PAM Pokrajinski arhiv Maribor

Moderna arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 2 (2019), št. 2 / Year 2 (2019), No. 2

Maribor, 2019

Moderna arhivistika

Časopis arhivske teorije in prakse

Journal of Archival Theory and Practice

Letnik 2 (2019), št. 2 / Year 2 (2019), No. 2

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. Pőlten, 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

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

Prejeto / Received: 17. 12. 2018

1.01 Izvirni znanstveni članek *1.01 Scientific Article*

CREATION OF AUTHENTIC DIGITAL COPIES IN THE FORM OF ORIGINAL BY THE PROCESS OF DIGITIZATION

Edvin BURŠIĆ, Ph. D.
Financial Agency, Zagreb, Croatia
Edvin.Bursic@fina.hr

Hrvoje STANČIĆ, Ph. D.
Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia
hstancic@ffzg.hr

Abstract:

In order to secure the usability of records over the retention period, the archives conduct digital preservation actions which may entail conversion, migration and other procedures. When using digitization for creating copies in the form of originals, the archives should consider the trustworthiness of the process – the involved professionals, equipment, software, and procedures. These are the critical points in the process of converting physical originals into digitized originals having legal validity. The authors discuss the authenticity of records converted from physical to digital form. They argue that a digitized record is a copy in the form of an original if it meets the requirements of completeness, reliability of the process, integrity and availability. The authors define 53 requirements to be met in the process of digitization.

Key words:

digitization process, copy in the form of original, digital evidence, trustworthiness, authentic records

Izvleček:

Ustvarjanje avtentičnih digitalnih kopij v obliki originala v procesu digitalizacije

Za zagotavljanje uporabnosti dokumentov po preteku njihovega roka hrambe izvajajo arhivi mnoge ukrepe, ki zajemajo tudi konverzijo in migracijo. Kadar gre pri izdelavi kopij v obliki originala za postopek digitalizacije, morajo arhivi poskrbeti za zaupanja vreden proces – od vključenih posameznikov do strojne in programske opreme ter postopkov. To so kritične točke v procesu konverzije fizičnih originalov v digitalizirane originale s pravno veljavo. Avtorja obravnavata avtentičnost dokumentov, ki so bili konvertirani iz fizične v digitalno obliko. Trdita, da je lahko digitaliziran dokument kopija v obliki originala, če izpolnjuje pogoje celovitosti, zanesljivosti postopka, avtentičnosti in dostopnosti. Avtorja določata 53 zahtev, ki jim je potrebno pri digitalizaciji slediti.

Ključne besede:

digitalizacija, kopija kot original, digitalni dokaz, zanesljivost, avtentični dokumenti

1. INTRODUCTION

Can a digitized copy substitute an original? If it can, what are the requirements? Does that depend on the quality of the scan or of the process? Is that legal? These and similar questions have been asked many times. In the era of mass digitization and constantly increasing requests for digital access, it is important to establish procedures, define requirements and set the minimum amount of quality needed in the process of (mass) digitization with the aim to create legally valid digital substitutes for the paper originals. The pre-process, not being elaborated in this paper, is the process of assurance that the originals are being submitted to the digitization process. If this pre-condition is satisfied, than the process of digitization, in our view, can be organized without a weak point in which someone might either intentionally or unintentionally make an unauthorized change and thus influence the trustworthiness of the digitized record.

The paper starts with the discussion of the authenticity and related concepts and proposes requirements a digitized document should meet to be regarded as an original. Next, the digitization process requirements and the necessary roles and responsibilities are detailed. The technical aspects are argued, followed by the conclusion.

2. CONCEPT OF AUTHENTICITY IN THE CONTEXT OF DIGITISATION

One of the basic reasons for keeping records is to provide evidence of activities, decisions and business processes (Public Record Office Victoria, 2010). The records should be **authentic**, **reliable**, and **accessible**. “Authentic” means that for a record it can be proven that it:

a) is what it purports to be;
b) has been created or sent by the agent purported to have created or sent it; and
c) has been created or sent when purported (ISO 15489-1, 2016).

“Reliable” means that the record trustworthily states facts, i.e. a trustworthy record is authentic and reliable. “Accessible” means the record can be read and interpreted whenever it is required. It should also be added that the record should be **accurate**. An “accurate” record has truthful content.

A paper document has context represented physically, but when paper is converted to digital form, the context of the document has to be represented by the appropriate metadata added in the capturing process. The metadata provide understanding and thus limit the vulnerability of reliability and authenticity. The digitized document is reliable if it is possible to show that the reproduction process constantly produces complete and accurate copies of originals.

To make a complete and accurate copy in the digitizing process means ensuring that all pages (including both sides, and where appropriate, additions in the form of loose notes or sticky notes) are reproduced.

The “integrity” of a record refers to the fact that the record is complete and unchanged, meaning it is secured from unauthorized access and modification. If the records need to be changed or there is a risk that the record may change, then a log file is required. Creating and keeping a change log should be a documented part of the document management process, and should be an automatic system function without involving the human factor. A change log record should include information on the circumstances that led to changes. The preservation of integrity can be achieved by well-defined quality control. Mason (Mason, 2006) suggests that it is necessary to document

every part of the process to ensure data retention. This may include authenticity, authorship, certification, and other requirements that may affect the validity of the document.

The above examples suggest that the following facts should be determined while assessing the authenticity of a digitized document:

- if the creator should be identified,
- which information and marks should be present on the document,
- whether all information and markings can be read and interpreted correctly,
- if there is an evidence that the document is produced and digitized as part of a prescribed business process,
- if the document is subject to enterprise management processes,
- if there is an evidence that the document has not been changed,
- if it is possible to access, reproduce and interpret documents correctly,
- if there is an evidence that all pages of the document have been reproduced (including both sides, and where appropriate, additions in the form of loose notes or sticky notes) and
- if all metadata describing the digitization process is recorded.

In other words, a digitized document can be determined to be a copy in the form of an original if the following requirements are met:

1. COMPLETENESS: the control data and the identification labels (e.g. bar coded) are present on the document and are linked to the document, all pages of the document are captured, including both sides and the accompanying connected papers (loose notes, sticky notes, etc.), and the creator of the document is identified,

2. PROCESS RELIABILITY: the process in which the document was created was prescribed, approved and verified prior to digitization, the process is monitored and verified during digitization and there are records proving that,

3. DOCUMENT INTEGRITY: the document and all control data and markings on the document and related connected documents are not changed from the moment the document was created or from the moment of import into the business system,

4. AVAILABILITY: the document and all control data and markings on the document and related connected documents can be properly accessed, interpreted and used.

A checklist consisting of 53 questions for checking a digitization process aiming to create digitized copies in the form of originals following the four requirements can be found in the Appendix.

This paper will not consider the authenticity of documents before digitization. The responsibility of every organization is to keep track of its records and preserve their authenticity. Also, all information supporting identity and integrity of records should be present and assessed at the entrance point of the digitization process.

However, this paper does take position regarding the documents after digitization and they are aligned with the findings of the InterPARES 1 project. The project's Authenticity Task Force differentiates between three types of copies: (1) copy in the form of an original, (2) imitative copy, and (3) simple copy. “The most reliable copy is a **copy in the form of an original**, which is identical to the original although generated subsequently. An **imitative copy** is a copy that reproduces both the content and form of the record, but in such a way that it is always possible to tell the copy from the original. A **simple copy** is a copy that only reproduces the content of the original. Any of these types of copy is authentic if attested to be so by the official preserver. By virtue of this attestation, the copy is deemed to conform to the record it reproduces until proof to the contrary is shown. Such attestation is supported by the preserver’s ability to demonstrate that it has satisfied the applicable baseline requirements for maintenance and all of the requirements for the production of authentic copies.” (Authenticity Task Force, 2005).

In this paper the authors consider digitization, i.e. conversion from paper to digitized record, as the process of creation of the copy in the form of an original and will argue how it can be achieved.

3. DOCUMENTARY UNIT AND DIGITIZED COPY

The document is part of a larger entity and should be well described and kept. This means that the documents that make up the documentary unit are arranged into technical units, such as a paper file folder, an archival box, and the like. For example, a decision to initiate procurement gets its unique classification number and is placed in a file that gets its unique classification number as well. This means that there are two unique numbers connecting the document and the technical unit. Or, for example, the received invoice is regularly placed in the appropriate paper file folder. The received invoice gets its unique number, and the paper file folder has its unique number. This relationship between the document, the technical unit, and the relevant description list can be depicted as a documentary unit shown in Figure 1.

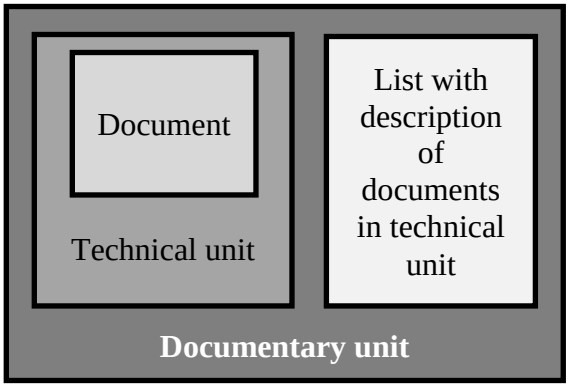

Figure 1. Documentary unit

During digitization, the documentary unit goes to the scanner and is being converted from paper to digital format. To preserve the original order of the documents in the documentary unit it is important to provide a link between the document and the documentary unit. Thus, metadata describing the digitized copy of the document become an integral part of the metadata that was captured during the document's lifecycle process prior to digitization. Additionally, for the correct interpretation of a digitized copy

of each image, it is important to associate other metadata, such as the data contained in the document: first name, last name, year, date, type of document, and the like.

From this we can conclude that the digitized document and the related metadata should be inseparable. This unit can be called a digitized copy as shown in Figure 2.

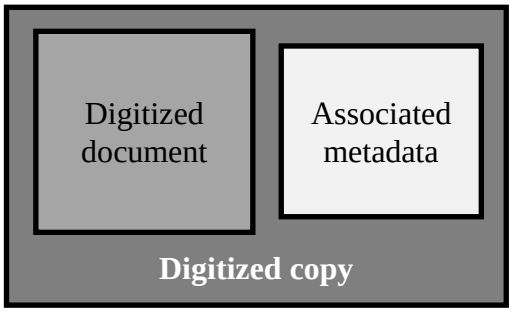

Figure 2. Digitized copy

We can conclude that by the process of digitizing, we are converting documentary units into digitized copies. If the digitized image is made according to the requirements for creation of master digitized images and if the corresponding associated metadata are present, then one can authenticate the copy as a **copy in the form of an original**.

4. DIGITIZATION PROCESS

The first step in the digitization process is the selection of documents to be digitized. It is carried out according to the prescribed procedures and the requirements of the creator whose documentation will be digitized. It is important to take into account normative and legal limitations, for example the Copyright and Related Rights Act or the General Data Protection Regulation.

The next step is an analysis of the chosen material as well as legal and normative environments in order to determine the digitization requirements and the system requirements that will ingest the materials and provide access. The Government of Alberta in Canada has prescribed a digitization standard to be applied by all ministries (Government of Alberta, 2013). The Washington D.C. Congress Library has digitized its materials since the 1980s and has well defined standards for that (Peterson, 2004). These two examples of good practice show that the steps in a digitization process include process analysis and system planning as well as the analysis of technical requirements, such as image bit depth, compression, file format, limitations in processing a digitized image etc.

Based on the results of the system analysis and planning, a process should be set up and tested on a test sample which represents documentation and also includes documents of inferior quality (ISO/TR 15801, 2017). The result of this test should be preserved for later verification of the established process. For example, the Digitization Procedures Manual specifies scanning in color so checks should be carried out whether all digitized images are in color. At this stage, a user acceptance test can also be performed (Government of Alberta, 2013).

It is further emphasized that there is a possibility of mixed sizes and weights of paper entering the digitization process (ISO/TR 15801, 2017). For example, a standard 80 gram paper that is used when printing on a printer may be mixed with a paper used in multi-sheet forms, the second and third pages being very thin. When scanning the thin

pages with the 80 gram paper settings, the light will most likely be too bright and the information on the other side of the document will be fed to the sensor. These occurrences should be monitored during the preparation of the material phase, and taken into account. The settings of the image capture program should be aligned with the characteristics of the original.

Naming the images is also important. The digitized images should be named using a unique identifier (ISO/TR 13028, 2010) and ideally be meaningful. Krogh (Krogh, 2009) argues that the file names are the foundation of the information structure and that special attention should also be given to naming of the folders since their names could be viewed as metadata.

If there is a need to preserve the integrity of the digitized image it is good practice to calculate a hash value of the digitized image at the point of scanning and at the same time record the hash onto the write-only data carrier that will be available for the comparison at a later time.

After the scan is completed, the digitized documentation, along with the calculated hash values is sent either by auxiliary sending software or image capture program to the centralized location. For example, this may be a dedicated file server that is set to accept digitized images.

In the digitization process the two-step control of the scanned document can be set up. The first step in the control process is usually performed by the operator during the scan. However, this could become a problem because of the high digitization throughput speeds achieved by the scanners For example, the Canon DR-X10C scanner has a declared scan speed of 130 ppm (pages per minute), averaging more than two images per second. Therefore, it is realistic to expect that the operator will not be able to control the quality of all the images by watching the images appearing on the screen during the digitization process.

The second step of control is usually carried out at the quality control stage (ISO/TR 13028, 2010). The documentation that does not satisfy control check at this stage is re-scanned and controlled again. Quality control should be carried out by the employees who are not directly involved in the digitization and indexing processes and quality control should be performed on at least 5-10% of total digitized pages (Government of Alberta, 2013).

There are two types of metadata to be added to the digitized images – the technical metadata about the process of digitization and the digital image itself, and the metadata describing the digitized materials. The first type, e.g. information about the type and model of equipment used, type and version of image capturing software etc., is usually written with the image in the same file automatically during the digitization. The second type is usually external to the file and, if deriving from the content of the scanned documents, can be added either manually or automatically. If it is done automatically it is possible to use the OCR technology in the process.

During the metadata control phase the predefined dictionaries of approved terms, like name dictionary from the classification scheme, should be used (ISO/TR 13028, 2010). If this is possible, the whole process will be accelerated.

After addition of the metadata all digitized copies, i.e. the digitized documents and the related metadata, are ingested to the digital archive which manages the digitized records taking into account the access rights over the entire retention period. It is therefore important to ensure the completeness of the digitized copies and their integrity.

Finally, if there is a requirement to preserve the authenticity of the digitized documents, the whole process should be accompanied by an audit log data, which includes technical requirements for systems, images and data (ISO/TR 15801, 2017).

5. ROLES AND RESPONSIBILITIES IN THE DIGITIZATION SYSTEM

In the digitization system, certain roles can be recognized and separated to support the process continuity that ensures the authenticity of digital copies in the form of an original during digitization. The roles can be divided in at least five categories: (1) process roles, (2) organizational roles, (3) supervisory roles, (4) control roles, and (5) maintenance roles. Process roles include operators in preparation, scanning and indexing. Organizational roles are roles responsible for development of manuals, their approval and control of execution of instructions. Supervisory roles are in authority of coordination of operators in the process of preparation, scanning and indexing. Control roles include the role of the coordinating operators in the process of preparation, scanning and indexing, together with the role of process control, and the role of image and metadata control. Maintenance roles include the roles of scanner servicer, auxiliary application developer, and the role of computer, computer network, database, and backup maintenance.

In order to successfully preserve the authenticity of documents during the digitization process, it is important that the responsibilities for these roles are divided without overlapping. Operators could be divided into operators for preparation, scanning and indexing. The operator for preparation prepares documentation for the scanner operator. During the scan, the scanner operator performs control of the preparation, which means that the scanner operator checks the existence of the required data on the lead document inserted to the top of each batch. Additionally (s)he has the obligation to perform visual control of the images of the digitized documentation during scanning time and mark the inappropriate image for re-scan.

The coordinating operator's responsibility is to control the process on day-to-day basis and write weekly reports of the process to role of process control. Image and metadata controller checks digitized images and associated metadata using tools to verify the correct input. All deviations should be compared to the information in the image, and, if needed, indicate the need for re-scan. When the quality and completeness of the batch is confirmed, the batch gets status “completed”.

At monthly level, all reports should be aggregated in one document prepared by the person responsible for the process, most suitably the person who conducted the test and made the manual. Twice a year, or at least two times during the digitization period, the supervisory body should carry out supervision of the entire system, checking all reports, testing sample documentation using the set up process, checking the set up process itself, checking roles and employees education performed. The audit result should be included in the report for the organization's management.

Daily scanner operation and maintenance should be carried out by the scanner operator by detecting improper operation and maintaining the scanner according to the manufacturer's instructions, for example, dusting or wiping glass that covers the lamp and sensor. Scanner control is mandatory once a month or more frequently depending on the manufacturer's recommendation and should be done by an independent expert service. This service needs to be backed up by the record of the lamp status and calibration of the device as well as records of all replaced parts.

All auxiliary applications are generated by the independent developer according to the instructions of the person responsible for the process. Applications requirements should be recorded. All applications should be tested and test results logged prior to using.

Maintenance of information infrastructure that includes computers, computer networks, databases, and backups should be prescribed and implemented in accordance with manufacturer's recommendations and implemented by the maintenance and service personnel. The installation and maintenance of the information structure should be documented. Information on the control and monitoring of information infrastructure should be available upon request.

Further, after the explanation of the process roles in a digitization system aiming at production of copies in the form of originals, the paper will focus on some more technical but nonetheless important aspects of the system covering scanner features, the role of drivers, capturing software, and system time.

6. TECHNICAL ASPECTS OF THE DIGITIZATION SYSTEM

6.1. Scanner features

The scanner features that are important in the context of creation of copies in the form of original cover types of sensors and their performance. It is a requirement that the optical resolution of the scanner sensor is higher than the scanning resolution, i.e. to use optical instead of interpolated resolution. CCD sensors should be used instead of CIS sensors because the latter have narrow depth of field which can produce loss of information for wrinkled originals. Scanners sensitivity to mechanical vibrations should be tested prior to usage by an authorized service personnel.

Light source should be appropriate. In order to properly capture all colors, white light should be used because it contains all wavelengths. The optical part of the process directs the reflected light to the sensor and should be of such a quality that minimally distorts the image or does not at all. During the conversion of light into the electronic signal there are various electrical interferences such as the influence of the magnetic and electric fields, so it is necessary to provide additional protection against these impacts.

6.2. Driver and Image Capturing Software

Between the sensor and the image capture software there is a driver program that can be either TWAIN or ISIS. Features used by any one of them in the capturing process should be clear at all periods of time. Special care should be given to functions that might result in the loss of information. This possibility has been confirmed (ISO/TR 15801, 2017; ISO/TR 13028, 2010; Government of Alberta, 2013). The general agreement is not to use image enhancement techniques that lead to data loss such as de-speckle, black border removal, form removal, sharpening, clipping, blurring, spotting and cropping.

It is suggested to make a master copy and a duplicate copy for use for other purposes (ISO/TR 13028, 2010). Additionally, it is recommended to save the master images to the folders representing scanned packages (batches) to enable traceability. Image capturing software should additionally report the number of scanned images on the package level (ISO/TR 13028, 2010).

6.3. System time and date

The use of time and date in the overall process should be from the single source (e.g. used at the system level instead of at the work-station level) in order to properly monitor the process of digitization (ISO/TR 15801,2017). The time and date could also be obtained from a timestamping service. Additionally, a blockchain-based service could be used in order to timestamp the calculated hash values of each document in order to confirm that the particular scan was created at certain time and that it has not been changed afterwards[1].

7. MASTER IMAGE SPECIFICATIONS

It is a common agreement to create PDF, PDF/A or TIFF files with lossless compression. Regarding the tonality, a black and white document should be converted to a black and white image, a color document should be converted to 24-bit color image, a black and white photography should be converted to an 8-bit grey-scale image, a color photography should be converted to 24-bit color image (Government of Alberta, 2013; The Library of Congress, 2007). It is important to mention that if a batch consists of documents in black and white as well as in color, the whole batch should be converted using higher requirements, i.e. 24-bit color.

Documents digitzed to become master images should be scanned in duplex mode in order to prove that all information was captured, i.e. that no information form the back side of the documents were omitted.

If master images are being created as TIFFs than the predefined tags should be used to store image metadata in EXIF during the scanning process. Populating these tags can be helpful later on in the process of confirming the authenticity of a master image, for example the ordinal number of the image and the unique number of batch being digitized could be added. These data could confirm that the scanning process was performed according to the manual.

Similarly, while digitizing directly to the PDF file format, metadata can be entered together with the image and can include, for example information determining the source of the image, technical metadata relating to the scanning of the document, operator's name etc.

The metadata associated with the document becomes an inseparable part of the master, either being in TIFF or PDF file format.

8. APPLICATIONS ORCHESTRATION

In the digitization system, several applications and system-level software which are required for the digitization process to be fully functional need to work together in coordination. These are:

- preparation application,
- scanner driver,
- image capturing software,
- hash calculating application or service,

[1] *An example of a blockchain-based timestamping service is the time: beat service by Enigio Time. Accessed 18.11.2018. from https://timebeat.com/*

- blockchain-based timestamping service,
- indexing application,
- image and metadata control application, and
- SIP creation application.

The preparation application is used for creating the leading document of each batch which contains information needed for later verification of the digitized batch of documents. The scanner driver is the system-level software bonding scanner and image capturing software. The image capturing software is used for configuration of digitization options and for image acquisition. The hash calculating application or service is used to calculate hash that is going to be used later on in establishing integrity of the scanned images. The blockchain-based timestamping service is used for combining calculated hash with the trusted source of time and registration in the blockchain. The indexing application is used by the indexing operators. It should be implemented in a way that controls data entry, i.e. to stop the process if a required metadata field is skipped. A controlled vocabulary can be used for each metadata field along with the validation mechanisms in order to standardize the metadata entry procedures. The image and metadata control application is used for quality control assuring overall correctness and completeness. The SIP (Submission Information Package) creation application is used for delivering captured images with the status “completed” and the associated metadata to the repository for long-term storage and access.

9. DIGITAL ARCHIVE PROCESSES

Digital archive is a repository holding the approved submitted information packages (SIP) transformed into the archival information packages (AIP). The digital archive could contain three versions of the same image – the preservation, master, copy (e.g. in TIFF), the user, derivate, copy (e.g. in JPG), and an indexing, thumbnail, copy.

The repository will also store log-files used as the audit trail. It is necessary that they contain information on batch preparation, batch scan, batch control, re-scanning of document, hash of digitized image, timestamp, hash of file containing metadata of digitized image included in the SIP, SIP ingest confirmation, results of periodic repository controls and master image database controls, system status results, including scanner maintenance, computer maintenance, and computer network, information on used manuals including title, version and manual number, and system maintenance information. The audit trail of the actions taken during the digitization, submission, ingest, long-term storage and digital preservation processes can be used for authenticating digital copies in the form of original.

The digital archive has to have regular backups scheduled following the 3-2-1 rule which requires the creation of at least 3 copies on at least 2 different media while at least 1 copy should be on an off-site location which is either 70 km away and/or at another tectonic plate.

10. CONCLUSION

The process of mass digitization needs to be set up in a way to enable the creation of digitized version of analogue originals which could be authenticated as digital copies in the form of originals. Such digital copies could than continue to function and be used as originals while the analogue originals in that case are no longer needed. The new Croatian Law on archival materials and archives (Zakon o arhivskom gradivu i arhivima, 2018) allows this procedure although the bylaws specifying the requirements have not been enacted yet. However, this paper detailed a proposal of a mass digitization process setup that could be used to create digital copies which can be legally used as evidence instead of the paper originals.

11. APPENDIX 1.

Checklist for a digitization process aiming to create digitized copies in the form of originals

The checklist is organized in four sections relating to the four requirements detailed under 2 and consists of 53 check points.

Requirement	Yes	No
1. COMPLETENESS		
1.1. Does the digitized document have control data and the identification labels present?	☐	☐
1.2. Is there a link between the control data and markings and the captured image?	☐	☐
1.3. Are all pages of the document digitized, including both sides?	☐	☐
1.4. Are the accompanying connected papers linked to the document?	☐	☐
1.5. Is the creator of the document identified?	☐	☐
2. PROCESS RELIABILITY		
2.1. Is the process of digitization prescribed, approved and verified prior to digitization?	☐	☐
2.2. Are all parts of the process documented:		
a) preparation	☐	☐
b) scanning	☐	☐
c) indexing	☐	☐
d) control	☐	☐
e) re-scan	☐	☐
f) hash calculation	☐	☐
g) timestamping	☐	☐
h) required metadata	☐	☐
i) SIP creation	☐	☐
j) log file creation	☐	☐
k) ingest	☐	☐
l) AIP creation	☐	☐
m) DIP creation	☐	☐
2.3. Has a test on a representative sample been performed and logged?	☐	☐
2.4. Are all technical aspects documented:		
a) preparation application	☐	☐

Requirement	Yes	No
b) scanner	☐	☐
c) scanner driver	☐	☐
d) image capturing software	☐	☐
e) master image quality	☐	☐
f) derivative image quality	☐	☐
g) hash calculating application or service	☐	☐
h) blockchain-based timestamping service	☐	☐
i) indexing application	☐	☐
j) image and metadata control application	☐	☐
k) SIP creation application	☐	☐
l) digital archive / repository	☐	☐
m) backup copy rules	☐	☐
n) digital archive / repository infrastructure	☐	☐
2.5. Are monitoring roles existent?	☐	☐
2.6. Is the process being monitored?	☐	☐
2.7. Are there any records of the monitoring process?	☐	☐
2.8. Are control roles existent?	☐	☐
2.9. Is the process being controlled?	☐	☐
2.10. Are there any records of the controlling process?	☐	☐
2.11. Are supervisory roles existent?	☐	☐
2.12. Is the process being supervised?	☐	☐
2.13. Are there any records of the supervision process?	☐	☐
2.14. Are there any records of non-scheduled inspections carried out?	☐	☐
2.15. Have the necessary education been carried out before the start of the digitization?	☐	☐
2.16. Is the necessary training organized after at each significant process change or used technology, applications or services?	☐	☐
2.17. Are the introduced process changes or changes of the used technology, applications or services reflected in the digitization manual?	☐	☐
2.18. Is the digitization manual keeping track of the changes by creating a new version of the manual each time a change is introduced?	☐	☐
2.19. Is there a digitization system components maintenance scheduled?	☐	☐
2.20. Is there a digitization system components maintenance regularly enforced?	☐	☐
2.21. Are there any records of the maintenance checks being performed?	☐	☐
3. DOCUMENT INTEGRITY		
3.1. Is there evidence that the document, its control data, markings and related connected documents were not changed from the moment they were either created or imported into the business system?	☐	☐
4. AVAILABILITY		
4.1. Is the document, its control data, markings and related connected documents accessible, interpretable and usable?	☐	☐

POVZETEK

USTVARJANJE AVTENTIČNIH DIGITALNIH KOPIJ V OBLIKI ORIGINALA S PROCESOM DIGITALIZACIJE

Dr. Edvin Buršić
Finančna Agencija, Zagreb, Hrvaška
edvin.bursic@fina.hr

Dr. Hrvoje Stančić
Filozofska fakulteta Zagreb, Unvierza v Zagrebu, Hrvaška
hstancic@ffzg.hr

Zagotavljanje uporabnosti zapisov po preteku njihovega roka hrambe v arhivu lahko zahteva tudi postopke, kot so konverzija, migracija, emulacija, virtualizacija itd. Avtorja si tako postavljata osnovna vprašanja – ali lahko digitalizirane kopije nadomestijo original? Če lahko, pod katerimi pogoji? Ali naj pretvorba temelji na kvaliteti skena ali procesa? Ali je to pravno veljavno? Standard ISO 15489-1:2016 sicer definira avtentičnost, toda ali jo lahko uporabimo v kontekstu digitaliziranih dokumentov? Glede na standard je potrebno ustvariti razširjen nabor zahtev za digitalizirane dokumente. Te zahteve morajo vključevati celovitost, zanesljivost procesa, integriteto dokumenta in dostopnost.

Avtorja predlagata izdelavo kontrolnega seznama za ustvarjanje digitalnih kopij, ki bi bile enakovredne izvirniku, s 53 vprašanji o procesu digitalizacije. Najpomembnejši element v procesu digitalizacije je pretvorba fizičnega zapisa v digitalnega, ki naj bi bil enak izvirniku. Na proces digitalizacije lahko gledamo tudi kot na niz različnih delov, ki so nujni za konverzijo. Te dele sestavljajo: izbira dokumentov za digitalizacijo, pri čemer je potrebno upoštevati tudi pravne omejitve, kot so avtorske in sorodne pravice, analiza izbranega gradiva, reprezentativen testni vzorec, poimenovanje datotek, izračun zgoščene vrednosti (t. i. hash) digitaliziranega posnetka, tipi metapodatkov, ki so dodani posnetku, ter podatki nadzornega dnevnika.

Proces mora vsebovati določene vloge in odgovornosti. Te vloge in odgovornosti morajo biti razdeljene tako, da podpirajo kontinuiteto procesa, ki zagotavlja avtentičnost digitalnih kopij med digitalizacijo. Vloge so lahko razdeljene v vsaj pet kategorij: (1) postopkovne, (2) organizacijske, (3) nadzorne, (4) kontrolne in (5) vzdrževalne. Drugi pomemben podporni del procesa je tehnični vidik sistema digitalizacije, ki vključuje lastnostni optičnega bralnika, gonilnik in programsko opremo za zajem slike in sistem kraja in časa. Lastnosti optičnega bralnika vključujejo primerno uporabo senzorjev, razmerje med optično in interpolarno resolucijo, občutljivost na mehanske tresljaje, primeren vir svetlobe in optični del procesa optičnega branja, električne motnje, kot so magnetna in električna polja.

Če naj digitaliziran dokument velja kot original, mora ustrezati postavljenim zahtevam, vključno s primerno kompresijo, tonaliteto, procesi barvne konverzije itd. Avtorja se dotikata tudi vprašanj, povezanih z enostranskimi ali obojestranskimi načini optičnega branja (t. i. simplex/duplex) ter uporabo predhodno definiranih ključnih besed za shranjevanje metapodatkov o digitalizatu.

Avtorja bosta izpostavila pomembne elemente, ki jih morajo arhivi upoštevati pri digitalizaciji dokumentov, katerih digitalna kopija bi veljala kot original. Pri tem poudarjata zaupanja vreden proces, ki vključuje strokovnjake, opremo, programsko opremo in

procese. To so kritične točke v procesu konverzije fizičnih originalov v pravno veljavne digitalizate.

Sources and literature:

***Digitization Standard Process*. (2013).** Government of Alberta. Retrieved 18. 11. 2018 from http://imtdocs.alberta.ca/A000015_Digitization_Standard_v3.pdf.

***Evidence and Electronic Public Records*. (2010).** Public Record Office Victoria. Retrieved 18. 11. 2018 from http://studylib.net/doc/18578486/evidence-and-electronic-public-records.

ISO 15489-1 Information and documentation – Records management – Part 1: Concepts and principles. (2016). Geneva: ISO.

ISO/TR 13028 Information and documentation – Implementation guidelines for digitization of records. (2010). Geneva: ISO.

ISO/TR 15801 Document management – Electronically stored information – Recommendations for trustworthiness and reliability. (2017). Geneva: ISO.

Krogh, P. (2009). *The DAM Book. Digital Asset Management for Photographers (2nd ed.).* Sebastopol CA: O'Reilly.

Mason, S. (2006). *Proof of the authenticity of a Document in electronic format Introduced as evidence*. ARMA International Educational Foundation. Retrieved 18. 11. 2018 from: http://www.mnhs.org/preserve/records/legislativerecords/docs_pdfs/Proof_of_authenticity_of_a_document.pdf.

Peterson, A. K. (2004). *Standards Related to Digital Imaging of Pictorial Materials*. Library of Congress, Washington, D.C. Retrieved 18. 11. 2018 from http://www.loc.gov/rr/print/tp/DigitizationStandardsPictorial.pdf.

The Library of Congress. (2007). *Technical Standards for Digital Conversion of Text and Graphic Materials.* Retrieved 18. 11. 2018 from https://memory.loc.gov/ammem/about/techStandards.pdf.

The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project – Authenticity Task Force Report. (2005). Retrieved 18. 11. 2018 from: http://www.interpares.org/book/index.htm.

Zakon o arhivskom gradivu i arhivima. (2018). Official Gazette of Republic of Croatia, no. 61.