Žarko Mijajlović; Zoran Ognjanović
Digitization of Mathematical Editions in Serbia
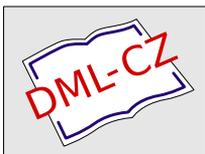
Persistent URL: http://dml.cz/dmlcz/702541

Digitization of Mathematical Editions in Serbia

Žarko Mijajlović[1] and Zoran Ognjanović[2]

[1] Faculty of Mathematics, Studentski Trg 16, Belgrade, Serbia
E-mail: zarkom@eunet.yu
[2] Mathematical Institute, Kneza Mihaila 35, Belgrade, Serbia
E-mail: zorano@mi.sanu.ac.yu

Abstract. We describe an ongoing project carried out by the Mathematical Institute of Serbian Academy of Sciences and Arts, and the Faculty of Mathematics, Belgrade. The project concerns building of electronic resources and presentations of electronic editions of mathematical works in Serbia, including retro-digitization of old books, articles and the other mathematical works, and development of the corresponding virtual library. The resources built in the project are freely accessible through Internet.

Key words: digital libraries, digitization, retro-digitization, mathematical journals, mathematical books, mathematical works in Serbia

1 Introduction

Due to the fast development of computer technologies and decrease of price of computer equipment and software, some digitization projects started in Serbia already in the middle of the nineties: PANDORA, an expert system for archaeology [4], Computer archiving and multimedia presentation of cultural values and national heritage [10], Old maps, engravings and photographs — Collection of the City museum of Belgrade [6], Presentation of Historical archive of town Kotor, Electronic catalog of cultural monuments [9], etc. A survey of these projects can be found in [8]. The first author of this paper managed all mentioned projects except the last one which was lead by the second author.

The principal participating institutions were Mathematical Institute of the Serbian Academy of Science and Arts, and Faculty of Mathematics, Belgrade, but the other institutions were involved too (Archaeological Institute, Institute of musicology, Institute for monument protection of Serbia, National library, National museum, etc.). The projects were partially financed by the Ministry of Science of Serbia, but the most of the job was voluntarily done by a group of enthusiasts.

The project *Computer archiving and multimedia presentation of cultural values and national heritage* was the most important and comprehensive project in our country in the area of digitization until now. The project consisted of two parts:

– infrastructure, standards and methodology of design and the architecture of this type of data, and

   – design and building of archive databases, together with the corresponding
    program implementation.

In the first part of the project the investigation was based on Standard
Generalized Markup Language, but we also experimented with the system
TeX, the well-known typesetting system. The second part of the project lead to
several derivatives of the main project, i.e. to sub-projects with very definitive
aims.

In this paper we describe one of these sub-projects. It concerns digitization
of mathematics-related books, theses, manuscripts, and mathematical journals
somehow related to Serbia or our region of South Eastern Europe. The
project's goal is to form digital archives, data bases and presentations of
digitized scientific editions in mathematical sciences (mathematics, mechanics,
astronomy, computer science, physics, mathematical geography, etc.) in order
to advance and offer easier on-line access both to old and recent mathematical
works. The corresponding presentations are freely available on Internet, see
`http://www.ncd.matf.bg.ac.yu`.

The rest of the paper is organized as follows. Our efforts in retro-digitization
of old mathematical editions are described in Section 2. Section 3 discusses
activities related to building an Internet data base and presentation of
mathematical journals printed in Serbia. In Section 4 some technological details
of the project are presented. We give concluding remarks and directions for
further investigations in Section 5.

2 Retro-Digitization of Mathematics — Related Books, Theses and Manuscripts

There are relatively large collections of rare editions in Serbian public and
semi-open libraries (by a semi-open library we mean any library of universities,
faculties, research institutions or their departments, and private collections).
For example, for the purpose of our project, we made an electronic catalog
according to librarian standards of almost 500 books (published before and
during the 19th century) in possession of the Faculty of Mathematics in Belgrade.
On the other hand, just a few of them can be found in the joint catalogue of the
network of the largest Serbian libraries (National Library, all university libraries,
etc.) which was one of the reasons to start with the project of retro-digitization
of mathematics-related works.

The first step was the work on transforming to digital form collected
papers of the prominent Serbian mathematician Bogdan Gavrilović (1864–1947).
B. Gavrilović obtained doctor's degree in mathematics in 1887 at the
Philosophical Faculty of the University in Budapest. He was appointed professor
at the University of Belgrade, three times elected president of the Serbian
Academy of Sciences (1931–1937), a member of Circolo matematico di Palermo,
and doctor honoris causa of the University of Athens. He published two
voluminous university textbooks which had the character of monographs:

Analytical Geometry (1896) on 900 pages and *Theory of Determinants* (1899). Those books, together with his thesis (in Hungarian), other papers and archive were digitized (1996–2001), put first in TEX frame, then in PDF, and published as a compact disk [7].

The further steps were related to development of a digital library as a comprehensive and semantically interconnected collection of retro-digitized material satisfying some of the following criteria:

– Books and manuscripts selected for digitization should be related with mathematical sciences: mathematics, mechanics, astronomy, physics, mathematical geography etc.
– Books considered for digitization had to be published before certain date in the past. We have chosen for this date the beginning of the World War II (1941).
– Preference will be given to Serbian authors, or to written works related in some way to the area of Balkan.

So far, more than 100 books, thesis and manuscripts have been digitized, including first two books on mathematics written in Serbian language:

– Vasilije Damjanović, *Aritmetika* (*Arithmetics*), Venice, 1767, and
– Jovan Došenović, *Čislenica* (*Arithmetics*), Budim, 1809.

These two textbooks are very elementary, and their authors did not pretend to be original. They are, however, important for the cultural history of the Serbian people. For example Čislenica was the only printed textbook used in the Big School in Belgrade, in the first state created by the Serbian Revolution at the beginning of the 19th century. Thus, it could be considered the first Serbian university textbook, because the Big School may be and is usually taken as the forerunner of the University of Belgrade.

Other digitized materials are works of Ruđer Bošković (1711–1787), two books of the famous Serbian scientist (mechanic, mathematician and astronomer) Milutin Milanković, including his celebrated work Kanon Der Erdbestrahlung (1935, 1941), all doctoral dissertations of Serbian mathematicians written before the First World War (8), including the oldest one:

– Dimitrije Danić, *Conforme Abbildung des Eliptishen Paraboloids auf die Ebene*, Inauguration Dissertation der Philosoph. Fakultaet zu Jena, printed in Belgrade, 1885.

Some of these theses were translated and whenever it was possible, the related documentation (biographies, archive materials, etc.) was also digitized and included in the presentation.

The first version of Internet presentation of retro-digitized books can be found at `http://elib.matf.bg.ac.yu:8080/virlib`, while a new version (still under development) which is based on a database is at `http://elib.matf.bg.ac.yu:8080/elib`.

At the same addresses a part of the presentation/data base contains almost 200 more or less recent (published and unpublished) doctoral and master theses defended at the Faculty of Mathematics in Belgrade. The council of the faculty decided that all submitted theses must be given also in the digital form, so we expect that this part of the data base will grow significantly in the next period.

3 Mathematical Journals

Digitization of mathematical journals printed in Serbia started in 1995 with the oldest (founded in 1932) and most important one — *Publications de l'Institut Mathématique*. Until the Second World War it was published by the Belgrade University, and afterwards by the Mathematical Institute. More then 2,000 articles appeared in almost 100 volumes. The scope of the journal in the beginning was broader, not only in mathematics, but articles referring to mechanics and astronomy were accepted in it as well. Almost every Serbian mathematician published there at least one paper, as well as many world leading mathematicians (H. Lebesgue, P. Montel, W. Sierpiński, P. Erdős, S. Shelah, etc.).

The first archiving technique was retyping articles using TEX system, starting with the most recent issues. About 25 volumes (published between 1982 and 1995) were electronically archived in that way. The archive was very compact, having less than 100 MB. It included source (tex) and output (device independent — DVI, and PDF) files. The archive has been permanently enlarged by adding all new volumes until these days. This archive is now included in the Electronic Library of Mathematics (ELibM) offered through EMIS (European Mathematical Information Service) [3].

However, since archiving using TEX system was expensive (about 2 Euro per page), it was decided that the remaining old volumes would be scanned. So, in the last two years 47 volumes (1932–1982) of *Publications de l'Institut Mathématique* were completely retro-digitized.

In the same period another journal, *Publications of the Faculty of Electrical Engineering*, Series Mathematics and Physics, was also retro-digitized (about 1,000 papers). It was founded in 1956. In the beginning, each contribution appeared separately bound and numbered consecutively, several times a year. Since 1959, the issues have been appearing collected in one or more volumes per year. In the first years, the journal had contributions from different fields apart from mathematics: physics, mechanics, and electrical engineering, but in the course of time, the journal focused almost exclusively on mathematics, especially convexity, functional equations and differential equations.

As Internet became more and more popular, in the year 2002 we decided to create an Internet database and the corresponding presentation of freely accessible full-text mathematical journals. The following journals have been involved so far:

– *Publications de l'Institut Mathematique* (100 volumes, since 1932),
– *Kragujevac Journal of Mathematics*, published by the Faculty of Sciences Kragujevac (8 volumes issued since the year 2000),

- *Matematički Vesnik*, published by Mathematical Society of Serbia (28 volumes, since 1993),
- *Nastava Matematike* (*Teaching Mathematics*), published (in Serbian) by Mathematical Society of Serbia (30 volumes, since 1992),
- *Review of the National Center for Digitization*, published by Faculty of Mathematics Belgrade and the National center for digitization (12 volumes, since 2002), and
- *The Teaching of Mathematics*, published by Mathematical Society of Serbia (16 volumes, since 1998).

The corresponding presentations of journals are dynamically generated from the database and can be searched (both in English and Serbian) by: authors' names, titles, titles of special sections within the journals, key words and words contained in abstracts, classification numbers, and downloaded and printed. The last version of the database and presentation of mathematical journals is given at the address `http://elib.mi.sanu.ac.yu/pages/browse_journals.php?`, while the retro-digitized journals can be found on `http://publication.mi.sanu.ac.yu` and `http://pefmath2.etf.bg.ac.yu/pages/main.php`.

Since 1999 Mathematical Institute publishes the journal Visual Mathematics [13] with the goal to show the beauty of mathematics in a broad artistic-scientific context. It was one of the first journals appearing in digital form only. All 33 volumes of the journal are available at `http://www.mi.sanu.ac.yu/vismath/`.

4 Technological Aspects of the Project

Digitization is heavily based on information technologies. But information technologies are one of the fastest changing parts of the contemporary world. For example, since the appearance of disk and tape based digital storage in the 1960s, we have witnessed evolution and proliferation of more than 200 different storage formats. Therefore we face a continuing challenge in maintaining files on currently supported storage hardware and media in currently supported file formats, but operating systems as well [1]. We had in mind all these issues when we placed in focus the following aspects of the project:

- rules for choosing material to be digitized,
- the mathematics of information storage (database architecture, data manipulation, data forms, etc.),
- information retrieval,
- metadata and rules for handling them,
- distribution of information to the wide audience using Internet,
- ability to search electronic files efficiently and retrieve information quickly,
- ability to reuse information in other documents and other formats,
- fast retrieving of digitized material through electronic transactions and data downloading,
- proper choice of data format and resolution.

There are many proposed standards for the database architecture, distributions of information and information retrieval. Common practice in database projecting is usage of so called Data Base Management Systems (DBMS) and especially Relational DBMS. These systems include variety of tools which allow users to create, update, and extract information from their databases.

We also note that metadata for digitized documents may differ significantly from data for printed books obeying some of the existent standards (e.g. Unimarc [5]). For example, metadata for a digitized document should include the following items, ordinarily not found in the mentioned standards: name of the editor (person and/or committee) who made decision for digitization, then resolution, file type, revision number and date, etc. Some of the new initiatives, like TEI [12] and Dublin core metadata initiative [11] can help us to obtain very expressive description methods.

In our approach to retro-digitization, documents are basically processed in four steps:

– scanning document pages,
– processing scanned pages,
– storing processed pages in appropriate file format, and
– collecting processed pages into an electronic document.

The future use of the digitized document determines how these steps will be performed.

In most cases the scanning is performed with 300 DPI resolution in gray-scale mode for the best quality/size trade off. We found that acceptable resolution range is 200–600 DPI, particularly if one wants to perform OCR on the scanned document. Even if the mentioned facts might sound trivial they may be useful, as examples show. There is a warning example when a recognized institution performed digitization in so called screen resolution (96 DPI). After some time they discovered that so scanned printed works are of small value. So when they turned to higher resolution, they decided to repeat the scanning of all documents.

Due to imperfections of the scanner, dust and document state, some improvements should be made to scanned images: impurities cleaning, contrast improving, right positioning, setting the same size for all pages, etc. We use Adobe Photoshop and various embedded filters in this software in order to increase the quality of the scanned pages.

We always archive our master copies of scanned images in TIFF format because it is the widely used and supported standard. Some of the main advantages of this format are the high images quality and possibility to easily convert TIFF files into other file formats. We use lossless — which means there is no quality loss due to a compression — LZW compression that is also more effective for gray-scale images than for color ones.

Scanned pages are stored into an electronic document by appropriate software. We use PDF and/or DjVu [2] file formats. Very often and whenever is appropriate, we are performing OCR, however without recognizing mathematical formulas. For this purpose we the *Adobe Capture* software. In this

case the final digital document consists of two layers. The first layer consists of the exact image of the original document, while the second layer consists of the recognized texts. The second layer enables one to search through the document for certain contents. In some cases further ramifications and structuring are done on documents: introducing hyperlinks on the content and the index of the original, embedding annotations, comments etc.

Our data bases rely on three types of metadata: descriptive, structural and administrative. Descriptive metadata follow data contained in librarian printed catalogs, i.e., they obey librarian standards. There are some special data important for scientific papers: keywords, scientific classification (Mathematics Subject Classification, UDC), numbers of reviewer reports in Math Reviews and Zbl, DOI numbers, and statistics of accessing and downloading papers. One problem was that old issues do not have standard descriptive tags (such as ISSN numbers) so classifying them needed particular solutions. 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, file type containing the digitized image, etc.

The software that realize our data bases supports all usual functions, browsing, searching under various criteria, examining and downloading papers. We have chosen MySQL server 2000 for a database since it supports UTF-8 encoding. The multilingual support is embedded into the model of data. Java programming language is used in developing the web application for administering and searching the database, especially Java advanced features like JavaBeans and Struts, which enable high performance web applications. Other used technologies are PHP and Apache as a web server.

5 Conclusion

We describe in this paper a digitization project in progress carried out by the Mathematical Institute of Serbian Academy of Sciences and Arts and the Faculty of Mathematics, Belgrade. The project goal is the building of digital resources and presentations of mathematics-related books, theses, manuscripts, journal etc. which are somehow related to Serbia and our region.

This project gathers about ten collaborators and volunteers: Žarko Mijajlović, Zoran Ognjanović, Dragan Blagojević, Nada Djorđević, Aleksandar Pejović, Tijana Zečević, Nikola Šćepančević, and some students from the Faculty of Mathematics Belgrade. The corresponding editorial board contains the following members: Žarko Mijajlović, Faculty of Mathematics Belgrade, Editor in chief, Simon Dragović, Belgrade Children's Hospital, Tamara Butigan, National Library of Serbia, and Miodrag Rašković, Rade Dacić and Zoran Ognjanović from the Mathematical Institute of the Serbian Academy of Sciences and Arts.

As it can be seen from Table 1, the digitized mathematical resources are included into some data bases and presented at different addresses. One of the

Table 1. Digitized mathematical resources freely available on Internet

Virtual library of old mathematical works	`http://elib.matf.bg.ac.yu:8080/virlib`, `http://elib.matf.bg.ac.yu:8080/elib`
eLibrary of mathematical journals	`http://elib.mi.sanu.ac.yu/pages/browse_journals.php?`
Retro-digitized journal *Publications de l'Institut Mathématique*	`http://publication.mi.sanu.ac.yu`
Retro-digitized journal *Publications of the Faculty of Electrical Engineering*	`http://pefmath2.etf.bg.ac.yu/pages/main.php`
Visual Mathematics	`http://www.mi.sanu.ac.yu/vismath/`

first goals of the future work will be to unify that material so that it can be searched and presented from one location.

In the next few months a new journal will be added to the journal data base: *Bulletin, Classe des Sciences Mathématiques et Naturelles, Sciences mathématiques Académie Serbe des Sciences et des Arts*, published by the Serbian Academy of Sciences and Arts. The next goal is to continue with retro-digitization with the other journals from the data base so that all their volumes will be presented in the digital form. There are also a few other mathematics-related journals printed in Serbia (*Theoretical and Applied Mechanics, Yugoslav Journal of Operations Research, Novi Sad Journal of Mathematics*). These journals have their particular Internet-presentations, and we plan to start negotiation with their publishers to involve them into our data base. Also, there is an idea to enlarge the data base with some other mathematical journals from the region, for example, from Bulgaria and Romania.

For the purpose of this project our group has very limited fund and modest equipment on its disposal. So, the aim of the project might seem as a hardly reachable aspire. However, we are aware that for some of the considered editions only a few copies are left. We think that it is very important to preserve them in some form, not only as a cultural or scientific heritage important for our local community, but also as a part of the World heritage. Therefore, we decided to preserve and present them to the wide audience in electronic-digitized form.

References

1. Council on Library and Information Resources, Washington D.C. and Library of Congress. *Building a national strategy for digital preservation: issues in digital media archiving*, a collection of papers, 2002. (Available at `http://www.digitalpreservation.gov/`).
2. DjVu. `www.djvuzone.org`.
3. European Mathematical Society. *The European Mathematical Information Service*. (Available at `http://www.emis.de/`).
4. Korać, M., Ognjanović, Z., Simić, V., Dugandžić, F. PANDORA – Expert system for dating artifacts. *Proceedings of theM ECAI06 Workshop on Intelligent Technologies for CulturalM Heritage Exploitation*, Riva del Garda, Italy, August 28 – September 1, 2006, pp. 40–44, 2006.

5. MARC standards. (Available at `http://www.loc.gov/marc/`).

6. Mijajlović, Ž., editor. Compact disk *Old maps, engravings and photographs from the collection of the City Museum of Belgrade*, 1997. (Available at `http://www.mi.sanu.ac.yu/muzej.beograd/`).

7. Mijajlović, Ž., editor. Compact disk *Collected works of Bogdan Gavrilović*. Mathematical Institute, Belgrade, 2001.

8. Mijajlović, Ž., Ognjanović, Z. *A survey of certain digitization projects in Serbia*. Review of the National Center for Digitization, 4 (Proceedings of the Symposium Digital Preservation of Cultural Heritage, 16–17 September 2003, Borovets, Bulgaria), 52–61, 2004. (Available at `http://elib.mi.sanu.ac.yu/`).

9. Ognjanović, Z., editor. *Electronic catalouge of cultural monuments in Serbia*. Funded by UNESCO, Participation Programme 2004–2005 (grant no. 27243305 YUG), 2005. (Available at `http://spomenicikulture.mi.sanu.ac.yu/`).

10. Project *Technology of Archiving and Multimedia Presentation of Data Texts and Cultural Objects on Electronic-Optic Media*. Funded by the Ministry of Sciences, Technology and Development 1995–1996. (Available at `http://www.mi.sanu.ac.yu/innovative_projects/iproject1.htm`).

11. *The Dublin Core Metadata Initiative*. (Available at `http://dublincore.org/`).

12. *Text Encoding Initiative*. (Available at `http://www.tei-c.org/`).

13. *Visual Mathematics*, an electronic journal. Mathematical institute, Belgrade. (Available at `http://www.mi.sanu.ac.yu/vismath/`).