DIGITISATION OF CULTURAL HERITAGE:
BETWEEN EU PRIORITIES AND BULGARIAN REALITIES[1]

Milena Dobreva, Nikola Ikonomov

Abstract: The paper presents the Bulgarian setting in digital preservation of and access to cultural and scientific heritage. It mentions key Bulgarian institutions, which take or should take part in digitisation endeavours. It also presents examples of building and adapting specialised tools in the field, and more specifically SPWC, ACT and XEditMan.

Keywords: digital preservation of and access to cultural and scientific heritage, legislative issues, SPWC, ACT, XEditMan

ACM Classification Keywords: I.3.3. Digitizing and scanning; I.3.6. Standards; D.2.1. Methodologies

Introduction

Before the current period of economic transition, until 1990, Bulgaria was the Eastern-European country with highest expertise in information technologies within the CMEA (Council for Mutual Economic Assistance). During this time, digitisation of cultural and scientific heritage was not a separate field of work – but domains such as library automation systems, databases of cultural heritage objects; corpora of ancient and mediaeval texts had been already developed for several decades. Such work was also done in Bulgaria but not on a large-scale systematic basis.

In the years after 1990, Bulgaria has been undergoing a period of structural changes and economic transition. The acquired technological excellence in information and communication technologies (ICT) had been transferred from huge institutions to small and medium-sized enterprises functioning in a highly competitive environment. Digitisation activities, which require large investments and are not bringing quick profit, have not become attractive for the companies from the ICT sector.

Additionally, culture, education and science sectors have been suffering from inadequate funding during the transition period (the share of gross national product spent for science for example in the last years is about 0,29% which is 10 times less than in the EC). This general setting was not favourable for the establishing of national and institutional digitisation programmes.

At the same time, Bulgarian collections house over 12,500 manuscripts of Slavonic, Greek, Latin, Ottoman Turkish and other origin. Another key example is the epigraphic inscriptions from the Antiquity period, which form the third largest collection in the world following those of Italy and Greece. Precious monuments of immovable heritage, nine objects in the UNESCO World Heritage List, numerous archæological findings, Old Bulgarian runic inscriptions — all these materials are of interest not only for the local community, but also for the wider European community of which Bulgaria is a cultural part and indeed for all of mankind globally by virtue of the shared meaning of the culture of the Other. Yet, electronic information on these resources is still hardly accessible in its fullness not only to foreign experts, but also to regional and local specialists.

In this paper we will present the key types of institutions playing different roles in digitization processes with a brief comment on their actual involvement in digitisation.

---

[1] The part of this article presenting institutional involvement is an extended and updated version of the respective part of a report on Bulgaria prepared by Milena Dobreva for a MINERVAPlus Global Report (**Coordinating digitisation in Europe**. Progress report of the National Representatives Group: coordination mechanisms for digitisation policies and programmes 2004, forthcoming).

Bulgarian Institutions and Their Level of Involvement in Digitisation Activities

Six types of organisations are potentially interested in digitisation of cultural heritage: government bodies, repositories, research and/or educational institutions, companies, and foundations.

Government bodies are entrusted with the supervision of such activities.

Three institutions should play the key role for establishing digitisation policy in Bulgaria but neither one is currently working in this direction:

1. *The Ministry of Culture and Tourism.*[1] The last structural change was done very recently, in the end of January 2005 when Tourism was added to the activities of the ministry in recognition of the fact that cultural tourism will be one of the basic specialisation sectors for the Bulgarian economy in the next years.

2. *ICT Development Agency at the Ministry of Transport and Communications*[2], Republic of Bulgaria. The agency was created in 1995 and currently is the body responsible for the development of the information and communication technologies in Bulgaria. In the last few years it provided funding for projects aimed at presentation of cultural heritage in electronic form. One example is the first XML repository of catalogue descriptions of Old Bulgarian manuscripts preserved in Bulgaria – a project carried out by the Institute of Mathematics and Informatics and funded by the Agency in 2004. However, a coherent strategy has not been created and respectively followed.

3. *Ministry of Education and Science.*[3] Digitisation is not a technical activity; it develops rapidly and involves new research results.

Repositories (libraries, archives and museums), which seem the most natural initiators of digitisation projects because of the close relationship between digitisation and preservation, are currently in the position of observers due to lack of funding on the one hand, and copyright issues for digital collections, on the other hand. There are about 7000 public, university, scientific, specialised libraries and information centres in the country. As most important institutions in this group we should mention:

1. *General Department of Archives at the Council of Ministers of Republic of Bulgaria.*[4] The General Department of Archives initiated pilot work in digitisation of archival documents with the publication of the documentary CD compendium "The Independence of Bulgaria and the Bulgarian Army" containing materials from the Central Military Archive in Veliko Turnovo in 2003. The vision on digitisation activities of the State Archives was presented recently [Markov 2004].

2. *The National Library "Saint Cyril and Saint Methodius"*[5] plays a leading role in the process of expert decision-making related to measures of digital cataloguing and publishing of mediæval manuscript heritage and early printed books. Its prescriptions in these fields are adopted in other libraries in the country, which have such collections. The National Library is also the basic driving force for digital cataloguing of modern books. Although the library experts have quite extensive experience in following the current practices, real digitisation work has not been planned [Moussakova, Dipchikova 2004].

3. *The National Museum of History*[6] does not seem to be currently involved in any digitisation-related work.

---

[1] http://www.culture.government.bg/ date of last visit 25.5.2005.

[2] http://www.ict.bg/, date of last visit 25.5.2005.

[3] http://www.minedu.government.bg, date of last visit 26.5.2005.

[4] http://www.archives.government.bg/index_en.html, date of last visit 25.5.2005.

[5] http://www.nationallibrary.bg/, page does not open on 25.5.2005.

[6] http://www.historymuseum.org/, date of last visit 25.5.2005.

Research and/or educational institutions are the most active initiators of small-scale digitisation projects in Bulgaria. They usually do not have the funds and resources for running mass digitisation projects, but are the most active promoters of this field of work.

1. *The Institute of Mathematics and Informatics*[1] of the Bulgarian Academy of Sciences (IMI) plays the leading role in this direction. Digitisation of Scientific Heritage department[2] was established in IMI in 2004. The institute took part in projects related to digitisation of mathematical heritage; cataloguing and electronic publishing of mediæval Slavonic manuscripts. In addition, IMI organised in the last years three summer schools and four specialised workshops related to digitisation of cultural and scientific heritage which were targeted at Central European countries' participants and have regional impact.

   The Institute produced the most extensive XML catalogue (over 800 catalogue records) of Old Bulgarian manuscripts stored in Bulgaria [Pavlov 2004] in cooperation with specialists from the Faculty of Mathematics and Informatics of the Sofia University "Kliment Ohrdiski" and the National Library "St Cyril and St Methodius" IMI is the coordinator of the international project Knowledge Transfer for the Digitisation of Cultural and Scientific Heritage in Bulgaria (KT-DigiCult-BG), supported by the Marie Curie programme, Framework Programme 6 of the European Commission which is implemented in 2004-2008.

   The Institute is a partner in the COMTOOCI project supported by eCulture program of the EC, which is coordinated by the Institute for computational linguistics in Pisa, Italy. In this project its role is to support the localisation and local implementation of a specialised software for philological and librarian work in the cultural institutions which was developed by the Italian institute.

   IMI also works on presentation of folklore archives in digital form in cooperation with the Institute for folklore of the Bulgarian Academy of Sciences.

2. *The Institute for Bulgarian Language*[3] (IBL) works on digital preservation and use of audio archives containing live recordings presenting various Bulgarian dialects. These records originally were collected in the 50s and 60s in the 20c, and their conversion in electronic form was absolutely necessary since the original tapes started to deteriorate.

3. Amongst educational institutions we should mention The State Library Institute,[4] which recently opened specialized programme Information funds *of the cultural and scientific heritage*. Sofia University offers a general programme on Library and information activities[5].

Companies are interested in presenting sections of cultural heritage to the world, which they believe will be easily realised on the market. Today it is rather difficult to establish customer interest. The Bulgarian market for such products is unsatisfactory. This is why their main market is abroad. As an example of a company, which specializes in digitisation services, we could mention BalkanData[6] - a US-owned company based in Bulgaria. This combination seeks to offer the winning combination of the local technological and intellectual excellence and the low labour costs in the country.

Non-governmental institutions (NGOs). One active organisation in the library field is The Union of Librarians and Information Services Officers (ULISO).[7] It produced in 1997 the National Program for the preservation of Library Collections.

---

[1] www.math.bas.bg, date of last visit 25.5.2005.

[2] http://www.math.bas.bg/digi/indexbg.html, date of last visit 25.5.2005.

[3] http://www.ibl.bas.bg, date of last visit 25.5.2005.

[4] http://www.svubit.org/, date of last visit 25.5.2005.

[5] http://forum.uni-sofia.bg/filo/display.php?page=bibliotekoznanie, date of last visit 25.5.2005.

[6] http://www.balkandata.net/, date of last visit 25.5.2005.

[7] http://www.lib.bg/act.htm, date of last visit 25.5.2005.

Funding bodies (foundations) rarely support projects undertaken in the field of digitisation. In addition, the scale of their support cannot meet the real costs of serious digitisation projects. In the last years the tendency is that such bodies are supporting basically dissemination activities (workshops, conferences, trainings).

Legislative Issues

The main cultural and scientific heritage collections in Bulgaria belong to the State and their maintenance is totally dependent on the State budget. One would expect that the development of a national policy for digitisation would be an easy task when most collections of the cultural heritage are State-owned. Unfortunately, most of the legislation in the cultural sphere does not cover any digitisation aspects. A brief presentation of key legal acts covering issues, which could be approached also in digitisation programmes follow.

The Law for Protection and Development of the Culture[1] (in force since 1 January 2001) defines the basic principles and functions of the national cultural policy and the cultural institutions. However, digitisation is not mentioned amongst the issues that are covered in it.

The Deposit Law[2] (last version in force as of 1 January 2001) addresses works on digital media (electronic documents). According to it, works published on digital media should be presented in three copies to the National Library within two weeks after the publication. The National Library stores these materials as physical copies, and is not seen as a body, which would include the electronic publications into a digital library.

The Regulation for Rendering and Saving Movable Cultural Monuments[3] addresses the matters of finding, collecting, and preserving of movable cultural heritage monuments and making scientific descriptions related to them. Its application is mandatory for all museums, art galleries, museum collections as well as individuals. According to Article 62, the basic form of record and scientific description is the inventory book. The detail and accuracy of records is the responsibility of the directors of the collections. The scientific descriptions of immovable objects are presented as "Scientific passports" of the objects (Article 79). This regulation is in force since 1 January 1974. Understandably, electronic records and links between documentation of various collections were not planned in that time, but changes, which would take into account the current state of technology, have not been made.

The Regulation N 26 of 10.04.1996 of the Development, Use and Management of an Automated Information System "An Archæological map of Bulgaria"[4] seems to be the only legislative act in Bulgaria which treats a matter of digital presentation and storage of data related to the cultural heritage. It addresses the development of a specialized information system. The feeding of the database is the responsibility of the Institute of Archæology of the Bulgarian Academy of Sciences and the National Institute for Cultural Monuments based on primary data supplied from specialists who worked *in situ*. Information can be obtained from this automated system only on the basis of a written request for a service fee. The collection of data and their use were adequate for the state of the technologies in 1996; now this is outdated but changes to adapt the collected data and to provide access via the Internet have not been done.

The Tariff of rates collected by State Cultural Institutions for Services and Provision of Documents and Copies[5], date of last update 5 January 2001 does not include any fees related to digital images despite the actuality of the update.

---

[1] http://www.culture.government.bg/docdetail.html?id=16, in Bulgarian, date of last visit 25.5.2005.

[2] http://www.culture.government.bg/docdetail.html?id=66, in Bulgarian, date of last visit 25.5.2005.

[3] http://www.culture.government.bg/docdetail.html?id=49, in Bulgarian, date of last visit 25.5.2005.

[4] http://www.culture.government.bg/docdetail.html?id=48, in Bulgarian, date of last visit 25.5.2005.

[5] http://www.culture.government.bg/docdetail.html?id=38, in Bulgarian, date of last visit 25.5.2005.

EU Cooperation and Current EU Priorities

Bulgarian institutions are active in searching for international cooperation possibilities. Within the trend of Digital culture (Access to and preservation of cultural heritage) in FP6, we can mention the following projects where Bulgarian institutions participate as members:

- CALIMERA (participant ULISO)
- EPOCH (participant New Bulgarian University)
- MINERVAPLUS (participant – IMI-BAS as an associated member)
- PRESTOSPACE (participant Sirma AI Ltd)
- KT-DigiCult-BG is a project coordinated by IMI-BAS.

IMI-BAS was an initiator of the creation of the South-Eastern European Network for Digitisation of Scientific and Cultural Heritage[1], constituted with the signing of the Borovets declaration of 17 September 2003.

The current priorities under IST 2.5.10 (Access to and preservation of cultural and scientific resources) are targeted to:

- Enriched conceptual representations
- Advanced access methods
- Long-term preservationThe presentation of current Bulgarian setting in the previous sections shows that some Bulgarian institutions are trying to be in line with current developments.

Development of Local Tools vs. Adaptation of Existing Tools

In the digitisation work one crucial matter is what tools will be applied for the practical work. In the last years IMI gained experience in two approaches: *developing local tools* for support of specific task and *localisation of existing platforms* to the Bulgarian environment.

As a home-made tool we could mention XEditMan [Pavlov 2004]. This is a tool, which combines an editor and visualisation component for preparing and studying manuscript description of mediaeval manuscripts. Its interface is in Bulgarian and follows the local practices in cataloguing work. The descriptions of manuscripts are produced in XML format following the TEI P4 guidelines.

The experience with development and use of this tool is very positive, since it supports the performance of a specific task and facilitates the preparation of large amount of data in digital form.

As an example of localised tool, we could mention SPWC, Software Platform for archivist, librarian and philological Work in Cultural Institution. The platform offers a set of tools for document management in cultural institution including digitizing, cataloguing and transcription of primary sources. ILC are active in building specialized workstations for philological work for decades [Bozzi, Corradini 2004] and constantly improve the capabilities and the spread of use of their specialized tools. IMI worked on localisation of the software (translation of the user interface and documentation in Bulgarian), identification of experimental materials (in the case of Bulgaria this are local DTDs – manuscript and archival records), and training and dissemination activities. All abovementioned endeavours were in the frame of the project COMTOOCI (COMputational TOOls for the librarian and philological work in Cultural Institution). The project, supported by the CULTURE 2000 program and coordinated by the Institute for Computational Linguistics (ILC) – Pisa, Italy started in September 2004. In the forthcoming months a pilot installation of SPWC in the General Department of Archives will be done.

---

[1] http://www.ncd.matf.bg.ac.yu/?page=news&lang=en&file=declaration.htm, date of last visit 25.5.2005.

At the same time, we are studying the possible application of ACT [Ribarov 2004]. This software combines the presentation of manuscript images and annotated texts. Because of the high level of variety in mediaeval Slavonic manuscripts, the author chose an approach where previous human annotation of word forms is used in subsequent annotation activities.

We present a comparison of the features of ACT and SPWC in Table 1.

FEATURES	ACT	SPWC
Image analysis module		✓
Image representation module	✓	✓
Cataloguing module		✓
Representation of texts	✓	✓
Representation of variants	✓	✓
Multilingual support	✓	✓
Interface in different languages		✓
Annotation of various levels, up to morphology	✓	
Morphological annotation supported on 'learning by example' basis	✓	

Conclusion

Under the described lack of national policy, the various institutions in the cultural and scientific heritage sector have the freedom to design their own policies. Unfortunately, this is combined with lack of methodological, financial, technological and human resources support. On this setting, the Digitisation of Scientific Heritage Department at IMI has as a core part of its mission to contribute to the improvement of human resources qualification and support memory institutions through joint activities, which would lead to a difference in the future.

Bibliography

[Bozzi, Corradini 2004] Bozzi A., Corradini M.S. (2004). Aspects and methods of computer-aided textual criticism. In Digital Technology and Philological Disciplines. *Linguistica Computazionale* XX-XXI, (2004). Pisa-Roma, IEPI. 49-66.

[Markov 2004] Markov N., National Archives. In: International Journal Information Theories and Applications (special issue: Proceedings of the International Seminar Digitisation of Cultural and Scientific Heritage, Bansko, 27 August — 3 September 2004). Vol. 11 (2004), No. 3, pp. 282-283.

[Moussakova, Dipchikova 2004] Moussakova E, A. Dipchikova, The Role of the National Library in Preserving National Written Heritage. In: International Journal Information Theories and Applications (special issue: Proceedings of the International Seminar Digitisation of Cultural and Scientific Heritage, Bansko, 27 August — 3 September 2004). Vol. 11 (2004), No. 3, pp. 284-287.

[Pavlov 2004] Pavlov P., XML Presentation of Catalogue Data on Medieval Slavonic Manuscripts: Experience and Perspectives, In: Proceedings of the 33rd Conference of the Union of Bulgarian Mathematicians, Borovets, 1 – 4 April 2004, pp.236–240.

[Ribarov 2004] K. Ribarov *et al.*, "We present the ACT Tool". In: Scripta & e-Scripta, Volume 2, Institute of Literature, Bulgarian Academy of Sciences, Sofia. 2004.

Authors' Information

Milena Dobreva — Chair of the Department *Digitisation of Scientific Heritage*, Institute of Mathematics and Informatics, BAS, Acad. G. Bonchev St., bl. 8, Sofia-1113, Bulgaria, e-mail: dobreva@math.bas.bg

Nikola Ikonomov — Chair of Laboratory on Phonetics and Speech Communication, Institute for Bulgarian Language, BAS, Shipchenski prohod 52, Sofia-1113, Bulgaria, e-mail: nikonomov@ibl.bas.bg.