

Žarko Mijajlović
(Matematički fakultet, Beograd)

Zoran Ognjanović
(Matematički institut SANU, Beograd)

A SURVEY OF CERTAIN DIGITIZATION PROJECTS IN SERBIA

Abstract. A survey of some digitization projects in Serbia in the last decade is presented. These projects were carried out mainly by the Mathematical Institute of the Serbian Academy of Science and Arts (SANU) and the Faculty of Mathematics in Belgrade. The projects concern standards for cultural and scientific heritage digitization, building of electronic archive bases, development of digitized heritage presentations, as well as research in the field of expert systems applied to archeology. The work on the foundation of the National Center for Digitization is described. The Center consists of the leading Serbian institutions involved in digitization. Some initiatives for collaboration with similar projects from abroad are presented.

Key words. Digitization, cultural heritage, scientific heritage

Archives, libraries, museums and other repositories in Serbia for organised storing of records of the national interest did not have any significant project in the field of digitization until the last decade. Due to the fast development of computer technologies and decrease of price of computer equipment and software, many projects were started recently in this field. The aim of this article is to give a survey of some of our most important undertakings in this field.

Let us mention first the following facts common to almost all projects:

- First projects in the field of digitization were started in early nineties of the last century.
- Many of the projects were established with the principal aim of presentation.
- The principal institutions participating in these projects were the Mathematical Institute SANU, the Faculty of Mathematics, the Archeological Institute SANU and the City Museum of Belgrade. Other institutions participated occasionally, e.g. the Musicology Institute SANU and the Historical Archive in Kotor. Except the last institution, all others are situated in Belgrade.
- The greatest part of funds for financing these projects was coming from government sources, mainly from the Ministry of science and technology of Serbia.

1. Project: Pandora

The system Pandora [15] was designed as an expert system for archeology. This project was the first one in the digitization area in Yugoslavia. The main parts of the system are:

- An electronic data base of types of Roman oil lamps.
- An inference engine supplemented by a set of rules for expertise.

The prototype of Pandora was built at the beginning of nineties of the last century with joint efforts of the Mathematical institute and Archeological institute SANU. The authors of the system are Miomir Korać and Zoran Ognjanović. Pandora was designed as a consulting system for dating excavated objects, but also for training of young archeologists. For this reason, the general rules and principles of archeological classification and dating were implemented in the system. The archeological knowledge contained in the system is dynamically organized, as it is changed and enlarged as the database is altered and increased. Database includes facts about our most important archeological sites: Viminacium [19], Sirmium, Nais, Mediana [11] and Diana.

The mechanism for reasoning, i.e. the inference engine, used in the expert system PANDORA is based on implicative rules while the reasoning and the knowledge space search are based on backtracking. The knowledge base in the system Pandora contains about 200 rules. The programming language PROLOG was used as the principal tool in the developing of the corresponding expert system shell.

2. Project: Computer archiving and multimedia presentation of cultural values and national heritage

This project was basic and the most comprehensive project in our country in this area until now. The project consisted of two parts:

- Infrastructure, standards and methodology of design and the architecture of this type of data.
- Design and building of electronic archive bases, together with mathematical and program implementation.

The project lasted for two years (1995–96), and four institutions participated: the Mathematical institute, the Mathematical Faculty, the Archeological institute, and the City Museum of Belgrade. There were about 15 collaborators. The head of the project was Žarko Mijajlović.

Concerning the first part of the project, our investigation was based on SGML (Standard Generalized Markup Language) which is now *de facto* the standard in this area (ISO 8879:1986) [16], and its derivative HTML (Hyper Text Markup Language) [2], then the main tool for making Internet presentations.

We also experimented with no to widespread machineries. One tool of this kind was the system TeX [8, 18], the well known typesetting system (mainly for mathematical texts). The advantages of this system are: it is freeware and allows very sophisticated actions on textual data (due to its characteristics of an universal

programming language), for example: automatic transliterations from Latin to Cyrillic and vice versa, word division (hyphenation), use of diacritical signs, independence of output devices and high resolution outputs, compact size of files, etc. The development of TeX was stopped in 1983, which caused this system to become a real standard. However, TeX is complicated for use for a casual user, handling images is difficult, etc. We found that some special uses (particularly in archiving of scientific texts), accompanied with OCR could be of the great interest. We note that some extensions which use TeX as their formatting engine were developed. One of them is LaTeX [9, 10] which is used by some major publishers as a standard (and not only for mathematical texts).

We also considered standards in other areas: image, cinema and audio formats, for various purposes: archiving, OCR, multimedia presentations etc. Of course, many of then established standards now are surpassed, due to the fast development of computer technology and hardware. For example, video records obeying today standards could not be played on computers from 1995's. This fact only proves that project of this kind should be permanent, especially in this period of rapid change of computer technologies. An attempt in that direction is the establishment of the National Center for Digitization which is described in Section 8.

The second part of the project was concerned with design and building of specific archive databases of participating institutions. In fact, this activity leads to several derivatives of the main project, to particular projects with very definite aims. We shall describe them in the next part of this article. Let us mention that this work was the compound of two activities. The first one was concerned with building of parent databases: collecting and electronic archiving of "raw data". We considered this activity as the most important part of the project. The second activity was concerned with making various presentations (compact disks, WEB presentations, etc) based on parent databases. Of course, these outputs of the project were (and are) the most interesting aspects of the project to the general public.

3. Project: Archiving the journal *Publications de l'Institute Mathématique*

Publications de l'Institute Mathématique is the most important and oldest (founded in 1932) Serbian mathematical journal. Until the Second World War it was published by the Belgrade University, afterwards by the Mathematical Institute SANU. The archive consists of 90 volumes and comprises about 20000 pages and 2000 articles. Almost every Serbian mathematician published there at least one paper, as well as many world leading mathematicians (W. Sierpinski, P. Erdős, S. Shelah, etc.).

Digitization of the journal started in 1995. The first archiving technique was retyping articles by use of TeX system, starting with most recent issues. So, about fifty volumes were electronically archived in this way (about 11000 pages and 1200 articles). The archive is very compact, having less than 100 MB. It includes source (tex) files and output, device independent (dvi) files. Resolution of output files comes in 300dpi and 600dpi (both present for all files). All files are converted into pdf format (Adobe portable document format), and so both TeX and pdf formats are publicly distributed [1,5].

This archive base is now included in the large European mathematical database EMIS (European Mathematical Information System) [1]. At this moment PIM is the

most complete and meeting highest standards sub archive of EMIS system, due to the fact that we started first electronic archiving, using the technique (TeX) of the highest quality. However, as archiving using TeX system is expensive (about 2€ per page, comparing to less than 1€ per scanned page), it was decided that the first 40 volumes would be scanned. Ž. Mijajlović started the project, and now it is handled by D. Blagojević.

4. Project: Collected works of Bogdan Gavrilović

The goal of this project was to put in digital form collected works of the prominent Serbian mathematician Bogdan Gavrilović. The project started in 1996 and was finished in 2001. More than 2000 pages were scanned and put first in TeX frame, afterwards in pdf format. The archive is published as a compact disk presentation. The editor and the author of CD was Ž. Mijajlović. Collaborators on the project were D. Blagojević, S. Kordić and A. Mijajlović. Here is a short biography of Bogdan Gavrilović. It explains why we decided to digitize his collected works.

Bogdan Gavrilović was born in Novi Sad on January 1, 1864. Top of the class in his generation, after completing secondary education he was sent by *Tekelium* (a Serbian educational institution) to study at the Philosophical Faculty of the University in Budapest, where he obtained doctor's degree in mathematics in 1887. In the same year he was appointed professor at the High School in Belgrade which in 1905 was promoted to the University of Belgrade. He lived in Belgrade until his death in 1947, active as university professor until 1941. Before the turn of the century he had published two voluminous university textbooks which had the character of monographs: *Analytical Geometry* (1896) of length 900 pages and *Theory of Determinants* (1899) on linear algebra. Both works may be considered as capital works in mathematics in Serbia. Academician Radivoj Kašanin thus wrote of the two books: *Both, especially the latter, would do honor to any nation, and many countries, at that time more powerful and luckier than us, could not boast of such works.*

For a whole decade at the beginning of the 20th century he was busy mostly with scientific subjects, publishing a score of excellent treatises, mainly in the periodical *Glas*, organ of the Serbian Academy of Sciences, and *Rad* of the Yugoslav Academy of sciences and Arts in Zagreb. His interest was in the fields of algebra (theory of numbers and linear algebra), analytical geometry and theory of functions. He was elected member of the Serbian Academy of Sciences in 1901 and of the Yugoslav Academy of Sciences (Zagreb) in 1906. He later expanded into foundations of mathematics. Professor Gavrilović was an excellent organizer of scientific work at the Belgrade University and in the Academy. He made a great contribution to the promotion of High School into the University of Belgrade, and subsequently as rector, to the raising of teaching levels and university's development. Also he was three times (1931–1937) elected president of the Serbian Academy of Sciences. In 1894 he founded the mathematical library of the Department of Mathematics which was unfortunately destroyed at the end of the Second World War. With Mihailo Petrović and Milutin Milanković, he takes credit for introducing modern mathematics in Serbia at the beginning of this century. Academician Gavrilović held many important scientific and social functions and received a number of high honors. He was one of the founders of the Mathematical Institute SANU in Belgrade (1946). Bogdan Gavrilović was an

eminent educationalist and cultural worker, who had strong views on many important questions concerning national life, politics, history and philosophy. As a member of an exclusive Belgrade intellectual circle, he helped to create a special atmosphere thanks to which Belgrade became one of the centers of scientific work.

All his works: two books, doctoral dissertation (written in Hungarian), twenty scientific papers, 19 papers on general matters (mainly his speeches addressed to general public), an extensive biography and a lot of archival material (mainly from Budapest Archive) are presented on CD.

5. Project: Old maps, engravings and photographs – Collection of the City Museum of Belgrade

The archive consists of two collections: the collection old maps and engravings, and the collection of Groman's photographs.

The City Museum of Belgrade owns one of the most valuable collections of old engravings in Serbia. Almost 200 engravings were selected for the project – single sheets and album engravings that are thematically related to the history of the country, i.e. to the history of Belgrade. A majority of engravings were donations of the industrialist Đorđe Vajfert, besides those of other donators and purchased ones. The engravings encompass a period between XVI and XIX century. A large number of authors like Resch, Ortelius, Mercator, Bodenehr, Jacob Alt, etc, are mainly represented by copper engravings and lithographs, and in lesser degree woodcuts. Engravings were printed in Amsterdam, Augsburg, Vienna, Nuremberg, Paris, London, and, since XIX century, in Belgrade. A smaller group of engravings relates to the Turkish conquest of Belgrade, and a larger group deals with Austrian-Turkish wars and the reprints of engravings depicting these events, as well as representations of the town when it passed into Christian hands. The engravings are valuable for many scholars, often being a decisive if not an only historical source about a period of their creation. Their artistic value must certainly not be neglected.

Another section consists of the photographs of I.V. Groman, Russian military photographer. He is not well known, but his arrival to Serbia after the outbreak of Serbian–Turkish war (1876–1878) has been ascertained. Photographs of the environment of Belgrade with many public buildings from the end of XIX century predominate, along with those made at numerous battlefields of Serbia, i.e. in the valleys of Timok, Morava, then at Jagodina Paraćin, etc. Groman's photographs are also preserved in the Military Museum and a small number is kept in the library of daily paper "Borba" in Belgrade. His work exists likewise in Moscow (in the former Central State historical Archive of USSR). The importance of Groman's photographs arises from a fact that these are the first photographs of Belgrade and Serbia. It also may be confirmed by the interest of learned and broader public for pictorial evidence about Belgrade and Serbia in XIX century.

The project was headed by Ž. Mijajlović, and lasted for two years (1995–96). Authors of CD are: Ž. Mijajlović, M. Korać, Z. Ognjanović, Ž. Novaković (the author of accompanying text) and M. Mijajlović. Musicologist K. Tomašević selected the music. This edition of CD was bilingual: Serbian and English. The presentation was written in HTML. All items are accompanied with text (about 300 pages) and selected music works of old Serbian composers (XIX century), some of them published for the

first time. Resolution of some images on CD is unusually high (up to 300dpi), so very good printouts can be obtained.

Compact disk received good attention and obtained fine critiques not only in main daily news papers, but in specialised journals as well. Compendium of the Archive and CD is displayed on the www [4].

6. Project: Presentation of Historical Archive of town Kotor

The main product of this project was the compact disk presenting cultural and historical heritage of Kotor and Historical Archive in Kotor. This institution is the oldest and richest archive in Montenegro and Serbia, and was the second one in former Yugoslavia after the Archive in Dubrovnik. The oldest documents in archive are dating from the beginning of XIV century. Some of the oldest documents were presented: the first State Notary Book (1326-1335) and the Statute of town Kotor (*Statuta et leges civitatis*, 1616). CD contains altogether 185 documents, and 87 historical and cultural sites in Kotor and vicinity are presented. The noted historical monument portrayed on CD is the mediaeval Catholic church Saint Tripun.

The compact disk was ordered by the Montenegro government for needs of Yugoslav pavilion on the World exhibition in Lisbon, EXPO '98. The content of CD was written in HTML and produced in the archiving laboratory of the mathematical Faculty in Belgrade. Editors were J. Antović and Ž. Mijajlović. Colaborators were S. Kordić, M. Milošević, Z. Čubrić and J. Katelan. It should be mentioned that Stevo Kordić by his work on this CD introduced into our country new and very high standards in use of digital photography and visual design in multimedia presentations.

7. Project: Memorial compact disk of the Faculty of Mathematics in Belgrade

This CD is related to the celebration of 125th anniversary of the Faculty (it was founded in 1873, as a division of Belgrade High School, a counterpart of the university). The CD has two parts: a. historical, b. appropriate content concerning the contemporary work and life at the Faculty.

The most interesting part is historical. It contains several fine archives, e.g. the list of all students who graduated mathematics (or astronomy, or mechanics). For example, we can see that Mihailo Banić was the first student who graduated (1875), and that 35 students graduated in the period 1875–1914. We can see also the names of all mathematicians, astronomers and mechanics who obtained their PhD at this Faculty. Besides of a lot of information of this kind, CD contains supplements consisting of several books in mathematical science written by Serbian authors of the 19th century. Also, there are first eight doctoral dissertations of Serbian mathematicians (obtained until First World War). This part of CD was in fact the beginning of a new project: *Elderly works of Serbian authors in mathematical sciences*. The project is now in progress.

The compact disk has about 7000 pages of printed text, several hundred pictures, and even some music. There were more than sixty authors with about 1000 pages of new and original text. The content is presented in HTML and pdf formats. The CD was

realized by Ž. Mijajlović (editor), Z. Ognjanović, M. Korać, S. Kordić, V. Vučković, K. Tomašević and U. Midić. As the disk was finished in 1999, two events that happened that year are depicted: Sun eclipse and the NATO attack on Yugoslavia [14].

8. National Center for Digitization (NCD)

One of the most important of our recent activities in the area of digitization of cultural and scientific heritage is the work on the foundation of the National Center for Digitization (NCD). The idea of this project is to form a consortium consisting of our leading cultural and research institutions involved in digitization. At the present state, the consortium includes the Mathematical institute SANU and the Mathematical faculty Belgrade (they give the technical support to NCD), National Library of Serbia, National Museum Belgrade, Archaeological Institute Belgrade, Archive of Republic Serbia, Serbian Institute for Monument Protection, Belgrade, and Yugoslav Film Archive. However, collaboration with other cultural and research institutions from Serbia and Montenegro, as well as with the similar projects from abroad, is in progress.

The main subjects of our cooperation are the following:

- I. Coordination of efforts of institutions involved in the cultural and scientific heritage digitization.
- II. Establishing and promoting a national strategy for the cultural and scientific heritage digitization.
- III. Exploring, adaptation and implementation of international standards and protocols for the cultural and scientific heritage digitization and preservation at the national level. Development of new standards in areas where they do not exist.
- IV. Launching the cultural and scientific heritage digitization and making plans for possible migration process to new formats and technologies for already digitized data.

These goals should be obtained through the activities of the NCD. Some of the activities are presented in Table 1. They illustrate the multidisciplinary character of the NCD. For example, the issue of standards in the area of digitization of cultural and scientific heritage concerns metadata, digital data storage formats and data structure for different kinds of objects (pictures, sound, video, electronic publications containing plain and mathematical texts, etc.), designing and managing local data bases of cultural and scientific institutions, protocols for access to the digitized data including some uniform middleware and end-user interfaces for all types of information resources, rights management, thesaurus, search procedures, tools for storage, processing and presentation of the digitized data, digitized data protection including access restrictions, data authentication, etc.

During the digitization process, models of a distributed information system for digitized cultural and scientific heritage and specific software which obey the above standards should be developed. The models should allow dissemination and wide access over Internet to cultural and scientific values that at the present are not easily accessible. These systems should be mutually compatible as much as possible. Collaborators from cultural and scientific institutions ought to use the software both as end-users and as experts who define meta-data for corresponding objects and enter information into databases.

1. Forming expert groups for specific fields of interest (groups for museums, libraries, archives, monument protection, standardization, data protection, etc.).
2. Establishing, at the national level, a basic set of standards in the area of digitization of cultural and scientific heritage.
3. Developing models of the appropriate distributed information systems and specific software for the realization of the digitization standards.
4. Connecting cultural and research institutions included in the digitization project into a particular Internet subnet and developing a network of information and WWW-servers.
5. Securing long-term storage and maintenance of the digitized heritage.
6. Organizing continuous, persistent and systematic work on the cultural and scientific heritage digitization.
7. Training collaborators from cultural and scientific institutions.
8. Collaboration with the similar institutions from abroad.

Table 1. Planned activities of the NCD

The main results achieved so far by the NCD are: the Committee consisting of representatives of involved institutions was formed, two national conferences under the name “New Technologies and Standards: Digitization of National Heritage” were organized in 2002 and 2003 [13], our journal, called “Review of the National Center for Digitization” was established and published in electronic and printed form [17] (the editor of the journal is Ž. Mijajlović), and the Internet presentation of the NCD containing the main information about NCD was developed [7]. We believe that in the near future both our digitization conference and the journal will obtain the international flavor.

9. Other projects and activities

There are several projects and activities that have been recently launched by our institutions, for example: the project on standards in digitization which started last year, and it may be considered as a continuation of the project Computer archiving and multimedia presentation of cultural values and national heritage described in this article, electronic edition of the journal *Visual Mathematics* [6] edited by Slavik Jablan, etc.

After the workshop “Digital Preservation of Cultural Heritage” [12] which was a part of the International Congress MASSEE’2003 [3], held in Borovets, Bulgaria, the participants agreed on the text concerning the future development of a network for digitization of scientific and cultural heritage in South-Eastern Europe. It was declared that “We are convinced that we face common problems and share common scientific and cultural heritage. The knowledge and experience of single institutions from our countries should not stay isolated. It is of great importance to take measures for increase of the communication and exchange of technological expertise, standards and practical skills within the region, taking into account the experience of colleagues outside the region.” As the first step in this direction, the proceedings of the workshop will be

published as a special volume of “Review of the NCD”, while the editorial board will be expanded by a number of prominent researches from abroad. The next meeting of the group will be held in Belgrade in the June, 2004, during the third conferences “New Technologies and Standards: Digitization of National Heritage”. We are sure that such an international cooperation will significantly enhance results of our activities in the field of digitization.

References

- [1] European Mathematical Information System - EMIS, <http://www.emis.de> [Date of last access: 2003-11-24]
- [2] HTML, HyperText Markup Language, The World Wide Web Consortium, <http://www.w3.org/MarkUp/> [Date of last access: 2003-11-24]
- [3] International Congress MASSEE'2003, 15-21. 9. 2003, Borovets, Bulgaria, <http://www.math.bas.bg/massee2003/index.html> [Date of last access: 2003-11-24]
- [4] Internet presentation of the CD “Old maps, engravings and photographs – Collection of the City Museum of Belgrade”, <http://www.mi.sanu.ac.yu/muzej.beograd/> [Date of last access: 2003-11-24]
- [5] Internet presentation of the journal “Publications de l’Institute Mathematique”, Mathematical Institute, Belgrade, <http://www.mi.sanu.ac.yu> [Date of last access: 2003-11-24]
- [6] Internet presentation of the journal “Visual Mathematics”, Mathematical Institute, Belgrade, <http://www.mi.sanu.ac.yu/vismath/> [Date of last access: 2003-11-24]
- [7] Internet presentation of the National Center for Digitization, <http://www.ncd.matf.bg.ac.yu> [Date of last access: 2003-11-24]
- [8] Donald E. Knuth, *The TeXbook*, Addison-Wesley, Reading, Massachusetts, ISBN 0-201-13448-9, 1986.
- [9] Leslie Lamport, *LaTeX – A Document Preparation system*, Addison- Wesley, Reading, Massachusetts, ISBN 0-201-15790-X, 1985.
- [10] LaTeX Project Home page, <http://www.latex-project.org/> [Date of last access: 2003-11-24]
- [11] Mediana, http://www.serbia-tourism.org/engleski/culture/cul_medi.htm [Date of last access: 2003-11-24]

- [12] Minisymposium "Digital Preservation of Cultural Heritage", International Congress MASSEE'2003, 16-17. 9. 2003, Borovets, Bulgaria, http://www.math.bas.bg/massee2003/BAL_conference.html [Date of last access: 2003-11-24]
- [13] National conferences "New Technologies and Standards: Digitization of National Heritage", <http://www.ncd.matf.bg.ac.yu/?page=conferences&lang=en> [Date of last access: 2003-11-24]
- [14] NATO Aggression Against Yugoslavia, Internet presentation of the Mathematical Institute, Belgrade, <http://www.mi.sanu.ac.yu/nato.htm> [Date of last access: 2003-11-24]
- [15] Zoran Ognjanović, Miomir Korać and Filip Dugandžić, *PANDORA – ekspertni sistem za datiranje iskopina*, chapter 11 in Miroslav Jocković, Zoran Ognjanović and Stevan Stankovski, *Veštačka inteligencija. Inteligentne mašine i sistemi*, Krug, Beograd, ISBN 86-7136-032-6, 1997.
- [16] Standard Generalized Markup Language (SGML), ISO 8879:1986, <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387> [Date of last access: 2003-11-24]
- [17] Review of the National Center for Digitization, Faculty of Mathematics, Belgrade, <http://www.ncd.matf.bg.ac.yu/?page=publications&lang=en> [Date of last access: 2003-11-24]
- [18] TeX Users' Group, <http://www.tug.org/> [Date of last access: 2003-11-24]
- [19] Viminacium, US-Serbian project, <http://www.viminacium.org.yu/> [Date of last access: 2003-11-24]

PREGLED NEKIH PROJEKATA DIGITALIZACIJE U SRBIJI

Sažetak. U radu je dat pregled nekih projekata digitalizacije koji su sprovedeni u Srbiji u protekloj deceniji. Te projekte su većinom realizovali Matematički institut Srpske akademije nauka i umetnosti (SANU) i Matematički fakultet u Beogradu. U projektima se razmatraju standardi u digitalizaciji kulturnog i naučnog nasleđa, kreiranje elektronskih arhiva, razvoj prezentacija digitalizovanog nasleđe, kao i istraživanja u oblasti primene ekspertnih sistema u arheologiji. Opisana su i nastojanja u vezi osnivanja Nacionalnog centra za digitalizaciju koji čine vodeće domaće institucije koje se bave digitalizacijom. Predstavljene su inicijative za saradnju sa srodnim inostranim projektima.

Ključne reči. Digitalizacija, kulturno nasleđe, naučno nasleđe

Žarko Mijajlović
Matematički fakultet
Studentski trg 16, 11000 Beograd
Srbija i Crna Gora
<mailto:zarkom@eunet.yu>

Zoran Ognjanović
Matematički institut SANU
Kneza Mihaila 35, 11000 Beograd
Srbija i Crna Gora
zorano@mi.sanu.ac.yu
<http://www.mi.sanu.ac.yu/~zorano>