Strategic, Methodological and Technical Solutions for the Creation of Seamless Cultural Heritage Content and Access: Lithuanian Approach

Assoc. Prof. Dr. Regina Varnienė-Janssen
Martynas Mažvydas National Library of Lithuania
and
Jonas Juškys
Sintagma Ltd

My acknowledgements are to:

Tomas Auškalnis
and
Liubovė Buckienė

for their assistance in the preparation of this report
This report aims to indicate the lessons learnt from initiatives by the Martynas Mažvydas National Library of Lithuania and nine other memory institutions on creation of seamless cultural heritage content. It also focuses on the legal basis for creation of such content, financing issues and creation of the Virtual Electronic Heritage System.

1. New paradigm in the Lithuanian legislation on cultural heritage

- A Digital Agenda for Europe - 7 priority areas for development of the information society and increasing social openness
- The Concept for Digitization of Lithuanian Cultural Heritage (2005)
- Implementation measure “Lithuanian culture in the information society” of the 3rd priority of Lithuania’s Action Programme for Economic Growth in 2007-2013
- Progress Strategy for Lithuania “Lithuania 2030”
The recent strategic document of the European Commission *Europe 2020: A European Strategy for Smart, Sustainable and Inclusive Growth* and one of its flagship initiatives *A Digital Agenda for Europe*, which identifies seven priority action areas for the development of the information society, give clear evidence that provisions of “The Strategy for Digitization of Lithuanian Cultural Heritage, Digital Content Preservation and Access” adopted by the Government of the Republic of Lithuania in 2009 have retained their currency up to the present day. “The Strategy for the Development of the Lithuanian Information Society in 2010-2015” defining one of its principle objectives as “fostering the Lithuanian language and culture by using information technologies” also confirms that creation of the digital content of the national cultural heritage and providing access to it remains one of the priority areas for the information society. This largely coincides with the measure “Lithuanian culture in the information society” of the 3rd priority of Lithuania’s *Action Programme for Economic Growth in 2007-2013. The Progress Strategy for Lithuania “Lithuania 2030”*, which defines priorities and their implementation policies until 2030, sets the goal to retain and promote the Lithuanian cultural identity, which consolidates the inherited and contemporary European cultural values, to ensure its sustainability, open access and competitiveness [κυκφυτοναφονο] (konkurencingumas) in the modern context of the cultural variety of the European Union and the rest of the world.

We could finally admit, with some satisfaction, that the initiative started by the National Library, which originated in a request for financial support from the European Structural Funds, has developed [διαχειρισμένη περιφερειακή στρατηγική και προγράμματα] into a common methodological approach to cultural heritage digitization and access with respect to modelling, standards, recommendations and practices.
Methodology for the creation of seamless heritage content

- Defined in the following strategic documents:
  - The Concept for Digitization of Lithuanian Cultural Heritage (2005)
  - Implementation Framework 2009-2013 for the Strategy
- Implemented within the Virtual Electronic Heritage System

This methodology has been:

- implemented within the Virtual Electronic Heritage System and the web portal of the cultural heritage;

The Concept for Digitization of Lithuanian Cultural Heritage

- Defines the goal, objectives and principles for digitization of cultural heritage;
- Sets common selection criteria for objects to be digitized.

The above-mentioned Concept initiated by the National Library was adopted by the Government of the Republic of Lithuania in January 2005. The Concept defines the goal, objectives and principles for
digitization of the cultural heritage, also common selection criteria for objects to be digitized, setting the foundation for consistent and purposeful activities by memory institutions in securing the nation’s cultural heritage, sustaining its currency and improving access to it.

Taking into account that the Concept does not define organization of digitization activities and funding of these activities, further improvement of the legislation on digitization and access of the national cultural heritage was undertaken. And on 20 May 2009, The Strategy for Digitization of Lithuanian Cultural Heritage, Digital Content Preservation and Access and its Implementation Framework 2009-2013 were approved and adopted by the Government of the Republic of Lithuania.

The Strategy pursues to:
- create a seamless system for digitization of cultural heritage and guarantee its reliable functionality;
- develop and expand the virtual system of the digital cultural heritage combining search, preservation and access features;
- standardize cultural heritage digitization, preservation and access workflows;
- digitize cultural heritage objects and deliver their content to the web portal of the cultural heritage;
- encourage initiatives on cultural heritage access.
We can see from the list that the eight projects cover several areas of the cultural heritage (archives, museums, libraries) and multiple types of cultural heritage objects.

2. Virtual Electronic Heritage System as the basis for seamless cultural heritage content

In order to achieve interoperability between national memory institutions, a virtual system has been created within the project “Development of the Virtual Electronic Cultural Heritage System (VEPS, \textit{Virtuali elektroninio paveldo sistema})“ launched on 3 February 2010 by the National Library together with nine partners from all over Lithuania.

This project, which is a follow-up of the project implemented in 2005-2008, has been financed from the EU Structural Funds (under the implementation measure “Lithuanian culture in the information
society” of the 3rd priority “Information society for all” of the Action Programme for Economic Growth in 2007-2013) and the State budget (EUR 3 250 000 and EUR 570 000 respectively).

**Development of the Virtual Electronic Heritage System (VEPS)**

**Goal of the project:**
- Ensuring sustainable creation of Lithuanian cultural heritage content and access.

**Project objectives:**
- Expansion of the network for the creation of digital content by establishing new digitization centres and units;
- Development and expansion of the virtual system for the digital cultural heritage;
- Ensuring digitization of cultural heritage objects and access to them.

With the involvement of other memory institutions, the number of metadata formats has increased as well (UNIMARC, MARC21, ESE, EAD, DC).

VEPS has been designed for preservation of descriptive metadata of digitized objects, and not metadata about the original works.

I would like to underline that along the development of the above-mentioned strategic documents, much consideration has been given to standardization of the digitization process, image quality and metadata.

During the implementation period of the above-mentioned Strategy, the Minister of Culture of the Republic of Lithuania adopted the Order No. IV-6 (7 January 2010) “Concerning the adoption of the lists of standards and normative documents for creation, preservation and access of digital content” specifying standards used for digitized objects delivered to VEPS. In accordance with this Order, Lithuanian archives prepare ISAD(G) compatible exhaustive descriptions and ISAAR(CPF) compatible authority records; the internationally adopted ISBDs, UNIMARC and MARC21 are applied for the description of objects from libraries, and bibliographic descriptions of the holdings of the Lithuanian Art Museum follow the specification for the use of CCO and CDWA Lite. This makes it possible to standardize presentation of metadata and its integration into a virtual system of the cultural heritage. The established Virtual Electronic Heritage System, which is still being refined, could serve as an example of successful interoperability between descriptive metadata.
As it is depicted in the scheme, various memory institutions apply different descriptive metadata formats.

This is the architecture of the VEPS infrastructure under development. As it is evident from the scheme, all the metadata is converted into the UNIMARC format.

We apply the CIDOC CRM semantic model and its extension CIDOC CRM$_{dig}$, which was developed by FORTH (The Foundation for Research and Technology – Hellas) within the framework of the European Integrated Projects CASPAR and 3D-COFORM. CIDOC CRM$_{dig}$ is used for documenting all stages of the production and reusing of 3D-models of cultural heritage objects and sites, including all kinds of digitization processes and the digitized objects themselves.
One of the goals of the VEPS project was to establish a universal method and database for presentation of personal names, geographical names and historical chronology, which would serve for all memory institutions of Lithuania. For this purpose, the CIDOC CRM-based integrated thesaurus of personal names, geographical names and historical chronology (BAVIC, Bendras asmenvardžių, vietovardžių ir istorinės chronologijos tezaurus) has been created.

Integrated thesaurus (BAVIC) (authors Dr. R. Laužikas and Dr. V. Vaitkevičius) serving for:

- Establishment of a universal model for presentation of personal names, geographical names and historical chronology
- Creation of an integrated database of personal names, geographical names and historical chronology

The BAVIC database is based on the model of the BAVIC thesaurus. This database will be more exhaustive in comparison with authority DBs of individual memory institutions, as it will integrate personal names, geographical names and historical chronology from multiple institutions.
This segment of the BAVIC DB may help to illustrate the pursued objectives. The slide shows the location of the city of Marijampolė within the map of Lithuania and the evolution of this geographical name during the history. All this data is from the name authority file of the National Library, the public enterprise Registrų centras (Register Centre) and the Institute of the Lithuanian Language.

Another segment of the database illustrates the variety of languages and scripts (Lithuanian, Russian, Polish and German) that geographical names were recorded during the 20th c.

After the BAVIC thesaurus has been created, further effort was undertaken to integrate the rich digital content into a seamless whole (e.g., link a geographical name with information objects or related artworks). A conclusion has been made that BAVIC, which comprises only part of CIDOC CRM classes and properties, is not sufficient – its entities do not provide for complete realization of the
intrinsic (vidinės) logics (semantics) of CIDOC CRM. This was also approved by the international expert Dr. Mirna Willer. Therefore the following approach has been adopted:

- applying CIDOC CRM for the VEPS software to a degree that is necessary for delivery to the virtual space of collections preserved at memory institutions;
- setting up a database aggregating personal, corporate body and place names, and subject headings from multiple memory and research institutions of Lithuania;
- supplementing this database with additional data, such as characteristics of a person (occupation), geographical type of a location, hierarchical subordination between locations, etc.

Though the functionality of the system is not yet operating in full, testing of its software started. For the sake of visualization, I will present images from a search. Let us take Steponas Batoras, King of Poland and Grand Duke of Lithuania (1533-1586), and a manuscript written by him – the order concerning the complaint by the resident of Šiaulėnai volost Statkevičius of 1582.

The CIDOC CRM semantic model allows retrieving some additional information related to Steponas Batoras.

This is a painting by the Lithuanian painter Vincentas Smakauskas (1796-1876) “Steponas Batoras in the Act of Founding Vilnius University”. On the 1st of April 1579 Steponas Batoras signed the charter of privileges which granted the Jesuit [δ εξφο τ] College of Vilnius (founded in 1570) the status of a university.
This is a 19th century anonymous oil painting “Steponas Batoras Welcomes Turkish Emissaries [ευ σ ρ ζ]” done on canvas.

This is the armorial [α: μ :ρ λ] (herbinis) seal of Steponas Batoras made of wax and depicting an eagle with an armorial shield on its breast.
Now we can see the manuscript order by Steponas Batoras (dated 1586) addressed to inspectors concerning the measurement of the land in the village of Dainiai and assignment of the right of its ownership to a certain Darata Šmetienė under the exchange law. The language is Ruthenian [ρυ‹ Šινˆ … ν].

This is the order by Steponas Batoras (dated 1581, written in Polish) addressed to his vicar [ơ κ] in Vilnius Jonas Abramavičius (Abramowicz) and all citizenry [ơ τ щр ] (gyventojams) concerning the pogrom [ơ γρ μ] against evangelicals [1:ơ ν δ еλ κλζ].
And, finally, we can see an article from the Lithuanian periodical “Aušrelė” of 02/03/1935 about Steponas Batoras praising (giriantį) him as an ingenious [ ν δ i:ν σ] (sumanus) sovereign [ σ ϝρ ν] and skillful military leader.

It is evident from this example that users are given the opportunity to select the information that is of interest to them and navigate as broadly and deeply as they wish.

Now I will present the interaction of the first two entities within CIDOC CRM and CIDOC CRM\textsubscript{dig} as it is realized within VEPS.
From the point of view of the CIDOC CRM model, the model of the information system of VEPS is fully in line with the CIDOC CRM model. Presentation of data of the IS of VEPS in the RDF form will comply with the official scheme of CIDOC CRM.

Within VEPS, CIDOC CRM has an extension of CIDOC CRM which has the designation „1“ reflecting the Lithuanian context of the CIDOC CRM entities.

3. Archiving of metadata and digital objects

Here is an example of digital content and metadata in METS within VEPS.

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 needed for successful management and use of a 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 software for workflow management allowing to aggregate digital objects into integrated sets, where quality control, return for redigitization of corrupt images and linking with the optically recognized full-text file and the bibliographical object is performed. The export of objects with added descriptive metadata to the central database of VEPS is performed with the application of METS. The METS package used for data export includes extensive information related to the work and links to external objects (digital image files). 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 METS are used in the system:
1. METS Document Header;
2. Descriptive Metadata;
3. Administrative Metadata;
4. File Section;
5. Structural Map.
This is an example from the current system. Descriptions for digitized objects will be created in accordance with the established methodological guidance for libraries and archives; there is no such ready guidance for museums yet.

The procedures are aggravated (sunkina) 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: UNIMARC/MARC21/DC/ESE/EAD. If necessary, the virtual system could be easily extended by additional encoding standards for bibliographic description.

4. Web portal of the cultural heritage

Text (4.5 million)
- books
- manuscripts
- periodicals
- printed music
- church documents

Images (85 thousand)
- paintings
- photographs
- negatives
- textile works
- maps

Audio (85 thousand)
- first shellac records
- vocal music
- church vocal music
- instrumental music
- music for theatre

Integrated portal - www.epaveldas.lt

- Integration of bibliographic, archival and museum information within VEPS
- Integration of digital content within VEPS
- Conclusion: supports end-user navigation from concept to particular items, artefacts and archival documents
The portal “epaveldas.lt” has been established for access to VEPS via OAI-PMH and SRU. Since VEPS has been designed for provision of audio and visual information, digital objects within this system also contain descriptive metadata. The user interface allows switching to a page with a full-scale description within partners’ systems for more detailed examination. For this purpose, a reciprocal linking system has been introduced, i.e. (that is) there is a link between each work and the relevant description in the source system, and vice versa. The linking mechanism to objects within VEPS has been established to retain permanent HTTP links: the links will remain intact in spite of changes occurring within VEPS.

Search within the web portal will be based on semantic relationships defined by the CIDOC CRM model.
The data of BAVIC will be used to populate the following focal, appellation and other CIDOC CRM classes with the Lithuanian context: Person (E21), Group (E74), Place (E53), Conceptual Object (E28), Period (E4). The metadata of digitized objects will populate the focal class Information Object (E73) and its subclasses, thus creating links between the information object and the above-mentioned classes of BAVIC.

Application of this semantic principal will enable the users of the web portal to browse the BAVIC records (in addition to searching for digital objects) due to the application of CIDOC CRM dig.

The following access standards and protocols will be implemented for automatic systems used for aggregation and retrieval of information:

- BAVIC DB elements in the RDF form compatible with the CIDOC CRM RDF schema;
- SPARQL interface for searching BAVIC data in RDF;
- the OAI-PMH protocol for harvesting digitized cultural heritage objects;
- the SRU protocol for searching digitized cultural heritage objects.

Devoting all our effort to the development of the VEPS system, we anticipate the much desired result – creating seamless digital content of Lithuanian cultural heritage and access to it.