



PAM Pokrajinski
arhiv
Maribor

Moderna arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 1 (2018), št. 1 / Year 1 (2018), No. 1

Maribor, 2018

Moderna arhivistika

Časopis arhivske teorije in prakse

Journal of Archival Theory and Practice

Letnik 1 (2018), št. 1 / Year 1 (2018), No. 1

ISSN 2591-0884 (online)

ISSN 2591-0876 (CD_ROM)

Izdaja / Published by:

Pokrajinski arhiv Maribor / Regional Archives Maribor

Glavni in odgovorni urednik / Chief and Responsible editor:

*Ivan Fras, prof., Pokrajinski arhiv Maribor, Glavni trg 7, SI-2000 Maribor,
telefon/ Phone: +386 2228 5017; e-pošta/e-mail: ivan.fras@pokarh-mb.si*

Glavna urednica / Editor in chief:

mag. Nina Gostenčnik

Uredniški odbor / editorial board:

- dr. Thomas Aigner, Diözesanarchiv St. Pölten, Avstrija
- dr. Borut Batagelj, Zgodovinski arhiv Celje, Slovenija
- dr. Bojan Cvelfar, Arhiv Republike Slovenije, Slovenija
- mag. Nada Čibej, Pokrajinski arhiv Koper, Slovenija
- Ivan Fras, Pokrajinski arhiv Maribor, Slovenija
- mag. Nina Gostenčnik, Pokrajinski arhiv Maribor, Slovenija
- dr. Joachim Kemper, Institut für Stadtgeschichte Frankfurt am Main, Nemčija
- Leopold Mikec Avberšek, Pokrajinski arhiv Maribor, Slovenija
- dr. Miroslav Novak, Pokrajinski arhiv Maribor, Slovenija
- dr. Rik Opsommer, Stadsarchief Ieper - Universiteit Gent, Belgija
- Darko Rubčić, Državni arhiv u Zagrebu, Hrvaška
- dr. Izet Šabotić, Filozofski fakultet Univerziteta u Tuzli, Bosna in Hercegovina
- mag. Boštjan Zajšek, Pokrajinski arhiv Maribor, Slovenija

Recenziranje / Peer review process:

Prispevki so recenzirani. Za objavo je potrebna pozitivna recenzija. Proces recenziranja je anonimen. / All articles for publication in the conference proceedings are peer-reviewed. A positive review is needed for publication. The review process is anonymous.

Lektoriranje / Proof-reading:

mag. Boštjan Zajšek, mag. Nina Gostenčnik

Prevajanje:

mag. Boštjan Zajšek (slovenščina), mag. Nina Gostenčnik (slovenščina, angleščina), Lučka Mlinarič (bosanščina, hrvaščina, srbščina)

Oblikovanje in prelom / Design and typesetting:

mag. Nina Gostenčnik

Objavljeni prispevki so prosto dostopni. Vse avtorske pravice ima izdajatelj Pokrajinski arhiv Maribor.

©Pokrajinski arhiv Maribor Za prijavo in objavo prispevkov ni potrebno plačilo. / The publication offers open access to whole texts of the published articles. ©Pokrajinski arhiv Maribor. All articles are published free of charge.

<http://www.pokarh-mb.si/si/p/3/49/moderna-arhivistika.html>

Prejeto / Received: 16. 02. 2018

1.04 Strokovni članek

1.04 Professional Article

DIGITIZATION OF ARCHIVAL RECORDS AT THE STATE AGENCY OF KOSOVO ARCHIVES

Aliu Bekim

State Agency of Kosovo Archives, Pristina, Kosovo

bekim.r.aliu@rks-gov.net

Abstract:

The paper focuses on the challenges and problems faced by the State Agency of Kosovo Archives in digitizing the "Economic Court-Pristina" fonds. The project team tackles broad and specific technical and intellectual challenges, from managing rights to an online access environment to the sustainability of used equipment. To better understand the challenges we have faced and what we are being faced with in digitization, we will first discuss in this paper the projects that the Agency itself had carried out as well as digital documents resulting from researches done in the similar archives. Then we will talk about the process of the currently digitized fonds, about technical challenges, personal data protection and finally the lessons learned so far from this pilot project.

Key words:

archives, digital records, projects, electronic systems, management.

Izvleček:

Digitalizacija arhivskega gradiva v Državni agenciji kosovskih arhivov

Prispevek se osredotoča na izzive in težave, na katere je agencija naletela pri digitalizaciji fonda Gospodarsko sodišče Priština. Projektna ekipa se je spopadala tako s splošnimi kot specifičnimi tehničnimi izzivi, od upravljanja s pravicami spletnega dostopa do vzdržnosti uporabljene opreme. Za boljše razumevanje izzivov, prispevek najprej podaja informacije o že izvedenih projektih in digitaliziranem gradivu, ki so rezultat raziskav narejenih v sorodnih arhivih. Prispevek predstavlja proces digitalizacije trenutno v teku, tehnične izzive, zaščito osebnih podatkov in podaja izkušnje, ki jih je projektna ekipa v toku projekta pridobila.

Ključne besede:

arhivi, digitalni zapisi, projekti, informacijskih sistemi, upravljanje, digitalizacija

1 History of projects

The State Agency of Kosovo Archives is the highest state institution in the Republic of Kosovo, which is responsible for collecting, protecting, preserving, arranging, processing, publishing, presenting and making available for use archival records for professional and scientific purposes (Law No. 04 / L-088 on Kosovo State Archives). Until February 2007, this institution failed to make the first steps to digitize its archival possessions due to the lack of financial resources. The Organization for Security and Co-operation in Europe (OSCE) has enabled the implementation of the first digitization project of documents from microfilms.

This four-year project enabled digitization of over 300,000 frames and physical storage of 295 microfilms endangered by destruction, since Kosovo after the last war inherited completely destroyed equipment. A professional Zeutschel OM 1200 scanner was used to digitize microfilms. Scanning was done according to the standards of NARA¹ for film scanning. A 1TB database was created with data according to the inventory entry book of microfilms. The digital copy of a microfilm is the same as the analogue one. However, the problem of indexing the documents quickly arose. The solution is made through a log file that scanner creates during scanning each frame. We used this to later index each document separately. The metadata standards used in this project are those defined in the Dublin Core² Initiative. The project was intended to maximize the use of information transmitted by microfilms for internal use in the reading room at the Archives, but not for online use.

Another project intended for internal use was carried out in the Inter-Municipal Archives of Gjakova, where the Urbanization Fonds were completely digitized. With this project, 10,000 case files were digitized over a period of 15 months, from 2014 to 2016. The software used to process scanned images was standard for Microsoft. TIFF format was used for a 300-dpi Master File and PDF format for use. Indexing was done by the chronological order by dates or by names, and the linking to a digital file in PDF format was performed only with the index name. This system managed to serve the parties successfully, and it took only 3 to 4 minutes. The project was performed by using a professional scanner (HP Digital Sender 9250c) and only one employee.

Two other digitization projects of archival records were accomplished within the period from 2013 to 2015. One project relates to the Ottoman documents fonds from 1806/1920, which consists of more than 3,000 records, while the other project was related to the collection of Photographs belonging to the period between 1900-1995, which consists of 1,000 photos. With these projects, we are entering into a new field of cooperation with the third parties. In these two projects the Archives was a determining factor of the criteria and parameters of digitization of archival records. All the work has been carried out at the premises of the State Archives Agency under the supervision and control of the archives' staff.

In these projects, metadata standards, defined by the Dublin Core Initiative, were used. Even these projects, like previous ones, due to the lack of financial resources, were intended to use digital documents only in the reading rooms and to provide a quicker service to the users at the archives' premises instead of putting them online.

A special place in the Archive's mission are also the research-scientific activity and cooperation with local, regional and international archives. As part of this activity, the State Agency of Kosovo Archives has been enriched with 236,481 digital documents since 2004. This large number of digital documents has caused great difficulties in the records management and providing services for the usage of these documents. The problem lies in the fact that these documents were brought without associated metadata, which would facilitate their management. Excluding the Ottoman Archives in Turkey, which has provided also the index files with the digital documents, no other archives was able to provide digital documents associated by standard metadata or even with any index of the scanned documents. All of these data and descriptions of digital documents were done by the team working on digitization of archival records.

¹ *More about the guidelines: Puglia, Reed, Rhodes, 2010*

² *More: Dublin Core (2012).*

2 Digitization Project of the "Economic Court" Fonds

Project preparation lasted for a year and encompassed the infrastructure adjustment, work environment, internet network expansion, server purchase, Software Assurance, and associated digitization technology such as scanners and computers. Funding and configuration of the systems was done by the State Agency of Kosovo Archives itself, while the arrangements of the workspace was done by the Ministry of Public Administration.

The fonds consist of 252 meters linear length of textual archival records. Archival materials date back to 1973 and continue until 1999. Until the 90's, the text of the documents is in two languages, in Albanian and Serbo-Croatian. During the violent measures, the text is only in one language, in Serbian, and in most cases, apart from Latin letters, it is also mixed with Cyrillic letters. Therefore, after a test, it was decided not to get the text recognition through the OCR software because during the testing there were problems with the text recognition.

The Access to Memory (AtoM) platform was selected for Internet access due to the capability of attaching digital records with standards based on archival records descriptions such as ISAD (G), ISAAR (CPF), ISDIAH, and also Dublin Core. So far, the project team has managed to process and input over 35,000 documents in the descriptor system of archival records.

Technical challenges

Technical challenges that this project initially faced have affected the image quality matters, resulting in the size of the electronic file, digital storage, digitization act and creation of access copies. Image quality parameters for digitizing resulted in large electronic folders, which created a massive amount of data for digital storage.

Scanning has become a technical issue that the project team had to deal with. At the beginning of the project, it was believed that the Automatic Document Feed Scanner (ADF) would speed up the scanning process. Uniform size sheets that were in good physical shape and not too fragile could be scanned faster with Automatic Document Feeder (ADF). During scanning of documents belonging to this fonds, up to 50 pages could be loaded on the automatic feeder. After a period of three months of using this scanner, pages started to tilt in the ADF. It was also noted that manufacturers of the automatic feeder use very little space for image capture. As a result, in this situation, even a small particle of dust caused lines in the image, which resulted in a waste of time, as it was often necessary to clean the scanning glass in order to provide high-quality scanning. Dust became an increasingly common issue. At the end, the use of an automatic feeder, including the time for cleaning the glass and rescanning of documents, reduced time for all other work processes. Therefore, the use of a flatbed scanner was found to be the preferred method of scanning.

The creation of a stable and secure connection between the scanner and the computer was another challenge. Often, to reach this connection, it was necessary to turn off the scanner to cool for a while, and later on to return to the system to start scanning. This is due to exceeding standards of equipment, that our project uses scanners of higher intensity than recommended by the manufacturer.

Technical specifications for this project, such as resolution, file formats, methods and storage facilities are based on international standards. Based on these standards, pages were scanned using the HP Digital Sender 9250c scanners - 300dpi color with 24bit, and stored in uncompressed TIFF format. These standards have provided high-

quality images that truly represent the material of the fonds in question. These digital duplicates ensure that if the color of the paper fades or any page is torn, ripped, stolen, replaced, or destroyed, there is a copy that can easily be replaced even for decades after it was scanned.

The specifications used for scanning in this project may raise potential critiques. In the mostly textual fonds, like our project "The Economical Court" fonds, there is a limited amount of colors. It may be suggested that full-color scanning is unjustified considering the increase of file size and space that can be saved by gray-scale scan. However, some documents in reality are gray on the paper, some are cream or in other nuances, and also colors can be blue, red or other colors. It was decided to scan everything in color to capture those elements without discussing which pages require color scanning and which ones did not deserve it, and this decision is also consistent with the idea of "scan once save forever".

Other major issues generated by the technical specifications relate to the absolute size of the digital material created. Each page of the size A4 paper resulted in an electronic document of up to 7 MB. This size is simply not practical for use and certainly not for online use. It was decided that the scanned pages would be converted to multi-page PDF format, which will present files according to the classification in the boxes of the respective fonds. This conversion was originally made from TIFF to PDF format, and resulted as unsuccessful in terms of space saving because even PDF files were large and difficult to manage. As a successful step of compressing these files, it turned out to be the conversion from TIFF format to JPEG format. This intermediate step enabled us to create manageable PDF formats without losing quality and at the same time saving space for storage.

A very crucial technical issue facing this project was the active preservation of scans during the course of the project. The first step in scanning was the computer to which the scanner was connected. This computer did not have enough storage capacity, therefore after scanning every four boxes, respectively 40 scanned files, these had to be transferred to an external hard drive without completing the conversion process, which caused difficulties in time management without mentioning the time it took to name the scanned documents.

Naming of scanned documents turned out to be an important task in the work process. It was initially believed that the automatic number, provided by the scanner during scanning was sufficient. However, that was not the case. A wise and logical naming should have been defined that would enable the identification of documents, even single ones. This led us to the question; to which fonds does the document belong? Which box? Which folder? And, where is the document placed within the file case? The answer to these questions brought to the naming that contained the Fonds' Inventory Entry number, then the box number and file number. To make the identification even easier, we added the file title and the document file number. For example, naming of a scanned document now looks like this: F574_K46_D615_Autorska Agencija Kosova _48.

However, regardless of the technical and intellectual challenges, the project team managed to create good quality images and so far has managed to process and put in the system for archival description over 40,000 documents.

Data Protection

Data protection challenges were encountered since the beginning of the project. The first concern that every project team needs to address is the selection of what needs to be digitized. As mentioned above, the fonds are large - about 252 meters linear and cannot be digitized completely within a year. By reviewing the materials, we found a variety of documents in the files of companies registered in this court (the Economical Court), which belong to a period of more than four decades. These files contain personal data of these enterprises' founders, including financial data, education, birth certificates, etc. After careful consideration, it was decided that the digitization process should be executed in two ways; for internal use and for introducing online data. The decision for selecting documents for scanning and deciding which ones should not be scanned because of data protection facilitated the whole digitization process. The online platform will show only data such as: title, respectively the company name; year of establishment or years of activity; the amount of documents and the headquarters of the enterprises.

What we have learned

This brings us to our last point with regards to the execution of massive archival digitization projects. The question arises what advice can we give in creating a balanced project that meets both access and storage goals? Three ideas are worth mentioning here: automation, documentation and risk assessment.

Firstly, the concept of automation. The number of steps that can be shortened in a project makes project tasks faster. A contribution would be an automatic format conversion and the transfer of digital documents to the long-term storage environment. This process would shorten two steps of work. However, quality assurance is important, so it is necessary to control automated processes to make sure that they are working properly and the results are at the standard level.

Secondly, it is important to record what is done and above all, when. The project has disclosed and made it clear that time spent on certain tasks can be completed by taking more or less time than it was initially planned. The technical and intellectual challenges significantly affected the workflow, and it was therefore necessary to track down what steps had been taken in specific work processes. In addition, this is useful for the fact that electronic document formats due to the technology development cannot last endlessly and the solution is the migration of electronic documents. Thus, recordkeeping of the time when archival records were scanned responds to the question "Is it time to migrate?". If a document was scanned two years ago or seven years ago, this information is invaluable in maintaining workflow. Knowing the time, quality, software, and file types, and without having to spend time finding them, we can save a lot of time and work in the future. For our team, the visit to the Ottoman Archives in Istanbul was of special importance. There, the team gained good knowledge of the work process, digitization formats, the migration period and the access and storage facilities of digital archival records.

Dealing with these challenges and lessons learned in detail about failures are important. Hiding these problems does not allow anyone else to learn from your mistakes, and it makes the institutions repeat each other's mistakes.

Thirdly, we learned that a balance should be achieved when it comes to data protection. We wanted to avoid a conservative approach, thus, for inserting online data we were limited to the basic ones in order to not violate the legal rights.

With this, we come to the end of the presentation of the digitization project of the Economic Court fonds. Hopefully, by sharing the lessons presented here, this project will

be useful for the archives interested in digitization, providing an overview of how we have balanced various intellectual and technical concerns. Our specific decisions are related to project objectives, and are retrieved from the resources, budget, support and technology available at the State Agency of Kosovo Archives. Our hope is that the work we have done in this project is a step forward for a common practical culture in promoting the archives.

POVZETEK

DIGITALIZACIJA ARHIVSKEGA GRADIVA V DRŽAVNI AGENCIJI KOSOVSKIH ARHIVOV

Aliu Bekim

Državna agencija arhivov Kosova, Priština, Kosovo

bekim.r.aliu@rks-gov.net

Projektov digitalizacije, ki so bili izvedeni v Državni agenciji kosovskih arhivov, je malo in so slabo finančno in kadrovsko podprti. Zato je bil namen prvih projektov digitalizacije interna raba digitaliziranih dokumentov oziroma zaščita originalnih dokumentov.

S povečanim povpraševanjem po uporabi arhivskega gradiva ter povečano količino digitaliziranih dokumentov (digitalizatov, ki jih ustvari arhiv sam ali jih ustvarijo uporabniki iz drugih arhivov), se je pojavila potreba po pridobitvi programske opreme za upravljanje z dokumenti, ki smo jo uspešno pridobili v letu 2016. Platforma AtoM je bila izbrana na podlagi dejstva, da je prosto dostopna in ustreza vsem kriterijem Mednarodnega arhivskega sveta za popisovanje arhivskega gradiva.

Pogosta raba gradiva fonda Gospodarsko sodišče je spodbudila odločitev za njegovo digitalizacijo. Fond obsega 252 tekočih metrov arhivskega gradiva, za digitalizacijo pa je bil uporabljen skener z avtomatičnim podajanjem, saj je bila količina gradiva, predvidena za digitalizacijo velika. Kmalu je postalo jasno, da novih digitalizatov ne bo mogoče povezati oz. pridružiti digitalizatom, ki so nastajali med letoma 1973 in 1999.

Projektna ekipa se je med digitalizacijo fonda soočala z mnogimi težavami, od neprimerne ureditve fonda in pravnih ovir (varovanje osebnih podatkov) do težav, ki so se pojavile ob uporabi tehnologije. Ob njihovem reševanju pa je ekipa pridobila neprecenljive izkušnje z digitalizacijo arhivskega gradiva, ki jih predstavlja v članku.

Reference list:

Dublin Core Metadata Element Set, Version 1.1: Reference Description. (2012).

Dublin Core Metadata Initiative. Retrieved 5. 1. 2018 from:
<http://dublincore.org/documents/dces/>.

Law No. 04 / L-088 on State Archives. (2012). Official Gazette of the Republic of Kosovo, Nr. 04.

Puglia, S., Reed, J., Rhodes, E. (2004). Technical Guidelines for Digitizing Archival Materials for Electronic Access - Creation of Production Master Files – Raster Images. Washington, DC: U.S. National Archives and Records Administration (NARA). Retrieved 5.1.2018 from:

<https://www.archives.gov/files/preservation/technical/guidelines.pdf>