Methodological and Organisational Aspects of Digitisation and Bibliographic Access of the Cultural Heritage: the Lithuanian Approach

by

Regina Varnienė-Janssen

Martynas Mažvydas National Library of Lithuania, Vilnius, Lithuania

Abstract

This article aims to indicate the lessons learnt from the digitisation initiative of the Lithuanian National Bibliographic Agency (the Centre of Bibliography and Book Science at Martynas Mažvydas National Library of Lithuania and other memory institutions — the Lithuanian Art Museum and the Lithuanian Archives Department, which provide authoritative bibliographic descriptions for the national collection. It describes experience in the development, application and updating of formats and standards used for recording bibliographic information and cultural content, provision of access to it, establishment of the Integrated Virtual Library Information System and the common Web portal of the cultural heritage http://jlmnwp.epaveldas.lt. The article treats the Lithuanian national bibliographic agency's efforts to update the legislation to reflect the partnership of memory institutions, to create a methodology for such partnership and to encourage the compilation of the national bibliography.

Methodology: a common approach to the cultural heritage

Preservation of cultural heritage objects and information about them is at the core of the mission of memory institutions — archives, libraries and museums. Sectoral boundaries between them often become blurred when we take into consideration memory institutions which, due to historical circumstances, sheltered cultural heritage assets that differed from the type of objects usually preserved by them. Though this "object type" factor still plays an important role in framing the organisational structures of memory institutions, it could hardly be argued, however, that a work of art, the author’s manuscript of this work or a review of this same work — all share the same cultural and historical context, and provide evidence of comparable cultural features. With the purpose to achieve consistency between the cultural heritage content and the presentation of information about this content, IFLA, CIDOC, ISO and other communities have started developing a common conceptualisation of information about the cultural heritage preserved at the archives, libraries and museums. In this context, the results achieved by the international Working Group on FRBR/CIDOC CRM Harmonisation are very encouraging as regards the implementation of the integrated shared ontology for the information accumulated by both libraries and museums. This, in turn, will ensure integrity of the cultural heritage content and convenient access in the Semantic Web environment.

We could finally admit, with some satisfaction, that the initiative started by the Centre of Bibliography and Book Science, which originated from a request for financial support from the European Structural Funds, has developed into a common methodological approach to cultural heritage digitisation and access with respect to modelling, standards, recommendations and practices. The methodology has been:
• implemented within the Integrated Virtual Library Information System and the portal of the cultural heritage http://www.epaveldas.lt;
• realised within the virtual interorganisational structure through coordination centres for digitisation.

Objectives

**New paradigm in the legislation on the cultural heritage: Concept, Strategy and Implementation Framework**

Lithuanian memory institutions collect and preserve the precious national heritage, which serves as a source for the nation's public spirit, self-awareness and self-esteem, its contemporary knowledge, as well as educational and leisure pursuits. With Lithuania joining the EU, the importance of cultural heritage actualisation, i.e. its presentation within a single digital space of Europe, has considerably increased. In 2006, the Ministry of Culture together with the Institute of Library and Information Science at Vilnius University's Faculty of Communication held a survey among Lithuanian memory institutions. It served as the basis for research into activities on digitisation of the preserved cultural heritage, the results of which were published in the annual report "Coordinating Digitisation in Europe" by the National Representatives Group for Coordination of Digitisation Policy and Programmes in 2006. The results of the research revealed a wide range of multiple memory institutions (libraries, museums, archives, research and educational institutions, other public bodies) that were involved in various cultural heritage digitisation projects. At that time, 58% of memory institutions were involved into digitisation activities, and 55% of them considered digitisation as their strategic priority. The largest project ever launched by a memory institution in Lithuania was "Creation of the Integrated Virtual Library Information System" carried out by the Martynas Mažvydas National Library of Lithuania together with partners — the Lithuanian Art Museum and the Lithuanian Archives Department — under the Government of the Republic of Lithuania in 2005-2008. The successful implementation of the project allowed setting up a data bank of digital data with the capacity to store over 2,800,000 digitised pages of original documents. It has enabled national and worldwide users to get acquainted with cultural assets preserved at these institutions.

The digitisation projects initiated by academic institutions and their international involvements mostly support research and educational needs. The Institute of Mathematics and Informatics, the Lithuanian Institute of History, the Institute of the Lithuanian Language, and the Institute of Lithuanian Literature and Folklore launched the scientific project for data digitisation Aruodai in 2003. In 2006, the project LDK Skaitmenų for digitisation of manuscript documents of the Grand Duchy of Lithuania was undertaken by the Institute of Mathematics and Informatics together with the Institute of the Lithuanian Language, the Faculty of Philology at Vilnius University and Vilnius University Library under the support of the Lithuanian Statė Science and Studies Foundation. The project "The Information System of Metadata of Archival Documents and Books of Vilnius Ecclesiastical Province (BARIS)" was launched by the Faculty of Communication at Vilnius University together with the Faculty of History, the Curia of Vilnius Archdiocese and the Lithuanian Museum of Ethnocosmology in 2006. The project "Digitisation of the Ancient Lithuanian Manuscripts of the Eastern Prussia. Danielius Kleimas' Hymnal and Prayer" was initiated by the Institute of the Lithuanian Language together with Berlin Humboldt University and financed by the European Union PHARE 2003 programme. This initiative is a continuation of efforts to set up a database of old manuscripts. Vilnius University Library, together with the Institute of Mathematics and Informatics, have joined the international project ENRIC, which aims to create a digital library database of the European cultural heritage by integrating the created digital content that still remains scattered. Vilnius University Library has been involved in a long-term project "Creation of the Full-Text Database of Manuscript Judicial Books of the Grand Duchy of Lithuania".

Lithuanian National Radio and Television has completed the project "Creation of the Virtual Television Product Library Enabling Public Electronic Access to the Lithuanian Audiovisual Heritage Recorded in Lithuanian Television (LTV) Programmes and Ensuring Continuous Digitisation, Preservation and Availability of This Heritage to the Public". Under this project, a public online library of television products providing access to the national audiovisual heritage recorded in television programmes was built.

After the adoption in 2005 of "A Concept for Lithuanian Cultural Heritage Digitisation" memory institutions continued to be involved in large-scale digitisation projects; cultural heritage digitisation activities, however, remained fragmentary throughout Lithuania. Project activities were not coordinated; initiatives were mostly focused on short-term goals, the only exception being the project of the National Library and its partners. Even the completed projects did not often guarantee preservation of digitised cultural heritage objects and access to them.

This inconsistency of digitisation activities has been the key incentive for further improvement of the legislation on national cultural heritage digitisation and access. Proper legislation is the only mechanism providing maximum security for successful strategic activity of memory institutions and their financial support, which, in turn, ensures sustainability of the process of national cultural heritage digitisation.

In the context of this article, it is important to highlight the role of the Lithuanian national bibliographic agency in the development of the legal basis for digitisation of the national cultural heritage and its access. In 2003-2004, the Centre of Bibliography and Book Science launched several projects for cultural heritage digitisation and integration of memory institutions. One of the core aims of these projects was to bring the attention of public authorities to the essential need for integration into the space of the European electronic content.

The above-mentioned Concept defines the objective, goals and principles regarding cultural heritage digitisation, and also common selection criteria for objects to be digitised. It sets the foundation for consistent and purposeful activities by memory institutions in safeguarding and actualising the national cultural heritage and improving access to it. On the basis of the Concept, coordination and monitoring of the support for
cultural heritage digitisation on a national level is carried out by the Ministry of Culture of the Republic of Lithuania together with the Ministry of Education and Science, the Information Society Development Committee, as well as the Lithuanian Archives Department, under the Government of the Republic of Lithuania. They also present suggestions for efficient realisation of this support to the Council on Digitisation of the Lithuanian Cultural Heritage.

The Concept, however, does not define the organisation of digitisation activities and their funding. This imperfection motivated the author to initiate further updating of the regulatory framework. The initiative group formed by the author developed a draft project for the strategy on digitisation and presented it to the Ministry of Culture. After a general debate by the team from the Ministry of Culture and the Council on Digitisation, and a public debate, "The Strategy on Digitisation of the Lithuanian Cultural Heritage, Preservation of Digital Content and Access" and its Implementation Framework for 2009-2013 were approved and adopted by the Government of the Republic of Lithuania on 20 May 2009. These strategic documents will help to expand the existing infrastructure in order to create a single digital space of the cultural heritage information, to extend the lifetime of cultural heritage objects, to provide current and authoritative information about the Lithuanian cultural heritage in the virtual environment, and also to facilitate preservation and promotion of the Lithuanian culture in a globalised world.

To give a concise description of the purpose of the Strategy, the main emphasis should be on provisions that define coordination of digitisation activities through digitisation centres. Their role is to be taken by institutions having the most prolific experience in digitisation and providing access to digitised materials — the Martynas Mažvydas National Library of Lithuania, the Lithuanian Art Museum and the Lithuanian Archives Department. The establishment of such competence centres within national memory institutions must facilitate the selection of cultural objects to be digitised and the compilation of lists of such objects, and also eliminate unnecessary overlap of procedures. Thus financial resources will be saved and consistency of the entire process will be achieved. The centres' remit should include provision of manifold methodological assistance and training services for the professional community, and invariable concern about financial support, which could be obtained in line with the guidelines provided in the above-mentioned strategic documents.

The vision of the Strategy manifests itself as a single digital information space of the Lithuanian cultural heritage content, which extends the lifetime of cultural heritage objects, provides current, comprehensive and authoritative information about Lithuanian cultural heritage to the European and worldwide public, and ensures permanent and reliable use of cultural heritage resources for research, and for educational and cultural purposes. The Strategy aims at promoting digitisation activities among memory institutions, ensuring long-term preservation of the Lithuanian cultural heritage, its integration into the global digital space, and worldwide access to it. The Strategy aspires to:

- create a seamless system of cultural heritage digitisation and guarantee its reliable functionality;
- develop and expand the virtual system of the digital cultural heritage combining search, preservation and access features;
- standardise cultural heritage digitisation, preservation and access workflows;
- digitise cultural heritage objects and transfer their content onto the Web portal of the cultural heritage;
- encourage initiatives on cultural heritage access.

It must also be taken into consideration that the adopted Implementation Framework for 2009-2013 will provide a foundation for seeking sustainable funding options both from the state budget and the EU Structural Funds. Priority should be given to projects that provide possibilities for participation in multi-partner activities of creating a single digital information space of the Lithuanian cultural heritage and integration of its objects into the digital system of the cultural heritage and the common Web portal.
Shared standards u s the basis for the Integrated Virtual Library Information System and Web portal

The expansion of digitisation activities has stimulated the development of an integrated shared ontology for information about the collections held by memory institutions. It was only a short time ago that shared standards started to be applied for most of the digital repositories and access systems set up by memory institutions. Prior to that, the absence of concern for application of shared standards prevented consistency in the access to the digital content and its retrieval.

That is why in the above-mentioned strategic documents on cultural heritage digitisation, special prominence is given to the necessity for shared standards, and the goal has been set to adopt shared digitisation standards in order to coordinate projects and initiatives undertaken by memory institutions, and ensure compatibility of cultural heritage repositories at a national level and their integration into the European digital space.

Currently, the National Library together with the project partners — the Lithuanian Archives Department and the Lithuanian Art Museum — are the only institutions adhering to shared metadata, digital object archiving, preservation and access standards within the Integrated Virtual Library Information System created under the joint project of 2005-2008.

Application of shared standards is essential, because they call for integrity of the digital content and more qualitative information resources, where objects can be browsable, searchable and identifiable within a hierarchy system.

To achieve interoperability with national memory institutions, the European Digital Library (EDL) (http://www.theeuropeanlibrary.org) and Europeana (http://www.europeana.eu/portal/), the following standards have been applied within IVBIS (see Figure 1 overleaf).

Figure 2 above illustrates the interaction of these standards within IVBIS.

Descriptive metadata solutions

The level of metadata used within IVBIS has been the principal issue calling for a most urgent solution.

There was no uncertainty over the presentation online of digital image metadata, since the Dublin Core (DC) metadata standard had been adopted as a national standard in Lithuania. An example of DC metadata within the portal of the cultural heritage is available at http://www.epaveldas.lt/vbspi/biDetails.do?biRecordId=12269

The DC metadata format is simple to manage: the creation of Dublin Core records for any kind of information presents no difficulties. However, some of the elements are capable of causing confusion, e.g. "creator data". Within the archives, a description of each archival item is relevant only when an
explicit description of the creator of the collections to which the object belongs is provided. It is only the history and structure of the creator that can help the user understand conditions in which the document was created or collected and appraise the content of the document. The relevance of this category of data for the description is very high. The Dublin Core lacks this category of data. In the Dublin Core, the “creator” element is defined as an entity primarily responsible for making the content of the resource. This definition will not pertain to archives. In ISAD (G), a creator is defined as "the corporate body, family or person that created, accumulated, and/or maintained records in the conduct of personal or corporate activity". Individuals responsible for the accumulation of the collections usually differ from creators. Though the qualified version of the Dublin Core has been chosen for IVBIS, we pursue creation of extensive records.

The creation of IVBIS was guided by the approach that descriptions of digital objects should have the same hierarchy structure as traditional descriptions.

The General International Standard Archival Description, ISAD (G), provides guidance for recording information in each of the 26 elements, which can be merged to constitute a description of an archival object. ISAD (G) ensures the creation of consistent, appropriate, and self-explanatory descriptions, facilitates retrieval and exchange of information about archival materials, and enables the integration of descriptions from different systems into a unified information system.

In case of a digitised physical archival object, it is necessary to provide access to its content. For this purpose, either the physical representation of the object or its content has to be described. In case the object is a collection, an archival description representing the architecture and component parts of the whole collection has to be created.

Bibliographic descriptions for digitised archival objects are constructed in compliance with the General International Archival Description ISAD (G), which makes it possible to provide a hierarchy structure, in other words, more complete and simplified descriptions of archival objects and databases in an electronic environment. This methodological solution was applied for the Integrated Virtual Library Information System.

For the bibliographic description of the holdings of the Lithuanian Art Museum, the specification for the use of 40 data elements based on ODOC CRM (ODOC Conceptual Reference Model), approved by the Resolution of the Ministry of Culture of the Republic of Lithuania of 1997, is followed. This set of rules specifies what data related to a collection item is mandatory and what is left at the discretion of the museum.

For the description of objects from the National Library, the internationally adopted ISBDs are applied. They have been translated into Lithuanian, and the translation of the Consolidated ISBD is underway.

When transferring records to the common system, some form of encoding is needed. For encoding descriptive metadata of archival digital objects transferred to the virtual cultural heritage system, the Lithuanian archives use the Encoded Archival Description (EAD).

For encoding records for digital objects from the National Library, UNMARC/B (for bibliographic records) and UNMARC/A (for authority records), which have been designated as national standards by a Resolution of the Minister of Culture, are applied.

For transferring digital images from the Art Museum to IVBIS, the DC metadata format is applied. The problem with compatibility between different metadata description schemes and different systems for data encoding has been solved by introducing crosswalks. They allow describing compatible objects by the use of different schemas and encoding formats.

Since the presented digital object includes very extensive metadata, the system supports a possibility to revert from the short record within the portal http://www.epaveldas.lt containing the key elements of the DC bibliographic description, to the source system hosting complete records. This is made possible by crosswalks.

The example below shows a descriptive metadata crosswalk indicating the content and encoding of descriptions for digital objects from the source systems.

<table>
<thead>
<tr>
<th>ISAD (G)</th>
<th>ISBD</th>
<th>ODOC CRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1. Name(s) of creator(s)</td>
<td>1.5 Statement of responsibility</td>
<td>Actor</td>
</tr>
<tr>
<td>EAD</td>
<td>UNMARC</td>
<td>Dublin Core</td>
</tr>
<tr>
<td>&lt;origination&gt;</td>
<td>7—Responsibility block</td>
<td>&lt;dcreator&gt;</td>
</tr>
</tbody>
</table>

Several crosswalks have been introduced and are currently used within IVBIS. They enable to uniquely represent descriptive and other elements essential for transferring objects from partner systems to the common data archive.

As regards harmonisation of descriptive metadata intended to be done during the follow-up of the project scheduled for late 2010-2013, the previous efforts by all IFLA working groups (on ISBD, RDA, FRBR/CIDOC CRM harmonisation, etc.) should be taken into consideration as being of very high importance for the integration of the digital content and simplification of retrieval.

Besides further enhancing the structural compatibility between descriptive metadata, the upcoming stages of the project imply the establishment of an integrated database of names, geographical names and chronological data of Lithuania. They will also contribute to laying the groundwork for drafting a standard. The benefit of these efforts is self-evident. After they are embraced by all Lithuanian institutions involved in digitisation of the national cultural heritage and provision of access to it, retrieval of digital cultural heritage objects will be made much more effective.

It is important to indicate the productive application of the project results for the retrospective bibliography retro-conversion started in 1991 and the digitisation of the documentary heritage.

When digitisation of books in Lithuanian (1547-1830), books in Polish published in Lithuania (1799-1830) and books in Hebrew published in Lithuania started in 2005, their bibliographic and authority records had been already completed, which, in turn, facilitated delivery of computerised records to HPB (Hand Press Book), TEL and the established Web portal of the cultural heritage.

Towards the second stage of the project, intense activities related to the preparation of the national bibliography for its conversion into digital form are underway. Taking into account the more express nature of the digitisation process in comparison with cataloguing, organisational changes must be applied in order that a maximum number of professionals could be involved into the process.
The initial stage of the project was marked by the active participation of the Lithuanian Art Museum: a great number of books and albums from its holdings were digitised. This fact had an impact on the decentralisation of the national bibliography.

The dynamics of presentation of the documentary heritage online will be increased by the involvement of the two major national academic and five public county libraries into the project in 2010-2013. As a result of the project activities, the dynamics of extensive records creation will also be strengthened for such parts of the collection as books and periodicals in Hebrew, Polish and other languages published in Lithuania. The involvement of Vilnius University Library and the Library of the Academy of Sciences will have a significant impact on the production of the national bibliography, because these institutions are our partners in the compilation of the national retrospective bibliography.

**Archiving of metadata and digital objects**

The Metadata Encoding and Transmission Standard (METS) has been chosen for archiving digital objects and metadata within the virtual system of the cultural heritage. It has been developed as a means of putting a virtual wrapper around digital files that need to be held together. Besides providing the structure for complex digital documents, METS also bundles the documents together with their descriptive metadata, such as Dublin Core, and administrative information about the documents (see Figure 3).

METS is the most convenient XML schema for creating XML document instances that express the structure of digital library objects, the associated descriptive and administrative metadata, and the names and locations of the files that comprise the digital object. The metadata necessary for the successful management and use of digital object is both more extensive than, and different from, the metadata used for managing collections of printed and art works or archival documents.

Project activities included creation of the software for workflow management permitting to aggregate digital objects into integrated sets, where quality control, return for redigitisation of corrupt images, linking with the recognised full-text file and linking with the bibliographical object within the database of the relevant catalogue, could be performed. The export of objects with added descriptive metadata to the central database of IVBIS is performed with the application of METS. Extensive information related to the work and links to external objects — digital image files — are included into the METS package used for the data export. The description of the object in METS serves as a linking element between different parts of the document and its different versions. The following 5 initial sections of the METS standard are used:

1. METS Document Header;
2. descriptive metadata;
3. administrative metadata;
4. file section;
5. structural sap.

The advantage of the software consists of its being operational both under the connection with IVBIS and under the absence of such connection; it also supports transferring of data in different media. Currently, data is transferred by the use of the ftp protocol and portable discs. To avoid failures in copying files and corrupt or incomplete images, the checksum algorithm of the MD5 standard is applied. Before the METS package is wrapped, all the MD5 sums of images related to the works are calculated and included into the package. Before uploading the data into the system, the sums are checked in order to avoid errors.

The procedures are complicated by the necessity to associate different descriptive metadata encoded in different systems. As mentioned above, this problem has been solved by the application
of crosswalks: UNMARC/DC/EAD.

Transferring METS-wrapped data with the application of UNMARC and Dublin Core standards has been fully achieved. If necessary, IVBIS could be easily extended by additional encoding standards for bibliographic description.

Access solutions

For access to IVBIS, the portal http://www.epaveldas.lt accessible via SRU / W and Z39.50 has been introduced. Since IVBIS has been designed for provision of visual information, digital objects within this system also contain descriptive metadata. The user interface permits switching to a page with a detailed description within UBIS, LAFSIS or RIS/IRIS for more in-depth examination. For this purpose, a reciprocal linking system has been introduced, i.e. there is a link between each work and the relevant description in the source system, and vice versa. The linking mechanism to objects within IVBIS has been established to retain permanent http links: with changes in realisation of the virtual system, the links will remain intact. The structure of the link is the following: http://www.epaveldas.lt/vbspibiDetails.do?libisId=CIR0000046342. Here http://www.epaveldas.lt/vbspibiDetails.do is the permanent part, and libisId=CIR0000046342 is the variable part, which corresponds to the value of field 001 of the UNIMARC record.

Interorganisational structure as a guarantee for interoperability between Lithuanian memory institutions

In scientific literature, the interorganisational form of a virtual system is defined as a specific form of collaboration, in which partnerships retain their legal and economic rights. Such an organisation is characterised as an alliance of individual, advanced and legally independent organisations sharing common professional background and possessing a modern technological framework, which remains operational until the common goal is achieved.

This goal is achieved without any special institutional supervision and control mechanisms, mutual trust playing the decisive role. The connective factor within a virtual organisation is knowledge and expertise.

It could be acknowledged that the joint project by the National Library and the partners, which was completed in 2008, resulted in the establishment of a virtual organisation, which did not end its activity after the project’s conclusion. Martynas Mažvydas National Library of Lithuania, the Lithuanian Art Museum and the Lithuanian Archives Department jointly proceeded with preparing the follow-up of the project and drafting the Implementation Framework for 2009-2013.

Conclusions

1. A common methodological approach to cultural heritage digitisation and access, irrespective of the type, provenance, content or structure of objects, has been established for the first time in Lithuania. It has been defined in strategic documents on the cultural policy: “A Concept for Lithuanian Cultural Heritage Digitisation”, “The Strategy on Digitisation of the Lithuanian Cultural Heritage, Digital Content Preservation and Access” and its Implementation Framework for 2009-2013.

2. Decisions taken on digital object description and use of these descriptions comply with the requirements for memory institutions on representation of objects in the digital environment. They also answer the users’ needs for acquiring more detailed information than a digital object within the VVeb portal could provide.

Building on the previous experience with the creation and development of LAFSIS, LIBIS and RIS/IRIS has enabled a next fresh step to be taken (to create a virtual digital heritage system integrating all the above-mentioned systems in order to provide national and worldwide users with as varied and efficient services as possible.

The decision to make extensive use of databases and bibliographic record catalogues of the project partners has enabled to avoid duplication of functions, had a cost-cutting effect on the process of transferring data into IVBIS and provided a safeguard against errors.

3. The established virtual interorganisational structure has provided a background for initiatives on further development of the legislation on national cultural heritage digitisation, and for obtaining financial support for future projects. The major achievement of the interinstitutional framework has been the adoption of strategic documents on digitisation of the national
cultural heritage, which shall ensure successful strategic activities of national memory institutions and other agencies concerned with long-term preservation and access of the national cultural heritage, and will provide a guarantee for sustainable financial support.

References