

Scientific Heritage in Bulgaria Makes First Digital Steps

Milena Dobрева¹; Nikola Ikonov^{1,2}

¹ Digitisation of Scientific Heritage Dept., Institute of Mathematics and Informatics
bl. 8, Acad. G. Bonchev St., Sofia 1113 Bulgaria

e-mail: dobreva@math.bas.bg

² Phonology and Speech Communications Lab, Institute for Bulgarian Language
Shipchenski Prohod 52, Sofia 1113 Bulgaria

e-mail: nikonov@ibl.bas.bg

Abstract

The paper presents recent initiatives in creation, delivery and management of scientific heritage digital resources in Bulgaria. The local and international tendencies will be sketched. Then the work of the Department for Digitization of Scientific Heritage at the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences and more specifically the joint projects with the State Department of Archives and the Central Library of the Bulgarian Academy of Sciences will be described. Finally, we will present a SWOT analysis of the local situation and suggestions for urgently needed actions.

Keywords: digitization; BulDML; musical periodicals; DIGMAP

1 Introduction

The high expectations about the digital libraries are seen from their listing amongst the three flagship initiatives of the strategy "i2010 – A European Information Society for growth and employment" (the other two areas are caring for people in an ageing society; and the intelligent car). Amongst the basic aims in the joint European efforts are to *avoid duplication*, to *cooperate in networking and standards*, as well as in *developing common and more cost-effective solutions* [1].

Without any doubt, the availability of high quality digital content is in the basis of humanitarian and social research. Most countries from the South-Eastern Europe are still far from moving in line with the European Commission (EC) guidelines in this area. The *EC Recommendation of 24 August 2006 on the digitization and online accessibility of cultural material and digital preservation* [2] emphasizes the importance of setting up of large and sustained digitization facilities, encouraging partnerships between cultural institutions and the private sector, solving the problems around orphan works as well as developing clear quantitative targets for digitization efforts. Furthermore, the *Council Conclusions on the Digitisation and Online Accessibility of Cultural Material and Digital Preservation (2006/C 297/01)* [3] suggest an action plan which can not be followed in countries where there is neither national strategy nor large scale digitisation facilities or recognized competence centres promoting digitization activities.

Recently at the closing of Conference on Scientific Publishing in the European Research Area Access, Dissemination and Preservation in the Digital Age (Brussels, 16 February 2007), Ms. Viviane Reding, EU Commissioner on IS & Media stressed:

*“... if we do not actively pursue the preservation of digital material now, we risk having a **gap in our intellectual record**. If you allow me another historical reference, we do not want to experience the digital equivalent of the destruction of the Alexandria Library. Scientific assets are just too valuable to be put at risk.”* [4].

Countries like Bulgaria, which still do not have a national framework for digitization of cultural and scientific heritage, are even in more danger of deeper digital divide. Not only the structured effort to digitize and preserve is missing, but there is yet another danger. The extensive brain drain causes a gap in the community of those who could work on content provision, the experts expected not only to take care of digitisation of cultural and scientific material, but also to place it in the local and wider European context. The experienced researchers

which are still active in their profession do not have to whom to transfer their knowledge because local research career is not attractive to the young generation.

2 The Current Situation in Bulgaria

As it was already mentioned, the concerns communicated on the top EC level are not exactly matching the local Bulgarian situation. However, common and more cost-effective solutions, cooperation in standards and practical work, and care to avoid duplication should be strictly followed in a country with a population of 7, 2 million which hosts over 5 million cultural and scientific heritage objects.

Another important issue of the cultural and scientific heritage institutions nowadays for Bulgaria is the adoption of brand-new IT applications in the sector. Especially new and emerging technologies which are presented in [5] if used at all in Bulgaria appear just in small demonstration projects. Amongst the current problems in Bulgaria, we should mention:

- The absence of a *national strategy*, which leads to lack of co-ordination between separate local initiatives;
- The lack of understanding and practical solutions on importance of such issues as *common quality standards* and *interoperability*;
- The gaps in the local laws and *legislative regulations* related to digitization lead to difficulties for the decision makers in the cultural and scientific heritage sector institutions;
- The need for better *co-operation on regional and European level*, since most of the cultural heritage is one we all share;
- The ambiguity of legal *copyright issues* which leads to serious problems in persuading researchers to share their knowledge in digitization projects affecting the level of presentation of materials, and restricting the depth of presentation. Copyright issues are related to the primary sources on the one hand; on the other hand the issues of legally using the results of research work during digitization are completely unclear.

If we would have to summarize the current situation in Bulgaria, we could draw the following basic conclusions:

- Bulgarian collections are of European importance but they still are not accessible in electronic form;
- Experience exists basically in the pre-digitisation stages of work such as cataloguing, and text encoding, but mass digitisation projects are just about to start in several libraries;
- Digitisation work per se has not been done, thus the country does not match current EC priorities;
- No regular governmental programme (respectively, funding) is available, digitization in Bulgaria strongly depends on external financial support. The Ministry of Education and Science had one stand-alone call for projects on cultural heritage in 2006. Through it several libraries in Bulgaria currently start their own digitisation projects which are not interoperable. Currently, the State Agency for Information and Communication Technology is working on a national strategy for the accelerated development of information society in Bulgaria in 2007-2010 and digitization is included amongst the priority areas as a general topic. This is a positive sign that the field of work becomes officially recognised, but this is not sufficient for the success of the efforts of various institutions;
- Regional cooperation in the field is realistic. SEEDI (South Eastern Europe Digitization Initiative, [6]) is a joint effort to develop awareness about digitization of cultural and scientific heritage in the region along the Lund Principles of the European Union. It is based on the acceptance that researchers and institutions from the region face common problems and share common scientific and cultural heritage, which still cannot be widely accessed in electronic form. The cooperation within SEEDI is bringing together researchers from regional and European centres with similar scientific and practical interest in digitization and by supporting cooperation between them. For that purpose core groups of specialists are created in order to consult, assist, monitor and develop innovative technologies and digitization projects in collaboration with the local cultural and scientific heritage institutions.

Thus, currently there is not only a niche but an urgent need for several interconnected efforts: creating a national framework, boosting wide-scale digitization work, promoting cooperation of local institutions and improving the excellence in the profession.

3 The Experience of the Digitization of Scientific Heritage Department

In 2004, the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences created the Digitization of Scientific Heritage department which hosts the first and still unique in the country Bulgarian digitization line for scanning books and archival documents. The department has set the ambitious goal not only to do research work but to serve as a competence centre, providing following basic activities:

- assistance to technology and content providers;
- development of state of the art workflows and best practices in various areas of digitisation of scientific and cultural heritage;
- implementation of new technologies related to the digitization of cultural and scientific heritage;
- organization of specialised trainings;
- active participation at the elaboration of a National strategy for digitization of cultural and scientific heritage;
- contribution to the international cooperation within regional and European initiatives;
- representation of the country at international fora;
- methodological guidance and practical implementation within participation in various projects.

4 Current Activities

4.1 Cooperation with the General Department of Archives

The Digitisation of Scientific Heritage department cooperates with the General Department of Archives at the Council of Ministers of Bulgaria (GDA) since 2004. GDA is contributing with defining the priorities for selecting materials for digitization; developing the strategy for preparing descriptions and metadata; preparing specification of the search tools and their future improvement. The Digitization of Scientific Heritage department is contributing with providing its know-how for scanning and optical character recognition, workflow choice, digital image processing and with ensuring the necessary equipment and qualified personnel.

The selection and preparation of documents for digitization (single documents, parts of the archival funds and complete archival funds) is based on the holdings of the Regional Unit “State Archive” – Sofia and includes interesting materials, related to the management of Sofia Municipality, the history of Sofia University, the archives of the former Bulgarian communist party, archival funds of the Monarchy Institute, The Parliament, the Council of Ministers, etc. Selected materials contain valuable manuscripts and printed documents, photographs, sketches, geographical maps and rare books, etc.

The joint work has already brought practical outcomes. In March 2005 both institutions released a multimedia disk “Sofia. Religious spaces”, containing items displayed during an exhibition of the same name, and including scanned documents, digital copies of canonical and dogmatic books, and photographs of paintings, ritual clothes and cult objects.

Recently GDA and the Digitization of Scientific Heritage department started another joint project aimed at the electronic publishing of archive documents related to the Temporary Russian Governance which was established after the liberation of Bulgaria and ruled in the period 1878–1879. This collection of documents is being prepared as a combination of digitised images and full text. It will be organised as a semantic web portal. Scientists and general public will benefit from the availability of full-text transcriptions and tools for semantic search of the archival documents, mostly hand-written sources in Bulgarian and Russian.

Another ongoing effort is aimed at building an electronic archive of documents issued by the Bulgarian Ministry of Education in the 40ies and 50ies of the 20th century [7]. The department provides the methodological guidance in this project. The collection of documents is stored in the Archive of the Ministry of the People’s Education within the State Archival Fund of the General Department of Archives. This is a mixed collection which contains quite diverse documents - official documentation which follows specific templates; letters; notes, certificates; photographs; newspapers, etc. The text documents are printed, typewritten or handwritten. The basic aim of this work is to provide access to different users (specialists in education, historians, and the citizens) to the educational documentation of this historical period. The long-term goal is to build a joint collection of such documents from Bulgaria and Greece.

4.2 Joint Work with the Scientific Archive of the Bulgarian Academy of Sciences

Both institutions cooperate since 2005. The Scientific Archive of the Bulgarian Academy of Sciences stores precious documents, related to the history of the Academy. A pilot project was initiated in December 2006, aimed at the digitization of personal archives of famous Bulgarian scientists. As a kick-off both institutions selected the archive of Marin Drinov, one of the founders of the Academy of Sciences. It contains valuable documents, letters, personal notes and pictures. All of them were digitized and then prepared for electronic publishing. Thus scientific archives will become easily accessible both for the wide public and the researchers.

4.3 Involvement in the Digsaw Project

DIGMAP (Discovering our Past World with Digitised Maps) is a project which is supported through the eCONTENTplus programme [8]. It proposes to develop solutions for geo-referenced digital libraries, especially focused on historical materials and in the promoting of our cultural scientific heritage. The final results of the project will consist in a set of service available in the Internet, and in open-source software solutions that will be able to be reused in other services. The main service will be a specialized digital library, reusing metadata from European national libraries, to provide discovery and access to contents. Also, relevant metadata from third party sources will be reused, as also descriptions and references to any other relevant external resource. Ultimately, DIGMAP will pursue the purpose to become the main international information source and reference service for old maps and related bibliography. DIGMAP will develop solutions for georeferenced digital libraries, especially focused on historical materials and in the promoting of our cultural and scientific heritage. The final results of the project will consist in a set of services available in the Internet, and in reusable open-source software solutions.

The project will make a proof of concept reusing and enriching the contents from the **National Library of Portugal (BNP)**, the **Royal Library of Belgium (KBR/BRB)**, the **National Library of Italy in Florence (BNCF)**, and the **National Library of Estonia (NLE)**. In a second phase, that will be complemented with contents and references from other libraries, archives and information sources, namely from other European national libraries members of TEL – The European Library [9]. DIGMAP might become an effective service integrated with TEL - in this sense the project is fully aligned with the vision “**European Digital Library**” as expressed in the “i2010 digital libraries” initiative of the European Commission.

The technology will be developed by the Department of Information Systems and Computer Engineering of the **Instituto Superior Técnico — Lisbon (IST)**, in cooperation with the Group MERCATOR of the **Polytechnic University of Madrid (UPM)**. The project started in October 2006, and will have the duration of 24 months. The project coordinator is Prof. José Borbinha, of the IST. The technical work will be co-ordinated by the IST, UPM and KBR. The evaluation of the results will be co-ordinated by the NLE. The liaisons with external entities and advising groups will be coordinated by the BNCF. The dissemination will be co-ordinated by the BNP.

The **Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences (IMI)** will provide assistance to the evaluation, liaison with the Southern East Europe, and to the dissemination. In particular, the Digitisation of Scientific Heritage department already discussed what are the available maps and books relevant to the project in the collections of the General department of archives, Central Library of the Bulgarian Academy of Sciences, and National Library ‘Ivan Vazov’ in Plovdiv. Electronic records on these objects are not available yet and the participation in this project would be a chance for exposure of local materials.

4.4 Digitisation of Bulgarian Mathematical Heritage

One of the principal activities of the department is the digitization of mathematical publications of Bulgarian mathematicians, with the aim to build BulDML – Bulgarian digital mathematical library. Following journals and documents are currently being digitised:

1. Physical-Mathematical Journal (in Bulgarian), 1958 – 1991, 1993;
2. Serdica (articles in different languages): 1975, 1995-2000 г. ;
3. Archive on the development of the Union of Bulgarian Mathematicians: Book with records of board meetings, 1905-1936. In addition to the digital images full texts are also entered. Research on obtaining photographs of the mathematicians whose names appear in the records is being done – this will allow us to offer a complete resource on the dawns of Bulgarian mathematics;

4. Full collection of books and publications of Nikola Obreshkov, a famous Bulgarian mathematician.

The works of the Department are not only limited to mathematical publications. Efforts have been made in order to enlarge the spectrum of digitization activities. As a result additional fields were covered by the Department, thus ensuring more versatility and flexibility:

4.5 Digitisation of Historical Musical Periodicals

This work aims to prepare digital copies and descriptions compatible with the Retrospective Index of Musical Periodicals (RIPM) [10]. Currently the following sources has been digitised and described:

- Gusla (1891), printed text, musical fragments, illustrations
- ASO (1934), printed text, illustrations
- Materials from the archive of Stoyan Kenderov (musicologist)

One specific difficulty is the identification of a repository which holds Bulgarian historical musical periodicals. For a variety of reasons, these publications are difficult to find.

4.6 Digitisation of Historic Newspapers

This is an effort started recently which aim is to produce collection of digitised old newspapers: Daga, Lampion, Mir, Dnes (selected issues from the period 1890-1930). This activity is a joint project with the Central Library of the Bulgarian Academy of Sciences. The idea is to offer digital images of newspaper pages as a whole, and access to the texts of the separate articles. Photographs which appear in the newspapers will also be included into a collection of images.

4.7 Electronic Records of Manuscripts

The department prepared descriptions of Old Bulgarian manuscripts in TEI conformant XML (the total number of these descriptions is 806). Manuscripts are not digitised because what to be done with them is a matter of library policies. However, the detailed description is an important preliminary activity.

4.8 Towards The Creation of a National Digitization Network

The “Digitization of Scientific Heritage department” is the initiator of the creation of a nation-wide network of institutions – museums, libraries, archives and research centers, which are intending to start mass digitization, thus avoiding the implementation of scattered non-effective small-scale projects.

There are a number of good examples of joint work with such institutions. We present here briefly only two case studies, which illustrate well the trend of creating synergies in the digitization field.

5 SWOT Analysis

The SWOT analysis is used to evaluate the *Strengths*, *Weaknesses*, *Opportunities*, and *Threats* involved in a process or more specifically in a project or in any other situation of an organization or individual requiring a decision in pursuit of an objective. It involves monitoring and analyzing the environment internal and external to the process in question. In order to provide a precise picture of the current situation we have tried to summarize all relevant conditions to the digitization of scientific and cultural heritage in Bulgaria.

Strengths	Weaknesses
<ul style="list-style-type: none"> • Experience already available. The positive influence is that some institutions already have the feeling what efforts are needed. • Good contacts with colleagues from the region and other EC countries. This is important for being in line with the current practices. • Trainings/specialists meetings done on a regular basis. The circle of specialists from the community of practice grows although this is quite slow process. • Established professional bodies. The existence of departments such as the Digitisation of scientific heritage in IMI is important since it is in contact with many institutions which will play the role of future content providers. 	<ul style="list-style-type: none"> • Lack of established and working national strategies in the field of digitization, online accessibility and preservation. This leaves all decisions on specific actions to the institutions which in fact would play the role of content providers. In most cases their ideas and vision on digitisation are quite simplified. • Strong dependence on external funding. This is in controversy with the need to set up national priorities. External funding is not reflecting the national vision on importance. • Scattered experience. The experience which exists is for small initiatives, not for large projects/programmes.
Opportunities	Threats
<ul style="list-style-type: none"> • Great amount of work to be done, space for creativity. Since digitisation and online accessibility are in the beginning here, this gives space for creative approaches and innovative solutions. • Local specifics may provide interesting cases. For organisations which seek extension of their activities to Bulgaria, the local cultural materials might be very interesting and enrich their vision on the work which they are doing. 	<ul style="list-style-type: none"> • Copyright issues. The unclearness on copyright issues and how they should be approached and solved may create tension for those who do digitisation work. • Various levels of relevant experience The vision of museums, libraries and archives differ. These institutions still do not have designated digitisation units. • Small projects, scattered efforts. The danger is that small project repeat similar efforts and choose solutions where operability is not guaranteed. • Lack of crosswalks. There is no responsible body which would collect data on the standards used in different institutions, respectively there are no crosswalks which could help to build a big shared resource. • Work in conditions where neither governmental nor institutional policies are well established. In most cases this will mean that institutions will reinvent the wheel.

Table 1: SWOT analysis: digitisation in Bulgaria

6 Conclusion

The analysis reveals the basic problems that put obstacles on the way to the mass digitization in Bulgaria. At the same time it contains as well the potential possibilities to improve the situation. On this basis we have formulated following basic tasks, that in our opinion, will boost the digitization activities in the country, and will put the overall process on European level:

- Creating a joint infrastructure for the key cultural and scientific heritage institutions work;
- Establishing a common methodological network for institutions which take care for different types of heritage;

- Finding common standards for encoding and data interchange for the locally-specific features and workflows assuring quality;
- Overcoming the practice of small scale isolated initiatives and promoting a trend to structured complementary activities;
- Introducing areas such as data protection and integrity and digital curation which are currently not used in the cultural heritage sector in Bulgaria;
- Affecting the training and educational gap in the digital preservation and access field, specialists learn from their own pitfalls, not from structured programs;
- Drawing a “map” of existing resources and expertise – this will facilitate the participation in further EU initiatives.

The cultural heritage which we have inherited from the past is quite rich – over 5 million objects in Bulgaria, comparable to its 7,2 million inhabitants. In the present this heritage is underrepresented in the digital space.

To change this, serious efforts are needed in the future. We wrote this paper with the intention to mark the common lines along which the digitisation of cultural and scientific heritage is developing in Bulgaria, and to contribute to future cooperation and exchange of experiences.

We should not forget that what happens in our country is part of the general development worldwide, which currently seems well manifested as follows:

Information technology is now so pervasive and so necessary in our society that we must find ways to effectively manage its costs and its impacts across multiple organizations. The best way to do this is to forge partnerships based on a set of common requirements that individual organizations can refine to meet specific business needs and mission priorities. In terms of implementation, this can take the form of a distributed network where organizations can draw from shared knowledge and leverage a technical infrastructure while operating independently. [11]

What is said relates completely to the digitisation of cultural and scientific heritage. And yet, there is another issue which we should not forget. The production of digital objects nowadays is a gigantic industry:

IDC estimates that the world had 185 exabytes of storage available last year and will have 601 exabytes in 2010. But the amount of stuff generated is expected to jump from 161 exabytes last year to 988 exabytes (closing in on one zettabyte) in 2010. Chuck Hollis, vice-president of technology alliances at EMC Corp., the data-management company that sponsored the IDC research... said the new report made him wonder whether enough is being done to save the digital data for posterity.

"Someone has to make a decision about what to store and what not," Hollis said. "How do we preserve our heritage? Who's responsible for keeping all of this stuff around so our kids can look at it, so historians can look at it? It's not clear." [12]

Such questions are raised in the countries which are already ahead in the digitisation work and whose heritage is much better exposed in the electronic space, compared to the Bulgarian one. Yet, we still should find better ways to digitise it and make it visible.

Acknowledgements

We would like to acknowledge the support of the project DIGMAP (Discovering our Past World with Digitised Maps, Programme eContentplus – Project: ECP-2005-CULT-038042)

Notes and References

- [1] The website of i2010 is http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm; see also REDING V. *The role of libraries in the information society*, speech at the CENL Conference, Luxembourg, 29 September 2005. <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/05/566&format=HTML&aged=1&language=EN&guiLanguage=en>.

- [2] *EC Recommendation of 24 August 2006 on the digitization and online accessibility of cultural material and digital preservation*
http://europa.eu.int/information_society/activities/digital_libraries/doc/recommendation/recommendation/en.pdf
- [3] Council Conclusions on the Digitisation and Online Accessibility of Cultural Material, and Digital Preservation, (2006/C 297/01, Official Journal of the European Union, 7.12.2006, 5 pp.
http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/c_297/c_29720061207en00010005.pdf
- [4] REDING V., *Scientific Information In The Digital Age: How Accessible Should Publicly Funded Research Be?*, Closing speech, Conference on Scientific Publishing in the European Research Area Access, Dissemination and Preservation in the Digital Age, Brussels, 16 February 2007.
<http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/07/90&format=HTML&aged=0&language=EN&guiLanguage=en>
- [5] ROSS, S; DONNELLY, M.; DOBREVA, M., *Core Technologies for the Cultural and Scientific Heritage Sector* (Technology Watch Report 3), European Commission, ISBN 92-894-5276-5, 2005.
- [6] Website of SEEDI, South-Eastern European Digitisation Initiative,
<http://www.ncd.matf.bg.ac.yu/seedi/>
- [7] DEVRENI-KOUTSOUKI, A., *Electronic Presentation of Bulgarian Educational Archives: an Ontology-Based Approach*. International Journal Information Theories and Knowledge 2007, 8 pp, (to appear).
- [8] DIGMAP project website
<http://digmap.eu/>
- [9] The European Library website
<http://www.theeuropeanlibrary.org/>
- [10] The Répertoire international de la presse musicale (RIPM)
<http://www.ripm.org/>
- [11] PARDO, T. et al., *Building State Government Digital Preservation Partnerships: A Capability Assessment and Planning Toolkit*. Center for Technology in Government, University at Albany, SUNY. 2005, p. 16.
http://www.ctg.albany.edu/publications/guides/digital_preservation_partnerships/digital_preservation_partnerships.pdf
- [12] http://blogs.warwick.ac.uk/hsirhan/entry/how_much_data/