Helping Archivists to Receive Records From ERMS Into the Digital Archive Using Semi-Automated Adjustable Software

Lauri LEHT, Dr.
Head of Digital Archives, National Archives of Estonia, J. Liivi 4, Tartu 50409, Estonia
Lauri.Leht@ra.ee

Helping Archivists to Receive Records From ERMS Into the Digital Archive Using Semi-Automated Adjustable Software

ABSTRACT

In practical digital preservation one common challenge is how to have good quality metadata with the electronic records that are transferred from electronic records management systems (ERMS) to the public digital archives. It is vital not to loose the descriptive and contextual metadata that is created in the ERMS, but it is not an easy task for public archives. The problem is that many different ERMSs are used in agencies, which has created a variety of technically different outlooks of exported metadata and records. The National Archives of Estonia has created and implemented a sophisticated semi-automated adjustable software tool for agencies to prepare records for archiving – the universal archiving module (UAM). This software can be installed into agency's archivist's computer and configured to use a certain ERMS. The main criterion for the ERMS is the ability to provide records and metadata in extensible markup language (XML) format. UAM can transform exported XML files with the built-in XSLT engine into the unified suitable format. The tool allows rearranging the structure of records and adding additional descriptions to them. It is also possible to check whether the data that is prepared for archiving meets the rules set by the archive institution or not. Transmission to the National Archives can be done manually (e.g. saving submission information packages to the DVDs) or using the national secure internet. The first version of UAM was introduced in 2008, but significant improvements have been made and in the beginning of 2011 a new version of UAM was launched, which is also available in English. UAM has been successfully tested and launched in a couple of electronic records transfer projects in Estonia and is being implemented in all of the ministries' offices of Estonia. It is also used extensively by the archivists in the National Archives for rearranging descriptions of existing archival materials and supporting transfers of paper records. In the presentation a deeper look is given to the functionalities of UAM and to the problems and possibilities of implementing such a revolutionary tool for archivists.

Aiutare gli archivisti a ricevere dati dal sistema ERMS nell'archivio digitale mediante l'utilizzo del software semi-automatico regolabile

SINTESI

Nella pratica della conservazione digitale una sfida usuale è ottenere metadati di buona qualità con i record elettronici trasferiti da sistemi di gestione di documenti elettronici (ERMS) ad archivi pubblici digitali. E' fondamentale non perdere i metadati descrittivi e contestuali che si creano nel sistema ERMS, ma non è un compito facile per gli archivi pubblici. Il problema è che le agenzie utilizzano ERMS differenti, il che ha creato una varietà di prospettive tecnicamente differenti di metadati esportati e documenti. L'Archivio Nazionale dell'Estonia ha creato e implementato un software sofisticato, semi-automatico ed adattabile per le agenzie che preparano i documenti per l'archiviazione - il modulo universale di archiviazione (UAM). Questo software può essere installato nel computer dell'archivista dell'agenzia e configurato per l'utilizzo di un determinato sistema ERMS. Il criterio principale del sistema ERMS è la capacità di fornire dati e metadati in Extensible Markup Language (XML). Lo UAM è in grado di trasformare i file esportati XML con il built-in del motore XSLT nell'adatto formato unificato. Lo strumento permette di riorganizzare la struttura dei documenti e di aggiungere loro ulteriori descrizioni. È anche possibile controllare che i dati preparati per l'archiviazione soddisfino o meno le regole stabilite dall'istituzione archivistica. La trasmissione agli archivi nazionali può essere effettuata manualmente (ad esempio salvando i pacchetti informativi su DVD) oppure utilizzando il sistema internet nazionale protetto. La prima versione di UAM è stata introdotta nel 2008, ma sono stati effettuati significativi miglioramenti ed all'inizio del 2011 è stata lanciata una nuova versione di UAM, disponibile anche in lingua inglese. UAM è stato testato con successo e lanciato in un paio di progetti di trasferimento di documenti elettronici in Estonia, ed è in corso di attuazione in tutti gli uffici dei ministeri estoni. È anche usato ampiamente

dagli archivisti dell'Archivio Nazionale per riorganizzare la descrizione del materiale d'archivio esistente e per supportare i trasferimenti di documenti cartacei. Nell'articolo viene dato uno sguardo approfondito alle funzionalità di UAM ed alle problematiche ed alle possibilità di sviluppo di tale rivoluzionario strumento per gli archivisti.

Pomoč arhivistom pri prenosu dokumentov iz ERMS v digitalni arhiv s pomočjo polavtomatske prilagodljive programske opreme

IZVLEČEK

Pri praksi digitalne hrambe predstavlja skupen izziv način, kako ohraniti kvalitetne metapodatke pri prenosu dokumentarnega gradiva iz elektronskih sistemov pisarniškega poslovanja (ERMS – Electronic Records Management System) v javne digitalne arhive. Pri tem je ključno, da se ne izgubi metapodatkov, ki se navezujejo na kontekst ter opis, in nastajajo v ERMS-ju, kar je seveda težak podvig za javne arhive. Problem je v tem, da institucije uporabljajo različne EMRS-je, kar predstavlja množico tehnično različnih pristopov pri izvozu metapodatkov in dokumentov. Estonski nacionalni arhiv je ustvaril ter uvedel prefinjeno polavtomatsko prilagodljivo programsko opremo za institucije, s pomočjo katere le-te pripravijo dokumente za arhiviranje – univerzalni modul za arhiviranje (UAM – Universal Archiving Module). Programsko opremo je mogoče naložiti na računalnik arhivarja določene institucije ter prilagoditi za uporabo določenega ERMS-ja. Glavni kriterij za ERMS je možnost posredovanja dokumentov ter metapodatkov v zapisu XML. UAM lahko pretvori uvožene XML zadeve z vgrajeno XSLT napravo v primeren enoten zapis. Orodje omogoča urejanje strukture dokumentov ter dodajanje popisnih enot. Prav tako je mogoče preveriti ali podatki, ki so pripravljeni za arhiviranje, ustrezajo pogojem, ki so jih postavili pristojni arhivi. Prenos podatkov v Nacionalni arhiv je možen v fizični obliki (npr. s pomočjo zapisa informacij na DVD) ali s pomočjo nacionalne varne povezave. Prva verzija UAM je bila izdana leta 2008, nova izpopolnjena verzija, ki je na voljo tudi v angleškem jeziku pa na začetku leta 2011. UAM je uspešno prestal preizkušnjo z uvedbo v nekaj projektov elektronskega prenosa podatkov v Estoniji, sedaj pa ga uvajajo v pisarniško poslovanje vseh estonskih ministrstev. Redno ga uporabljajo tudi arhivisti v Nacionalnem arhivu pri urejanju popisov obstoječega arhivskega gradiva ter kot podporni program pri prenosu dokumentov v fizični obliki. V predstavitvi sledi večji poudarek na funkcionalnosti UAM-ja ter problemih in možnostih uvajanja takšnega revolucionarnega orodja za arhiviste.

Dokumentide kandmine EDHSist digitaalarhiivi arhivaaride kasutatava poolautomaatse kohandatava tarkvara abil

ÜLEVAADE

Digitaalse arhiveerimise üks olulisi väljakutseid on tagada elektroonilistest dokumendihaldussüsteemidest (EDHS) koos dokumentidega ka hea kvaliteediga metaandmete avalikku digitaalarhiivi üleandmine. Probleem seisneb selles, et riigiasutustes kasutatakse paljusid erinevaid EDHSe, milles on dokumentide ja metaandmete väljundid tehniliselt erineval kujul. Eesti Rahvusarhiiv on loonud poolautomaatse kohandatava tarkvara asutustele dokumentide arhiveerimise ettevalmistamiseks – universaalse arhiveerimismooduli (UAM). See on abivahend arhivaaridele dokumentide EDHSist eraldamiseks, ümber korrastamiseks, puuduvate metaandmete lisamiseks, andmete valideerimiseks ja XML-kujul avalikku arhiivi saatmiseks. UAMi juurutatakse kõigis Eesti ministeeriumites ja see tarkvara on muutnud ettekujutust praktilisest digitaalarhiivindusest Eestis.

In Estonian government offices electronic records management systems (ERMS) were widely taken into use in the end of 1990s and shortly after year 2000. At the time it was said that it is a matter of free competition between IT-companies and every ministry can choose an ERMS that it wants. Therefore no central ERMS was acquired for the entire public sector, although it would have been probably economically feasible considering the size of Estonia (1.3 million people) and its government. But it would have been most reasonable considering the future needs of digital archiving of the records. Anyway, more than 10 different ERMSs with more or less different metadata schemas are in use in the public offices currently. And even if there is the same ERMS in use in two agencies, most probably the system is installed and introduced using not identical metadata schemas which makes extracting and archiving records a difficult task.

From 2007 the National Archives of Estonia started preparing itself for the task of archiving digitally born records. In Estonia there are real digital original records, as the national electronic signature was launched in 2002 and since 2008 public agencies can only exchange digitally signed documents between each other. It is possible as all Estonian citizens must have the national ID-card which has a chip inside and is used for giving the electronic signature using the computer's card reader.

In most other countries the problem of unstandardised extractions from ERMSs that are unsuitable for digital archiving by the central public archives is solved by an imperative that allows only transferring records with standardised structure and metadata schema. How each of the agencies' ERMSs deals with it, is not the problem of the central archives. But actually it is very much the problem of the archives as this way the central archives probably receives very few electronic records or often records with not compliable quality that have to be redone by the agency several times.

The digital preservation team of the National Archives of Estonia invented a different approach that is more efficient in addressing the problem of receiving good quality digital records from public agencies to the central digital archive. For that a new desktop software was developed which tackles several aspects of the issue. First, the software is meant to be used by the agency's archivist who actually has the responsibility of transferring records to the public archives. Second, the software has a configurable import component which means it can handle all kinds of different export variations from ERMSs. Third, it exports data into the national standardised record transfer capsule and sends it directly to the National Archives digital archives' ingest module via a secured public internet route. Fourth, it allows the responsible agency's archivist to check and rearrange records' metadata and structure before transfer. This software was called the **Universal Archiving Module** or **UAM** for short.

Technically speaking, UAM is a desktop software tool which is programmed in the .NET programming language. It should be installed at a computer in the public agency, most suitably to the archivist's or record keeper's computer. It is free for use and is available in Estonian and in English. All the necessary download and installation files and information about UAM is available on the web page of the National Archives of Estonia: http://www.ra.ee/en/universal-archiving-module/.

UAM has a configurable import engine which enables agencies to create mapping tables from almost any ERMS export structure by using XML technologies (XSLT, XPath). There is a native input XML schema in use inside UAM which is a de facto national standard set by the National Archives. All the exported records and their metadata from the ERMS are transformed into the standard XML schema. The only prerequisite is that the ERMS has to be able to export data in XML, one way or another.

In the core functions of UAM there are all the relevant technical and archival requirements that are necessary to successfully prepare records and their metadata for a formal transfer of archival records from an agency to a public archives. The user of UAM can

- find gaps in imported metadata,
- input missing mandatory elements,
- identify imported records' components (computer files) and compare them against a list allowed archival file formats,
- automatically create technical metadata,
- convert non-conforming computer files into archival formats,
- etc.

All actions on metadata and file formats are logged and it is possible for users to create reports on the current status of works at all times. If all necessary requirements on archival and technical metadata and file formats are met it is possible to create the submission information packages (SIPs) for transfer to the long-term digital archives.

The standard installation of UAM creates SIPs in the national XML structure which is defined by the National Archives. However, if a user wants to use UAM to transfer its records to any other digital archives with a different XML schema it is possible to use the UAM XSLT engine to transform the standard SIP into other structures (e.g. METS or other national transfer package standards).

Two external components are used in the UAM. In order to carry out computer file identification and to extract technical metadata a good use is made of the JHove tool. To convert files from common office formats into preservation formats (currently PDF) the migration component of OpenOffice is used.

It is important to notice that UAM itself is not a long-term archive where records can stay. It is only an interim environment after export records from the ERMS and before transferring to the long-term archives.

UAM is only meant to be used for records in XML structure from ERMSs. It is not a tool for managing and transferring extractions from databases.

There are several aspects why the Universal Archiving Module is a very useful tool for the agencies' archivists for transferring records to the public archives:

- it enables the archivist to have a full control of the transfer process of technically complicated digital records, although he or she is usually not a technical person,
- it can be used also for transfers of paper or hybrid archives as there can be only metadata of the records without any existing computer files,
- it allows automatic validation of data for discovering errors before transfer,
- it allows automatic creation of inventory list, archival schema and other official documents,
- it allows adding upper-level data to the transfer project which usually does not come from the ERMS.

For the National Archives UAM is the first component in the set of its long-term digital preservation software. It is the pre-ingest tool that allows the archives to receive transfers from public agencies of as good quality as possible. For the archivists of the National Archives UAM has the following impact:

transfers with good quality of data from the agencies are received,

- from the point of view of the national archives records are described and metadata checked at the earliest realistically possible time when the archives is getting control of it,
- UAM is also used by the archivists of the National Archives for rearranging and additional descriptions of the archival fonds that are already in the archives or are received outside UAM,
- descriptive metadata coming in from the UAM will be re-used in the public catalogue system of the archives (archival information system) and no more manual entering of data is necessary.

In practice, UAM is used for several practical transfer projects since 2010. Legal acts from the 1990s were transferred from the Tartu City Government. Some parts of the archives of the Government Office of Estonia are transferred. Several transfers are being prepared. The National Archives is in constant contact with all the ministries and is helping them to finalise the UAM import interface. Transfers of digital records have been of ad hoc nature so far as the transfer period of both paper and digital records to the National Archives was 20 years for a long time.

The flow of transfers of digital records from agencies is about to begin within a couple of years as with the new Archives Act of 2011 the transfer period of records of archival value to the National Archives was shortened to 10 years. The National Archives' digital archives were also given the obligation to serve as a mid-term digital archive for government agencies' long-term digital records (documents with a retention period of over 10 years). The National Archives of Estonia is ready for the challenge with the Universal Archiving Module being the most notable part of the digital preservation solution for all the Estonian government agencies.

SUMMARY

In practical digital preservation one common challenge is how to have good quality metadata with the electronic records that are transferred from electronic records management systems (ERMS) to the public digital archives. It is vital not to loose the descriptive and contextual metadata that is created in the ERMS, but it is not an easy task for public archives. The problem is that many different ERMSs are used in agencies, which has created a variety of technically different outlooks of exported metadata and records. The National Archives of Estonia has created and implemented a sophisticated semi-automated adjustable software tool for agencies to prepare records for archiving – the universal archiving module (UAM). This software can be installed into agency's archivist's computer and configured to use a certain ERMS. The main criterion for the ERMS is the ability to provide records and metadata in extensible markup language (XML) format. UAM can transform exported XML files with the built-in XSLT engine into the unified suitable format. The tool allows rearranging the structure of records and adding additional descriptions to them. It is also possible to check whether the data that is prepared for archiving meets the rules set by the archive institution or not. UAM has been successfully tested and laun-

ched in a couple of electronic records transfer projects in Estonia and is being implemented in all of the ministries' offices of Estonia. It is also used extensively by the archivists in the National Archives for rearranging descriptions of existing archival materials and supporting transfers of paper records.

Original scientific article
Submitting date: 03.04.2012
Acceptance date: 30.06.2012