of records creators—the Government Office (Riigikantselei) and the Office of the Minister for Population and Ethnic Affairs (Rahvastikuministri büroo)—were created using the Estonian customisation of ISAAR (CPF) elements (International Council on Archives 2004) (Fig. 3).

As the archival description process required research, namely working through series of regulations and manual data entry into the UAM by archivists, it took a lengthy period of time. Fortunately, descriptions of record creators change rarely, and if they do, only some updates are needed as it is possible to reuse earlier created descriptions in subsequent transfers of archival records.

The next step was the description of an archival project (step B, Fig. 2). It was a fairly simple task as the archival project function in the UAM is intended to group records that are transferred to the public archive at one time, and description consists of only a few elements. After creation of the archival project, it was possible to start building an archival classification scheme based on the appraisal decision (step C, Fig. 2). The provisional classification scheme was created

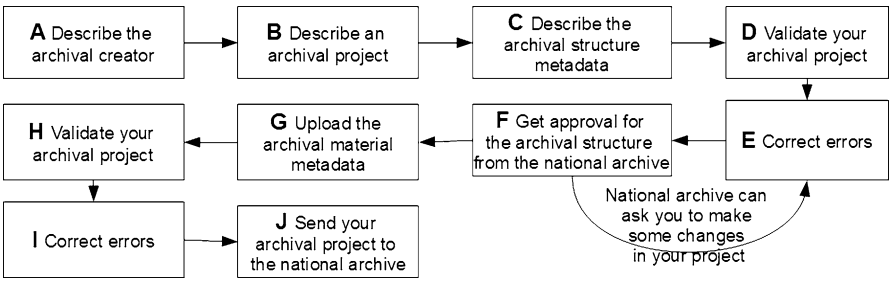

Fig. 2 Project's workflow

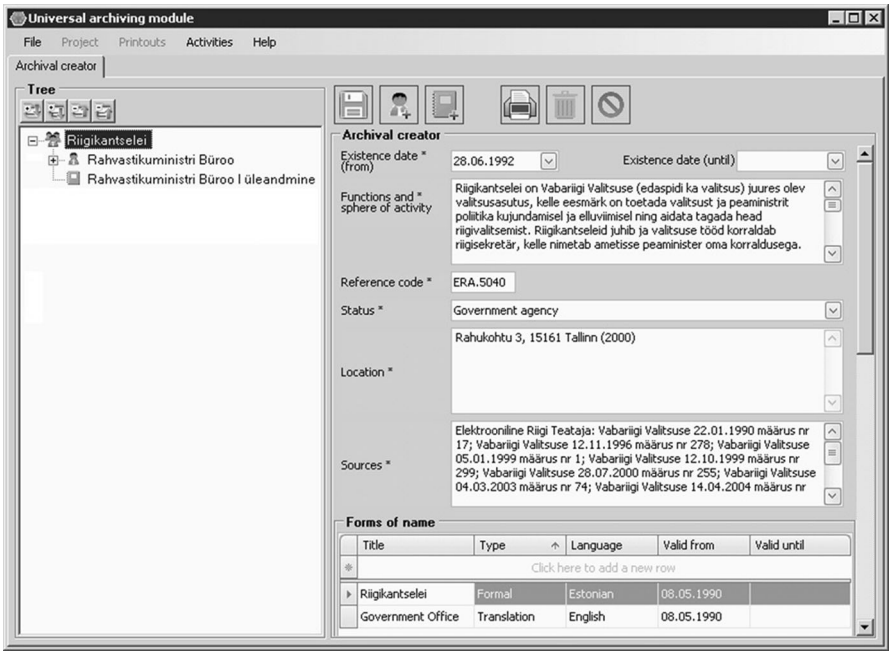

Fig. 3 A screenshot of the UAM

automatically by importing the metadata and structure from the ERMS. Some changes were made in the classification scheme's structure by merging series and adding subseries in cooperation with the NAE. Detailed descriptions can be seen in log files as all actions were automatically logged by the UAM. Some additional metadata were also identified and added. For example, in the ERMS, functions were described, but not mandates. The ERMS metadata did not include the "scope and content" element suggested by ISAD(G) (International Council on Archives 2000). This metadata therefore was manually added to the UAM by the Government Office's archivist.

The final version of the classification scheme encompassed correspondence, the Government Commission's records and surveys series that belonged to the jurisdiction of the Minister for Population and Ethnic Affairs. After dealing with all issues, the archival project was successfully validated using the built-in UAM validator (steps D&E, Fig. 2). The scheme was approved by the NAE, and permission for organising records was given (step F, Fig. 2).

It was agreed in step G (Fig. 2) that all records must be arranged into dossiers. Even digital records that were not in the ERMS were loaded into the UAM in order to guarantee their completeness. Some paper records were digitised and added into the appropriate dossiers as PDF files. Although hybrid series were allowed, creating mixed (digital records with paper) dossiers was strongly prohibited. In the interests of providing better usability to future users, only fully digital or completely paper-based dossiers were permitted. Each paper record description was added to the

UAM by importing it from the ERMS if available or creating manually. It is important to note that all the description elements were based on the general international standard archival description ISAD(G) standard enriched with some national-specific elements (e.g. digital signature metadata). It would have been possible to import some additional non-standard elements as the UAM includes a special "xs:anyType" schema element, but it was not necessary in this case.

In the import process, many different file formats were encountered (e.g. DDOC—a digital signature format specific to Estonia, DOC, PDF, CSV, TXT, XML, and RTF). The UAM selected the appropriate action by using predefined migration rules. The migration of files was done automatically and both versions were added to the UAM's project hierarchy. The main rule declared that the file that was present at the migration starting point and the last result of the migrations were always retained. No results from the middle of the migration path were kept. When a file or a group of files were digitally signed, the UAM read the information of the signer(s), role(s), and signing date(s). This information was automatically placed in the record's signer metadata section. The digitally signed capsule was encapsulated and the file(s) inside migrated into the archival format if needed.

The computer file's metadata was automatically extracted by JHOVE, so no additional descriptions were needed on that level. Also the information that a file is migrated or that it is a result of migration was added to the computer file's metadata. Technically speaking, every computer file always had three additional Boolean flags: Original, Final, In archival format. All non-migrated files were marked "Original = True" and the rest were assigned the value "False". Files that were in archival format were flagged "In archival format = True" and the rest "False". If the file migration process failed, the file was marked "Final = True", but "In archival format = False". This meant that the UAM conversion capabilities were insufficiently amended, and there was a need to repeat the migration process.

After the automated validation in step H (Fig. 2) and some error corrections in step I (Fig. 2), the final archival list was generated and sent to the NAE. The list of files received official approval, and on December 10, 2010 properly arranged and described records were sent through X-Road over DEC to the NAE followed by controls and the verification process on the NAE's side (step J, Fig. 2).

Discussion

The UAM has successfully carried out the digital preservation project as a total of 302 dossiers with 493 computer files of records of the Office of Minister for Population and Ethnic Affairs were transferred to the NAE through the UAM. There were 110 digital and 192 paper-based dossiers. All descriptions were in digital standardised form.

This project successfully confirmed that it is possible to reuse metadata from ERMS or content creation systems in archival description and achieve more detailed archival descriptions than before. The concept of using information packaging and standardising according to the NAE's metadata scheme, public legislation, and the

national data exchange layer rules was successfully demonstrated as it provided a good opportunity to validate the SIPs and perform the transfer very smoothly and securely.

The project showed that the proposed workflow is sufficiently universal to be used for hybrid material transfer as there were hybrid series. This case study established that there is no need for more than 10 steps to transfer records. The most time-consuming steps were, as predicted, steps A and G. Processing step A will not be necessary for subsequent transfers as the agency description has already been created.

This case study showed that the input part of the pre-ingest tool must be very flexible to provide detailed configuration. There was intensive cooperation between the Government Office and NAE's archivists, and in addition, intensive interaction between information technology units as the import XSLT configuration of the UAM needed some editing during the project to improve accuracy and efficiency of mapping and transformation. It is important to note that initial transfer projects will require more time.

The Government Office's archivists also provided ideas for the improvement of the UAM's user interface, validation rules, etc. Subsequent transfers will be less problematic; the records creator's metadata and classification scheme can be sent to the UAM in XML format and reused. The agency's archivist will only have to load records and their descriptions into the UAM, validate them, and correct any errors and send the packages to the NAE or other repository in future transfers. Implementation of the UAM has commenced in several other Estonian public institutions because of the success of this project.

Conclusions

Innovative technologies may have learning curves. Although archivists mastered the use of UAM, it took a significant amount of explaining and guiding at the initial phase of the project.

A successful digital transfer requires both "traditional" archivists and information technology specialists. Traditional archivists (archivists who had previous experience mainly with paper-based documents) were now faced with a slightly different situation—on the one hand, they had records which were digitally described, but on the other hand, there were new regulations which needed to be followed. The UAM was designed to support the arrangement model described by ISAD(G) (Fig. 4), and it needed to be mapped with the classification used in ERMS.

In the active phase of records management, the records/items were filed directly into (sub)series. If items were created within one case, they were grouped in a case-file before filing in (sub)series. Archivists decided to unify the archival arrangement by creating new dossiers with more precise descriptions (as file titles were not very informative e.g. "Letter") and treating case-files as dossiers, so that all items could be grouped into dossiers. As the upper levels of archival hierarchy lacked some archival-specific descriptions (e.g. scope and content of series), there was a need to add missing descriptions as well.

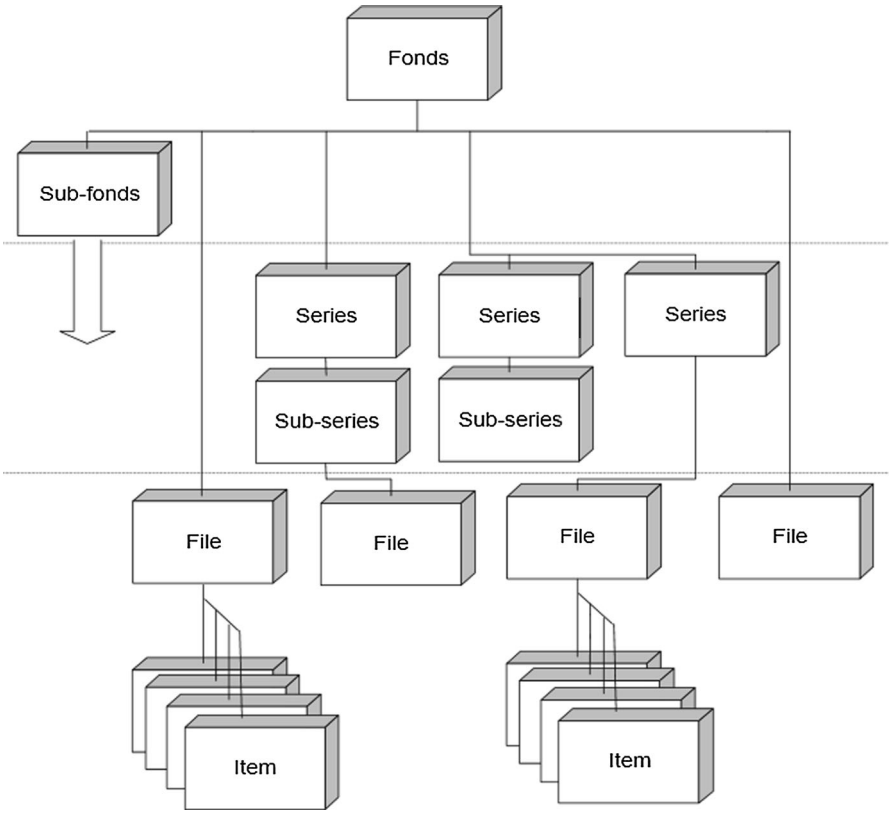

Fig. 4 Model of the levels of arrangement of a fonds (International Council on Archives 2000)

As configuring ERMS and UAM required information technology skills, this was carried out by the information technology specialists. Some of the configurations changed the maximum sizes of metadata fields (e.g. scope and content field was too small for description which was about 1.5 pages of text), designing support for various date and time formats (e.g. DD.MM.YYYY, YYYY-MM-DD, DD.MM.YYYY HH:MM) used all over the ERMS, converting computer files to archival formats if it was not possible to do it with UAM during import. The cooperation with information technology specialists was also important when some technical issues (e.g. OpenOffice converter "hanged" in converting more than 72 pages of a DOC file, and C# method failed to show some nodes of the classification tree when there were many objects) became apparent.

Having detailed understanding about the organisation's ERMS metadata set reduces the interface creation time. As the ERMS output lacked detailed documentation, it was difficult to create an interface between UAM and ERMS in a short time as getting information as to what elements existed and how they were used required a significant amount of time.

Real complexity lies in providing the mapping quality and elements granularity. The metadata mapping was relatively time-consuming because of its low-level complexity as well. As ERMS used many records forms and they had significantly different metadata sets, it was necessary to analyse each form separately.

Metadata mapping may need improvements during the project. The metadata mapping was completed at the initial phase of the transfer project, but some important updates were still required during the project. The most significant update was related to e-mail content. As some e-mails contained HTML formatting (e.g. different font styles, sizes, spaces), their representation was not preserved when they were imported to the content field in UAM as this was a UTF-8 text field. There were discussions about saving the content in PDF format, but that option was rejected as being too complicated solution for the current ERMS. The content of those e-mails, where formatting was important, were transferred to archives preserving HTML tags in UTF-8 text, so it could be rendered as HTML later if needed.

The second update related to an element describing the size of a dossier. The sum of digital records in a dossier was automatically calculated by UAM, but the amount of pages of a paper-based dossier was not included in the ERMS and needed to be added by the archivist. As the amount field did not allow manual insertion, the sum of pages was indicated in the comments field and extracted automatically into the right place during the transfer. In the next version of UAM, it is reasonable to allow the import of specific elements from ERMS to UAM by defining xs: any Type element in XSD (XML Schema Definition) of a dossier level as well. The XSD is a recommendation of the World Wide Web Consortium (W3C), specifying how to formally describe the elements and express constraints in XML documents.[1]

The third major update related to describing access restrictions. The access restrictions metadata elements of ERMS were mapped with elements defined in UAM, but it was not considered that the content of access restriction type element is a classification element. ERMS contained access restrictions types AIA (public, but contains private data—can be accessed only by request), AK (only for internal use), but UAM used different values, namely “contains sensitive and private data”, “restriction set by the transferral”, and “internal use”. The values were translated in XSLT accordingly AK-> “internal use” and AIA-> “contains sensitive and private data”.

So, it does not matter if metadata mapping is performed only for the records set which is intended to be archived in the first transfer or to all relevant metadata sets in ERMS, the interface may need to be configured on subsequent transfers as well. This is because ERMS is an active system and some new records forms may be created or already existing ones modified. Fortunately, the mapping can be updated by modifying the XSLT configuration file.

It is possible to use semi-automated tools in the pre-ingest phase. Using UAM in the pre-ingest phase was quite automated as it provided archivists with a very flexible and controlled way to transfer records and their metadata to archives with minimal human intervention.

[1] http://www.w3.org/standards/xml/schema.

Subsequent transfers can be more automated and efficient. As the Government Office used UAM for their first transfer, they can use the same classification schema and agency's description for future transfers.

Producers could use the UAM to fulfil submission requirements. UAM successfully supported fulfilling requirements described in Estonian archives legislation and criteria set by the memory institutions by validation of the project at various stages.

The project of archiving the records of the Office of Minister for Population and Ethnic Affairs is an important milestone in the mission of the NAE, because it was the first official transfer project that contained digitally born archival records and was coordinated by using the UAM. Despite some issues, it could definitely be said that the project was successful and provided valuable experience to both parties. This project was not only a practical proof of concept, but also gave the opportunity to analyse issues that may arise in the pre-ingest phase of digital preservation. It is also important to continue the theoretical work and practical experiments to reach future researchers' needs by providing them with archived codified knowledge, which is achieved through (inter)national semantic interoperability and by giving comprehensive structure and detailed context to the information. Using pre-ingest tools like UAM makes it possible to raise digital information "to the next level" (to codified knowledge) and prevent those digital resources from becoming "information islands" as the knowledge is dynamic and can acquire new relations with other repositories if they share for example common ontologies. This means that taking the right steps at the pre-ingest stage helps broaden the context of digital resources and allows multi-faceted access as the archival discovery environments capabilities will become much more comprehensive.

Acknowledgments The author gratefully thanks the Government Office, the National Archives of Estonia and Mr Lauri Leht, the Director of the Digital Archives at the National Archives of Estonia. The author gives special thanks to editors in Chief Gillian Oliver, PhD and Elizabeth Yakel, PhD for their valuable support and anonymous reviewers for their constructive comments.

References

Ball A (2006) Briefing paper: the OAIS model. Accessed 05 January 2013 at: http://www.ukoln.ac.uk/projects/grand-challenge/papers/oaisBriefing.pdf

Beedham H, Missen J, Palmer M, Ruusalepp R (2005) The assessment of UKDA and TNA compliance with OAIS and METS standards. Accessed 05 January 2013 at: http://www.jisc.ac.uk/media/documents/programmes/preservation/oaismets.pdf

CCCDS (2012) Reference model for an open archival information system (OAIS), consultative committee for space data systems, Magenta Book. Accessed 05 January 2013 at: http://public.ccsds.org/publications/archive/650x0m2.pdf

Deromedi N (2011) The next generation archives: digital curation at the Bentley historical library, University of Michigan. Accessed 05 January 2013 at: http://bentley.umich.edu/dchome/papers/20111011deromedi_paper_final.docx

Estonian Information System's Authority (2012a) About document exchange centre. Accessed 05 January 2013 at: https://www.ria.ee/dec/

Estonian Information System's Authority (2012b) Data exchange layer X-Road. Accessed 05 January 2013 at: https://www.ria.ee/x-road/

EVS-ISO15489-1 (2004) Information and documentation—Records management. Part 1: general. Accessed 05 January 2013 at: http://www.evs.ee/products/evs-iso-15489-1-2004

EVS-ISO 23081-1 (2006) Information and documentation. Records management processes. Metadata for records. Part 1: principles. Accessed 05 January 2013 at: http://www.evs.ee/products/evs-iso-23081-1-2006

International Council on Archives (2000) ISAD(G): general international standard archival description: adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19–22 September 1999 (2000) 2nd edn. ICA, Ottawa. Accessed 05 January 2013 at: http://www.icacds.org.uk/eng/ISAD%28G%29.pdf

International Council on Archives (2004) ISAAR(CPF): international standard archival authority record for corporate bodies, persons and families Accessed 05 January 2013 at: http://www.icacds.org.uk/eng/ISAAR%28CPF%29.pdf

National Archives of Estonia—NAE (2008) Digitaaldokumentide arhiveerimise nõuded [Archives management requirements for digital records]. Accessed 08 January 2013 at: http://www.ra.ee/public/Juhised/digidok_arhiveerimine.pdf

National Archives of Estonia—NAE (2010) Briefing paper: value of software agents in digital preservation. Accessed 05 July 2012 at: http://www.protage.eu/files/Potential%20of%20agents%20in%20DP.pdf

National Archives of Estonia—NAE (2013) Millised on universaalse arhiveerimismooduli tehnilised parameetrid [Technical parameters of universal archiving module]. Accessed 05 January 2013 at: http://www.ra.ee/et/universaalne-arhiveerimismoodul-2/&i=#256

Pullonen K, Unt R, Uuet L (2009) Arhivaalide üleandmine [Transfer of records]. ISBN 978-9985-858-64-6

RIA, MKM (2011) Ülevaade avaliku sektori toimimisest digitaalse dokumenditöö tõhustamiseks. Uuringu lõpparuanne [Overview of the activities of intensifying digital records management in public sector. The final report]. Accessed 05 July 2012 at: http://www.ria.ee/public/Programm/ddoktoo_uuringu_aruanne.pdf

Strodl S, Petrov P, Rauber A (2011) Research on digital preservation within projects co-funded by the European Union in the ICT programme. Accessed 05 July 2012 at: http://cordis.europa.eu/fp7/ict/telearn-digicult/report-research-digital-preservation_en.pdf

TMB (2011) Key components documented for output of D3.5 e.g. BHL-Europe Portal, OCR demonstrators, distributed storage model, etc. BHL-Europe. Accessed 05 July 2012 at: http://pro.europeana.eu/documents/862740/0/BHL-Europe+D3.7+Key+Components.pdf

UK Data Archive (2011) Preservation policy. Accessed 05 July 2012 at: http://www.data-archive.ac.uk/media/54776/ukda062-dps-preservationpolicy.pdf

Waugh A (2007) The design and implementation of an ingest function to a digital archive. D-Lib Mag, 1: 11/12. Accessed 28 May 2013 at: http://www.dlib.org/dlib/november07/waugh/11waugh.html

Author Biography

Tarvo Kärberg is a PhD student at the University of Tartu where he teaches a course in digital archival holdings and access. He has a Masters degree in information technology. Over the past 5 years he has worked as project manager at the National Archives of Estonia. His research focuses on the theory and methods of collecting and preserving digital codified knowledge.