DIGITAL PRESERVATION AND ACCESS TO CULTURAL AND SCIENTIFIC HERITAGE: PRESENTATION OF THE KT-DIGICULT-BG PROJECT

Milena Dobreva and Nikola Ikonomov

Abstract: The fast development and wide application of digital methods, combined with broadened access to the Internet and falling computing costs, have created intense interest in electronic presentation and access to cultural and scientific heritage resources. Information technologies have offered cultural institutions new opportunities for the presentation of their holdings, which are now made accessible not only to the specialists, but also to the citizens and interested parties worldwide.

The paper presents an overview of the Bulgarian experience in the field of digital preservation and access and on-going work on the project "Knowledge Transfer for the Digitisation of Scientific and Cultural Heritage to Bulgaria" (MTKD-CT-2004-509754) supported by the Marie Curie programme of the FP6 of the EC.

Keywords: digitisation, cultural and scientific heritage.

Introduction

The field of digitisation of cultural and scientific heritage is within the priority areas for the European Union as an inevitable part of the understanding of our common grounds and local similarities and diversities. While in the previous decades the exposure and discussion over scientific and cultural heritage were a privilege to small scientific communities, now this heritage reveals greater interest and can be exposed easily to the general public with the help of modern information technologies.

This is an important development stimulus. However, the practical work in this field requires substantial efforts and much specialised expertise both to prepare the resources in electronic form and to present them properly to various audiences. The project presented in this paper aims to remove currently existing research and practical work gap for Bulgaria. It in fact serves as a pilot effort for a country within the Balkan region, and has all chances to multiply the transfer knowledge via know-how transfer to other countries with similar needs.

We witness for over one decade now that while the interest to methods and tools for presentation of cultural heritage in Bulgaria grows, the real efforts usually end up with small demonstration projects. To move the current state from this point, we applied to the Marie Curie programme (action Transfer of Knowledge), which enables us to bring on a regular basis experienced researchers to the Bulgarian host (Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences). The design of the programme should contribute to the development of synergies between various institutions, both on national and international scope. The ratio of experienced and more experienced researchers is 1:1 and this actually should help to avoid the generation gaps which cover the field in an unbalanced way concerning experience and practical knowledge.

Even in the countries from the European Union, the expertise of specialists working in the field of digitisation is gained mostly by individual practice. When a young researcher has the chance to become a part of an experienced group, he/she would gain necessary skills to work on a good quality. Otherwise, many institutions establish small-scale projects learning basically from pitfalls. This tendency is not a positive one, because it results in the production of scattered resources which could not be interconnected.

Overview of the Bulgarian Experience

Collections in Bulgaria

In Bulgarian repositories over 12,500 manuscripts of Slavonic, Greek, Latin, Islamic and other origin are preserved. Bulgarian institutions also keep the third largest collection in the world (following Italy and Greece) of

epigraphic inscriptions from Antiquity. It is obvious that such materials are of interest not only for the local community but also on a European and global scale.

This set of resources is still hardly accessible in its fullness not only to foreign experts but also to regional experts. Electronic cataloguing and digital preservation are still not popular in the region.

Brief History of Digitisation Initiatives

The first initiatives in this field were launched about 15 years ago by research institutions and companies. A national strategy and funding for digitisation programmes has not been available since the very beginning of works. Actual activities on digitisation have not been done on large-scale basis yet.

The first field where computers were implemented in the late 80s and early 90s of the last century was *cataloguing*. The ISIS library cataloguing software was introduced in Bulgaria and tested in the National Library "St. Cyril and St. Methodius". However, major cataloguing effort has not been started in that time, because of the limitations of the model. Already in this first introduction of computers the expectations that the new technologies would assist the work of specialists in medieval manuscripts, had not overcome the fact that these specialists did not like the model built in the system.

A next effort in the field of electronic presentation of manuscript data was the project called The Repertorium of Old Bulgarian Literature and Letters. It started as "an archival repository capable of encoding and preserving in SGML (and, subsequently, XML) format of archeographic, palæographic, codicological, textological, and literary-historical data concerning original and translated medieval texts represented in Balkan Cyrillic manuscripts" [Repertorium]. The project grew out of an initiative of D. Birnbaum (University of Pittsburgh), A. Bojadžiev (University of Sofia), M. Dobreva (Institute of Mathematics and Informatics, Bulgarian Academy of Sciences), and A. Miltenova (Institute of Literature, Bulgarian Academy of Sciences) in 1994. The computer model is discussed in [Dobreva 00]. Currently there are 300 manuscript descriptions available in the Institute of Literature of the Bulgarian Academy of Sciences. The basic publication on this endeavour is [MB 00].

Adoption of MASTER-Manuscript Access through Standards for Electronic Records for Cataloguing of Old Bulgarian Manuscripts. MASTER was a European project funded under the Framework IV Telematics for Libraries programme [MASTER]. It developed a TEI[1]-conformant DTD[2] for medieval manuscripts with the ambition to serve the needs of all repositories in Europe, and software for making and visualising records on manuscripts. The MASTER standard (with some small revisions) was adopted by the TEI in 2003. The National Library "St. Cyril and St. Methodius" and the Institute of Mathematics and Informatics joined MASTER as associated members. They made 30 descriptions of manuscripts with two basic aims: to test descriptions providing data both in English and Bulgarian, and to supply examples coming from the non-Latin literary tradition.

Although the *Repertorium* and *Master* project are oriented to the same standard framework, the models underlying both projects are not similar (i.e. they contain different sets of elements structured in different ways).

In addition to these cataloguing efforts, some companies created CD-ROMs (four CD-ROMs already exist, two of manuscripts from the National Library "St. Cyril and St. Methodius", one of Macedonian coins and one of Bulgarian Iconography).

Beyond the manuscript field, an interesting project was launched to present one of the most interesting historic buildings in Bulgaria – the Boyana Church in 2000-2002 [Boyana]. This is the first computer-based 3D model of a Bulgarian monument of culture. Author of this project is Trifon A. Trifonov.

Experience of IMI-BAS

Specialists from IMI-BAS took part in the Repertorium and MASTER project mentioned above thus gaining expertise in presentation of medieval written heritage.

---

[1] Text Encoding Initiative, see [TEI].
[2] Document Type Definition

In addition, one of the obvious interests of this community is related to **digitisation of mathematical heritage**. The Institute of Mathematics and Informatics has been hosting a growing group of people accomplishing various related activities with mathematical texts. These activities include the whole preprint process of the key Bulgarian mathematical journals:

– Serdica Mathematical Journal Pliska Studia Mathematica Bulgarica Proceedings of the Annual Conference of the Union of the Bulgarian Mathematicians and also of the interdisciplinary review

– Comptes rendus de l'Académie Bulgare des Sciences In collaboration with Lefkowitz & Co.

The group took part in the digitisation of the *Jahrbuch Für Mathematic*. Other fields of recent work of specialists from IMI include **study of chronological distribution of historical artefacts**, **edutainment**, **encoding of Early Cyrillic and Glagolitic alphabets**, **presentation of immovable heritage**.

In addition, IMI hosted or contributed to the organisation of a number of international events related to digitisation of cultural heritage. This places it in a very important position for **disseminating and raising awareness activities**. Here we should list The First International *Conference Computer Processing of Medieval Slavonic Manuscripts* was held in Blagoevgrad in 1995 (see [BBDM 95]). The UNESCO workshop *Text Variety in the Witnesses of Medieval Texts* was held in Sofia in 1997 (see [Dobreva 98]). The Institute of Mathematics and Informatics organised a series of summer schools in 1998, 1999 and 2002 aimed at enlarging the international community of young specialists who would share their experience, and form co-operation ties, as well as a workshop on Digitisation of Cultural Heritage within the frameworks of the 1st International Congress of the Mathematical Society of South-Eastern Europe [NCD 2004].

Some Conclusions

From this brief presentation of digitisation-related activities, it is obvious that the major field where experience was gained is electronic cataloguing of manuscripts. Yet, we cannot speak about a complete consensus between different teams, using separate catalogue descriptions. This fact causes diversity in approaches, but it also illustrates the lack of efforts for integration, which on the long term run leads to dissolved results. Such disagreement is also contradictory to the European recommendations expressed in the Lund principles.

The KT-DigiCult-BG Project

The fast development and wide application of digital methods, combined with broadened access to the Internet and falling computing costs, have created intense interest in electronic presentation and access to cultural and scientific heritage resources: original manuscripts, early printed books, epigraphic inscriptions, etc. Information technologies (IT) have offered cultural institutions new opportunities for the electronic presentation of their holdings, which are now made accessible not only to the specialists, but also to the citizens and interested parties worldwide.

On this setting, the electronic resources available for the Slavonic countries, first of which became members of the European Union in 2004, are still scarce. In a small country like Bulgaria it is impossible, due to the lack of specialists, and economically not efficient to form digitisation groups attached to the various cultural and scientific heritage institutions. The work on this project will strengthen the experience gained by the Institute of Mathematics and Informatics in the digitisation field and develop it further through knowledge acquisition and transfer measures. Thus the Institute will develop as a national centre of best practice in the field, and will be able to support on-going initiatives in the digitisation sphere.

The basic activities envisaged in the project will contribute to change the current state in the digitisation field in Bulgaria through:

1. Designing an integrated approach for the presentation of the material, which is large in volume, rich in language variation and multimodal as a computer presentation;

2. Implementing IT framework suitable for appropriate presentation of the local cultural heritage within the European electronic space;

3. Establishing the bases for a cost-effective and fast semi-automatic and automatic content-sensitive annotation of the word mass of the written cultural sources;

4. Integrating the experience of EC partners and accession countries, thus decreasing the gap between the state-of-the-art and real work in Bulgaria and the rest of Europe.

Digitisation field is combining knowledge from several different specialised fields (the project considers digitisation in the broad sense, including presentation of a variety of data on a cultural artefact in a computer form, not just digital imaging, i.e. texts, structured data, audio, etc.). To fulfil project goals, the host will cooperate with partners who had gained expertise in different fields. This will help to build a well-balanced team within the host which is not concentrating on one of the problems in the field, but is able to approach it in creative way, taking into account the methods and techniques applied for digital image processing, digital document archiving and cataloguing, information retrieval, distributed systems, classical and historical lexicography, encoding and document type definitions.

Providing digital access to cultural heritage has an important influence on the actual preservation of the originals. To these economic and societal impacts of digitisation we could add the effect of improved visibility of cultural artefacts for the citizens. The profound comprehension of the current settings and historical reasons for the present status quo lays in the better understanding of where our sources are stated in the Lund principles of 2001.

Our project fits most to the following major trends envisaged in the Lund action plan (ordered according to their importance):

– Action 4b: *Sustainable access to content* – it will be assured through the framework offered.

– Action 3a: *Good practice examples and guidelines* – they are per se incorporated into KT-DigiCult-BG and disseminated widely through the manuals which will be created.

– Action 3b: *Competence centres* – the host organisation from Bulgaria will have the real opportunity to grow as such centres which will 'spread the word' further. In addition, the close ties with partners will provide to them valuable feedback and probably would boost their own development as competence centres.

– Action 1d: *Supporting coordination activities* – the project defines a stable inter-cooperation, a core of a network, which did not exist in the past in this field.

Through the KT-DigiCult-BG project we intend to overcome well-known barriers present in the associated countries community and identified in the Lund Principles:

– *Fragmentation of approach* – this is what we have been witnessing in the last decade in Bulgaria, and the project will contribute to change this tendency;

– *Obsolescence* – we are targeting at learning from the best experience in the EC, which guarantees the state-of-the-art in this transfer of knowledge project;

– *Lack of simple, common forms of access for the citizen* – this is currently a fact for the Slavonic heritage, and also for the Bulgarian environment;

– *Data Protection and Intellectual property rights* – they are recognised in our effort and could serve as a real life example in the future endeavours in this field.

Our strongest contribution is in the following areas endorsed by the European experts in the Lund Principles:

– *An accessible and sustainable heritage* – through developing state-of-the-art framework and tools which would place in the e-European space the mediæval Slavonic written heritage which is now available as small-size and scattered resources;

– *Support for cultural diversity, education and content industries* – by exposing a cultural heritage, which is now missing in the electronic space;

– *Digitised resources of great variety and richness* – adding to the existing resources one more significant group.

This project attracts as participants key organisations from EC member countries (Charles University Prague, Trinity College Dublin, Copenhagen University, Institute of Informatics and Telecommunications at NCSR Demokritos Institute in Athens). Some of them already cooperated, basically in training activities, which contributed to raise the awareness on the importance of the digitisation of cultural heritage in Bulgaria. Therefore, KT-DigiCult-BG improves the maturity and the high quality approach in the filed. The previous experience contributes to transfer of knowledge of highest possible quality. Up till now, specialists in Bulgaria mastered in this field if they had the chance to work with a leading specialist abroad, or devised their knowledge from own experience. The project gives a chance in future to young people to receive structured and well-balanced theoretical and practical framework and will boost real work in the field. The development of local centres of such high quality in Bulgaria is one of the measures to prevent brain-drain.

Basic fields of work which will be supported through visits of incoming researchers include but are not limited to:

– General methodology and practical setting for digitisation of cultural and scientific heritage.
– Digitisation of medieval manuscripts (incl. digital imaging, cataloguing, text representation, electronic publishing).
– Digitisation of mathematical texts and building digital mathematical library of works of Bulgarian mathematicians.
– Virtual reality applications for presentation of immovable cultural heritage.
– Audio archives: methods for digitisation and restoration.
– Application of quantitative methods for the study of data related to the cultural heritage.
– Applications of edutainment to cultural heritage studies.

During the first project year, incoming researchers included Dr. Matthew Driscoll from Copenhagen University who worked together with project team members on an XML editor for cataloguing mediaeval Bulgarian manuscripts; Boris Shishkov from Delft University of Technology, the Netherlands, who will develop an electronic brokerage system for sites presenting cultural and scientific heritage, and Philip Zrantchev from the University of Reading, UK, who develops an Old Cyrillic UNICODE font based on Codex Suprasliensis script.

Thus our project already contributes to improve the quality and contents of future national and international work in digitisation of cultural and scientific heritage by very intensive and well-designed transfer of knowledge, which not only brings knowledge to Bulgaria, but is aimed at its best integration according to the local needs. The project will have important impact on the future developments of electronic presentation/publishing/preservation of cultural heritage in Bulgaria, but also will serve as an example for future work in countries with similar economic and cultural settings.

Conclusion

In the period between the submission of the project and its start the priorities in the field of digital preservation of and access to cultural and scientific heritage resources of the Information Society Technologies thematic area of FP6 changed. Dealing with complex and dynamic objects and new knowledge technologies, visualisation and virtual reality are now in the focus of EC support.

On this setting, Bulgaria still has to solve numerous problems related to making available cultural and scientific heritage resources in digital form. We hope that one of the feasible outcomes of this project will be to produce resources in digital form at least in the field of mathematical heritage and archival collections.

Bibliography

[Boyana] http://www.boyanachurch.org/ – website of the Boyana Church

[BBDM 95] Birnbaum, D., A. Bojadjiev, M. Dobreva, A. Miltenova, ed. *Computer Processing of Medieval Slavic Mauscripts. Proceedings of the First International Conference.* 24–28 July 1995. Blagoevgrad, Bulgaria. Sofia: Professor Marin Drinov Academic Publishing House. 1995. ISBN 954-430-417-7.

[Dobreva 98] Dobreva, M. (Ed). *Text Variety in the Witnesses of Medieval Texts. Proceedings of the International Workshop.* Institute of Mathematics and Informatics. Sofia, 21–23 September, 1997. Sofia: Institute of Mathematics and Informatics. 1998. ISBN 954-9650-02-2.

[Dobreva 00] M. Dobreva, *A Repertory of the Old Bulgarian Literature: Problems Concerning the Design and Use of a Computer Supported Model*, In: A. Miltenova, D. Birnbaum (eds.), *Medieval Slavic Manuscripts and SGML: Problems and Perspectives*, Sofia, Academic Publishing House, 2000, pp.. 91-98.

[Lund Principles 01] http://www.cordis.lu/ist/ka3/digicult/lund_principles.htm – *eEurope: creating cooperation for digitisation (Lund Principles)*

[MASTER] http://www.cta.dmu.ac.uk/projects/master/, website of the MASTER project.

[MB 00]A. Miltenova, D. Birnbaum (eds.), *Medieval Slavic Manuscripts and SGML: Problems and Perspectives*, Sofia, Academic Publishing House, 2000, 372 pp.

[NCD 2004] Review of the National Center for Digitisation, No 4(2004), Serbia and Montenegro, a special issue with papers presented at the minisymposium Digitisation of Cultural Heritage, Borovets, 2003.

[Ognjanović 02] *National Center for Digitization,* In: *Review of the National Center for Digitization*, 1/2002.

[Repertorium] http://clover.slavic.pitt.edu/~repertorium/index.html – website of the *Repertorium of Old Bulgarian Literature and Letters*

[Ross et al. 03] S. Ross, M. Donnelly, M. Dobreva, *New Technologies for the Cultural and Scientific Heritage Sector* (DigiCULT, Technology Watch Report 1), European Commission, 2003, 196 pp. ISBN 92-894-5275-7.

[TEI] http://www.tei-c.org/ – *Text Encoding Initiative Website*

[Tutorial] http://www.library.cornell.edu/preservation/tutorial/ – *Moving Theory into Practice: Digital Imaging Tutorial*

Web resources

http://palimpsest.stanford.edu/ – CoOL, Conservation OnLine, incl. http://palimpsest.stanford.edu/bytopic/imaging/ – Digital Imaging

http://www.bl.uk/gabriel/services/lists_generated/services_digital_en.html – Gabriel (links to collections of Europe's national libraries that have been digitised)

http://www.hatii.arts.gla.ac.uk/SumProg/DigiSS03/urls.htm#DigiHerAssets – The Humanities Advanced Technology and Information Institute, University of Glasgow, Links to digitisation resources and sites

http://www.isos.dcu.ie – Irish Script on Screen (ISOS)

http://www.kb.nl/kb/resources/frameset_kb.html?/kb/sbo/digi/verhanen.html – Research and development of electronic access to Medieval Illuminated Manuscripts from the Koninklijke Bibliotheek, The Netherlands

http://www.memss.arts.gla.ac.uk/ – The Digitisation of Middle English Manuscripts

Authors′ Information

Milena Dobreva — Chair of Dept. on 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.