

Electronic Editions of Mathematical Works in Serbia

Žarko Mijajlović¹, Zoran Ognjanović², Dragan Blagojević²

¹Faculty of Mathematics, Belgrade, zarkom@eunet.yu

²Mathematical Institute, Belgrade {zorano | blagoje}@mi.sanu.ac.yu

Abstract

The paper describes two ongoing projects carried out by the Mathematical Institute of Serbian Academy of Science and Arts, Belgrade [5] and the Faculty of Mathematics, Belgrade [4]. The projects concern building of electronic resources and presentations of electronic editions of mathematical works in Serbia. The first project is related to retro digitization of old books, articles and the other mathematical works and development of the corresponding virtual library. In the second project a database and www-presentations of some Serbian mathematical journals were created. The resources built in the projects are freely accessible through Internet [2, 4]. These projects are included in a recent initiative on the foundation and development of the National Center for Digitization [6, 9, 11].

1 Virtual Library

1.1 Goals of the Project

Objectives of the project include electronic archiving of old manuscripts, their publishing in electronic phototype form and presentation to general audience. Our proposal is mathematically inclined, more specifically the project is concerned with digitization of old books and other manuscripts satisfying the following criteria:

1. Books and other manuscripts contained in Serbian public and semi-open libraries. By semi-open libraries we mean departmental libraries and private collections.
2. Books and manuscripts selected for digitization should be mainly with topics in mathematical sciences: mathematics, mechanics, astronomy, physics and mathematical geography.
3. Books considered for digitization should be published before certain date in the past, for example the Second World War.
4. Preference will be given to Serbian authors, or to works somehow related to Balkan area.

There are relatively large collections of rare mathematical books in Serbian libraries. For example, the Mathematical Faculty in Belgrade has about 500 mathematical books published in the 19th century or before. On the other hand, just few of them can be found in the joint catalogs of the network of largest Serbian libraries (that include the National Library and all university libraries). Some of these books are rather rare and it is known that only few copies of them survived. We think that it is important to preserve them in some form. Not only as a cultural or scientific heritage of our local community but also as a part of the World scientific heritage. About 100 books, articles, doctoral dissertations are digitized, and now they are the part of our Virtual library, which can be seen at <http://alas.matf.bg.ac.yu/biblioteka/home.jsp>. The project gathers about ten collaborators and volunteers and it is directed by Žarko Mijajlović.

1.2 Technical Issues

Most of the technical issues concerning our virtual library are common to other types of digital archives. Besides cultural and historical, the library has proficient and scientific nature. Obviously these have the influence on accompanying files and metadata to digitized manuscripts (references, keywords, biographies, etc.). There are many old works that should to be digitized. Therefore a selection should be done what to digitized first. Specialists are usually best acquainted with this problem, so it is not strange that professional mathematicians are primarily involved in the work around the library. Many classical libraries are required to devote most of their cataloging resources to published materials, for circulating collections and printed books used daily. There are still library catalogs not yet transferred into electronic forms. The full scope of

preservation needs can be realized only if libraries and archives can devote more resources to cataloging unique and unpublished holdings. However this is not always the case. For example, mathematicians-volunteers working for our virtual library made the electronic catalog of 500 old books in possession of the Faculty of Mathematics in Belgrade (see [10], Addendum 1: *Old Serbian books on logic*, Addendum 2: *Elderly Mathematical books of the Library of the Mathematical Faculty*). In fact, these books did not exist in the (printed) catalog of the Faculty's library at all!

Digital media and format. Recently archivists begun to accept digitization as a mean to preserve manuscripts that are at risk of deterioration. In the past librarians and archivists distrusted digital media as a format for saving important holdings. No medium has proved stable enough to be called permanent. At this moment the only way to preserve them is to keep several copies and occasional making of new copies. It should be mentioned that most preservationists believe that resources spent to identify and develop a permanent medium are better spent building systems that acknowledge impermanence and exploit the potential of available technology. There are mass-storage systems that ensure the persistence of data by validating their integrity as they are copied periodically. But such systems are expensive and complex in design and inherently dependent upon sophisticated technology that must be maintained in continuity. This problem can be overcome only if a strong institution supports such projects. In small countries probably this could be only the government. At this level we see the support of various European institutions and organizations very important.

Storing scanned images in an appropriate file format is an important issue concerning preservation of digitized works. We are always archiving our master copies of scanned images in tiff format because it is the widely used and supported standard. But the main advantage of this format is the high images quality. Using tiff is not a question; only some options were taken in mind when we faced photo images since tiff photo images are pretty large. In that case we are using tiff with LZW compression that is more effective for gray-scale images than color (most of our scans are in gray-scale mode), but it is lossless – which means there is no quality loss due to compression. Also, tiff can be easily converted into other file formats and the choice depends on what we intend to do further. Usually it is Adobe pdf format but recently we also started to use DjVu format. This format has some advantages against pdf format, in particular in the case of digital images of colored documents.

Digital objects and metadata. The digital object in the virtual library usually consists of several components: digitized image of the manuscript, some graphic components (as the cover of the book), and metadata. The border between components is not always clear. For example, where belongs the file containing the extended biography of the author (sometimes it is included in the printed version of the book). In our approach we consider three types of metadata: descriptive, structural and administrative. Descriptive metadata usually follow data contained in librarian printed catalogs, i.e., they obey librarian standards. Particular problem is that old books do not have ISBN and ISSN numbers so classifying them is rather hard. Structural data explain how the components of the digitized object are interconnected. Administrative data describe exactly how an item is preserved: resolution, rate of compression, type of file containing the digitized image, revision number, whether OCR is performed, specifics of hardware used, rights information, editor's name (person who made decision for making digital copy), various dates, etc. It is clear that the success of digital preservation efforts will rest to a significant degree on the scope and reliability of the metadata records. For example, metadata make possible the asset-management systems that back up and periodically duplicate digital records. As cataloging information they enable one to locate what they are looking for in the library. Metadata help to make various Internet presentations. It is obvious that full repository system require hundred of metadata elements for each digitized item.

Building such database systems and populating them is very labor-intensive and expensive.

Some trade off should be found. For resolving these issues, cooperation between institutions working in the field of digitization is very important, in particular exchange and agreement of metadata formats. We are still in development of metadata formats.

Standards. The digitization standards are by no means connected to metadata. Among all standards probably the most important are resolution of scanned documents, file formats for images and file format in which the digitized manuscript is kept. There are still debates among archivists what should be the resolution of scanned sources. It varies between 200 dpi and 600 dpi.

Most of our scans are 600 dpi, some of them are 400 dpi, and the early ones are 300 dpi.

Digitized works on display in Virtual library are compressed files in 300 or 400 dpi; the uncompressed scans of higher resolution are kept for archive purpose and possible future processing. Particular scans are kept in tiff or pcx bitmapped file format. As already stated, digitized manuscripts are presented in pdf or DjVu file format.

Implementation. We have chosen SQL server 2000 for a database. JAVA programming language is used in developing a web application for administering and searching the database, especially JAVA advanced features like java beans and struts, which enable a high performance web application. Also, Tomcat as an application server and Apache as a web server are used. The application is developed by Nada Đorđević and Tijana Zečević. At this moment the Virtual library is still in development. The presentation is static and the search engine does not exist. We recently started very seriously to consider Fedora system as a programming platform for administrating and accessing the database of digitized manuscripts. Fedora is a system specifically designed for handling digitized values, and among other benefits, it is open source software.

1.3 The Virtual Library Itself

We have digitized, by now, more than 100 books and manuscripts in mentioned fields:

1. One of the finest digital collections is the electronic edition of collected works of the prominent Serbian mathematician, Bogdan Gavrilović (1863–1947). This collection includes about 50 items: books and articles (about 2000 printed pages).
2. Mathematical works of Ruđer Bošković (1711–1787).
3. 12 doctoral dissertations of Serbian mathematicians, including all written before the First World War (8).
4. Two books of the famous Serbian scientist (mechanics, mathematician and astronomer) Milutin Milanković, including his celebrated book *Kanon Der Erdbestrahlung* (The Canon of Earth Irradiation).
5. All books of the 19th Century mathematician and astronomer Milan Andonović (on the Probability theory and Astronomy).
6. Books of other Serbian mathematicians from the 19th Century and the beginning of 20th Century: Kosta Stojanović, Dimitrije Nešić and Mijalko Ćirić. Stojanović's book (in Serbian) is on Bošković's atomistic theory of matter.

Editor of items 1 and 3–6 is Žarko Mijajlović. Editor of item 2 is Zoran Ognjanović. Electronic editions of these retro digitized works were published since 1995 on several compact disks.

Further plans for Virtual library include digitization of a Kepler's book on astronomy, published in the 17th century, then unique Nikola Tesla's collection of newspaper clips, from the end of 19th and the beginning of 20th century (the collection is in the possession of Museum "Nikola Tesla" in Belgrade), digitization of 500 hundred old books in possession of the Faculty of mathematics, build of digital archive base of doctoral dissertations in mathematics, mechanics and astronomy written at the Belgrade university (about 400 items, the first one dated 1912).

The editorial board consists of Ž. Mijajlović (editor-in-chief), M. Rašković, R. Dacić, Z. Ognjanović, T. Butigan and S. Dragović. It can be accessed at the address <http://alas.matf.bg.ac.yu/biblioteka/home.jsp>

2 Internet Presentation of Serbian Mathematical Journals

An Internet presentation of some of the mathematical journals printed in Serbia is freely accessible in the full-text format [2]. This project is cooperation between a non-governmental organization from Belgrade, called "Communication" [1], and the Mathematical institute. "Communication" started a project of organizing and creating a database and www presentations of leading theoretical journals in Serbia and Montenegro, in 2000. The main objectives of the project were to present to interested Internet users scientific-theoretical output of the country, and to improve communication and exchange of ideas in the region—without linguistic obstacles, and customs barriers for importing journals. Improving cooperation in the region characterized by a high possibility of linguistic communication was (and is) especially important after the destruction brought during the 1990s by wars, ethnic and confessional conflicts. The leader of the whole project is Draško Grbić.

Collaborators of the Mathematical Institute (Zoran Ognjanović, the leader of this part of the project, Dragan Blagojević, Zoran Kadelburg and Tatjana Timotijević) organized the activities related to mathematical journals. Currently, there are 6 journals with 116 volumes and more than 13000 pages and 1500 articles in the database:

1. Publications de l'Institut Mathématique (45 volumes in English, from 1980)
2. Matematički vesnik (24 volumes in English, from 1993)
3. Kragujevac Journal of Mathematics (5 volumes in English, from 2000)
4. Nastava matematike (25 volumes in Serbian, from 1992)
5. The Teaching of Mathematics (11 volumes in English, from 1998)
6. The Review of the National Center for Digitization (6 volumes in Serbian and English, from 2002).

The fact that two of these journals (Publications de l'Institut Mathématique and Matematički vesnik) were also in the journal section of the Electronic Library of Mathematics [3] supported by the European Mathematical Society (EMS) helped a lot in defining formats for the corresponding (meta)data. Also, at least for these two journals, most of the material was already prepared (bibliographical data, TeX-source files and/or PostScript files of articles) following the requirements of EMS.

The database has been developed using Zope [8], an open source object-oriented environment for building web applications, under Linux operating system. Zope features a transactional object database, which can store not only content and custom data, but also dynamic HTML templates, scripts, a search engine, and relational database connections and code. The presentations of the journals are dynamically generated from the database.

The structure of the representation of a journal is of the following form:

- journal home page,
- separate page for every volume, and
- article representation.

A journal's home page contains a list of available volumes starting from the most recent issue. The metadata related to a journal are:

- the name of the journal,
- the corresponding ISSN, and
- the name of the publisher.

A volume's page contains the special subtitles of that issue (if they exist) and a table of contents, which may have internal named subdivisions. For example, a special subtitle of an issue can explain that the issue contains papers presented at a conference, while the table of contents can have subdivisions for different sections of that conference. A table of contents contains the usual bibliographic data for articles (names of authors, titles) together with links to abstracts and full-texts. The metadata related to an issue are:

- the number,
- the year of publishing,
- the name of the editor, and
- the number of pages of that issue.

To every issue two files containing data about the editorial board and the instruction for authors are attached.

The description of an article contains:

- title,
- names of authors (with possible links to web-presentations of authors),
- the numbers of the first and last pages,
- keywords,
- the corresponding Mathematics Subject Classification (MSC), and
- the links to the abstract and full-text.

Abstracts and full-texts can be downloaded and printed or directly read from the screen. The suitable metadata (for example titles, keywords) can be given both in English and in the original language of the journal.

In the database, the files containing abstracts are in HTML, while the full-text files can be in HTML and/or PDF. For mathematical journals we have chosen that full-text files are stored in PDF only, since we have no satisfactory way to represent mathematical text in HTML.

From the journal home pages and the pages of volumes a user can perform search. Papers included in the database could be searched (both in English and Serbian) by:

- authors' names,
- titles,
- titles of subdivisions of tables of contents,
- key words and
- words contained in abstracts.

It is worth mentioning that Publications de l'Institut Mathématique (the oldest purely mathematical journal in Serbia—founded in 1932) is retro digitized – all the volumes between 1980 and 1990. From 1990 it has been regularly typeset electronically. At the time of starting retro digitization we decided to re-typeset all the texts in TeX, since no standards were developed. After establishing cooperation with EMS we prepared all the necessary metadata for all (retro)digitized volumes to include them in the e-library of EMS. So, in this moment we cover the largest time span among the journals in the EMS e-library. Retro digitization in TeX was organized and metadata were prepared by Dragan Blagojević.

3 Conclusion

The goal of the mentioned projects is to advance and offer easier on-line access both to old and recent mathematical works in Serbia. The projects can be seen as activities included in the development of the regional mathematical information center [11] in the framework of the DAAD-Sonderprogramm Akademischer Neuaufbau Südosteuropa in the framework of the Stability Pact for South Eastern Europe project “Multimedia Technology for Mathematics and Computer Science Education” [12]. Beside the mentioned resources, the center also offers access to the local mirrors of the European Mathematical Information Service-presentation [13] and Zentralblatt MATH [14] and MATHDI [15] databases.

The realization of the described projects is a continuous task with some open problems. For example, one of them is related to defining an efficient and comprehensive standard format for enlarging HTML to allow presentation of mathematical texts, full-text searching etc. Also, we would like to mention that one of the latest activities aims to make the database accessible in European alphabets and languages. A program package, which enables translation of the interface and records, was developed. It will be possible to enter in the database the original or translated texts, abstracts, key words, titles, etc. in local languages and not only in Serbian and English. We hope that it makes the database even more attractive for editorial boards of the other regional mathematical journals and that more journals will be included in the database. Also, establishing connections with other similar projects in the world, mutual linking and referencing is a promising direction for further development of our database and presentation which will allow mathematicians to have benefit from the fast development of IT.

Finally, we would like to mention the importance of collaborative approaches. Until recently, there was virtually no coordination between various Serbian institutions active in the field of digitization. While this might be not uninspected given their different missions, collaboration could be mutually beneficial for many reasons. Still, no central databases – digital archives exist. Thus, the National Center for Digitization, (coordinated by Z. Ognjanović), and the ongoing project "Digitization of National Heritage" supported by the Serbian Ministry of Science and Technology (leded by Ž. Mijajlović) are important steps toward building such archives and introducing general standards Serbia in this area. Such collaborative projects would not be easy to undertake. For example, copyright laws, particularly concerning author rights and digital piracy, might prohibit legitimate and necessary preservation functions. All digital preservation is expensive. While scanning itself is relatively cheap, equipment for mass scanning and the time required for post processing (cleaning, OCR, binding, handling digital archives), averages three to ten times the length of source recordings. We also face an extraordinary situation. Today printed books and journals once will become old manuscripts. Most, if not all publications are prepared by use of computers, and so their digital copies exist at some stage of their production. However, most of them are not available for general archiving purpose; even these ideal forms of digital records are destroyed in some cases! Obstacles lie in copyright laws, commercial reasons and still unrecognized importance of digital preservation. Two presented projects in this article should help to resolve mentioned issues on digital archives and presentations of electronic editions, old and new, of mathematical works in Serbia.

References

- [1] Communication non-governmental organization, Belgrade, <http://www.komunikacija.org.yu>
- [2] Electronic editions of Serbian mathematical journals, http://www.mi.sanu.ac.yu/biblioteka/elect_pub.htm
- [3] Electronic Library of Mathematics, <http://www.emis.mi.sanu.ac.yu/EMIS/ELibM.html>
- [4] Faculty of Mathematics, Belgrade, <http://www.matf.bg.ac.yu>
- [5] Mathematical Institute of Serbian Academy of sciences and Arts, <http://www.mi.sanu.ac.yu>
- [6] National Center for Digitization, Belgrade, <http://www.ncd.matf.bg.ac.yu>

- [7] Virtual library of old mathematical works, <http://alas.matf.bg.ac.yu/biblioteka/home.jsp>
- [8] Zope, <http://www.zope.org>
- [9] Žarko Mijajlović, Zoran Ognjanović, A survey of certain digitization projects in Serbia, Proceedings of the Symposium Digital Preservation of Cultural Heritage, 16–17 September 2003, Borovetz, Bulgaria, *Review of the National Center for Digitization* 4, 52–61, 2004.
- [10] Ž. Mijajlović, Z. Ognjanović, N. Đorđević, T. Zečević, *Virtual library—data base of textual data*, Review of the national Center for digitization 5 (2004), 42–48, <http://www.nacd.matf.bg.ac.yu>
- [11] Zoran Ognjanović, Žarko Mijajlović, Digitization projects carried out by the mathematical institute Belgrade, Proceedings of the International conference "Digitization of Cultural and Scientific Heritage", 27 August – 3 September 2004, Bansko, Bulgaria, International Journal "Information Theories & Applications" 11(3), 275–278, 2004.
- [12] Regional mathematical information center, http://www.mi.sanu.ac.yu/main_pages/regional.htm
- [13] Multimedia Technology for Mathematics and Computer Science Education, <http://www.matf.bg.ac.yu/~daad/>
- [14] European Mathematical Information Service, <http://www.emis.mi.sanu.ac.yu/EMIS/>
- [15] Zentralblatt MATH database, <http://www.emis.mi.sanu.ac.yu/ZMATH/>
- [16] MATHDI database, <http://www.emis.de/MATH/DI.html>
- [17] Dale Flecker, *Preserving Digital Periodicals*, in *Building a National Strategy for Digital Preservation Issues in Digital Media Archiving*, copublished by CLIR, Washington, www.clir.org, Library of Congress, Washington, www.loc.gov, April 2002, pp. 10–22.
- [18] Frank Romano, *E-Books and Challenge of Preservation*, *ibid.* pp. 23–37.