New Publishing Models for the Slavonic Mediaeval Manuscripts

M. Dobreva, K. Sotirova, and A. Sameva

Institute of Mathematics and Computer Science
bl. 8, Acad. G. Bonchev St., 1113 Sofia, Bulgaria

Abstract. The paper presents the experience of our team in using methods for electronic presentation of mediaeval manuscripts in Bulgaria.
The advantages of the electronic edition are:

- *providing high quality* of the images
- *facilitating and widening the access* (Slavonic manuscripts and early printed books are spread in different libraries in Europe and former USSR. Because of the scattered holdings of the manu-scripts, danger of additional harm, and specific conservation conditions of the fragile and valuable manuscripts, the access to them is strictly limited and sometimes even impossible for specialists).
- *providing presentation which can be used for comparative research and analysis*
- *presenting material for distance learning*

Therefore the electronic presentation of the cultural heritage is the only way to make it accessible from everywhere without being harmed. We hope that the experience from this project will will boost future endeavours.

1 Introduction: One Example

The latest computer technologies give new opportunities for presentation and research of the mediaeval cultural heritage of Bulgaria, one of the oldest Europe countries, which hosts 8 500 manuscripts and over 35 000 early printed books.

The company Verbatim made a 44 thousand compact disks donation to the Russian State Library last year. This amount of CDs will be sufficient for the work of specialists working on electronic archive in the next three-years. About 250 specialists of the library are involved in scanning of the holdings of the library,

which are not a subject of copyright. Other 180 people are processing the digital images. Thus, 500 to 1000 books are being digitized in one-month term [1].

This is a typical example of a current large-scale project in the field of digitization of a library collection. It illustrates several important questions: what are the expected outcomes, how the project is organized, and how the copyright issues are approached.

The expected outcomes are thousands of compact disks containing images. The organization of this extensive digital archive is an interesting task on its own. These images could be consulted by specialists instead of the originals, and could be made available for the general public on the Internet or as a CD edition. Nothing is said about any effort to transcribe the texts (although image quality is a subject of a special concern).

This brief information presents a specific approach to project organization. It illustrates how many people could be needed in a collection to be digitized and a speed of work (the framework here is quite vague, and instead of being measured in pages, gives a rough estimation of 500 to 1000 books).

The copyright issues are put aside in this particular case, targeting the project to books, which are not a subject of copyright.

In addition this example demonstrates that large-scale projects in this field often rely on external support from sources different from cultural heritage repositories.

Although the large scale of work is clear, the digitized collection should undergo several further steps before being effectively used:

1. Embedding text transcription into the electronic presentation (to make the collection usable by specialists who are not familiar with the original language of the publication, translation to other languages might be considered).
2. Making the digital archive searchable under various criteria.
3. Making the digital records available via Internet services (as a complete content, or at least as a set of bibliographic descriptions).

The example discussed above is not directed to mediaeval texts, which are even in a greater need to be digitized. This need is caused by the restricted access to originals and small amount of published mediaeval manuscripts.

The work with old texts requires greater effort because the presentation of texts is not so easy.

2   Necessity of work in the field in Central and Eastern Europe and local specifics

2.1   Overview

The Central and Eastern European (CEE) region is extremely rich in cultural heritage resources, which are not stored properly in many cases and often are a subject

of limited access. At the same time, the unique mixture of cultures in the region makes its historical heritage important on a global scale. In the recent years, a general tendency in the field of cultural heritage is its *digital preservation* and the wide provision of *electronic access* to it. This is a field whose introduction throughout the CEE region is met with a significant interest. In the current situation with preservation programs reduced to a minimum, the possibility to combine a new method contributing to the protection of the originals with immensely larger dissemination opportunities is considered often as a panacea. However, the successful endeavours in this field in the region are still an exception and direct application of foreign solutions is not possible, because the used models should be refined to answer the local needs.

What is necessary to run a successful national policy in this field in the CEE region? What are the differences in the introduction of digital preservation and access to cultural heritage in the region compared to Western Europe and North America? The answers of these questions are still to be found.

Even in the Western countries where state-of-the-art practice are being introduced into life rather quickly and the financing of projects is not causing the same difficulties as in the CEE countries, some specific digitisation issues (such as choosing collections for digitisation, making digital collections available to different target audiences, introducing digitisation as a preservation measure, etc.) are still a subject of discussion and study [2]. This can be seen from such major international fora in the field as the European Conference 'Preservation Management: Between Policy and Practice', held in The Hague in 1999 and the preceding conference held in 1996 [3].

The improved access and presentation of these specific materials will contribute to the development of the idea for the open society where the access to specialized collections is still a matter of special policy. The current development of the global information world endangers the proper placement of cultural heritage resources of nations where the adequate information technology solutions are not being used properly. In the long run this might result in an underrepresentation of important pieces of the global picture.

2.2   Collections in Bulgaria

Only in Bulgaria, over 12,500 manuscripts of Slavonic, Greek, Latin, Ottoman Turkish and other origin are being stored. Bulgarian institutions also keep the third largest collection in the world following Italy and Greece of epigraphic inscriptions from the Antiquity period. 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.

2.3   Current state of work on digital access to local cultural heritage collections

Digital access to cultural heritage is an important track of the activities in the major repositories in the world. In addition to the benefits from the improvement of the access to cultural heritage items, work in the field contributes to the advance of the studies in mediæval history, literature and linguistics, which are important for the cultural identification.

The current state in the CEE region is characterized by isolated small-scale projects and lack of co-operation between the institutions. Moreover, most institutions do not have coherent policy for electronic presentation of mediæval and ancient materials. Most often, the issues of interest are considered only on a national scale [4] while reports presenting the picture in the region are missing.

Digitisation is not a universal remedy, replacing the problems of designing and implementing a preservation strategy. It is meant as a measure which can contribute to *improve access* to library materials and to *meet the challenge of broadening its audience*. Special attention is paid to the issues which are specific to digitisation projects in countries going through an economic transition.

2.4   Reasons for digitisation

The *typical reasons* for launching digitisation work in libraries, archives and other organisations responsible for cultural heritage preservation vary considerably depending on the size of the collection, the basic goals of the staff, the available funding. The most obvious reasons to digitise are to:

- reduce handling of the originals
- broaden access
- enhance the possibilities for research through innovative study aids (e.g., search tools; indexes; concordances (*an alphabetical index of the words in a text with their immediate contexts*), hypermedia-oriented tools, etc.)
- increase and improve use (e.g., by providing tools for manipulation of texts and images, application of specialised techniques for image enhancement, etc.)
- contribute to the creation of *"virtual collections/libraries"*.

While this is a general setting the reality brings additional concern to curators of manuscript collections in Central and Eastern European countries:

1. The lack of sufficient funds to maintain preservation programmes in their institutions, not to speak about starting large-scale digital preservation projects.
2. The necessity to develop skills and standards to present local written sources adequately. The available models created for cultural heritage of other regions cannot be applied directly. This option should be considered with a special care, due to the fact that specific features of local materials could be misrepresented.

3. The difficulty to present local cultural heritage to a wide audience: in the past specialists who consulted local material were presumed experts in their respective subject areas. The possibility of exposing local materials using the new electronic media like Internet implies a shift in the audience. The new audience consists of people who are not necessarily experts in Mediæval Slavonic language and culture and because of that need access to the material itself, but also to a wide range of explanatory information.
4. The necessity to involve specialised professionals in such endeavours and train cultural institutions' staff having in general little specialised skills in this area.

As a general tendency, we could say that in the CEE region there is too much awareness and real-life attempts to discuss the specific collections settings and to build a programme for digitisation on the long run.

3 Previous Experience in Bulgaria

The first initiatives in this field were launched by research institutions and companies under the conditions of an absence of a national strategy and funding for digitisation programmes. Libraries and museums basically provided access to their collections instead of initiating their own programmes.

The work of specialists from research institutions is basically directed towards entering data on available resources. Actual work on digitisation has not yet been done on large-scale basis, because of the high costs of such projects. Amongst the projects describing available resources we could mention:

- The Repertorium of Old Bulgarian Literature (co-ordinated by the Institute of literature at the Bulgarian Academy of Sciences);
- The Corpus of Epigraphic Inscriptions in Greek and Latin (co-ordinated by the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences),

The basic work done by companies is oriented towards creating CD-ROMs (four CD-ROMs already exist, two of manuscripts from the National Library 'St. Cyril and Methodius', one of Macedonian coins and one of Bulgarian Iconography).

A most important problem in the field of digitisation is connected with the copyright on materials for digitisation.

Copyright issues in such a complex field have to be clearly defined for two different situations: a. when primarily sources are being digitised, and b. when publications of research of different specialists are being incorporated into the final product.

The second case is very important and even more complicated than the first one, because scholarly annotations and commentaries are important components

of any digitised collections. Since this work is done in teams, the contribution of each member of the team has to be clearly defined and protected.

An overview of the state-of-the-art in the field of computer-supported studies of Slavonic manuscripts in the Eastern and Central Europe based on 90 publications shows that most of the attention and practical experience was directed at the text encoding and processing issues of (66% of all described projects), followed by data base applications (14%) and image processing (11%). Contributions involving multimedia are an exception (only one) which shows that the representation of mediæval manuscripts is still far from what current technology has to offer.

This is an important motivation to improve the current practices in future endeavours.

4   Towards an electronic edition of the Enina Apostle

In an attempt to change the situation, a collaboration of experts from the Institute of Mathematics and Computer Science of the Bulgarian Academy of Sciences and the Manuscript Department of the National Library "St. Cyril and St. Methodius" in Bulgaria was started with the aim to prepare an electronic edition of the *Enina Apostle*. This is a seriously damaged manuscript dated 11 century. Only 39 folios of the manuscript survived in a lime-pit. The access to the manuscript is highly restricted. This was a good reason to choose it for this experimental edition. In addition, this is the oldest manuscript stored in the National Library "St. Cyril and St. Methodius" in Sofia.

The project was started with the intention to illustrate how a feasible digitisation work could be done with little or no external funding based on mutual co-operation of a library willing to promote activities in this field with a team of researchers who were able to help with expertise and supply the necessary equipment to fulfil the task. This joint team includes four specialists from the Institute of mathematics and computer science and two members from the manuscript department of the National library.

The electronic edition of Enina Apostle will include digitised images, its text in Old Bulgarian language with references to the current Biblical texts in Bulgarian and English, and analytical description of the manuscript. Currently we are collecting the contents, trying at the same time to solve copyright issues which would made possible scanning of the original.

4.1   Images

The electronic images are archival quality copies of the manuscript originals, which has a complex nature. It is important to consider the specifics and the nature of the

material. These specifics refer to different computer presentation of the original: image, text and structured data.

The copyright issues are still under discussion and only two folia were scanned from the original. We show here an illustration of a scanned image of folio 6r. During scanning (made at 600 dpi) Kodak color control patches and a ruler were added with the idea to provide the user with a tool to adjust display and printer settings to produce a copy which presents the original in a best possible way.
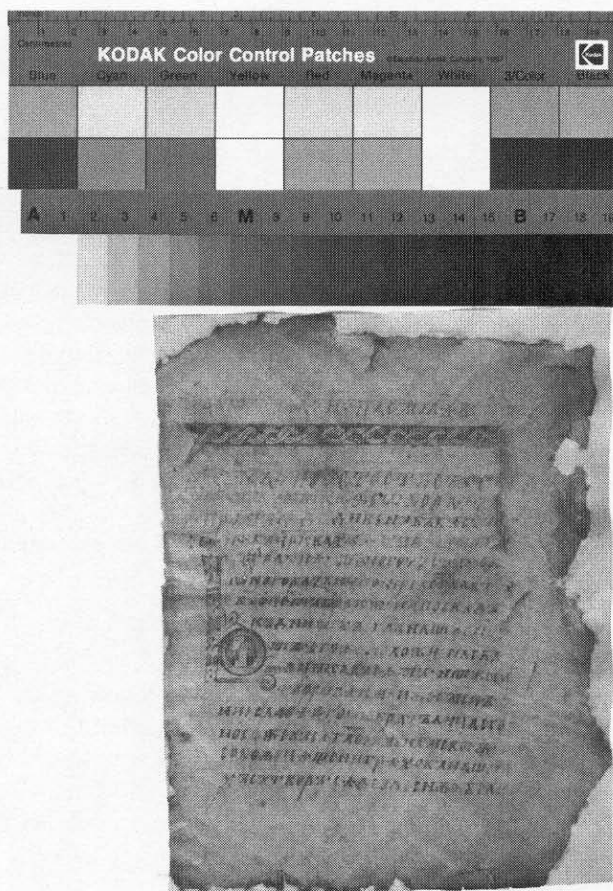

Fig. 1.

4.2   Texts

The complete text of the manuscript was entered following all its specifics (superscript letters, abbreviations, titles). Since the texts present in the manuscript are excerpts from the Epistles of the Apostles, the current versions of the texts in English and Bulgarian will be included in the digital edition. These modern texts are not a translation of the manuscript, but will give an idea what is the content of the manuscript.
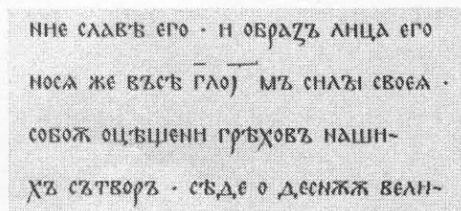

Fig. 2.

Working on an electronic edition is a serious challenge to the presentation of a digital content. It could be done in two different ways - presenting just the object (in this case we would have only images and texts), or accompanying it with various additional pieces of information which would help the readers to learn more about the manuscript. We are trying to accompany the content with such useful additional information, because if in the previous decades a traditional edition would go basically to readers who are specialists in the field, now the situation is completely different and potentially anyone who has access to Internet could have a look. There are not any recipies how to prepare a good electronic edition of such a specialised object.

4.3   Analytical Description

The analytical description of the manuscript is done following the MASTER [5] project template for manuscript description. This description is quite long, here we give an example which shows that we are describing the manuscript both in English and Bulgarian which makes the result usable by local and foreign specialists. The illustration presents part of the description which gives details on the contents of the manuscript.

5   Conclusions

Our example on the Enina Apostle is a typical case of a project targeted to one particular manuscript, which still is stuck with clarification of copyright issues.

```
<msContents>
<msItem n="1">
<locus from="1r" to="20r" supplied="yes">19</locus>
<title type="uniform" lang="BUL">липсва в текста</title>
<title type="uniform" lang="ENG">missing</title>
<rubric lang="OBG">Съб. 36 пред Месопустна</rubric>
<note lang="BUL">Това е първата запазена рубрика в текста; целият
фрагмент започва с 2 Тим. II, 4-10.1
<bibl>Мирчев, К., Кодов, Х. Енински апостол ..., с. 9</bibl>
</note>
<note lang="ENG">This is the first extant rubric in the MS; the whole fragment
begins with 2 Tim. II, 4-10.1 <bibl>
Мирчев, К., Кодов, Х. Енински апостол ..., с. 9</bibl>
</note>
<incipit lang="OBG">fragm: [воинъ] пъвла овıа[з]аетъ сıа коѵпиами
житинсками</incipit>
<note lang="BUL">В състава на ръкописа се отбелязват само тези главни
дялове; не се отбелязват отделните четива от Апостолските послания и
съпровождащите ги допълнителни текстови части като антифони,
прокимени, алилуярии и причастни, съставени от псалтирни стихове,
както и молитвите и типикарските указания.</note>
<note lang="ENG">Here are not listed the separate lessons derived from the Epistles,
nor the additional texts such as antiphons, prokimena, hallelujaria and koinonika, all
composed of Psalm verses accompanying the lessons, nor the prayers or indications of
the Typika.
</note>
</msItem>
...
</msContents>
```

Fig. 3.

Such projects can serve only as an illustration of the work which has to be done in a considerably larger scale. To achieve this, cooperation of various types of institutions is crucial in the CEE countries.

Five types of organisations are potentially interested in digitisation of cultural heritage: government bodies, repositories, research and/or educational institutions, companies and foundations. These organisations with a different profile have significantly different approaches in the field of digitisation of cultural heritage due to their different aims and needs.

Government bodies (the Ministry of culture) are entrusted with the supervision of such activities. A project on networking of the museums is currently underway, however it does not contain any official statement or plan for digitisation of cultural heritage collections in the wide sense.

Repositories (libraries 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. The Union of Librarians and Information Services Officers produced in 1997 a National Program for the preservation of Library Collections [6], which was adopted by the Library Council at the Ministry of Culture. Unfortunately, this interesting program is adopted only formally, without any real work on its implementation into the practice.

Research and/or educational institutions are the most active initiators of digitisation projects in Bulgaria as centres of study of the cultural heritage and the impact which digitisation could have on:

- routine work
- potential for large-scale comparative studies
- application of new research methods.

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 of such products is unsatisfactory. This is why their main market is abroad.

Funding bodies (foundations) supported practically all projects undertaken in the field of digitisation. However, the scale of their support cannot meet the real costs of serious digitisation projects.
This situation has led to several important issues:

- The absence of a national strategy has resulted in the lack of co-ordination between separate local initiatives which usually do not contribute to each other.
- The work of separate teams in the same field has lead to the application of many different *ad hoc* solutions, instead of a search for a general *ad modum* strategy.
- There is a clear need for international co-operation in the fields of Slavic and Balkan Studies, because of the wide spread of primarily sources throughout the whole of Europe. The lack of a national strategy does not support such co-operation in spite of its importance for real large-scale comparative studies.
- Ambiguity of legal copyright issues has lead to serious problems in persuading researchers to share their knowledge in digitisation projects affecting the level of presentation of materials, and restricting depth of presentation.

- In order not only to present materials, but to support research work in the field of Ancient and Medieval Studies, all data should be properly organised and processed. It is not sufficient to have collections of digital images, or text corpora. The application of information technologies in the Humanities is complicated by the specificity of the models in the subject domain. If we consider the example of Medieval Slavic studies, a commonly accepted model of which elements should be included in a formal model of the subject domain of knowledge can not be decided. For this reason, the creation of a generally accepted model is more a wish than a reality even after having the experience of several projects and organising a wide scientific discussion [7]. The creation of a specialised workbench for Slavic Medieval studies which will be sufficiently flexible to support those views and materials, which serve the needs of the concrete specialist is one of the possible solutions for this problem. I would like to stress that existing workbenches for the study for example of Latin manuscripts [8], will not match the needs of specialists in Slavic studies because of the impressive variety in Medieval Slavic texts for which computer presentation is still a subject of wide discussion (see the papers presented at the International workshop on Text Variety modelling [9]).

The paper deals with the initial Bulgarian experience in the field of developing electronic resources in the presentation of cultural heritage. The first projects in the field fall into two categories: research and commercial.

Bulgarian specialists encountered problems related to:

- The lack of a national strategy and co-ordination amongst institutions in the field of digitisation;
- Copyright issues (both for primarily sources and results of their scientific examination);
- Difficulties with the setting up of adequate workbenches for specific research tasks like Medieval Slavic manuscript studies.

The solution of those problems will contribute to the development of real digital resources in the field of cultural heritage. Probably the basic problem for countries in transition is whether they will be able to set-up their own programs and start work on them meeting the quality criteria of the European Community.

There are two strategies which can be followed under these conditions:

1. Waiting for better economic conditions and for guidance of more experienced countries in the field of cultural heritage digitisation. The dangers in this approach come from the poor conditions for the preservation of our cultural heritage.

2. Searching for *ad hoc* solutions, which will not lead to qualitative preservation of the whole cultural heritage, but at least will partially preserve it. The danger in this approach comes from the serious differences in the quality standards in the field of digitisation. Is it worth spending money on digitisation projects with insufficient budgets?

These decisions are very difficult. But they should be taken, and sooner the better.

References

1. http://www.dialog-21.ru/full_digest.asp?digest_id=8217
2. D. Hazen, J. Horrell, J. Merrill-Oldham, *Selecting Research Collections for Digitization*, CLIR, 1998, 18 pp.
3. Yula de Lusenet (ed.) *Choosing to Preserve: Towards a Cooperative Strategy for Long-term Access to the Intellectual Heritage*, Papers of the international conference, Frankfurt am Main, March 29-30 1996, European Commission on Preservation and Access, Amsterdam, 1997, 165 pp.
4. G. Kislovskaya, *Preservation Challenges in a Changing Political Climate: A Report from Russia*, The Commission on Preservation and Access, 1996, 20 pp.
5. http://www.cta.dmu.ac.uk/projects/master/
6. National Program for the Preservation of Library Collections, Sofia, 1997.
7. D.J. Birnbaum, A.T. Bojadzhiev, M.P. Dobreva, A.L. Miltenova (eds.), Proceedings of the first Int. Conference "Computer Processing of Medieval Slavic Manuscripts", Blagoevgrad, 24-28 July 1995, 336 pp.
8. S. Calabretto, B. Rumpler, *Distributed Multimedia Workstation for Medieval Manuscripts*, In: F. Rowland and J. Smith (eds.), *Electronic Publishing'98: Towards the Information-Rich Society*, Proceedings of an ICCC/IFIP Conference, Budapest, Hungary, 20-22 April 1998, pp. 166-178.
9. M. Dobreva (ed.) Text variety in the Witnesses of Medieval Texts: Study from Cooperative Writing Perspective, Int. workshop proceedings, Sofia, 1998.
10. M. Dobreva, S. Ivanov, *Issues in Electronic Publishing on the Medieval Slavic and Byzantine World*, In: F. Rowland and J. Smith (eds.), *Electronic Publishing'98: Towards the Information-Rich Society*, Proceedings of an ICCC/IFIP Conference, Budapest, Hungary, 20-22 April 1998, pp. 55-64.

Some Useful Web Resources

* European Commission on Preservation and Access: www.knaw.nl/ecpa
* Conservation Information Network: www.chin.gc.ca
* The Bodleian Library, Oxford, Preservation and Conservation Dept: http://www.bodley.ox.ac.uk/dept/prescons/world/links.htm
* Conservation OnLine: Resources for Conservation Professionals: http://palimpsest.stanford.edu/
* Master project: http://www.hcu.ox.ac.uk/TEI/Master/Reference/ms.htm