

1.09 Objavljen strokovni prispevek na konferenci
1.09 Published Professional Conference Contribution

Flóra ORTHMAYR*

THE ELECTRONIC ARCHIVES PROJECT IN BUDAPEST CITY ARCHIVES

Abstract:

National Infocommunication Services, National Archives of Hungary, and Budapest City Archives created a consortium for the realisation of the Electronic Archives project. This EU funded project, aimed at constructing a complex e-archives system in Hungary for the complete archival registry and the long term preservation of digital documents, was concluded in September 2013. The paper reports on this system and its introduction in Budapest City Archives.

Key words:

e-archives, preservation

Izvleček:

Projekt elektronskega arhiva v Mestnem arhivu Budimpešta

Državna informacijsko-komunikacijska služba, Državni madžarski arhiv in Mestni arhiv Budimpešta so ustanovile konzorcij za realizacijo projekta elektronskega arhiva. Projekt, financiran s strani Evropske unije in s ciljem zgraditi kompleksen sistem elektronskega arhiva na Madžarskem za celotno arhivsko in dolgoročno hrambo digitalnih dokumentov, se je zaključil v septembru leta 2013. Prispevek podaja poročilo o tem sistemu in njegovi implementaciji v Mestnem arhivu Budimpešta.

Key words:

e-arhivi, hramba

1 THE ELECTRONIC ARCHIVES PROJECT

The goals of the EU funded Electronic Archives project in Hungary were to improve the archival service by information technology support, to make electronic documents in archival keeping widely available, and to create an infrastructure for the long-term preservation of electronic documents, especially those generated by public administration. (*Tervezési felhívás és...*, 2013) The latter goal requires not only saving documents from destruction, alteration, and unauthorized access but also making them easily available and legible for authorized persons and, later, for the general public, even if by then the digital environment will have considerably changed. These functions should be also available to public institutions for the keeping of their own records before they hand them over to the assigned archives. Making available the born digital documents also entails improving the availability of digitalised documents.

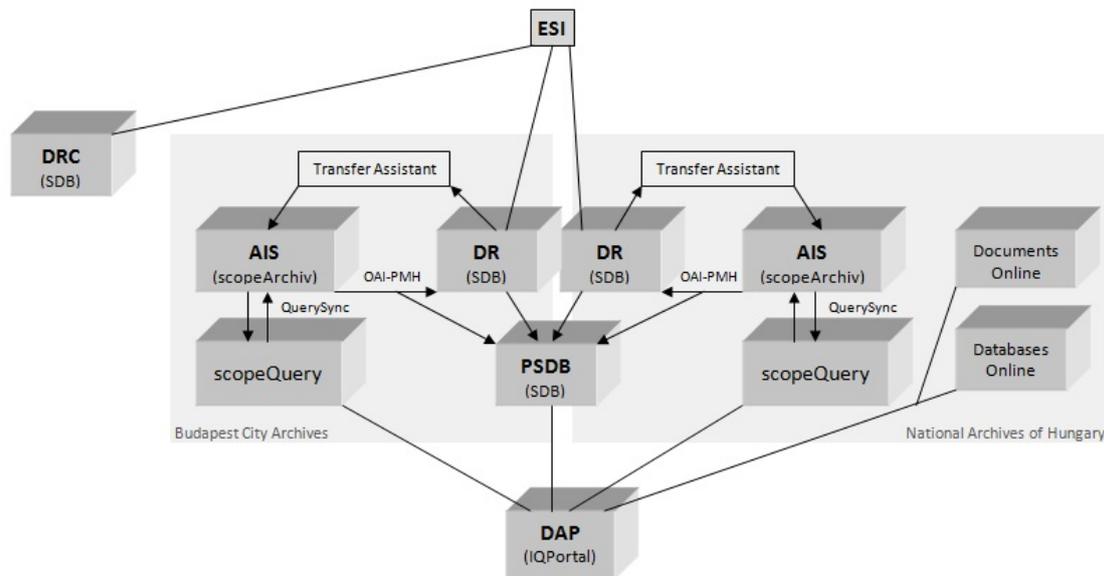
* Flóra Orthmayr, archival information technologist, Budapest City Archives, Teve u. 3-5, H-1139 Budapest, Hungary, contact: orthmayrf@bparchiv.hu.

KopintDatorg Zrt. (now called Nemzeti Infokommunikációs Szolgáltató Zrt., National Infocommunication Services, NIS), National Archives of Hungary (NAH) and Budapest City Archives (BCA) created a consortium for the realisation of this project in 2009. (Budapest City Archives, 2013) The project involved the construction of a complex e-archives system for the complete archival registry and long term preservation of digital documents with different interfaces for record generating organs and researchers. The official Hungarian guidelines for the handling of electronic documents also had to be designed. After lengthy preparations and two unsuccessful general supply procedures, in October 2012 the consortium contracted T-System Magyarország Zrt. to implement the Electronic Archives in Hungary by combining scopeArchiv and Tessella's SDB, complemented by several minor additions. Devices purchased for the project, but not directly connected to the e-archival system like special archiving scanners and high-capacity storage devices, were procured by the archives in separate procedures (*Press release, 2012*).

As the result of the project, technical conditions described below are present in both archives. The specific examples of usage I refer to are, however, taken from the practice of Budapest City Archives.

2 THE STRUCTURE OF THE ELECTRONIC ARCHIVES

The main components of this structure are the Archival Information System (AIS); the storing of digital documents in the Digital Repository (DR) and of public documents also in a separate Public SDB; the Digital Records Centre (DRC) service and the External System Interface (ESI) for record generating organs to connect to the DRC and to submit digital documents to the archives; and the Digital Archives Portal (DAP) where all public documents and metadata from both archives can be accessed by the general public.



Picture 1: The Structure of the Hungarian Electronic Archives

2.1 Archival Information System

The Archival Information System is implemented by scopeArchiv, release version 5.1. This archival information management software is the main product of scope solutions ag. ScopeArchiv is widely used by local and national archives as well as by international organisations and companies, and it has been improved by their feedback. Since it applies the same Central European archivist logic as the Hungarian system does, it provides a good base for the functions expected by Hungarian archivists.

The main characteristic of the scopeArchiv is its modular structure. Different functions are connected to distinct modules, and the customer can choose which modules to buy, though most of the processes use more than one module. For example, this project did not need the Ingest module because the SDB is used for its functions. Apart from that, most of the archival processes are already or will soon be administrated by scopeArchiv.

2.1.1 Units of description

The core of the AIS is the Units of Description module where any level of the archive plan hierarchy can and should be described, from the archives down to the items. The predecessor of this function in the Budapest City Archives was the Registrum AIS (locally developed in 2002 and used until it was replaced by the scopeArchiv in 2013), where the descriptions reached down only to series level (registry information) and locations finding aid was attached in table form to the lowest available level. The registry information has been fully migrated with a default ISAD(G) form to the scopeArchiv. Further forms can be defined for the diverse records on lower levels. So far two types of data have been migrated to the lower levels of the Units of Description: those from the databases based on the archival material and the containers from the rows of the locations finding aid tables. The locations finding aid combines the properties of a logical and a physical register of documents used both as finding aid by researchers and as internal registration. A location finding aid is attached to the lowest level of the registry (fonds, sub-fonds, or series) and comes in a table form organised according to the numbered list of containers (physical units), but divided by the logical units within them (so one container can appear in several successive rows of the table). The migration of these physical-logical units as parts of the hierarchy is only a temporary solution, but for now necessary for the registration of research orders.

2.1.2 Researchers and orders

Budapest City Archives did not use any purposefully dedicated program for the registration of researchers and orders before. This was all done using Excel tables and paper-based documentation. Since in this respect the implementation did not have to adapt to any previous data structure, it followed only practical needs and legal instructions. With the new possibility of sending orders via the Digital Archives Portal, the administration of orders is the first process in the Electronic Archives system which has to be used in everyday practice, representing every step from the receipt of the order through issuing and then returning the ordered documents.

2.1.3 Other functions

In the Partners module, not only the researchers, but the record generating organs can also be registered. Their data was also migrated from the previously used database and the registration of organisations can be continued in the Partners,

Dossiers and Events modules of the scopeArchiv.

Inventory and location management in the scopeArchiv is separate from the Units of Description hierarchy, each element of which has to be attached to its container. Since the above mentioned temporary solution made containers part of the hierarchy, now all containers as logical units are attached to themselves as physical units.

2.2 Safety Deposit Box

In this system, Tessella's Safety Deposit Box (SDB) is used for keeping digital documents safe, both as Digital Records Centre for the record generating organs and as Digital Repository for the archives. SDB was constructed by the request of, and in partnership with The National Archives of the UK expressly for the preservation of digital objects (Tessella plc., 2009).

SDB is based on the Open Archives Information System (OAIS) reference model, thus documents can be ingested only in Submission Information Packages. As part of the ingest workflow SDB checks the integrity of the package and the connection between the files and the metadata, which will be stored together. Once ingested, the documents can be easily downloaded one file at a time or as part of a Dissemination Information Package (DIP), but they cannot be altered in the system. Authorized users can edit the metadata, change the structure or delete items using predefined workflows, but the program keeps detailed record of all such actions. It is possible, though, to create various 'manifestations' of the documents (e.g. smaller size for online publication) which will be stored in the same folder, available in a different view. SDB also creates new manifestations by migrating files from obsolete formats using the tools of the 'active preservation' system (Tessella plc., 2010). Using the flexibility of the SDB, most of the workflows were customised for the needs of the archives.

In the e-archives system there are three distinct SDBs for three different functions.

2.2.1 The Digital Repository

In the Digital Repository (DR) the two archives have distinct 'tenants'. From the user's point of view, they seem to be two distinct systems, each with its own data structure and access privileges. This SDB stores all the electronic documents and their metadata in the keeping of the archives in their full, original form - both born digital and digitalised documents.

2.2.2 The Public SDB

There is another ('public') SDB (PSDB) as the content provider of the Digital Archives Portal (DAP). As the inside storage has to be clearly separated from the publicly available, all public materials from the DR have to be copied to the PSDB. This happens by a 'publication' workflow in the DR with built-in filters for keeping the private data from publication, continued by an 'ingest' to the PSDB. The contents of the two archives are not separated into different platforms here, though each document has its place in one of the two archival hierarchies placed next to each other. Users can search in both combined, as can any researcher on the internet via DAP.

2.2.3 The Digital Records Centre

The Digital Records Centre (DRC) is also implemented by a distinct SDB that is not accessible to either archive, but can be used by the record generating organs on the basis of transfer agreements. They can ingest their document packages there for safe-keeping and access them whenever needed. When they submit such documents to the archives, they can easily transport the SIP from the DRC to the DR of the assigned archive.

2.3 Synchronisation

To make a complex system, the AIS, the DR and the PSDB have to transfer data among themselves. This proved a much more complicated task than initially expected by the archives. SDB can use Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for metadata synchronisation, and in this way the hierarchy and the connected metadata can be synchronised from the AIS to the DR and PSDB down until series level. This process can take place automatically at a prefixed time. Now the defined subset of metadata from the ISAD(G) forms used in the scopeArchiv will appear in the DR and PSDB on the next day after this metadata have been filled in the AIS. In this way 'Collections' can be made in the SDB, while the files are stored in lower level 'Deliverable Units' following the file structure of the ingested SIPs. To transfer the metadata from these SIPs and the direct link of the digital contents to the scopeArchiv, an XML export from the SDB has to be made suitable for import through the scope's TransferAssistant.

3 MIGRATED DATABASES

Several databases created in the Budapest City Archives had been accessible previously from the archive's webpage and through the Hungarian Archives Portal. Those with digital content were ingested to the SDB, for example, the architectural designs by Miklós Ybl. The rest were migrated to the Units of Description module of the scopeArchiv: if the records could be fitted into the archival plan hierarchy, they were placed on levels below the migrated registry; if not, then they were put into a separate branch created for them next to the archival hierarchy. An example for the former case is the database of penal records, for the latter the database of the citizens of Buda and Pest. Thesaurus type data were also migrated into the Descriptors module of scopeArchiv. As a rule, the SDB or scopeArchiv are not used for building such extensive databases and, indeed, migration took much more time than initially expected. In the SDB, special workflows were defined for faster ingest and publication in the migration process.

4 INTERFACES

4.1 The External System Interface

Born digital documents will be transferred to the assigned archives through the External System Interface (ESI), whether sent by the creator organ through a client program, or handed over to the archives offline and then uploaded by archivists. The same client program can be used by the organ to access the Digital Records Centre (DRC) service. In order to submit digital documents to an SDB system, they have to be transformed into a SIP. Since few records management programs can create and

automatically forward to the ESI such a package, a SIP creator program (different from the one that belongs to the SDB) has been developed together with the proposed Hungarian regulation of this process. This program can be downloaded from the DAP and used without special informatics knowledge to produce the expected SIP format.

4.2 The Digital Archives Portal

Researchers and other internet visitors can connect to the Electronic Archives system through the Digital Archives Portal. Here they can search either simultaneously or separately the contents (all metadata and full text readable files) of the PSDB, the AIS of the two archives (AIS refers here only to the contents of the Units of Description module of the scopeArchiv) and two further databases of the National Archives of Hungary. Search results are categorised by their sources. They can either be listed in separate tabs for each source or, up to 500 results, together in a combined list. Clicking on an item in the list of the digital records of PSDB, its digital content and metadata can be accessed. Contents in the PSDB can also be reached by browsing in the hierarchy. Choosing an item from the search results received from the AIS, users can switch to the scopeQuery of the archives in question. Registered users of the DAP will be automatically registered in the system of the Query when they first click on any function available only for registered users. Researchers already registered in the archive and on the portal can send their orders from here, and they can also check the status of their previous orders. Registered researchers allotted two different storage spaces on DAP, where the archives can send them unpublished digital content or digital copies of paper-based documents in a Dissemination Information Package (DIP) directly from the DR by a dedicated workflow. Public materials are sent to the External Storage Space which is accessible on the DAP from anywhere. Documents under research limitations can be sent to the Internal Storage Space of researchers authorized to use them, but this storage can be accessed only from the research rooms of the archives. Authorized researchers with Hungarian Client Gate registration can receive these restricted use documents also in their Client Gate storage space.

Many unexpected difficulties have arisen and found solution in the course of less than a year left for the actual implementation of the system after lengthy preparations and unsuccessful supply procedures. Since the project was concluded in September 2013, its innovations have just begun to transform the everyday functioning of the archives. Though the system described is already working, there is as yet not much to report on practical experiences with it.

SOURCES AND LITERATURE

- Budapest City Archives (2013). *City gate to client gate*, compiled by A. J. Horváth & T. Csáki. Budapest: Budapest City Archives.
- *Press release: Implementation of the Electronic Archives in Hungary*. (2012, November 6). Retrieved 30. 11. 2013 from <http://www.digital-preservation.com/wp-content/uploads/Implementation-of-the-Electronic-Archives-in-Hungary.pdf>.
- *Tervezési felhívás és útmutató az Elektronikus közigazgatás operatív program keretében megvalósuló Elektronikus levéltár című kiemelt projekt támogatásához*. (2013, November 5). Retrieved 30. 11. 2013 from http://www.nfu.hu/download/49874/Tervezesi_felhivasi_es_utmutato.pdf.

- Tessella plc. (2009, September). *Case Study: The National Archives*. Retrieved 30. 11. 2013 from <http://www.digital-preservation.com/wp-content/uploads/TNA.pdf>.
- Tessella plc. (2010, November). *Safety Deposit Box: Digital Archiving*. Retrieved 30. 11. 2012 from <http://www.digital-preservation.com/wp-content/uploads/SDB4.pdf>.
- Webpage of scope solutions ag at <http://www.scope.ch>.

POVZETEK

Flóra ORTHMAYR*

PROJEKT ELEKTRONSKEGA ARHIVA V MESTNEM ARHIVU BUDIMPEŠTA

Cilji projekta elektronskega arhiva na Madžarskem, financiranega s strani Evropske unije, so bili izboljšati arhivsko službo s podporo informacijske tehnologije, napraviti elektronske dokumente v arhivski hrambi širše dostopne in ustvariti infrastrukturo za dolgoročno hrambo elektronskih dokumentov, še posebej tistih, ki nastajajo v javni upravi. Zadnji cilj zahteva ne samo reševanje dokumentov pred uničenjem, spreminjanjem ali neavtoriziranim dostopom, ampak tudi njihovo lažjo dostopnost in berljivost za upravljalce in kasneje za splošno javnost, tudi če se bo do takrat digitalno okolje močno spremenilo. Te funkcije naj bi bile dostopne tudi javnim institucijam za hrambo njihovih lastnih dokumentov, preden jih predajo pristojnemu arhivu.

Državna informacijsko-komunikacijska služba, Državni madžarski arhiv in Mestni arhiv Budimpešta so leta 2009 ustanovili konzorcij za realizacijo tega projekta. Slednji je zahteval zgraditev kompleksnega sistema e-arhiva za celotno arhivsko delo in dolgoročno hrambo digitalnih dokumentov z različnimi vmesniki za ustvarjalce in raziskovalce. Prav tako so morale biti pripravljene tudi uradne madžarske smernice za upravljanje z elektronskimi dokumenti.

Rezultat projekta so tehnični pogoji, predstavljeni v prispevku, ki so v uporabi v Državnem madžarskem arhivu in Mestnem arhivu Budimpešte. Posebni procesi, ki so opisani v prispevku, so praksa Mestnega arhiva Budimpešte.

Glavne komponente strukture so arhivski informacijski sistem, hramba digitalnih dokumentov v digitalnem skladišču in javnih dokumentov v posebnem javnem digitalnem skladišču, center za digitalne zapise in zunanji sistemski vmesnik za ustvarjalce, s katerim se povežejo s centrom za digitalne zapise in predajo digitalne dokumente arhivu, ter digitalni arhivski portal, na katerem lahko javnost dostopa do vseh javnih dokumentov in metapodatkov iz obeh arhivov.

Arhivski informacijski sistem je scopeArchiv. Njegovo jedro je modul Popisne enote, kjer so lahko oz. bi se naj popisali vsi nivoji od arhiva do dokumenta. Podatki iz registra so bili v scopeArchiv v celoti preneseni iz prej uporabljenih zbirk podatkov, in to v skladu s standardom ISAD(G).

Center za digitalne zapise za ustvarjalce, digitalno skladišče za arhive in generator podatkov digitalnega arhivskega portala implementira Tessellin Sef, zgrajen posebej za hrambo digitalnih objektov. Sef temelji na referenčnem modelu

* Flóra Orthmayr, arhivski informacijski tehnolog, Mestni arhiv Budimpešta, Teve u. 3-5, H-1139 Budimpešta, Madžarska, kontakt: orthmayrf@bparchiv.hu.

odprtega arhivskega informacijskega sistema (OAI), tako da so lahko dokumenti zajeti samo preko Sprejemnega informacijskega paketa.

Za diferencirano uporabo sta na voljo dva vmesnika: digitalni arhivski portal za raziskovalce, stranke in druge internetne obiskovalce ter zunanji sistemski vmesnik za ustvajalce, ki imajo sklenjen dogovor o prenosu.

V okviru projekta se je pojavilo mnogo nepričakovanih težav, najdene rešitve pa so bile implementirane v sistem po dolgih pripravah. Ker je bil projekt zaključen komaj pred kratkim (septembra 2013), so njegove novosti šele začele vplivati na vsakodnevno delovanje arhivov.