Digitization of the Court Books Collection of Grand Duchy of Lithuania: Challenges and Decisions

Tatjana Timčenko, Dr. Marija Prokopčik
tatjana.timcenko@mb.vu.lt, marija.prokopcik@mb.vu.lt

Цель статьи – представить новый проект Библиотеки Вильнюсского университета и обсудить проблемы, связанные с планированием и осуществлением проектов оцифровки памятников письменности. В статье рассматриваются: критерии отбора коллекции для дигитализации; проблемы подготовки документов к оцифровке (вопросы реставрации и консервации); некоторые аспекты подготовки метаданных для мультилингвистического описания коллекции.

Keywords: Manuscripts, Digitization, Metadata.

Selection Criteria

- **Collection’s Significance**

Court Books collection of the Great Duchy of Lithuania (thereinafter GDL) is one of the most interesting historical collections of VUL. It consists of more than 600 items dated back to XVI – XIX centuries written mainly in Office Slavic and Polish languages in different court institutions in the territories of GDL (XVI–XVIII c.) and Russian Empire (XIX). This collection could be used for various synchronical and diachronical researches on the history of GDL reflecting a broad political, social and economic panorama of GDL history, mode of life of the past generations, public and individual relations and psychology, changes of life standard. Court Books are important for research in history of language, historical geography and genealogy.

- **Intensive Use of the Collection Materials**

This collection is highly demanded. Users’ statistics show that Collection items each year are requested approx. 450 times. Scientists from different Lithuanian research centers, as well as researches from other countries are among the main users of Court Books. Reading of court books texts is included into a student curriculum.
In spite of a high demand and interest, Court Books are not included into publishing programmes neither of Institute of Lithuanian History, no of the faculty of History of VU. These are institutions involved into Lithuanian Metrica publishing (for detailed information please refer to periodical edition of the Institute of Lithuanian History “News of Lithuanian Metrica” [1]). In this situation digitization is not only an optimal way to present the content of the collection to the wide international research community, it is also a means to enrich the complex of already published and prepared for publishing GDL documents.

- **Poor Condition of Collection Materials**
  
  Due to the complicated acquisition history a significant part of the collection requires conservation, approx. 20 percent of items are inaccessible for users. Full conservation of the objects is a long and highly expensive process. Additional difficulty is a volume of damaged objects – 500–2000 leaves.

- **Continuing of the Digitization Process**
  
  Digitization of the collection started at 2004, when 18 books were digitized, Digital images are published at VUL Internet site [3]. The project, set up as a pilot one, was financed by the Open Society Fund and Open Society Institute. Due to limited funding and lack of metadata standards only small number of books were digitized; descriptions were brief and far from being exhaustive.

  It should be said, that VUL was the first library in Lithuania that started digitization of its cultural heritage collections in 1998 in the framework of the UNESCO program “Memory of the World”. It resulted in the CD “Historical collections of the Vilnius University Library” [4]. However majority of VUL activities were of rather occasional, fragmental nature that pursued short-term goals of various one time projects. Typically, only limited metadata were provided.

- **Continuation of Work on Collection Description**
  
  Historically VUL Collection of Court Books was part of a rich holding of Vilnius Central Archive of Ancient Documents1. Part of the documents were described before the Archive was formed, work on the records was then continued by director of the Archive N.I. Gorbachevski [5] and, later, his successor I.Y. Sprogis [6]. These records include data on the content of the documents comprising Court Books, and are still relevant today.

---

1 177 67 Court Books in 1863.
Therefore now we are facing two main tasks: to check old records and to supplement them with codicological and palaeographical data and to describe books, which are not included into old catalogues. Since size of some Court Books is rather impressive (up to 2000 leafs) one book could comprise hundreds of separate documents, so detailed content description will add value through providing in-depth information on different aspects of specific documents, which will increase research interest in some relatively unknown material.

Separate aim of the project is to create a database of scribes’ hands. Description of scribe’s hand is a part of each record, providing important information for studies on orthographical systems. Research in orthography of GDL documents is still insufficient.

- **Possibility to Re-create Holdings in Digital Space**

Collections of Court Books are kept not only in VUL. The Library keeps 600 Court Books, 10000 items are at the National Archives of Lithuanian History, and possibly related documents or their fragments are kept in other institutions in Lithuania and/or abroad. In this respect digitization is a means of rebuilding of the collection of Vilnius Central Archive of Ancient Documents in the digital space.

- **Digitization vs. Traditional Edition**

Regarding possibilities of publication of historical documents collections, we consider that digitization has a lot of advantages comparing with primary source transcription. Today’s practice of transcription of historical texts is rather ambiguous as it is impossible to speak about one publishing pattern. For example, texts of Lithuanian Metrica that are of similar nature to aforementioned documents, are delivered in a different way in the publications prepared both in Lithuania and abroad [7]. So transcription of the same original texts by different publishers will have different results. Nowadays common feature of all text transcriptions for historical publications is the problem of impossibility of reversible conversion (translation) of the text, because a modernization of orthography causes the conflation of several original graphemes into one. Therefore it is impossible to restore original orthography. This considerably limits circle of published texts’ researchers and eliminates significant part of potentially possible linguistic research.

Certainly in principle it is possible to produce “flexible electronic publication”, that provides both: a version of text transcription following rules of linguistic publication of manuscripts and a version/versions with simplified orthography created on its basis. Similar
practice already exists (For example, technology of mixed transcription (технология смешанного набора проекта «Фонд знаменныых песнопений» [8]) or decisions of TEI [9]), however this is a task for a separate big project.

Digitized document, enriched with sufficient description that enables orientation in the text, turn-up as reliable and handy source of data for professionals from different fields. Furthermore, results of the project could be implemented into a curriculum.

And last but not least are financial implications: transcription is much more expensive. Turning the presentation of digitized documents into a virtual space with free access considerably simplifies preparation of publications and enhances geography of publishing centres.

**Collection Preparation for Digitization**

The preparation of documents for digitization is an important element of digitization projects, however it is often disregarded. This stage of digitization is extremely important; it often determines smooth course of the whole project, safety of both documents and the staff, sometimes providing the possibility to reveal additional important information on the documents’ history, even to introduce new documents or their fragments into a scholar circulation.

- It is important to evaluate codicological structure of documents and make sure that object will not be damaged during digitization. The aim is to ensure that available scanning facilities are suitable for the selected documents and/or to select the documents accordingly.

- All the items, where micromycetes (mould) were discovered must be disinfected before scanning and/or metadata preparation in order not to endanger documents and the staff.

- A simple procedure of dust cleaning doesn’t require many efforts, but, firstly, helps to stop dissemination of micromycetes and dangerous chemical elements, secondly, adds aesthetic value to the digitized objects, and, thirdly, saves scanners and other equipment.

- The level of conservation depends on variety of factors: duration of the project, technical facilities, and skills of the staff [10]. In case of the Court Book digitization
the decision to stabilize existing condition of the object was taken. In some cases
documents will be fully restored.

At this stage staff could face serious and interesting problems, related to the authenticity and
use of the additional information revealed during conservation.

In our case two such problems are worth mentioning. First is a notarial binding of the
documents, which, on one side is a significant part of the documents history, on the other —
a big challenge for anyone wishing to read the document, as well as for anyone going to
digitize it (fig.1). So after numerous discussions it was decided that in cases, when it is
impossible to read the text in full, to cut the notarial binding safeguarding the seal.

Another problem is related to the authentic binding elements, which at the same moment are
the source of historical and linguistic information. Structure of Samogitian Court Books is
very special. Old manuscripts were used to bind the books (fig. 2).

In this case we need to choose: a) to infringe upon the authenticity and to keep documents,
which originally were used for bindings, as separate items and to include them into the
collection, using implants of other materials for re-binding the original, or b) to digitize
these “newly founded” documents and to place them back to the bindings. In each case
conservation could give additional information and enhance the value of the project.

**Metadata Presentation Strategies for Multilingual Environment**

Having in mind languages relevant for the researches in the fields of GDL history and
written heritage, it was decided to present medatada in Lithuanian, Polish and Russian
language, considering that at least one of the mentioned three languages is a working language for any specialist on the field of GDL written heritage.

The aim of the project is to present as exhaustive descriptions as possible regarding contents, structure and acquisition history, including data on restoration. We have already got some experience in using TEI manuscript description scheme [11] and consider this format the best for our tasks solution. This scheme provides us with the possibilities to describe huge number of elements and semantic relations between different entities.

Creation of digital content and improvement of access to it is one of the priority activities of Lithuanian cultural heritage institutions included into National digitization strategy [12] and implemented through variety of respective projects [13] of cultural heritage digitization resulting in an abundance of cultural information (books, manuscripts, films, audio records, maps, archival documents, art objects, etc.) in digital format. Research in accessibility focuses mostly on interoperability and integration, taking into account that even though cultural heritage comes in various forms it also has many similarities and is semantically related [14]. In this context one of the main tasks is to create and develop integral system of cultural heritage retrieval, preservation and promotion. Expertise and experience acquired while working with different objects and institutions and necessity to meet requirements stipulated by users stimulated implementation of CIDOC CRM [15] ontology. From the multi-semantic terminology viewpoint it is important that this standard is based on the objective attitude towards reality, i.e. it is oriented towards real objects not conventional concepts that depend on different linguistic, cultural and historical tradition.

Another important issue which is waiting for a solution is metadata presentation strategies for cultural heritage and historical multilingual environment. This issue is of great importance because of permanent change of place-names, change of administrative divisions and jurisdiction, great number of similar place-names, disappearance of certain names, necessity to link old/extinct locality with modern place, multinational and multilingual environment, etc. The same applies to personal names: use of different names (or pseudonyms) by the same person, multilingual environment and thus different notation. Separate issue is a presentation of historic chronology data in information systems: different dating and time reading systems were used in Lithuania and often historical dating is based solely on researcher’s personal interpretation. The use of the three languages – Russian,
Lithuanian and Polish descriptions necessitates the creation of controlled lists of topic/subject names, to be used in search. A special linguistic tool could enable move towards semantic integration of knowledge coming from different fields and institutions (libraries, archives, museums, etc.) and creation of a system defining interaction between location and person, location and topic/object, person and topic/object, etc. This requires a high level of systems knowledge.

In our case there is one obvious solution – use and active participation in further development of BAVIC (Integrated thesaurus of personal and place names and historical chronology) that could provide for comfortable access to information and enable semantic search. This thesaurus aggregates several already existing data bases of controlled terms, namely National Library authority records data base integrating variety of forms and scripts of local, personal names and subject headings, structure and contents of personal and place names from IS “Aruodai”, data base of place names assembled by Institute of Lithuanian language.

BAVIC thesaurus is based on interoperability approach towards cultural heritage related information when each system retains its independence however due to networked communication information created separately in each institutions interacts in virtual space to the semantic level of interaction inclusive.

**Some Conclusions**

The work on preservation of old and valuable documents at VUL is being carried out in two directions – digitization and restoration, which could enrich one another during the implementation of digitization projects. In this respect we consider the whole project both as the object and the source for research.

Apart from direct results and benefits of this project, it provides us with a possibility to pursue other goals: increase visibility within research and library community; possibilities to involve specialists from other branches (history, philology) and institutions to work on library collections (would be impossible because of the lack of funding); better relations/understanding between staff of different library departments; enhancement of VUL scholar research and implementation of goals defined by setting up of the new Research and cultural heritage collections centre.
References


2. VUL Court Book Collection. Режим доступа: http://gluosnis.vu.lt/biblio/dperziura.sarasas


4. Горбачевский, Н.И. Каталог древним книгам губерний: Виленской, Гродненской, Минской и Ковенской, также книгам некоторых судов губерний Могилевской и Смоленской, хранящимся ныне в Центральном архиве в Вильне, Вильна, 1872.


7. Фонд знаменных песнопений. Режим доступа: http://znamen.ru


