

Digitization of Old Mathematical Periodicals Published by the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

Vania Grigorova¹, Kalina Sotirova¹, Viktoria Naoumova¹, Anna Sameva¹,
Milena Dobрева², Krassimira Ivanova¹, Peter Stanchev^{1,3}

¹Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
{vanya.ang, kalina, vnaoumova, sameva, kivanova}@math.bas.bg

²University of Malta, Malta
milena.dobreva@gmail.com

³Kettering University, Flint, USA
pstanche@kettering.edu

Abstract. The digitization practice for retro-converting of the mathematical periodicals, published by the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences (IMI-BAS) and the followed benefits for long-term preserving and assuring open access to these materials is discussed in the article.

Keywords: Digitization Cycle, Open Access, IMI-BAS

1 Introduction

The age of access [18] introduces a new type of division in the world – digital divide, divide of people being into the global net and those who are out of it, i.e. online and offline. The age of access influences the world of science also. The paradigm of scientific publishing (publications and research data) and access to it has changed recently [11]. Moving towards an open access to scientific information is a worldwide endeavour where UNESCO, Organisation for Economic Co-operation and Development (OECD), and European Commission (EC) are working in agreement. Currently databases, such as ProQuest [17], ScienceDirect [19], Scopus [20] give full or restricted access to scientific papers depending on type of subscription and rarely give open access though. The definition of Budapest Open Access Initiative for open access is: "... free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited" [2].

The European Commission (EC) in its Recommendation on access to and preservation of scientific information (July 2012/417/EU) states that open access should become a priority in national research policy: "Open access policies aim to provide readers with access to peer-reviewed scientific publications and research data free of charge as early as possible in the dissemination process, and enable the use and reuse of scientific research results. Such policies should be implemented taking into account the challenge of intellectual property rights" [4]. Further, EC gives five main recommendations altogether with the deadline end of 2012 where national point of reference should be created. The five points are [4]:

1. Open access to scientific publications for publicly funded research, including financial planning and licensing systems.
2. Open access to research data with "concrete objectives and indicators to measure the progress".
3. Long-term preservation and reuse of scientific information by clear policies, effective deposit system and keeping hardware and software needed to read the information in the future.
4. e-Infrastructures, which are understood to be "an environment where research resources (hardware, software and content) can be readily shared and accessed wherever this is necessary to promote better and more effective research".
5. Multi-stakeholder dialogue at national, European and international level.

Recently EU introduces an open access to publicly funded scientific papers as a widespread policy through the FP7 projects OpenAIRE and OpenAIRE+ [16] and EuDML especially for mathematical publications [7].

Currently, the biggest online repository for digitized periodicals, proceedings and selection of monographs is JSTOR [12]. It covers 50 fields of human knowledge. Another one is NUMDAM (Numérisation de documents anciens mathématiques) [15], a repository for old mathematical resources. Euclid project [6] also has a mission to advance scholarly communication in the field of theoretical and applied mathematics and statistics through a collaborative partnership between low-cost independent and society journals. Altogether with ScienceDirect and JSTOR give e-access to all currently available periodicals in Mathematics. World Digital Mathematical Library (WDML) [22] and the European Digital Mathematical Library (EuDML) [7] have the ambitious goal to accumulate all digital resources in Math science.

The article gives an overview on publishing activity in IMI-BAS and on the digitization cycle, provided in IMI-BAS for retro-converting and providing digital access to the old mathematical periodicals, published by the Institute.

2 Publishing Activity in IMI-BAS

Publishing activity of IMI-BAS is connected with research periodicals in the area of pure and applied mathematics, informatics, and education in mathematics and informatics.

IMI-BAS became main publisher of the following research periodicals:

1. Physical-Mathematical Journal from 1958 to 1993 as a successor of the journal with the same name, issued by Bulgarian Physical-Mathematical Society from 1904 to 1950. This was the first Bulgarian research journal in the field of physics and mathematics.
2. Bulletin of the Mathematical Institute from 1953 to 1974.
3. Serdica Mathematical Journal (shortly Serdica) since 1975.
4. Pliska Studia Mathematica Bulgarica (shortly Pliska), which first issue was launched in 1977.
5. Serdica Journal of Computing. Since its start in 2007, Serdica Journal of Computing is published both on paper and online.

All journal issues have been collected and archived in the library of IMI-BAS. Deposit of and access to articles, published in IMI-BAS periodicals, is being preserved in our library and in the institutional e-repository. Majority of negotiations for authors' rights to deposit into open access repositories are exercised in the context of authors' (or publication) agreements signed by authors at the moment the article is submitted for publication.

According to the recommendations of IFLA (International Federation of Library Associations): "issues involved in the selection of material for digitization is examined from two perspectives – principal reasons for digitalization (to enhance access and/or preservation) and criteria for selections (based on content or based on demand)" [9].

From the mentioned five journals, Serdica Journal of Computing is from the new generation digitally born journals, while Physical-Mathematical Journal and Bulletin of the Mathematical Institute are too old and have more or less archival treasure. Most interesting for digitizing became old issues of Serdica and Pliska. Serdica is published once in three months. Pliska is an annual journal; each issue has a theme in particular area in mathematics. The articles in Pliska and Serdica in earlier years were published in Russian, English and German. Since 90ties, the both journals have published articles in English language only.

Each publication is peer-reviewed [3] and referred by Referral Journal of VINITI [21], Zentralblatt für Mathematik (Zentralblatt MATH) [23], and Mathematical Reviews (MathSciNet) [14].

Currently Serdica and Pliska can be received by subscription and are the main source for book exchange for libraries in BAS. According to statistics in the book-exchanges, 260 libraries in the world have Pliska in their collections. All above-mentioned characteristics of IMI-BAS periodicals justify their role as a foundation for the Bulgarian digital repository for Mathematics – as part of European and world mathematical heritage.

3 Digitization in IMI-BAS

The start of digitization in IMI-BAS was launched in 2004 during the sixth FP project "KT-DigiCULT-BG" (Knowledge Transfer for Digitalization of Cultural and Scientific Heritage in Bulgaria). For carry out the task a special temporary unit was created,

entitled "Digitization of Scientific Heritage", which in 2006 became department "IT applications in Humanities". Currently it functions as Laboratory of Digitization of Scientific and Cultural Heritage in the frame of Information Systems Department.

IMI-BAS is partner of projects EuDML, OpenAIRE, and OpenAIRE+. As a partner in these projects, IMI-BAS fostered digitization of mathematical periodicals. The digitization of these periodicals started as institutional project in 2006, shortly after creation of Digitization unit at IMI-BAS. Currently, digitized resources could be found online on a digital repository of IMI-BAS [10]. Open source platform DSpace [5] is based upon OAI-PMH and offers tools for managing various digital resources in institutional repositories. It has been installed in more than 800 institutions worldwide, and has flexible web interface and offers software optimization and personalization. IMI-BAS repository, based on DSpace, includes mainly publications of the institute' journals. They are organized in collections, and title, author, keyword, and chronology can do the search. The metadata schema is based on the format of Qualified Dublin Core. Collections are available online and could be linked to the online public access catalogue of BAS. The catalogue of BAS uses international cataloguing format for librarian resources – MARC21 [13]. It was integrated into information system Aleph 500 [1], where each resource has own bibliographical record.

4 Digitization Cycle in IMI-BAS

The adopted digitization cycle for mathematical periodicals in IMI-BAS includes following steps:

1. Scanning.
2. Metadata input.
3. Quality control and image processing.
4. Storage (long-term preservation strategy).
5. Public access.

Scanning is done by professional book scanner Zeutschel OS System 12 000. The scanning software is Omniscan under operation system Windows XP Professional. Omniscan has many technical options and filters for image improvement during scanning. The colour profiles to choose from are: Color, Gray4, Gray8, Black and White. The resolution is up to 600 dpi, 24 bits depth. Auto-focus option of the scanner allows high-speed of the scanning. When scanning without glass plate the option Perfect Book allows auto-removal of the fingers of the operator from the pages and fixing of the vertical/horizontal position of the journal. Omniscan software allows defining of scanning areas by clips (rectangular frames). The operator can choose the number of the clips to use, especially when there are images on the page.

The sequence of the scanning steps is as follows: turning on the computer system, fixing the scanner settings and putting the document on the scanner-cradle, creating a separate folder and naming the files, setting the file formats, colour profile, and resolution, initial metadata input, setting the clips, scanning and initial quality control,

ordering/reordering the pages, saving all the TIFF files in the concrete folder, and finalizing the process.

File organization, naming, and metadata input are linked to indexing process. The most important criteria here are hierarchical ordering in collections, while each file in a collection consists year of publishing, volume, issue, content page of the issues, author, title, number of pages for each article, and additional (optional) elements such as annotation, keywords, MSC classification, link to the whole PDF document.

Quality control of the TIFF files is to be done while and after scanning process. Quality control consists of checking some inconsistencies of the pages, spots, wrong angles of the clips, filter fixing, etc. If necessary, the re-scanning is done. Consequent software processing with specialized image software includes: adjustments of print area for each TIFF file, additional tuning of image/contrast, cropping and rotating. This way each article of the journal has equal margins, print areas and uniform colour gamma. The goal of this image processing is to make the reading easier of the e article.

TIFF file format is kept as highest quality format and as a masterfile since the start of DigiLab in IMI-BAS. This decision was taken after the in situ analysis of digitization practices in other European digitization laboratories. TIFF file format is kept as archival digital copy of the original, allows quick re-formatting and includes the metadata information, while protecting the high quality of the original. Masterfiles are archived on the department server and on DVD.

In the final stage, TIFF file is converted into JPG, the front page for each article is added, and all JPG files (plus the front page) are merged into one PDF file. This process is not automated. Each step requires human work and quality control. There are requirements for the size of the PDF file, since it needed to be downloaded quickly online. All articles are published on the web sites of the journals www.math.bas.bg/serdica and www.math.bas.bg/~pliska.

5 Conclusion

The current article discusses mathematical journals and related digitization activities at IMI-BAS. As a result following conclusions can be formulated [8]:

- the digitization of analogue media increased opportunities for access to their content and its use by consumers;
- digitized research resources offset some of the disadvantages of analogue formats;
- the digitization saves time for the consumers;
- it allows to be collected in a common repository retro-converted as well as digitally-born resources of one edition;
- the digitisation and open access assures accessibility of the scientific knowledge at any time and from any location;
- the digitization fits into the overall process of scientific exchange, contributes to saving time and extending, completeness, availability, functionality, efficiency of scientific knowledge.

References

1. Aleph 500, <http://www.exlibrisgroup.com/category/Aleph>, last accessed 25.05.2012
2. Budapest Open Access Initiative, <http://www.soros.org/openaccess/>, last accessed 25.05.2012
3. Christov D., V. Todorov, E. Brankova, etc., The Bulgarian contributions to mathematics, physics and chemistry 1889-1939, BAS, Bulgaria, 1999, ISBN 954-8854-06-6.
4. Commission Recommendation on access to and presentation of scientific information. In: Official Journal of the European Union 194/39, 21.7.2012
5. DSpace, <http://www.dspace.org/>, last accessed 25.05.2012
6. Euclid project, <http://projecteuclid.org/>, last accessed 25.05.2012
7. EuDML, <http://eudml.org/>, last accessed 25.05.2012
8. Grigorova V., From the Book Shelves to Zeros and Ones: the Digitization of the Mathematical Periodicals in Institute of Mathematics and Informatics, Bulgarian Academy of Sciences. Master thesis in Sofia University "St. Kl. Ohridsky", Sofia, 2012
9. IFLA, ICA – Preservation and Conservation Section: Guidelines for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives, March 2002, <http://www.ifap.ru/library/book126a.pdf>
10. IMI-BAS Repository, <http://sci-gems.math.bas.bg:8080/jspui/> Last accessed 7.08.2012
11. Johnson I., Libraries and Publishing: Time for New Paradigms? Papers from the Int. Conf. "Libraries, Globalisation and Cooperation", Sofia, Bulgaria, 2004, St. Kl. Ohridski Univ. Press, 2005. ISBN 0-934068-15-1
12. JSTOR, <http://www.jstor.org/>, last accessed 25.05.2012
13. MARC21, <http://www.loc.gov/marc/>, last accessed 25.05.2012
14. MathSciNet, <http://www.ams.org/mathscinet/>, last accessed 25.05.2012
15. NUMDAM, <http://www.numdam.org/>, last accessed 25.05.2012
16. Open AIRE, <http://www.openaire.eu/>, last accessed 25.05.2012
17. ProQuest, <http://www.proquest.com/>, last accessed 25.05.2012
18. Rifkin, J.: Age of Access. <http://www.techsoc.com/access.htm>, last accessed 25.05.2012
19. ScienceDirect, <http://www.sciencedirect.com/>, last accessed 25.05.2012
20. Scopus, <http://www.info.sciverse.com/scopus>, last accessed 25.05.2012
21. VINITI (World Institute of Scientific and Technical Information), <http://www2.viniti.ru/>, last accessed 25.05.2012
22. WDML, <http://www.mathunion.org/WDML/dml/index.shtml>, last accessed 25.05.2012
23. Zentralblatt MATH, <http://www.zentralblatt-math.org/>, last accessed 25.05.2012