Преглед НЦД 37 (2020), 28–45

Marija Šegan-Radonjić
The Mathematical Institute SANU

CONTEMPORARY MODEL FOR THE DIGITIZATION OF ARCHIVAL RECORDS OF SCIENTIFIC INSTITUTIONS

Abstract. The topic of thispaperis the development of a framework for digital archiving in order to preserve, present and enable the availability of digitized and digital content for the purpose of historical and other research. The proposed framework is based on the concept of "thematic collections" and is intended for researchers who want to create their own digital collections of historical sources and texts in order to acquaint the wider scientific community with their research, connect them with a wider context and create conditions for networking and cooperation. On the example of digitization of archival material of the Mathematical Institute SANU,and in accordance with current recommendations and regulations for digitization of cultural heritage in the Republic of Serbia, the proposed framework offers guidelines for: 1) economical digital translation process to obtain operational copies for web presentation, 2) cataloguing and description of digital documents using the *Dublin Core* set of elements, 3) creating a digital archive using the *Omeka Classic* platform, 4) creating guidelines for archival research on specific historical topics, and 5) compiling historical essays in a digital environment. The expected results are: 1) proposal for a framework for digital archiving of digital and digitized content for the purpose of historical and other research, and 2) digital collection dedicated to the history of mathematics and related sciences in Serbia and Southeastern Europe.

Keywords.: Digitization of cultural heritage, digital history, digital thematic collections, history of mathematics.

1. Introduction

The topic of this paper is the development of a contemporary model for the presentation of historical and other research, based on a comprehensive process of digitization of cultural and scientific heritage and the concepts of "digital history" and "digital thematic collections". The choice of topic stemmed from the author's interest in the application of new technologies and methods in the presentation and interpretation of the past, especially the history of the development of the Mathematical Institute SANU. The issue of the development of institutional mathematics in Serbia and Southeastern Europe draws attention in the last twenty years, but it is difficult to analyze it in depth, as a whole, due to the fact that archival material is scattered within various "hidden" collections and fonds and is mostly unpublished or unavailable to the wider scientific community.

The concept of "thematic collections" (Thematic Research Collection), to which the term "invented archives"likwise corresponds, was introducedat the beginning of the 21st century to mark the pioneering endeavors of individual researchers in creating digital collections of historical sources and texts dedicated to a particular research question or topic (Palmer, 2004; Cohen, 2005). The development and wider use of personal computers, the Internet and web browsers have encouraged these researchers not only to create their own collections of data, but also to publicly present and share them on the web. What distinguished these collections from the ones preserved in

traditional heritage institutions (archives, libraries, museums and institutes for the protection of cultural heritage) was the fact that they contained contributions in the form of historical interpretations, discussions and arguments. Accordingly, this paper aims to offer a contemporary solution for the creation of digital thematic collections and their application in research.

The proposed framework is intended for "small" projects of digitization of archival material undertaken by researchers. It envisages that researchers themselves select the material they want to digitize and present, which is often not recognized by traditional archives as material of general importance, thusly remaining"out of sight". Hence, it also promotes cooperation with heritage institutions in order to responsibly preserve, present and make available the content under consideration. Furthermore, it presupposes networking and cooperation with fellow researchers, enabling knowledge transfer and enrichment. It is based on the use of existing and affordable solutions, and in case of the lack thereof, it analyzes the potential of new technologies and concepts.

In line with current recommendations and regulations for digitization of cultural heritage in the Republic of Serbia, the framework offers guidelines for: 1) economical procedure of translating analogue material into digital form in order to obtain operational copies for presentation on the web, 2) cataloguing and description of digital documents using the *Dublin Core* set of elements, 3) creating a digital archive using the *Omeka Classic* platform, 4) creating guidelines for archival research of certain historical topics, and 5) compiling historical essays in a digital environment.

For the purpose of making this framework, the Archives of the Mathematical Institute SANU, consisting mostly of archival material preserved due to the efforts of its members, was taken as a case study. It includes written, typed, photographed and printed documentary material covering a period of one hundred years – from the nineties of the nineteenth century to the nineties of the twentieth century – which has not been acquired by the state archives and which, in all likelihood, has not yet been assessed as a cultural heritage of general interest. Officially, it does not have the property of archival material, given that so far it has not been recorded, classified or described, there are no strict guidelines for its preservation and management, and it is kept in the premises of the Mathematical Institute SANU. The wider scientific community is generally unaware of the existence of this archive, which makes a significant contribution to the study of the history of mathematics and related disciplines in Serbia and Southeastern Europe. Therefore, it is necessary to offer a solution for its identification, cataloguing and presentation, which would provide a more precise insight into the material, facilitate the creation of thematic collections and enable a more accurate assessment of its general significance.

One of the solutions for preserving and presenting the archival material of the Mathematical Institute SANU (or some other "modest" archive / collection) to the general public lies in the process of digitization of cultural and scientific heritage. Given the fact that this is the material that still does not enjoy the status of cultural heritage, and that at this moment it is difficult to justify the demand for traditional methods of storage, preservation and protection, the process of digitalization seems to be an adequate solution.

Digitization of cultural heritage is a relatively new field of research and due to its dynamic development there is still no general definition. It is a multidisciplinary procedure of cultural heritage management in a technological environment, encompassing philosophical, social, cultural and economic aspects and consequences (Manžuch, 2005, 37).With regard to the choice of objectives, technologies, standards or funding models, this process varies from organization to organization, from country to country. When it comes to the Republic of Serbia, it is defined as a comprehensive

procedure for managing cultural heritage in the digital environment, which includes, but is not limited to, translation from analogue to digital form; establishing a metadata system and a description of digitized and digital material; development of tools, electronic catalogues and information systems and long-term preservation, presentation and providing access to data (Ministry of Culture and Information of the RS, 2019). This procedure is recognized as part of a complex system of preservation and management of cultural heritage in cultural institutions in Serbia (by the term "cultural institutions" we refer to museums, archives, libraries and institutes for the protection of cultural monuments) (Ibid.).

The rest of the paper is organized as follows. Guidelines for translating selected archival material into digital form are given in §2. In §3 the process of cataloguing and describing digital documents is discussed. §4 analyzes the proposed procedure for creating digital archive, whereas §5 focuses on the development of instructions for archival research. §6 provides an example of formulating a historical text in a digital environment. §7 contains concluding remarks of the paper, followed by a list of sources and literature.

2. Translation into Digital Form

The focus of this paper is the archival material related to the activities of the Mathematical Institute in the period from its founding as part of the Serbian Academy of Sciences in 1946 to its transformation into an independent institution in 1961. The digitization of this material is important for several reasons: 1) as a historical source, it testifies to the crucial years for establishing the first scientific institute of the Academy, with reference to the circumstances of science and the scientific community in Serbia and Yugoslavia after World War II; 2) among the authors, we find the names of eminent mathematicians and scientists, such as, among others, MilutinMilanković, Vojislav Mišković, Anton Bilimović and Jovan Karamata; 3) the wider scientific community is generally unfamiliar with its existence and content, and 4) part of this material represents the only preserved original. In that sense, the process of digitization would provide an appropriate insight into the archival material, facilitate future research and be of assistance when making decisions about its protection and preservation.

Around 100 units of archival material or about 500 sheets related to the activities of this institution covering the period from 1946 to 1961 are kept in the Archives of the Mathematical Institute. For their translation into digital form, the Guidelines of the Ministry of Culture and Information on the Digitization of Cultural Heritage (Ministry of Culture and Information of the RS, September 2017) were followed. Since the idea is to preserve and present their content for use in research purposes, the question of materiality has been neglected in this case. So, instead of creating master digital copies intended for long-term storage, only operational copies for publishing and exchange were created. The procedure for translating into digital form is described below on the example of an administrative manuscript book, as well as on the example of a scientific text containing mathematical formulas and charts.

a) Example 1. *Book of Minutes of Sessions of the Council of the Mathematical Institute SAN 1948 – 1954* is a manuscript notebook with lines, hardcover, size 296 x 210 x 15mm, which is kept in the Archives of the Mathematical Institute (Figure 1). Its content includes 53 sessions of the Council of the Mathematical Institute SAN, at which the scientific, administrative and financial activities of the Institute in the period between 1948 and 1954 were discussed. Each session was conducted according to a template: the header begins with the ordinal number, date and names of the participants in the session; the main text consists of the issues discussed, and at the very end are the

signatures of the minute taker and the chairman (Figure 1). The book distinguishes four Cyrillic manuscripts of four minute takers, who at that time acted as secretaries of the Council: Jovan Karamata, Milan Vrečko, RadivojeKašanin and TatomirAnđelić.

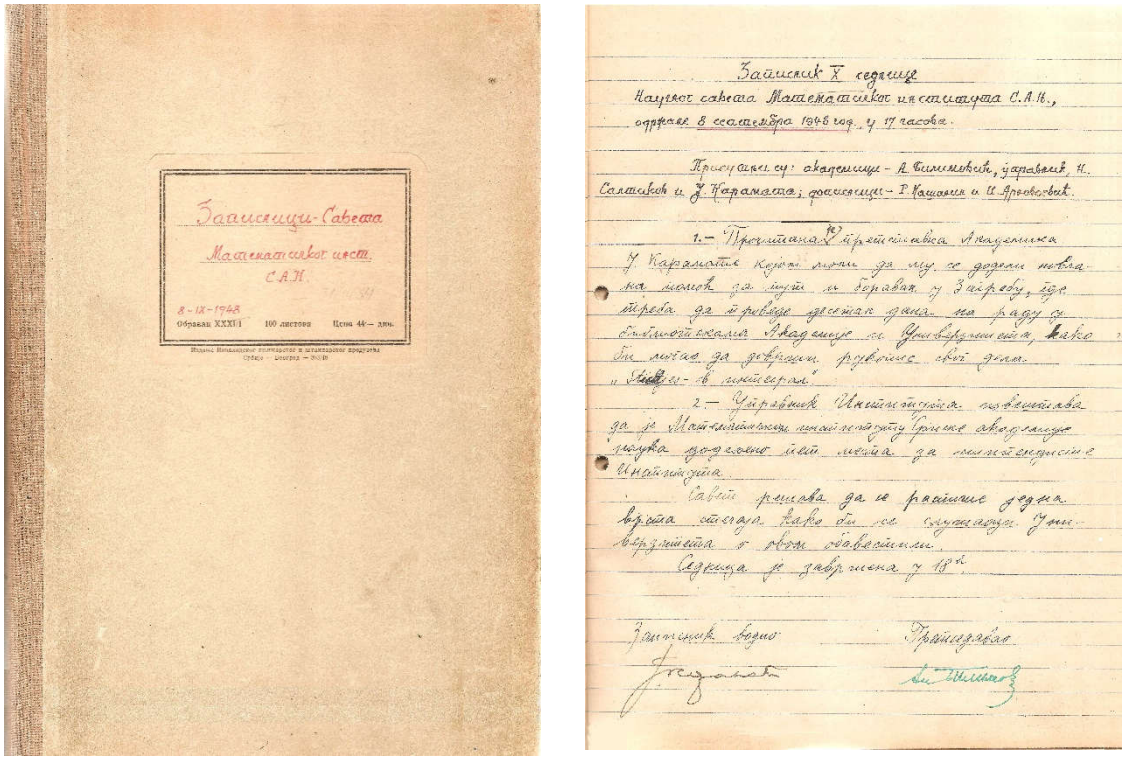


Figure 1. *Book of Minutes of Sessions of the Council of the Mathematical Institute SAN 1948 – 1954.* (Archives of theMathematical Institute SANU)

The need to translate this book of minutes into digital form arises from the following reasons: 1) as a historical primary source, it testifies to the pivotal years for the establishment and activities of one of the scientific institutes of the Serbian Academy of Sciences, as well as to the circumstances of science following World War II; 2) historians of science and the wider scientific community are not familiar with the existence and content of this source, and 3) only one copy has been preserved, which has not been published and presented to the public so far. It is worthwhile noting that it does not currently enjoy the status of archival material, which is why demanding methods of conservation and archiving are not justifiable in this case. Therefore, digital translation is an accessible and transitional solution for preserving and making it available until conditions are favorable for appropriate long-term storage.

The translation process was realized through the process of scanning analogue material in order to obtain a digital copy in the form of a raster image. Where necessary, image processing was performed to remove deformations. Affordable technology and open or free software were used to make the process as accessible as possible to the researcher (Table 1).

The process is not solely limited to obtaining a raster image, but also electronic text, since it does not require a large storage space, offering the possibility of further processing and allowing easier search of content. Translating handwritten text into electronic text poses a challenge because Handwriting recognition programs (HWR) are still under development. The recognition of different Cyrillic manuscripts in the Serbian language, as is the case here, makes the problem even more difficult. There is a

possibility that some Optical character recognition programs (OCR), such as free *Tesseract* software, may be "adjusted" to recognize the appropriate Cyrillic handwriting, since similar attempts – in the case of Chinese and Bengali scripts – have met with success (See Li, 2016 and Hasnat, 2009). In addition, it has lately been recommended that artificial neural networks be used in sequence recognition problems, such as hadwriting (Gavran, 2016, 41; Scheidl, 2018). In the meantime, in anticipation of advanced solutions for handwriting recognition, the author opted for manual typing of the text, and the translation results are presented within a special website available at: http://www.mi.sanu.ac.rs/~msegan/teza/convert/convert.html#title2.

Process	Technology	Resolution	Format	Expenditures
Scanning	HP DeskJet F370	72 dots per inch(dpi)[1]	Compressed: JPEG	cca. 60 pages per hour[2]
Raster Image Processing	GIMP-GNU Image Manipulation Program		JPEG	Open source software
	Image to PDF or XPS		PDF	Open source software
Translation into electronic text	Retyping			cca. 10 pages per hour[3]
Electronic text processing	Apache OpenOffice		Structured text: ODT	Free software
	Notepad++		Unstructured text: TXT	Free software

Table 1.. Procedure for converting an analogue unit into electronic form: Example 1

b) Example 2. Apart from texts of an administrative nature, the Archives of the Mathematical Institute SANU houses scientific manuscripts of its members. An example of translating a scientific text into digital form is shown below. It is the manuscript of the second part of the *Collection of Solved Problems in General Astronomy*, compiled by Academician Vojislav Mišković during the period between 1954 and 1961 when he headed the Astronomical and Numerical Section of the Mathematical Institute. Although there was a plan to publish this manuscript a year after the first part of the *Collection* was launched in 1956, unfortunately it never saw print (Pejović, 2011, 460).It was first digitized and presented to the public in 2010 within the "Virtual Library of the Faculty of Mathematics" (see Pejović, 2011), and in 2018 it was entrusted to the Mathematical Institute SANU until further notice. The manuscript is kept in a file which includes the first printed edition of the first part of the *Collection*, a draft of the second part, a plan for assembling both volumes into single whole and original drawings by Milan Čavčić (Figure 2). For the purpose of this paper, only one chapter of the manuscript has been digitized, as well as the drawings belonging to that

[1] Since the goal was to create operational copies to be published on websites, the author decided to use the resolution of 72 dpi. However, for master copies and long term preservation, it is recommended to use at least 300 dpi (Ministry of Culture and Information of the RS, September 2017, 18).

[2] Note that this estimation depends on how sensitive is a material being scanned, the scanning speed, etc. The given estimation applies to the examples 1 and 2.

[3] Note that this estimation depends on number of character per page, typing speed, readability of handwritten text, etc.The given estimation applies to the examples 1 and 2.

chapter to illustrate the text conversion process that includes mathematical formulas and charts.

Process	Technology	Resolution	Format	Expenditures
Scanning	HP DeskJet F370	300 dots per inch(dpi)	Compressed: JPEG	cca. 60 pages per hour
Raster Image Processing	AutoCAD Raster Design 2017		DWG	Commercial software
	Image to PDF or XPS		PDF	Open source software
Translation into electronic text	Optical character recognition		NewOCR.com	free OCR “online”service
	Retyping			cca. 10 pages per hour
Electronic text processing	TeX		Structured text: TEX	Free software

Table 2.. Procedure for converting an analogue unit into electronic form: Example 2

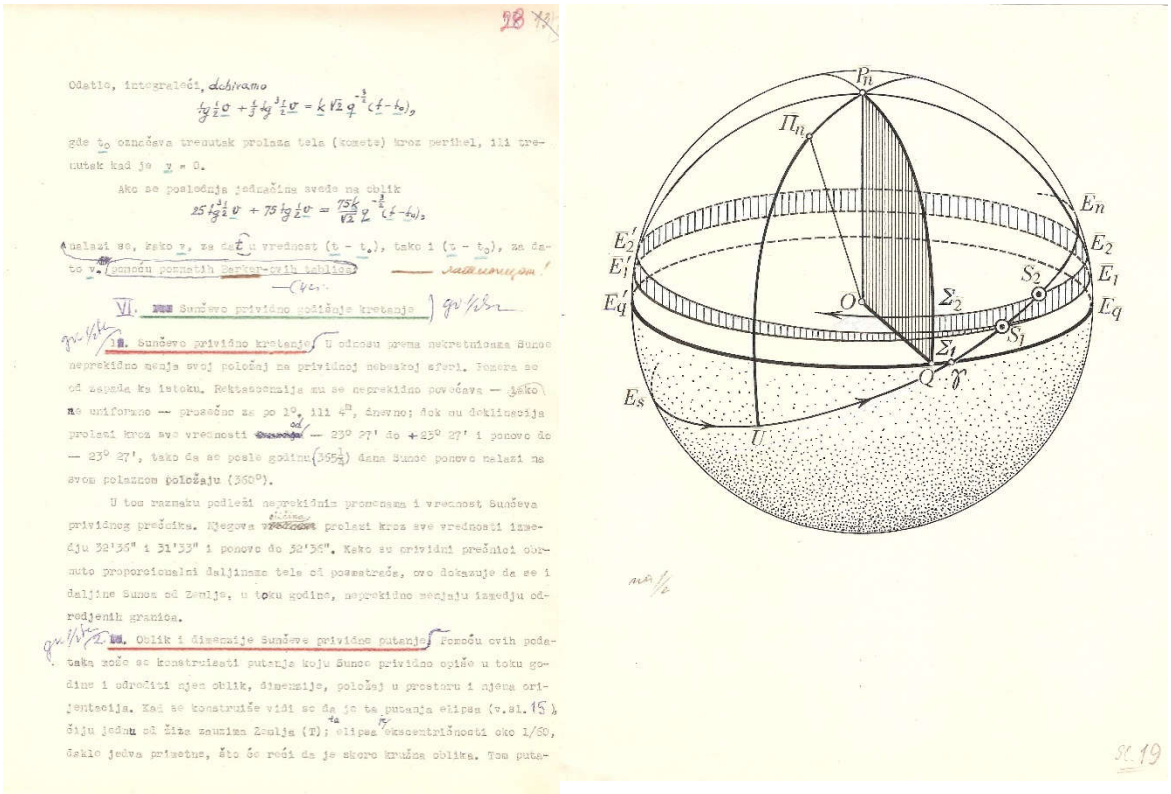


Figure 2. Manuscript of the second part of *the Collection of Solved Problems in General Astronomy*. (Archives of the Mathematical Institute SANU)

The manuscript of the sixth chapter, “The Annual Apparent Movementof the Sun”, is particularly noteworthy because it contains a typedtextin the Latin script,

handwritten interpolations in both Cyrillic and Latin alphabets, and formulas written by hand on papers of various formats (Figure 2). Therefore, apart from the scanning procedure for obtaining a digital copy in the form of a raster image, the procedure of optical character recognition and manual typing was applied in order to obtain electronic text. Since it is a material that contains mathematical formulas and functions, TeX free software for processing scientific texts was used to create electronic structured text.

The sixth chapter is supplemented by six cards with ten drawings created in 1962 by Milan Čavčić, who was employed as a calculator at the Mathematical Institute since 1954. In addition to the scanning procedure for obtaining raster images of these drawings, the vectorization procedure of the raster image of drawing no. 18 was also applied using appropriate vector graphics programs. After inserting the raster image of this drawing into the vector graphics program, the vectorization procedure was done manually, taking into account the scale factor so that the dimensions of the vector image correspond to its raster original. The obtained vector image is presented in Descartes' coordinate system as a series of objects, i.e. points and straight lines (Figure *3*). The vectorization procedure minimizes the consumption of storage resources without compromising information. In addition, it is possible to seemingly separate the segments of the drawing, in this case straight lines, as well as to select their building elements, in this case points. The coordinates of the selected (measured) points can be stored in special meta-files, which can be processed depending on the needs. For example, it is possible to automatically reproduce (graphically illustrate) drawing no. 18 solely by using metadata with point coordinates (i.e., without raster image as starting point) in vector graphics programs. The commercial software package *AutoCAD Raster Design 2017* was used for the vectorization process, and the results were presented within a special website available at:
http://www.mi.sanu.ac.rs/~msegan/teza/convert/convert.html#title3.

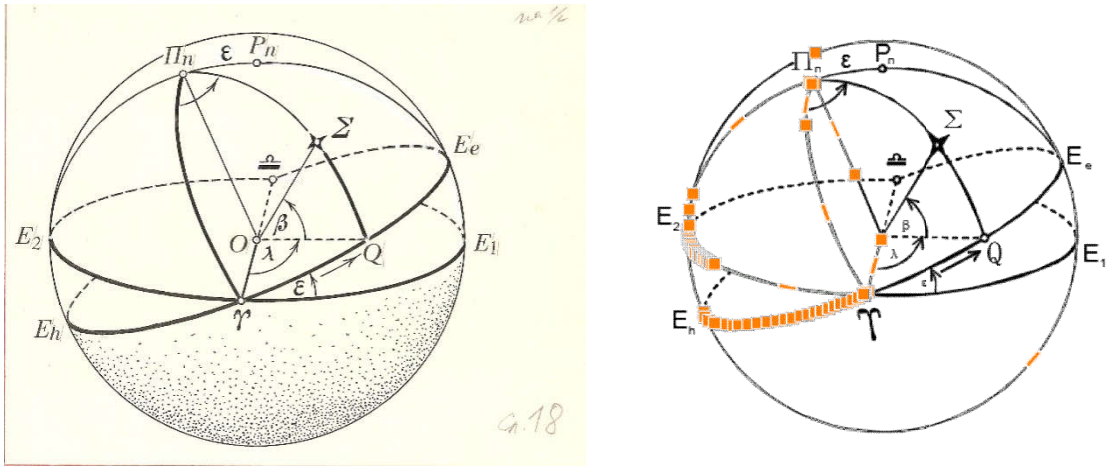


Figure 3. Manual vectorization procedure (right) based on raster image (left). (Archives of the Mathematical Institute SANU)

After translating the archival material into digital form, it is possible to move on to the next step of the digitization process, which includes cataloguing the material and establishing a metadata format.

3. Cataloguing and providing metadata formats

Cataloguing is an organized process of recording information possessed and emitted by objects and units of heritage(Maroević, 1993, 190). For this purpose, there are numerous information languages for analyzing, indexing and storing information with the goal of their subsequent successful locating, transfer and exchange (Crvčanin, 1983, 9). Therefore, the choice of appropriate metadata format represents one of the challenges in the cataloguing process. The term "metadata" has been widely used since 1983 to denote data that provides information about other data(„Metadata", n.d.*).* These are "structured information used to describe, explain, locate, or otherwise make it easier to retrieve, use, or manage a source of information" (Hodge, 2001, 3). Heritage institutions have a long tradition in creating structured data and, in this regard, they have developed different descriptive formats depending on current needs and regulations.

The development and wider use of personal computers, the Internet and web browsers have transformed the environment in which data is created and searched. There is a need for a simpler cataloguing practice that will enable the identification, linking and exchange of data without losing important information (Vukadin, 2014, 323). In the field of heritage preservation, various initiatives for establishing and improving metadata schemes and standards have emerged for the purpose of either describing the source for its identification or providing information on resource management (Öberg, 2008, 224; Hodge, 2001, 3).It is known that the most widely accepted formats are "Dublin Core metadata", which is mainly used in librarianship, "EAD" as the American standard for the description of archival material, or "CIDOC CRM"for the description of cultural heritage. Detailed reviews of several of the most widely used schemes and standards in heritage institutions are presented in (Sotirova, 2012, 32-41; Riley, 2017).

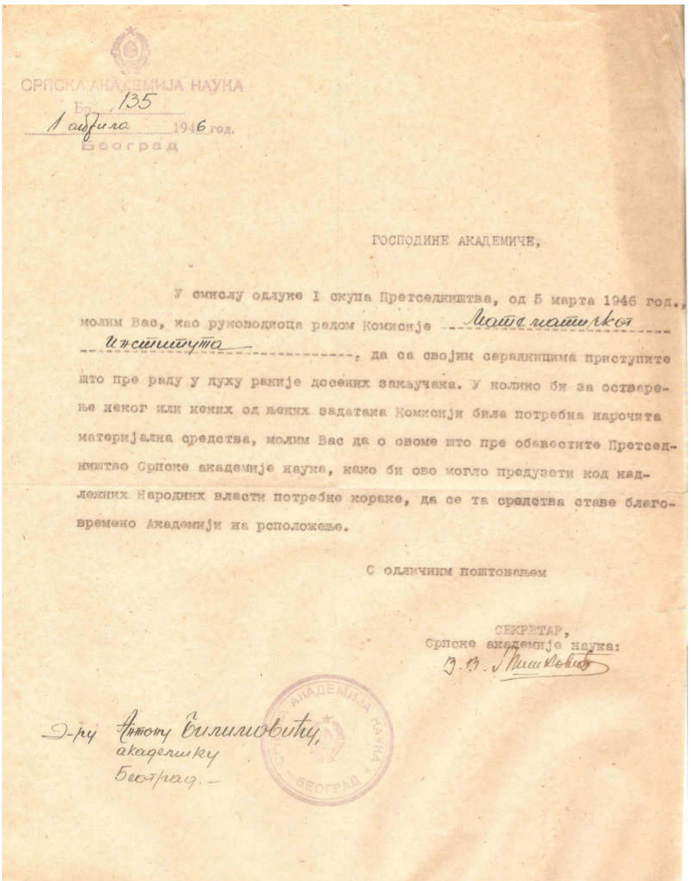
For the cataloguing of the archival material of the Mathematical Institute SANU, the choice fell on the mentioned "Dublin Core" (DC) set of descriptive elements, because it is the starting point of many metadata formats that are used nowadays. It was created as a result of collaboration between librarians and programmers, who gathered in 1995 at a workshop in Dublin (Ohio, USA) to improve existing standards and specialized dictionaries for describing material units and sources (The Dublin Core Metadata Initiative (DCMI), 2014). This metadata format became very popular in a short time, because it offered the possibility of describing all types of cultural goods, while ensuring interoperability (Marinković, 2016, 72). Also noteworthy is that it is simple to operate and intuitively understandable to the user, so even those who do not have experience in cataloguing can create a quality description. The Simple DC contains only fifteen optional elements, which can be repeated and which provide a basic idea of a certain unit (Table 2). It is a stable, somewhat restrictive set, because different information must be grouped within one element. However, there is a possibility to expand the basic set with new elements by switching from Simple DC to Qualified DC.

In order to illustrate the applicability of this format, a description of a digitized document is given below (Table 3), the original of which is kept in the Archives of the Mathematical Institute SANU, being part of a group of documents stored in a file for the year 1946. This is a letter of Vojislav Mišković, the secretary of the Serbian Academy of Sciences, to Anton Bilimović, the head of the Commission of the Mathematical Institute, regarding the organization of activities of the Institute. The letter is typed on an A4 paper, containing the official seal of the Academy with the ordinal number and date of registration, as well as Mišković's signature (Figure 4). It is a primarysource,particularlyrelevant with respect to determining the exact date of the establishment of the Mathematical Institute.

Element	Description
Title	Letter of Vojislav Mišković, Secretary of the Serbian Academy of Sciences, to Anton Bilimović, Head of the Commission of the Mathematical Institute
Creator	Vojislav Mišković
Subject	Commission of the Mathematical Institute
	Mathematical Institute
	Serbian Academy of Sciences
	Anton Bilimović
	Vojislav Mišković
Description	Letter of Vojislav Mišković, Secretary of the Serbian Academy of Sciences, to Anton Bilimović, Head of the Commission of the Mathematical Institute regarding the organization of activities of the Institute
Publisher	x
Contributor	Scan: MarijaŠegan-Radonjić
Date	1946-04-01
Type	Document
Format	PDF
Identifier	x
Source	Archives of MISANU, F1946
Language	Serbian=srp
Relation	X
Coverage	Serbia
	Yugoslavia
	1948 – 1954
Rights	Mathematical InstituteSANU

Table 3. Example of a description of an archival document of the Mathematical Institute using the basic DC set.

An example of a list of around 100 documents from the Archives of the Mathematical Institute, covering the period from 1946 to 1984, was created using the *Apache OpenOffice Calc* program. This example is a simple database compiled using an open source spreadsheet program, which is based on elements of the basic DC set. It leaves the possibility to be supplemented, through further research, with new findings, as well as to be "translated" into a machine-readable form. The example is available at: http://www.mi.sanu.ac.rs/~msegan/teza/katalogizacija_1946_1984.ods


СРПСКА АКАДЕМИЈА НАУКА
Бр. 135
1 априла 1946 год.
Београд

ГОСПОДИНЕ АКАДЕМИЧЕ,

У смислу одлуке I скупа Претседништва, од 5 марта 1946 год., молим Вас, као руководиоца радом Комисије Математичког Института, да са својим сарадницима приступите што пре раду у духу раније донених закључака. У колико би за остварење неког или неких од њених задатака Комисији била потребна нарочита материјална средства, молим Вас да о овоме што пре обавестите Претседништво Српске академије наука, како би ово могло предузети код надлежних Народних власти потребне кораке, да се та средства ставе благовремено Академији на расположење.

С одличним поштовањем

СЕКРЕТАР,
Српске академије наука:

Д-ру Антону Билимовићу,
академику
Београд

Figure 4. Letter of Vojislav Mišković, Secretary of the Serbian Academy of Sciences, to Anton Bilimović, Head of the Commission of the Mathematical Institute regarding the organization of activities of the Institute, April 1, 1946.(Archives of the Mathematical Institute SANU, F1946)

4. Establishing digital objects and creating a digital archive

After translating into digital form and cataloguing the selected archive material of the Mathematical Institute SANU, the next step includes the establishment of digital objects (consisting of digital documents and their associated metadata) and the creation of digital archive. As mentioned earlier, the process of creating this archive is based on the concept of “thematic research collection”, for which the term “invented archives” can also be used (Cohen, 2005). Lately, this concept has become an increasingly common genre of digital scientific production (Fenlon, 2017, 523). These are digital collections of the primary sources and related material intended for a specific research question or topic (Palmer, 2004, 348). Their purpose is to support current research and to encourage new knowledge, and their creators are mostly researchers (Ibid.). What distinguishes them, among other things, is that they are electronic in format, heterogeneous in the sense that they contain different types of material, thematically coherent, structured for easier analysis and search and open in order to provide the possibility for changes and additions (Unsworth, 2000). They can be extracted from or added to a larger collection or fund (Palmer, 2004, 348). One of the first collections of this type is *The Valley of the Shadow*, dedicated to the American Civil War, which was first introduced online in 1993 (Cohen, 2005). From that moment until today, hundreds of collections of this type have been published on the Internet, in various stages of development (Fenlon, 2017, 523).

The free open source platform *Omeka Classic*, developed by the Roy Rosenzweig Center for History and New Media, was chosen for creating a digital thematic collection or archive dedicated to studying the development and impact of the Mathematical Institute. It should be borne in mind that there are many other platforms for publishing digital objects on the Web, such as *DSpace*, but the advantage of *Omeka Classic* is that it was initially designed for individual projects by archivists, librarians and researchers, who have limited resources at their disposal (Puckett, 2016, 376).[4] Furthermore, it supports various types of digital documents, including image, text, audio and video, using the aforementioned *Dublin Core* metadata format for their description. Therefore, the use of this platform does not require special technical knowledge and can be quickly mastered after consulting the user guidelines available within the official *Omeka Classic* website (see „Omeka Classic User Manual“) or a large community of users active in various Internet forums (see „Omeka Forum“).

Although there is a possibility of free download of the platform and its installation on your own server, the author has chosen the most favorable subscription package for the hosted version available within the *Omeka.net* service, thus avoiding the effort of installing and maintaining his own system, which requires certain technology and technical knowledge (see „Omeka.net“). After opening an account and selecting one of the offered web presentation designs, several special plug-ins (e.g. option allowing the user to make web presentation in Serbian language) were activated – in addition to the basic function offered by the platform (publishing digital objects on the web) –for the purpose of expanding the functionality of the platform. The plug-ins that will be activated depend on the offer of the selected package, as well as the needs of the creators and users of the digital repository.

The procedure for creating a digital collection on the example of digital and digitized material on Professor VeljkoVujičić is described below. These are various digital documents – newspaper clippings, photographs, texts and videos – whose originals are kept in the Mathematical Institute SANU– testifying to his life and work, especially from the moment of his appointment in 1961 as assistant director of the newly formed independent Mathematical Institute (Figure 5).

The first step in creating a digital collection involves uploading digital documents to an online location, in this case *Omeka.net*. Metadata is then added to the uploaded documents to establish digital objects. It is possible to import metadata automatically using a metadata collection plug-in (OAI-PMH Harvester) or CSV format plug-in. For the collection of over 100 digital documents, it is recommendable to use the benefits of automatic import, provided that there is already a source of metadata (Jakovich, 2018).

In addition to the Dublin Core set of elements, one digital object contains metadata about the materiality of the object, its original format and dimensions, as well as an unlimited number of keywords (“tags”) that can be defined by the object creator or taken from controlled dictionaries. For the purpose of this paper, the key terms are mostly defined by the author herself.

4 There is also the *Omeka S* platform intended primarily for institutions that maintain multiple websites and are interested in their simpler updating, as well as in sharing resources using the Linked Open Data method (see Tour of Omeka S). The functionality provided by *Omeka S* exceeds the needs of this paper, hence the author opted for the Omeka Classic platform.

Figure 5. Examples of various archival material (Archives of the Mathematical Institute SANU)

Multiple files can be imported within each object, with a file size limit of 128 MB. Thus, for example, the object concerning the first session of the staff of the Mathematical Institute, held on September 5, 1962, at which Vujičić was elected a member of the Council, contains two files: a raster image of the original document and an unstructured text of the document.[5]

Digital objects can be organized into digital thematic collections. Thus, the objects on Vujičić are gathered within a special collection (see „VeljkoVujičić Collection"). The challenge is the *Omeka* platform condition that one digital object can belong to only one collection. For example, the object regarding the tenth session of the Council of the Mathematical Institute, held on September 13, 1962, which confirmed the election of Vujičić as a member of the Council, may belong to both the collection dedicated to Vujičić and the collection dedicated to the Council of the Mathematical Institute. However, due to the said condition, the author opted for one of the two collections.[6]

Following the establishment of the digital objects and collections, there is a possibility that they will be publicly or partially published or shared only with certain persons. The digital archive is available at the internet address: https://misanuarc.omeka.net/. It should be taken into account that this presentation is a temporary solution, since the *Omeka* platform within the *Omeka.net* service carries a number of challenges: among other things, it is not fully translated and localized into Serbian, it has limited storage space and there is no possibility for more adaptability of design and functionality. However, it provides an opportunity to present research and materials in an efficient and effective way, which can encourage greater visibility, cooperation and support for further development.

5. Creating guides for archival research

The archive research guides seek to provide users with an insight into archival, library, or museum collections dedicated to specific historiographical issues, making the search process more efficient. It is mainly compiled by the owners of collections, such as The

5 See*Minutes of the 1st Session of the Collective of the Mathematical Institute* (September 5, 1962). Digital Archive of the Mathematical Institute SANU. Available Online: https://misanuarc.omeka.net/items/show/79.

6 See Minutes of the 10th Session of the Council of the Mathematical Institute (September 13, 1962). Digital archive of the Mathematical Institute SANU. Available Online: https://misanuarc.omeka.net/items/show/80.

National Archive UK, which has compiled and published over 350 guides for research on various historical topics (see „Research Guides“, n.d.) or the Archives of Yugoslavia, which prepared a guide through the fonds of the Kingdom of Yugoslavia (see Pijevac, 2000). The development and wider use of the global computer network and new technologies have motivated researchers to take action. Thus, for example, a group of historians created guides for the study of the First World War and the Middle Ages and made them publicly available within the “Collaborative European Digital Archive Infrastructure” (CENDARI) platform (see „Public Research Guides“, n.d.).

One of the features of guides for archival research is that they offer the possibility of linking research in traditional institutions with research in the digital environment, thus overcoming the current limitations of historiographical research („Position Paper on ARGs“, 2013). It makes the transition from “traditional” to “digital” researcher easier; each researcher can use or create guides for archival research and then make them publicly available for sharing, networking, and collaboration. It also provides insight into collections of “hidden” archives, such as the Archives of the Mathematical Institute. Finally, it implies a *process*, which corresponds to a new paradigm of research work according to which the research does not end with the first publication of results, but is constantly supplemented and transformed by new findings (Fitzpatrick, 2011, 8-14).

In this paper, the CENDARI model was used and adapted as a concept for creating guides for archival research.[7] Apart from standard elements such as title, summary, and introduction, the guides may include an overview of the countries and institutions where the desired material is stored. Within the element “institution”, the author of the guides can provide a description of collections and fonds of the institution where the desired material is kept, conditions of access and use of materials, subject determinants that put his research in a broader context and review of sources and literature (Table 4).

Element (Eng)	Description
Title	Guide title
Author	Guide Author
Abstract	Short description of the contents
Introduction	Introduction to the research topic, determining the historiographical context and methodology for creatingresearch guides
Major Research Questions	
Primary Sources	Sources that authentically testify to the event described
Secondary Sources	Sources based on primary sources
Additional Sources	Sources that do not fall into any of the two categories mentioned above
References	

Table 4. Structure of the CENDARI model for creating guides for archival research[8]

7 The author had the opportunity to get acquainted with this concept as a student of the second CENDARI summer school “Researching the First World War in a Digital Environment”, held at the Friedrich Meinecke Institute in Berlin (Germany) from July 21–25, 2014. For further information about CENDARI, see: Boukhelifa, N. et al., 2018.

8 The chart and description are based according to: „Intro to Thematic Research Guides“, n.d.

For the purpose of compiling guides, which will be publicly available, CENDARI proposes the use of the tool "Note Taking Environment" (NTE), which is part of its virtual environment for researchers(„Intro to Thematic Research Guides", n.d.;„Working with the NTE", n.d.). Interesting benefits provided by this tool are, among other things, the recognition of named entities, the possibility for the user to mark persons, organizations, places and events so that the system can recognize and connect them to appropriate databases, as well as group entities into clusters or visually present their data. Since the environment is intended primarily for researchers, in order to be able to use it, it is necessary to obtain the authorization of the appropriate body for authentication and authorization. It should be borne in mind that the "online" guide can be created using other tools and programs. In this paper, for example, the guide is compiled as a simple web page using the descriptive languages HTML and CSS so that it can be imported into most other HTML-enabled environments, such as NTE. In doing so, efforts were made not to disregard the basic guidelines for creating websites for historical research, such as the guidelines of the Roy Rosenzweig Center for History and New Media (RRCHNM). (Table 5).

Questions	
General goal	Guidelines for archival research of a certain historical topic
Target group	Researchers/students, teachers/students
Sections	SeeTable3
Contents	Text, images
Language	Serbian
	English (as soon as conditions are met)
Tools	Notepad++ Editor

Table 5. Guidelines for creating web page for archival research.[9]

In accordance with the guidelines given in Tables 4 and 5, a web guide was created specifically for archival research on the role of academician Mihailo Petrović (1868–1943) in the establishment of institutional mathematics in Serbia and Yugoslavia, which is available at: http://www.mi.sanu.ac.rs/~msegan/teza/arg/arg.html. The proposed guide contains a list of heritage institutions in which the material necessary for the research of this topic is kept, such as the State Archives of Serbia, Archives of SANU, Archives of Yugoslavia, Archives of the Mathematical Institute SANU, etc. Then, an overview of fonds and collections with the desired material is given, for example, if we are referring to the State Archives of Serbia, these include the following administrative fonds: *Kraljevsko-srpsko Ministarstvo prosvete i crkvenih poslova* [Ministry of Education and Church Affairs of the Kingdom of Serbia], *Ministarstvo narodne privrede* [Ministry of National Economy], *Filozofski fakultet* [Faculty of Philosophy] and *Univerzitet u Beogradu* [University of Belgrade]. In addition, within each fond, there is a list of all the documents related to the research topic (see „Review of Documents on Mihailo Petrović in Archive of Serbia", 2018). Finally, the index of all the individuals and organizations appearing in these documents can link this research topic to a broader context, for example, the history of science and education in

9 The Table 4 is based on: Site Planning Tips, n.d.

Southeastern Europe. The advantage of web guides is that they can be constantly updated with new findings. They connect the previous two phases of the work: cataloguing and thematic digital collection with the next phase, i.e. historiographical research.

6. Historiographical research and production of digital essays

The development of the Internet and web technology has enabled the development of "digital" history, i.e. a new way of presenting and interpreting the past. In previous chapters, it has been shown how a researcher, in addition to publishing scientific papers in a digital environment, can take on the role of an archivist, create a digital collection of historical sources and texts, and publicly present and share them. Whether this material will be what historian Daniel J. Cohen calls "raw"or "cooked" depends on whether or not it is organized and structured (Cohen, 2004, 337). The author of this paper decided to structure the archive material of the Mathematical Institute using the *Dublin Core* set of elements, to organize it into thematic collections using the *Omeka* service, as well as to offer "online" instructions for archival research on a certain topic related to the history of mathematics in Serbia. The next step may be the use of structured and organized material in historiographical research and the compilation of historical papers and essays.

As previously mentioned, the first web presentations appeared at the end of the 20th century, intended for the widest audience, which brought together the presentation of archival material and accompanying historical texts (Cohen, 2005). The Internet and the web have provided researchers with the opportunity to publish all the collected material in one place and to give an accompanying interpretation, which printed publications could not often support. The possibilities of the new medium were especially attractive for the presentation of materials and topics dedicated to a narrow field of research, which at that time did not enjoy popularity or were considered significant, and therefore were not intended for publication and printing. Since it is difficult to predict the interests of future generations, and history is full of examples showing that what may seem unimportant in one period may be revolutionary for later generations, it is easy to understand the decision of a growing number of researchers to share their material and work in an environment free of central authority and reviews, as well as to contribute to the further development of digital standards, tools and methods (see *CENDARI* project).

Similar to the process of combining historical text with historical sources in a digital environment, the author has compiled a digital essay on how the person and the work of Mihailo Petrović were perceived at the Mathematical Institute back in the 1950s. This essay includes the author's text, digitized archival documents with metadata and instructions for archival research. It was presented as a digital exhibition within the *Omeka* service, and is available at: https://misanuarc.omeka.net/exhibits. The user has at his / her disposal both the original resource and the research text. He / She can offer his / her interpretation of the resource, as well as the original resource at the same time. For example, there is a statement that in 1947, only one year after the Institute was established, there was a proposition to the Government to allocate the birth house of Petrović to the Institute with the goal to create a museum and mathematical library (https://misanuarc.omeka.net/exhibits/show/https---misanuarc-omeka-net-mp/mp_doc_2). In the traditional publication, the reference will be put in text to back that statement. In a digital environment, however, we can provide the document itself. The advantage of this model is that the digital essay can be changed and supplemented with new sources and findings, and therefore represents a process rather than a research

result.

7. Conclusion and future work

The paper has provided a framework for digital archiving in order to preserve, present and enable the availability of digitized and digital content for the purpose of historical and other research. On the example of the digitization of so far unpublished archival material of the Mathematical Institute SANU, the paper has offered guidelines for translating analogue material into digital form, cataloguing and description of digital documents, creating a digital archive, creating guides for archival research and compiling digital essays. The conclusion is that the application of the proposed framework, apart from enabling interoperability, translation into other standards, as well as migration to other media, likewise provides the possibility of economical and quality presentation of digitized and digital content on the web and possible post-systems, as well as its availability for the long term.

The paper recommends the researchers to implement the practice of heritage institutions and compile and make available guides for archival research during the process of research and gathering material. The paper has presented a proposal for creating web guide in accordance with the CENDARI model, and an example has been created using the descriptive languages HTML and CSS. The conclusion is that this procedure provides insight into "hidden" collections and fonds, enabling easier search and connecting research in traditional archives with research on the web.

As a result of the application of the proposed framework, the paper cites the creation of the Digital Archive of the Mathematical Institute SANU, which contains digital and digitized archival material of this institution, as well as an example of a digital essay. The significance of this digital repository is that it makes a new way of presenting and interpreting the past of the Mathematical Institute, as well as the history of mathematics and related sciences in Serbia and Southeastern Europe. It has been concluded that its advantage is that it can be changed and supplemented with new sources and findings, thus representing a long-term process rather than the result of research.

Although initially this was a minor research project, the Mathematical Institute SANU has recognized the potential and expressed its readiness to support further digitalization of its archive. Hence, its developers, in cooperation with the author of this paper, are currently working on the development of an independent digital repository and migration of material from the *Omeka* platform, as well as the translation of metadata from DC to NCD standard, i.e. to national metadata format(seeNCD standard).

References

1. Boukhelifa, N. et al. The CENDARI Infrastructure. *Journal on Computing and Cultural Heritage*, *11*:2(2018). doi:10.1145/3092906
2. Cohen, D.J. Digital History: the Raw and the Cooked. *Rethinking History, 8*:2(2004), 337–340.
3. Cohen, D., Rosenzweig, R. *Digital History A Guide to Gathering, Preserving, and Presenting the Past on the Web*. University of Pennsylvania Press.Available. 2005. Online: http://chnm.gmu.edu/digitalhistory/ (accessed on 30 July 2020).
4. Crvčanin, M. *Fine Arts Thesaurus - Methodology of Design*. Belgrade: National Library of Serbia, 1983 (In Serbian)
5. Fitzpatrick, K. The Digital Future of Authorship. Rethinking Originality. *Culture Machine*, 12(2011), 1–26. Available Online:

https://culturemachine.net/wp-content/uploads/2019/01/6-The-Digital-433-889-1-PB.pdf
(accessed on 30 July 2020).
6. Fenlon, K. Thematic Research Collections: Libraries and the Evolution of Alternative Digital Publishing in the Humanities. *Library Trends*, *65*:4(2017), 523–539. doi:10.1353/lib.2017.0016
7. Gavran, S..*Artificial Neural Networks in Data Mining: A Review and Application* (Master Thesis). University of Belgrade, Faculty of Mathematics. 2016. Available Online:
http://elibrary.matf.bg.ac.rs/bitstream/handle/123456789/4237/Gavran_Sava(1).pdf?sequence=1
(accessed on 30 July 2020). (In Serbian)
8. Hasnat, M., Chowdhury, M., & Khan, M. An Open Source Tesseract Based Optical Character Recognizer for Bangla Script. *Proceedings of the International Conference on Document Analysis and Recognition,* 2009, 671–675. doi: 10.1109/ICDAR.2009.62
9. Hodge, G. *Metadata made simpler*. NISO Press: Bethesda, MD. 2001
10. Intro to Thematic Research Guides. (n.d.). CENDARI. Available Online:
http://www.cendari.eu/thematic-research-guides/intro-thematic-research-guides
(accessed on 30 July 2020).
11. Jakovich, K.. On Omeka: the Good, the Bad, the Ugly. 22 February 2018, Available Online:
https://battlehall.lib.utexas.edu/2018/02/22/on-omeka/
(accessed on 30 July 2020).
12. Manžuch, Z., Huvila, I., Aparac-Jelušić, T. Digitization of Cultural Heritage. У L. Kajberg, L. Lørring (Eds.)*European Curriculum Reflections on Library and Information Science Education*, 2005, 37–65. Copenhagen: The Royal School of Library and Information Science.
13. Marinković, B. Guidelines for metadata format for describing cultural heritage and the ways of their storage. In M. Stijović (Ed.) *Open Access to Museum Documentation in Serbia: Experiences, Challenges and Potentials, 2016, 43–52,* Central Institute for Conservation: Belgrade, ISBN 978-86-6179-042-3. (In Serbian)
14. Maroević, I. *Introduction to Museology*. Zagreb: Department of Information Sciences (Faculty of Humanities and Social Sciences, University of Zagreb), 1993 (In Croatian)
15. Metadata. [Def. 1] (n.d.). Merriam-Webster Online. Available Online:
https://www.merriam-webster.com/dictionary/metadata
(accessed on 30 July 2020). (In Serbian).
16. Ministry of Culture and Information of the RS. Guidelines for Digitization of Cultural Heritage in the Republic of Serbia. (September, 2017). Available Online:
http://www.kultura.gov.rs/docs/dokumenti/propisi-iz-oblasti-kulture/smernice-za-digitalizaciju-kulturnog-nasledja-u-republici-srbiji.pdf
(accessed on 30 July 2020). (In Serbian)
17. Ministry of Culture and Information of the RS. *Draft of the Strategy for Cultural Development in the Republic of Serbia for the 2017–2027 Period*. (May, 2017). Available Online:
http://www.kultura.gov.rs/docs/dokumenti/nacrt-strategije-razvoja-kulture-republike-srbije-od-2017--do-2027-/-nacrt-strategije-razvoja-kulture-republike-srbije-od-2017--do-2027-.pdf
(accessed on 10 April 2020). (In Serbian)
18. Öberg, L.M. Metadata use in document management systems, which support business processes. *Archives & Social Studies: A Journal of Interdisciplinary Research*, 2 :1(2008), 217–249.
19. Ognjanović, Z., Butigan-Vučaj, T., Marinković, B.. NCD Recommendation for the National Standard for Describing Digitized Heritage in Serbia. In M.A. Sicilia, L.D., Miltiadis (Eds.) *Metadata and Semantics*, 2009, 45–54,. Boston: Springer.
20. Omeka Classic User Manual. (n.d.). Available Online:
https://omeka.org/classic/docs/ (accessed on 30 July 2020).
21. Omeka.net. (n.d.). Available Online:
https://www.omeka.net/ (accessed on 30 July 2020).
22. Omeka Forum. (n.d.). Available Online: https://forum.omeka.org/ (accessed on 30 July 2020).
23. Palmer, C.L. Thematic Research Collections. In: S. Schreibman et al. (Eds.) *A Companion to Digital Humanities,* 2004, 348–365. Wiley-Blackwell.

24. Pejović, N., Mijajlović, Ž. (2011). Books by Vojislav Mišković in the Virtual Library. In: M.S. Dimitrijević (Ed.) *Proceeding of the Conference – Development of Astronomy Among Serbs VI*, 2011, 451–465. Belgrade: Astronomical Society Ruđer Bošković. (In Serbian)
25. Pijevac, K., Milošević, M., & Boričić, V. *Guide through the funds of the Kingdom of Yugoslavia*. Belgrade: Archives of Yugoslavia, 2000 (In Serbian)
26. Position Paper on ARGs. (12 November 2013). UOB.
27. Public Research Guides [Database]. (n.d.). CENDARI. Available Online: http://www.cendari.eu/thematic-research-guides/available-research-guides (accessed on 30 July 2020).
28. Puckett, J., & Leslie, S. Omeka. *Journal of the Medical Library Association*,104(2016), 374-376. doi:10.3163/1536-5050.104.4.030
29. Research Guides [Database]. (n.d.). *National Archive UK.*Available Online: http://www.nationalarchives.gov.uk/help-with-your-research/research-guides/ (accessed on 30 July 2020).
30. Review of Documents on MihailoPetrović in Archive of Serbia (2018). Available Online: http://www.mi.sanu.ac.rs/~msegan/teza/asrbije_mpetrovic_pregled.pdf (accessed on 24 December 2020).
31. Riley, J. *Understanding metadata. What is metadata, and what is it for*? Baltimore: National Information Standards Organization. 2017, Available Online: https://groups.niso.org/apps/group_public/download.php/17443/understanding-metadata (accessed on 30 July 2020).
32. Site Planning Tips. (n.d.). OMEKA. Available Online: https://omeka.org/codex/Site_Planning_Tips (accessed on 30 July 2020).
33. Scheidl, H. (2018). *Build a Handwritten Text Rcognition System using TensorFlow*. Available Online: https://towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5 (accessed on 30 July 2020).
34. Sotirova, K., Peneva, J., Ivanov, S., Doneva, R., &Dobreva, M. Digitization of Cultural Heritage – Standards, Institutions, Initiatives. In: K. Ivanova, M. Dobreva, P. Stanchev, G. Totkov (Eds.) *Access to Digital Cultural Heritage: Innovative Applications of Automated Metadata Generation*, 2012, 25–67. Available Online: http://eprints.nbu.bg/1479/1/Ivanov-308-v7-a5.pdf (accessed on 30 July 2020).
35. The Dublin Core Metadata Initiative (DCMI). (5 December 2017). In: IFLA. Available Online: http://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/8822 (accessed on 30 July 2020).
36. Tour of Omeka S. (n.d.). Available Online: https://player.vimeo.com/video/241702586 (accessed on 30 July 2020).
37. Unsworth, J. Thematic Research Collections. Paper presented at the Conference *Modern Language Association Annual Conference* on December 28th, 2000 in Washington D.C. Available Online: http://www.people.virginia.edu/~jmu2m/MLA.00/ (accessed on 30 July 2020).
38. VeljkoVujičić Collection (2019). Available Online: https://misanuarc.omeka.net/collections/show/2 (accessed on 24December 2020).
39. Vukadin, A. Named Entities in Subject Indexing: LCSH and NuovoSoggettario. *Vjesnik bibliotekara Hrvatske*, 57:1(2014), 3rd ser., 319–342.
40. Working with the NTE. (n.d.). CENDARI. Available Online: https://docs.cendari.dariah.eu/user/writing_arg.html (accessed on 30 July 2020).

msegan@mi.sanu.ac.rs