Digital Preservation and Access to Cultural and Scientific Heritage: Presentation of the KT-DigiCult-Bg project

Dr. Nikola IKONOMOV
Institute for Bulgarian Language, Bulgarian Academy of Sciences, Sofia, Bulgaria
nikonomov@ibl.bas.bg

Dr. Milena DOBREVA
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
dobreva@ufal.ms.mff.cuni.cz

Abstract. *The paper presents the Bulgarian experience in the field of digital preservation and access. It provides a general overview of the project "Knowledge Transfer for the Digitization of Cultural and Scientific Heritage to Bulgaria" supported by the Marie Curie Program of the FP6 of the EC. Some examples and practical results achieved during the first project year are reported. Considerations for the future work and directions of research are also presented.*

Introduction

Digitisation of cultural and scientific heritage is among the EU priorities for two basic reasons. On one hand this heritage consists of tangible and intangible values, accumulated during the centuries, and which can remain beyond time only if we take now the responsibility for their preservation. It is necessary to prepare beforehand physical copies of the most valuable products of human intellect in order to save them for the future generations. On the other hand, the new information technologies imply new understanding of access to cultural and scientific heritage material. While new possibilities to present this heritage appear and improve, many regions of the world are underrepresented in the electronic space. In this respect, South-Eastern Europe is a typical example: the cultural and scientific heritage available in this region is not represented proportionally in the electronic space. This situation was recently discussed in a regional UNESCO meeting on digitisation policy for South-Eastern Europe (14-17 February, Ohrid, FYROM).

On the other hand before the recent emergence of the Digital Age, characterized and typified by the World Wide Web, cultural heritage was available only for a restricted community. Nowadays information technologies can offer an easy access of this heritage to the general public, making it part of the common world cultural space.

Digitisation of cultural and scientific heritage is not an easy task. It requires significant efforts, specialised know-how and considerable financial resources. In small countries like Bulgaria, despite the growing interest to methods and tools for presentation of cultural heritage in the last decade, the results are far from being satisfactory. The current situation can be briefly

characterized by scattered experience, small demonstration projects and lack of mass digitisation. 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 basic aim of the project presented in this paper is to overcome the currently existing research and practical work gap for the country. It can be considered also as a pilot project within the Balkan region, and has all chances to provide know-how transfer to other countries with similar problems.

Previous Bulgarian Experience

Bulgarian repositories store over 12,500 manuscripts of Slavonic, Greek, Latin, Islamic and other origin. Bulgarian cultural institutions also keep the third largest collection in the world (following Italy and Greece) of epigraphic inscriptions from Late Antiquity. Such materials are of interest not only for the local community but also on a European and global scale.

First initiatives in the field were launched about 15 years ago by research institutions and companies and consisted basically of cataloguing activities. The *ISIS* library cataloguing software was introduced in the National Library "St. Cyril and St. Methodius". However, major cataloguing effort has not been started, due to limitations of the model. A next step was a project called *The Repertorium of Old Bulgarian Literature and Letters*. It was aimed at the creation of an archival repository capable of encoding and preserving archeographic, palæographic, codicological, textological, and literary-historical data related to original and translated medieval texts in Cyrillic manuscripts. Currently there are 300 descriptions available in the Institute of Literature of the Bulgarian Academy of Sciences [Repertorium]. In 2003 the National Library "St. Cyril and St. Methodius" and the Institute of Mathematics and Informatics joined MASTER (*Manuscript Access through Standards for Electronic Records)* as associated members [MASTER]. As a result 30 descriptions of manuscripts were prepared in order to test descriptions providing data both in English and Bulgarian, and to supply examples coming from a non-Latin literary tradition.

In addition to these cataloguing efforts, some companies created CD-ROMs—two of manuscripts from the National Library "St. Cyril and St. Methodius", one of Macedonian coins and one of Bulgarian Iconography. It is worth mentioning that in 2000-2002 an interesting project was launched in order to present one of the most fascinating historic buildings in Bulgaria—the Boyana Church [Boyana]. This was the first computer-based 3D model of a Bulgarian monument of culture. In addition, Bulgaria hosted a number of international events related to digitisation of cultural heritage.

The KT-DigiCult-Bg project

The project *Knowledge Transfer for the Digitisation of Cultural and Scientific Heritage in Bulgaria* (KT-DigiCULT-BG) is a four year project supported by the Marie Curie programme. It is coordinated by the Institute of Mathematics and Informatics (Bulgaria). Project partners are Det Arnamagnæanske Institut (Köbenhavns Universitet, Denmark), Trinity College (Dublin, Ireland), Charles University (Prague, Czech Republic), and the Institute of Informatics and

Telecommunications, National Centre for Scientific Research "Demokritos" (Athens, Greece). The project is a typical Transfer of Knowledge action, based on researchers' visits to the Bulgarian host (Institute of Mathematics and Informatics) and to the partners' institutions. Thus the development of international synergies between these institutions can be expected.

The project work plan addresses identification of key research topics, forums for sharing policies and best practice, usability requirements, dissemination and training, as well as management, coordination and evaluation. It is structured as two-phase development scheme. In the knowledge acquisition phase, state-of the-art knowledge in digitisation of cultural and scientific heritage is acquired, including general methodology, digital imaging, text corpora annotation, cataloguing systems and web publishing. The second phase, knowledge transfer envisages visits to the host. It is structured in two steps: direct guidance and consolidation of the transferred knowledge so that the host institution starts serving as a national digitisation centre in Bulgaria offering methodological, educational and practical assistance to organisations starting digitisation projects.

The ongoing digitisation-related endeavours in Bulgaria were not systematic, but adapted to the needs of the preservation, conservation, and presentation of the national cultural heritage or to the increased users requests for digital materials. The basic goals of the project are to find the best ways to systematize all those actual digitisation processes, to apply new international standards, to prepare recommendations on the best metadata form from the list of recommended standards, and to select what content to be digitised. At the same time the project should offer ways to overcome the actual problems of the digitisation in the country, such as the lack of the necessary equipment, the lack of educated staff, the lack of understanding within the national and local authorities, the lack of adequate tools and best practices for the management of those new digital collections, and the unsatisfactory coordination of activities at national level.

In a small country like Bulgaria it is impossible, due to the lack of specialists, and economically not efficient to build digitisation groups belonging 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 mature as a national centre of best practice in the field, and will be able to support on-going initiatives and to co-ordinate various national activities in the digitisation sphere.

Basic fields of work which include but are not limited to:
- General methodology and practical setting for digitisation of cultural and scientific heritage.
- Digitisation of mediaeval 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.

The KT-DigiCULT-BG project started one year ago. After the first project year, following basic results can be reported:

- The Institute of Mathematics and Informatics started the digitisation of its scientific heritage. The activities included scanning of mathematical texts and journals and subsequent hosting of the images on a dedicated server. In order to improve the coordination of the works, a new Department of Digitisation of Scientific Heritage was created.
- A compact disk "Sofia. Religious spaces" was created. It was prepared and realized by a team from the Sofia State Archives at the General Department of Archives at the Council of Ministers and specialists from the new Department of Digitisation of Scientific Heritage at the Institute of Mathematics and Informatics. The CD presents archives and photographs taken from various funds kept at the Sofia State Archives. At the same time different objects were also presented: paintings, photographs, cult objects, canonical and dogmatic books kindly provided by various museums and institutions. The creation of the CD can be considered as a pilot project of the two institutions, which have signed recently a contract for joint activities in the digitisation of archival collections.
- An Old Cyrillic UNICODE font based on Codex Suprasliensis script was designed and should be implemented during the second project year.
- A specialized editor XEditMan (XML Editor for Manuscript Data) which is an XML-oriented tool for editing and browsing catalogue descriptions of mediaeval manuscripts was developed [Pavlov 2004]. It offers a friendly interface for entering data on mediaeval manuscripts; visualisation and queries to the descriptions already available. The descriptions are compatible with the document type definition (DTD) structure suggested by the project MASTER (Manuscript Access through Standards for Electronic Records) and adopted by the Text Encoding Initiative. During the data entry the elements which are filled in appear in a sequence which is adopted in the manuscript cataloguing practice. The interface is in Bulgarian and this facilitates preparing electronic descriptions by people who are not acquainted in detail with the DTD structure. The tool can be used also for visualization of single descriptions in two modes: complete descriptions or user-selected group of elements. Comparative study of multiple descriptions is achieved through database queries. Currently, a collection of 800 descriptions on Mediaeval Bulgarian manuscripts is already prepared.
- As a result of a successful visit at the Demokritos Institute in Athens, Greece, a methodology of using wavelets for character recognition was elaborated and implemented.

Conclusions

The rich heritage of Bulgaria is still expecting to take its legitimate place in the world cultural space. We hope that through the project we will raise the interest of scientific and cultural heritage institutions in the country and will push them to start real work.

At the same time the project gives opportunities to broaden the international cooperation in the field of digitisation of cultural and scientific heritage. The acquired experience will boost the implementation of the ambitious goals, set by the project. We highly appreciate the experience of other small countries in Europe, and in particular of institutions from the Czech Republic. The similarity of territory and population, the richness of the cultural and scientific heritage, and the common historical fate in the last century rise comparable problems. The way they are or have been solved in the Czech Republic and the visible progress in the field is a certain sign for a successful Bulgarian digitisation.

References

[Boyana]   http://www.boyanachurch.org/—website of the Boyana Church, date of last visit 24 April 2005.

[MASTER]   http://www.cta.dmu.ac.uk/projects/master/—website of the MASTER project, date of last visit 24 April 2005.

[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.

[Repertorium]   http://clover.slavic.pitt.edu/~repertorium/index.html—website of the *Repertorium of Old Bulgarian Literature and Letters*, date of last visit 24 April 2005.