

**STRATEGIC, METHODOLOGICAL AND  
TECHNICAL SOLUTIONS FOR THE  
CREATION OF SEAMLESS CONTENT OF  
THE DIGITAL CULTURAL HERITAGE  
LITHUANIAN APPROACH**

**REGINA VARNIENĖ-JANSSEN**

*Martynas Mažvydas National Library of Lithuania, Vilnius, Lithuania*

**JONAS JUŠKYS**

*ASSECO Lietuva Ltd, Vilnius, Lithuania*

**ABSTRACT**

This paper elaborates on the issue of framing a conceptual approach for the creation of seamless content of the digital cultural heritage in Lithuania.

The first part of the paper focuses on a new paradigm in the Lithuanian legislation on cultural heritage as well as on financing issues. The second part describes the lessons learnt from the initiatives by the Martynas Mažvydas National Library of Lithuania and nine other memory institutions related to the creation of seamless content of the cultural heritage. It further outlines the application of the formal reference ontology CIDOC CRM and its extension CRM<sub>dig</sub> for different data formats, in particular, to support the implementation of automatic data transformation algorithms from local to national data structures (i.e. the Virtual Electronic Heritage System) without loss of meaning. It also discusses semantic querying of digital objects, their provenance and metadata by applying the Common Thesaurus of Personal Names, Geographical Names and Historical Chronology of Lithuania within the Virtual Electronic Heritage System.

**KEYWORDS**

legislation on cultural heritage, CIDOC CRM, CRM<sub>dig</sub>, RDF, metadata, digital objects, provenance modelling, BAVIC (Common Thesaurus of Personal Names, Geographical Names and Historical Chronology of Lithuania)

## Introduction

The recent strategic document of the European Commission “Europe 2020: A European Strategy for Smart, Sustainable and Inclusive Growth”<sup>1</sup> gives clear evidence that successful development of the EU countries will not be determined by purely economic factors, efficient working abilities, enterprise or premium service quality. The vision of the successful EU is built on changes in culture, thinking and behaviour as well as society’s values. One of the flagship initiatives of this strategic document, A Digital Agenda for Europe,<sup>2</sup> identifies seven priority action areas for building the information society and increasing social openness: development of a single digital market; enhanced interoperability; Internet trust and security; faster broadband; more sustainable investments in research and development; enhancing digital literacy, skills and inclusion; and increasing ICT benefits to society. It is obvious that the creation of digital content and access to it remain among the EU priorities in the development of the information society. The creation of digital content is a priority action area in Lithuania as well, which is reflected in the strategy for the future development of Lithuania and in programmes under preparation. The first part of the article describes the experience gained during the six years of successful effort in the above-mentioned areas.

To satisfy the user needs and those arising from the new paradigm in the Lithuanian cultural policy, namely, creation of digital content of the cultural heritage, the CIDOC CRM ontology has been applied for the first time in Lithuania. It allows linking metadata schemas used by diverse memory institutions, thus creating user-convenient virtual environment. As a formal reference ontology, CIDOC CRM particularly focuses on cultural heritage information and documentation.

The CIDOC CRM model and its extension CRM<sub>dig</sub> within the Virtual Electronic Heritage System (VEPS, *Virtuali elektroninio paveldo sistema*) serves as a generic background ontology for application modelling and interoperability tool, also for implementation of automatic

1 European Commission. Communication from the Commission : Europe 2020 : a strategy for smart, sustainable and inclusive growth : COM(2010) 2020. Brussels, 3.3.2010 [cited: 2011-10-20]. Available at: [http://europa.eu/press\\_room/pdf/complet\\_en\\_barroso\\_\\_007\\_-\\_europe\\_2020\\_-\\_en\\_version.pdf](http://europa.eu/press_room/pdf/complet_en_barroso__007_-_europe_2020_-_en_version.pdf)

2 European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions : a digital agenda for Europe : COM(2010) 245 final/2. Brussels, 26.8.2010. Pp. 7-8 [cited: 2011-10-20]. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF>

data transformation algorithms from local to national data structures (i.e. VEPS) without loss of meaning and supporting semantic queries against integrated resources. Though the CIDOC CRM model has been designed for universal application, the specific feature of its use in Lithuania is the fact that, in addition to the software based on CIDOC CRM and its extension CRM<sub>dig</sub>, the Common Thesaurus of Personal Names, Geographical Names and Historical Chronology of Lithuania (BAVIC, *Bendras Lietuvos asmenvardžių, vietovardžių ir istorinės chronologijos tezasauras*) has been created for the use by all memory institutions for semantic search.

### **A new paradigm in the Lithuanian legislation on cultural heritage**

The world being swept over by globalisation, issues related to the preservation of nationhood and national identity have become increasingly urgent. One of the major priorities of the EU cultural policy is fostering and promotion of the cultural heritage as a value system of the nation's living culture, which helps to maintain the national identity and provides possibilities for comprehensive expression of the national culture within the global community of cultures. This is confirmed by EU communications of the recent years and programmes encouraging digitisation of the cultural and scientific heritage and its integration into the European digital space. It is a new step towards consolidation of the assembly of European countries based on the existential right of every nation, which has the status of *lex naturale* and is realised through the nation's customs, traditions, arts and religion. It is by careful fostering and respecting of this right that harmonious coexistence of civilised nations is made possible. The new EU cultural policy, which acknowledges uniqueness, exceptional qualities and significance of every nation, is taking precisely this direction. The key EU documents on cultural policy also specify types of open access to the cultural heritage and knowledge without time and space limitations. They indicate that transferring cultural heritage into the digital form is an essential step providing for unrestricted access to and knowledge of the global variety of nations and cultures.

Archives, libraries and museums collect and preserve the precious national cultural heritage, which serves as the source for the nation's public spirit, self-awareness and self-esteem, and its educational, knowledge and leisure pursuits. Sectorial boundaries between these memory domains often become blurred when we take into consid-

eration institutions which, due to historical circumstances, sheltered cultural heritage assets that differed from the type of objects usually preserved by them. Though the “object type” factor still remains important in framing organisational structures of the memory institutions, it could hardly be argued, however, that a work of art, authorial manuscript of this work or review of this same work – all of them share the same cultural and historical context and provide evidence of comparable cultural features. To diminish this factor and encourage creation of cultural heritage content and the information about it, the Martynas Mažvydas National Library of Lithuania launched an initiative for establishing a consolidation approach for memory institutions in 2004, which was favoured by the Ministry of Culture of Lithuania. The outcome of this effort was adoption of the following strategic documents: the Concept for Digitisation of Lithuanian Cultural Heritage,<sup>3</sup> the Strategy for Digitisation of Lithuanian Cultural Heritage, Digital Content Preservation and Access and the Implementation Framework for the Strategy for Digitisation of Lithuanian Cultural Heritage, Digital Content Preservation and Access 2009–2013.<sup>4</sup> These documents as well as the project launched by the National Library together with the partners, the Lithuanian Art Museum and the Office of the Chief Archivist of Lithuania (until 1<sup>st</sup> January 2011, the Lithuanian Archives Department), resulted in origination of a new paradigm in the Lithuanian cultural policy – interaction of memory institutions in creating seamless content of the cultural heritage.

The Concept for Digitisation of Lithuanian Cultural Heritage provides for consistent and purposeful activities by public authorities targeted at preserving the national cultural heritage, improving access to it and retaining its currency by the use of modern information technologies. The Concept defines the cultural heritage as spiritual and material property and an authentic testimony to the beginnings and evolution of history, traditional culture, arts, print and other cultural domains. It also defines the scope of cultural heritage objects: “...various material and non-material products of social and individual activity; archaeo-

3 Lietuvos kultūros paveldo skaitmeninimo koncepcija : patvirtinta Lietuvos Respublikos Vyriausybės 2005 m. rugpjūčio 25 d. nutarimu Nr. 933. // Valstybės žinios 30, 105(2005), 5-6. Available also at:

[http://www3.lrs.lt/pls/inter2/dokpaieska.showdoc\\_l?p\\_id=260975](http://www3.lrs.lt/pls/inter2/dokpaieska.showdoc_l?p_id=260975) [cited: 2011-10-20].

4 Lietuvos kultūros paveldo skaitmeninimo, skaitmeninio turinio saugojimo ir prieigos strategija : patvirtinta Lietuvos Respublikos Vyriausybės 2009 m. gegužės 20 d. nutarimu Nr. 493. // Valstybės žinios 4, 66(2009), 19-29. Available also at: [http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc\\_l?p\\_id=389331](http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc_l?p_id=389331) [cited: 2011-10-20].

logical findings; artworks; manuscripts; published documents; objects of numismatic, sfragistic, heraldic or philatelic value; documents registering cultural phenomena (written sources; photographic, video and audio materials; and objects belonging to the National Documentary Fund); traditions; customs; dialects; onomastics; folklore; and other objects which are significant from the point of view of the cultural heritage.”<sup>5</sup> Natural heritage items of cultural and scientific significance are also considered as objects of the cultural heritage to be digitised. The Concept defines universal criteria for selecting cultural heritage objects for digitisation drawn up according to recommendations of the MINERVA programme. Age, uniqueness, content, scientific value and physical state of the cultural heritage object are considered the principal selection criteria. It also provides that coordination of cultural heritage digitisation is performed by the Ministry of Culture of the Republic of Lithuania together with the Ministry of Education and Science, the Information Society Development Committee under the Ministry of Transport and Communications, and the Office of the Chief Archivist of Lithuania. Initiation and coordination of digitisation programmes is the responsibility of the Council on Digitisation of the Lithuanian Cultural Heritage organised by the Minister of Culture. However, the Concept is not sufficient for the practical realisation of digitisation solutions. Therefore it was decided to prepare a strategy, special programmes and projects.

The principal purpose of the Strategy for Digitisation of Lithuanian Cultural Heritage, Digital Content Preservation and Access and its Implementation Framework 2009–2013 adopted by the Government of the Republic of Lithuania on 20<sup>th</sup> May 2009 is the 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 the digitised materials – the Martynas Mažvydas National Library of Lithuania, the Lithuanian Art Museum and the Office of the Chief Archivist of Lithuania. The establishment of such competence centres within national memory institutions must facilitate selection of cultural objects to be digitised and compilation of lists of such objects, thus eliminating overlap of procedures, saving financial resources and ensuring consistency of the entire digitisation process. The centres’ remit should include provision of manifold methodological assistance and training services for the professional commu-

5 Lietuvos kultūros paveldo skaitmeninimo koncepcija. Op. cit., p. 6.

nity, and concern about financial support. The vision of the Strategy is a single space of digital information about the Lithuanian cultural heritage extending the lifetime of cultural heritage objects; providing current, comprehensive and authoritative information about Lithuanian cultural heritage to the European and worldwide public; and ensuring permanent and reliable use of cultural heritage resources for research, educational and cultural purposes. The Strategy pursues the following goals and objectives:

- creation of digitisation infrastructure ensuring preservation and access of digitised cultural heritage objects and their integration into a single European digital space;
- creation and development of a seamless search, preservation and access system for the digital cultural heritage;
- encouraging qualifications development for digitisation professionals working at memory institutions;
- standardisation of digitisation processes, access and preservation of the cultural heritage;
- encouraging digitisation of Lithuanian cultural heritage and ensuring public access to digitised objects by supporting initiatives on the promotion of the cultural heritage.

It is important for institutions involved in digitisation of the cultural heritage to be aware of the perspective concerning their activities. The Programme for the Development of the Lithuanian Information Society in 2011–2019 sets as one of its objectives to ensure “...that the Lithuanian language survives in the global information society, in which the English language predominates. ICT could serve as a powerful means for disseminating information about the Lithuanian culture, thus contributing to preservation and support of the European cultural variety and promotion of the national self-expression. ICT provides memory institutions (libraries, archives, museums and other agencies preserving Lithuanian cultural heritage) with new possibilities for preservation, integration into the electronic space of the cultural heritage, and worldwide access of significant research, educational and art resources which are subject to deterioration with time”,<sup>6</sup> thus confirming that creation of the digital content of the national cultural heritage

6 Lietuvos informacinės visuomenės plėtros 2011–2019 metų programa : patvirtinta Lietuvos Respublikos Vyriausybės 2011 m. kovo 16 d. nutarimu Nr. 301. // Valstybės žinios 19, 33(2011), 20. Available also at: [http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc\\_l?p\\_id=394457&p\\_query=&p\\_tr2=](http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc_l?p_id=394457&p_query=&p_tr2=) [cited: 2011-10-20].

and providing access to it remains a priority for the national cultural policy. This strategic document could be considered as a follow-up of the investment area “Lithuanian language and culture” of the 3<sup>rd</sup> priority “Information society for all” of the Action Programme for Economic Growth in 2007–2013.<sup>7</sup>

These national documents provide strategic preconditions for creation of and access to the digital content and provide sustainable background for practical realisation of the Strategy for Digitisation. For example, in line with the Implementation Framework 2009–2013 of this Strategy, the following eight projects covering several cultural heritage domains (archives, museums and libraries) and multiple types of cultural heritage objects have been financed from the EU Structural Funds and the State budget:

- “Development of the Virtual Electronic Heritage System (VEPS)” (Martynas Mažvydas National Library of Lithuania): EUR 3 600 000;
- “Lithuanian Documentary Cinema Online (“e-kinas”)” (Lithuanian State Archives): EUR 3 000 000;
- “Historical Heritage of the Lithuanian Statehood Online” (Office of the Seimas of the Republic of Lithuania): EUR 900 000;
- “Virtual Historical Lithuania: The Grand Duchy of Lithuania” (Vilnius University): EUR 1 880 000;
- “Virtual Library for the Blind” (Lithuanian Library for the Blind): EUR 1 000 000;
- “Creation of the Virtual Audio Library of the Lithuanian Radio” (Lithuanian National Radio and Television) EUR 1 800 000;
- “Implementation of the Integrated Museum Information System (LIMIS) in Lithuanian Museums” (Ministry of Culture of the Republic of Lithuania): EUR 2 415 000;
- “Presenting Works of Lithuanian Classical Literature Online (“e-klasika”)” (Martynas Mažvydas National Library of Lithuania) EUR 580 000.

7 2007–2013 metų ekonomikos augimo veiksmų programa, Vilnius, 2007 m. liepos 5 d. P. 82 [cited: 2011-10-20]. Available at: [http://www.esparama.lt/ES\\_Paramas/strukturines\\_paramos\\_2007\\_1013m.\\_medis/titulinis/files/2VP\\_EA\\_2007-07-05.pdf](http://www.esparama.lt/ES_Paramas/strukturines_paramos_2007_1013m._medis/titulinis/files/2VP_EA_2007-07-05.pdf)

### **The Virtual Electronic Heritage System as the basis for seamless digital content of the cultural heritage**

In order to achieve interoperability between national memory institutions, a virtual system has been created within the project “Development of the Virtual Electronic Heritage System (VEPS)” launched on 3<sup>rd</sup> 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 and the State budget (EUR 3 600 000).

#### *Standardisation of metadata of cultural heritage objects*

The choice of different metadata schemas to be used within the framework of a digital library depends on the nature of data, treatment of objects in institutional collections and other factors.

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 began to be applied for most of the digital repositories and access systems within memory institutions in Lithuania. Prior to that, the absence of concern for application of shared standards had prevented consistency in the access to the digital content and its retrieval. Therefore the above-mentioned strategic documents on cultural heritage digitisation give special prominence to the necessity for shared standards. The goal has been set to have shared digitisation standards adopted in order to enable coordination of projects and initiatives undertaken by memory institutions and ensure compatibility between cultural heritage repositories on the national level and their integration into the European digital space. Currently the National Library and its nine project partners from different memory domains are the only institutions adhering to shared metadata, digital object archiving, preservation and access standards within the integrated system. With the involvement of other memory institutions, the number of metadata formats increased as well – at present UNIMARC, MARC21, ESE, EAD, CDWA Lite and DC are applied. The innovation at this stage of the project consists in the fact that VEPS has been designed for preservation of descriptive metadata of digitised objects, and not of the original works. Within this project, special guidelines defining a strict set of elements have been developed for descriptive metadata of digital objects.

During the implementation of the Strategy for Digitisation, the Minister of Culture of the Republic of Lithuania issued the Order No.IV-6 (7<sup>th</sup> January 2010) “Concerning the adoption of the lists of standards and normative documents for creation, preservation and access of the digital content”, which prescribes what standards must be applied for digitised objects and metadata: ISAD(G) for exhaustive descriptive metadata and ISAAR(CPF) for authority records (archives); ISBDs, UNIMARC and MARC21 for the description of digitised objects (libraries); and specifications for the use of CCO and CDWA Lite for descriptive metadata (museums).

*Semantic model for digital content of the cultural heritage*

Among the goals of the project “Development of the Virtual Electronic Heritage System (VEPS)” is creating an efficient ontology-based methodology for integrating the wealth of composite cultural heritage content and making it possible for users to retrieve the resources they need from different domains. For this purpose we apply CIDOC CRM, “... which is intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information”.<sup>8</sup> CIDOC CRM enhances accessibility to cultural heritage related information and knowledge, and provides an important information standard and reference model for Semantic Web initiatives. The specific aims of the application of CIDOC CRM and its extension CRM<sub>dig</sub><sup>9</sup> are:

- implementing automatic data transformation algorithms from local to national data structures (i.e. VEPS) without loss of meaning;
- applying a model for identifying provenance of digital objects, which in CRM<sub>dig</sub> is defined as metadata that establish chain-of-custody information needed for users to take trust decisions about digital data within VEPS;
- semantic querying of digital objects using the knowledge base of BAVIC and Semantic Web technologies.

8 Definition of the CIDOC Conceptual Reference Model / produced by the ICOM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group. Version 5.0.3, May 2011. P. i [cited: 2011-10-20]. Available at: [http://www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_5.0.3.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.3.pdf)

9 Doerr, Martin. CRM digital : a digital provenance ontology [cited: 2011-10-20]. Available at: [http://www.tpd12011.org/images/stories/tpdl/CRMDig\\_for\\_TPD1.pdf](http://www.tpd12011.org/images/stories/tpdl/CRMDig_for_TPD1.pdf); [http://www.ics.forth.gr/isl/rdfs/3D-COFORM\\_CRMdig\\_v2.5.rdfs](http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig_v2.5.rdfs)

The architecture of VEPS relies on the approach that it is more feasible to have mappings from many metadata schemas to a single core ontology than to apply mappings between numerous schemas. Figure 1 illustrates the centralised architecture of VEPS, which includes the manager and providers of VEPS. The Martynas Mažvydas National Library of Lithuania, which holds the main repository of digital content and the digitised objects of the highest resolution, is the manager and one of the providers of VEPS. The rest of the providers are memory institutions that are the project partners contributing digitised objects and their descriptive metadata, and institutions providing historical geographical names, personal names and historical chronology data to VEPS in accordance with the procedure laid down by cooperation agreements with the manager of VEPS.

Regarding the modelling methodology, all schemas are integrated by the use of CIDOC CRM, which functions as a universal schema allowing aggregation of digital objects and information related to them. For realisation of these processes, a module for the conversion of VEPS files and their metadata formats has been introduced, which ensures semantic interoperability between metadata from different domains. The integration occurs not only at the object level, but also at the level of the object-related information. To reinforce the semantic interoperability provided by the CIDOC CRM model, a syntactic interoperability tool for linking controlled vocabularies has been implemented (Figure 1).

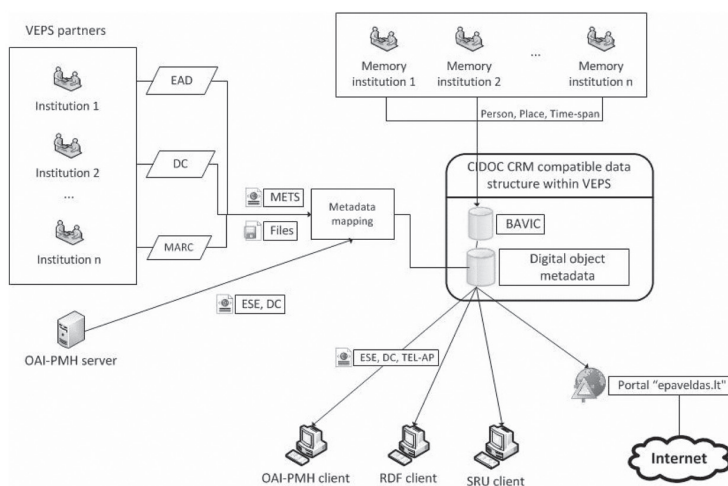


FIGURE 1.  
Architecture of the Virtual Electronic Heritage System

The extension of the CIDOC CRM ontology called CRM<sub>dig</sub> 2.5 is used for describing all stages of the production of digital cultural heritage objects, also capturing and modelling the query requirements regarding the provenance of digital objects. The following scheme presents an example of application of CIDOC CRM and CRM<sub>dig</sub> within VEPS (Figure 2).

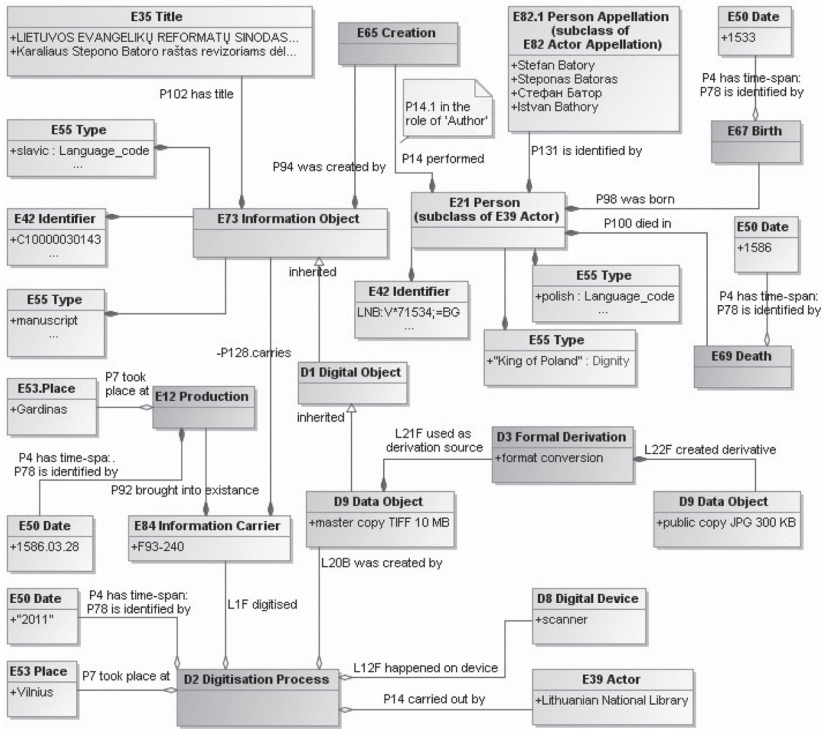


FIGURE 2. An instance of implementation of the simplified CIDOC CRM / CRM<sub>dig</sub> model within VEPS

**E73 Information Object** has the title (property P102) **E35 Title** *LIETUVOS EVANGELIKŲ REFORMATŲ SINODAS. Lietuvos ir Baltarusijos evangelikų reormatų vietovės. Dainiai. 240 : [Karaliaus Stepono Batoro raštas revizoriams dėl Dainių kaimo (Raseinių pav.) žemės apmatavimo ir perdavimo valdyti Daratai Šemetienei (Szemetowa) pagal*

*pasikeitimo teisę*]<sup>10</sup> and is carried by (property P128) **E84 Information Carrier**, which was brought into existence (property P92) by the event **E12 Production**, which comprises activities that are designed to and succeed in creating the item, took place at (property P7) **E53 Place Gardinas** and has time-span **E50 Date** (property P4) *1586*. **E84 Information Carrier** has the preferred identifier (property P48) **E42 Identifier**, which comprises the call number *F93-240* of the Manuscript Collection assigned to this instance of **E84 Information Carrier** in order to uniquely and permanently identify it within the context of the National Library. **E73 Information Object** was created by the event (property P94) **E65 Creation** that resulted in the creation of an immaterial product, which was carried out (performed) by (property P14) **E21 Person Stefan Batory** in the role of *Author* (property P14.1). The property **P98 was born** links the **E67 Birth** event to **E21 Person**, which is identified by **E42 Identifier** *LNB:V\*71534;=BG* and has the type (property P2 *King of Poland and Grand Duke of Lithuania : Dignity*) **E55 Type**. The event **D2 Digitisation Process** comprises events that result in the creation of instances of **D9 Data Object** that represent appearance and/or form of an instance of **E84 Information Carrier** (manuscript). The class **D9 Data Object** comprises instances of **D1 Digital Object** that are the direct result of a digital measurement, or a formal derivative of it, containing quantitative properties of some physical things (e.g. *Master copy*). In our example, the manuscript **E84 Information Carrier** is digitised by the event **D2 Digitisation Process**, which happens on a physical device (property L12F) **D8 Digital Device** *Digital camera Canon EOS-1 Ds Mark*, is carried out by (property P14) **E39 Actor** *Martynas Mažvydas National Library of Lithuania* and creates the file **D9 Digital Object** *Master copy 10 TIFF MB* stored within VEPS for long-term preservation. **D13 Digital Information Carrier** comprises all instances of **E84 Information Carrier** that are explicitly designed to be used as persistent digital physical carriers of instances of **D1 Digital Object**.

In addition to creating the instance of **D9 Digital Object** *Master copy*, the VEPS system automatically triggers resizing and format conversion event **D3 Formal Derivation** (subclass of **D7 Digital Machine Event**), which creates another **D9 Data Object** instance following a

10 THE SYNOD OF LITHUANIAN REFORMED EVANGELICALS. Locations of Lithuanian and Belorussian Reformed Evangelicals. *Dainiai*. 240 : [Steponas Batoras' written message to inspectors concerning the measurement of the land in the village of Dainiai (in the Raseiniai Powiat) and assignment of the right of its ownership to Darata Šemetienė (Szemetowa) under the exchange law].

deterministic algorithm, such that the resulting instance *Public copy 300KB JPEG* of the original digital object shares properties which represent the original object.

To reflect the semantics of entities in VEPS, some additional properties are used associated with an additional property to simulate specialisation of their parent property through the use of property subtypes declared as instances of **E55 Type**. According to the CRM documentation, they are designated with the extension “1”, e.g. **P14 carried out by (performed) by E39 Actor (P14.1 in the role of E55 Type)**. In our example, **E7 Activity** carried out by (property **P14**) **E21 Person Stefan Batory** is described by **P14.1** in the role of **E55 Type Author**. Distinction is made between Person appellations and Group appellations because of their properties extending the CRM class **E82 Appellation** with subclasses **E82.1 Person Appellation** and **E82.1 Group Appellation**.

#### *Archiving of metadata and digital objects*

It is globally recognised that METS<sup>11</sup> is the most convenient XML schema for creating XML document instances, which express the structure of digital library objects, 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 digital objects is both more extensive than and different from the metadata used for managing collections of printed and artworks or archival documents. The project activities included creation of software for workflow management allowing to aggregate digital objects into integrated sets, where quality control, return for redigitisation of corrupt images and linking with the optically recognised full-text file and 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 object and links to external objects (digital image or audio 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 five sections of METS are used in the system: Document Header, Descriptive Metadata, Administrative Metadata, File Section and Structural Map (Figure 3).

11 Metadata Encoding and Transmission Standard (METS). Official Web site [cited: 2011-10-20]. Available at: <http://www.loc.gov/standards/mets/>

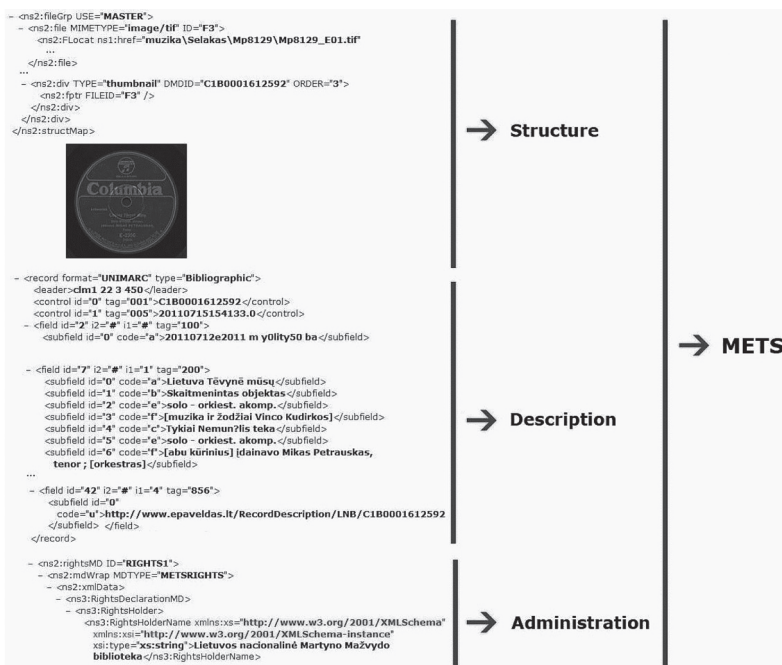


FIGURE 3.

An example of a digital object and its metadata in METS within VEPS

Descriptive metadata of digitised objects are created in accordance with the guidelines developed for libraries and archives. The guidelines for museums will be developed in the nearest future.

Preparation of separate records for digital objects enables obtaining more precise information about access to digitised objects, responsibility for their long-term preservation, storage location of the originals, etc. from management metadata.

### *Semantic querying of digital objects and metadata by using the BAVIC thesaurus*

Considering that query formulation and its resolving largely depends on the use of compatible thesauri, among the objectives of the VEPS project was the establishment of a universal method and thesaurus for presentation of Lithuanian personal names, geographical names and historical chronology, which would serve for all memory institutions of Lithuania. For this purpose, BAVIC is applied within the CIDOC

CRM compatible VEPS data structure. The BAVIC thesaurus brings together authority records created for different authority files. It is more exhaustive in comparison with authority databases of individual memory institutions, as it integrates data about persons, corporate bodies, historical and contemporary geographical names, administrative divisions, subjects, and chronology from multiple institutions. The benefit of the BAVIC thesaurus and its database lies in the representation of all names forms, which thus become equally eligible as access points. This database will be freely accessible for all project partners and other memory and research institutions.

At the first stage of the VEPS project, the BAVIC thesaurus was conceived exclusively as a model for personal names, geographical names and historical chronology. Further effort was undertaken to merge the rich digital content into a seamless whole (e.g. link a geographical name with information objects or related artworks). A conclusion was made that a thesaurus model comprising only some of the CIDOC CRM classes and properties is not sufficient – its data do not enable complete implementation of the intrinsic logic (semantics) of CIDOC CRM. Implementation of all the main classes of VEPS in compliance with CIDOC CRM was considered as a possible solution to this problem. It was also approved by the international expert Professor Dr. Mirna Willer. Therefore the following approach has been adopted:

- applying the whole of the CIDOC CRM model for describing within the VEPS databank collections preserved at memory institutions and presenting them online by using the VEPS software;
- supplementing BAVIC with personal, corporate body and geographical names as well as chronology and subject headings from multiple memory and research institutions of Lithuania with various data, such as characteristics of a person (occupation), geographical type of location, hierarchical relation between locations, etc.

The BAVIC data are used to populate the following classes of the CIDOC CRM ontology: **Person (E21)**, **Group (E74)**, **Place (E53)**, **Conceptual Object (E28)** and **Period (E4)**. The metadata of digitised objects populates the class **Information Object (E73)** and its related classes, thus creating links between the information object and the above-mentioned BAVIC classes. Provenance data of the digital object are also incorporated into the CIDOC CRM compatible data structure of VEPS.

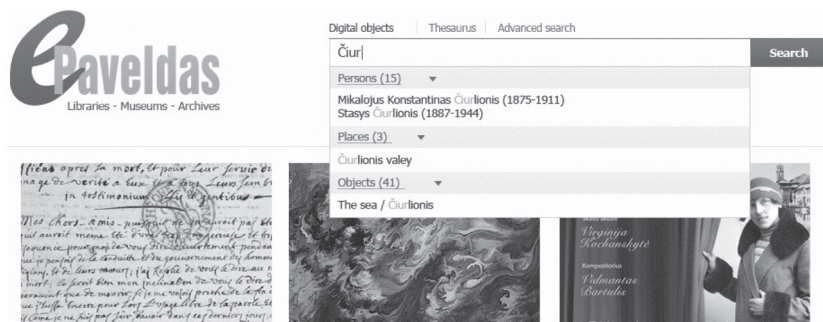


FIGURE 4.

An instance of search results within the portal "epaveldas.lt"

BAVIC has a pathfinder function and operates as a search aid within the portal "epaveldas.lt" (Figure 4). The search function includes pull-down menu options appearing instantly while the user types in the search box – persons, geographical names, groups and time periods. The resulting set is organised into subcategories of the selected category matching CIDOC CRM classes. For example, subcategories of the Place category are the following: administrative and geographical subdivisions, types of place (town, lake, etc.) and countries. Subcategories of the Person category include countries, a list of occupations from the controlled vocabulary, languages, dates of birth and death grouped by centuries, etc. The subcategories also match the CIDOC CRM classes intended to be used in searching, e.g. the occupation of the person (E21 Person) is the subclass E74.1 Occupation of the class E74 Group; the link between a person and occupation is defined in CIDOC CRM through the event E85 Joining: E21 Person is linked (property P143 joined (joined by)) by the event E85 Joining, which links a person with (property P144 joined with (gained member by)) the group E74.1 Occupation.

Another feature of the search within the portal is variant searching. Homonymous names of different entities worsen the results. To diminish these issues, two approaches have been developed. In the first place, selecting the category and its subcategory by the user minimises the occurrence of homonyms. Secondly, we use an internal relevance ranking mechanism based on the comprehensive structure of the extended CRM classes E82.1 Person Appellation and E82.1 Group Appellation, thus enabling the most relevant items to come first.

The developed model serves as a highway for the flow of digitised cultural assets and information about them from local to national systems and eventually to the global community.

### Conclusions

The cooperative effort by Lithuanian memory institutions resulted in the adoption of the following strategic documents obliging archives, libraries and museums to ensure the integrity of cultural heritage and information about it by digitisation and provision of convenient access: the Concept, the Strategy and its Implementation Framework 2009–2013. These documents as well as the VEPS project launched by the National Library and partners resulted in a shared methodological approach to cultural heritage digitisation and access irrespective of the type, provenance, content or structure of the object.

The Virtual Electronic Heritage System (VEPS) is an extended digital library supporting traditional functions of long-term preservation of heritage objects and providing assistance in retrieval of these objects. It also is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic instrument that any cultural information can be mapped to.

The CIDOC CRM model and its extension CRM<sub>dig</sub>, as well as Semantic Web technologies and the BAVIC thesaurus, are applied for aggregating and online access of the diverse digital content within VEPS, implementing automatic data transformation algorithms from local to national data structures (i.e. the Virtual Electronic Heritage System) without loss of meaning, and semantic querying of digital objects, provenance and metadata as well as their long-term preservation.

### References

- 2007–2013 metų ekonomikos augimo veiksmų programa, Vilnius, 2007 m. liepos 5 d. [cited: 2011-10-20]. Available at: [http://www.esparama.lt/ES\\_Paramas/strukturines\\_paramos\\_2007\\_1013m.\\_medis/titulinis/files/2VP\\_EA\\_2007-07-05.pdf](http://www.esparama.lt/ES_Paramas/strukturines_paramos_2007_1013m._medis/titulinis/files/2VP_EA_2007-07-05.pdf)
- Definition of the CIDOC Conceptual Reference Model / produced by the ICOM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group. Version 5.0.3, May 2011 [cited: 2011-10-20]. Available at: [http://www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_5.0.3.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.3.pdf)
- Doerr, Martin. CRM digital : a digital provenance ontology [cited: 2011-10-20]. Available at: [http://www.tpd12011.org/images/stories/tpdl/CRMDig\\_for\\_TPD1.pdf](http://www.tpd12011.org/images/stories/tpdl/CRMDig_for_TPD1.pdf); [http://www.ics.forth.gr/isl/rdfs/3D-COFORM\\_CRMDig\\_v2.5.rdfs](http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMDig_v2.5.rdfs)

- European Commission. Communication from the Commission : Europe 2020 : a strategy for smart, sustainable and inclusive growth : COM(2010) 2020. Brussels, 3.3.2010 [cited: 2011-10-20]. Available at: [http://europa.eu/press\\_room/pdf/complet\\_en\\_barroso\\_\\_007\\_-\\_europe\\_2020\\_-\\_en\\_version.pdf](http://europa.eu/press_room/pdf/complet_en_barroso__007_-_europe_2020_-_en_version.pdf)
- European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions : a Digital Agenda for Europe : COM(2010) 245 final/2. Brussels, 26.8.2010 [cited: 2011-10-20]. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF>
- Lietuvos informacinės visuomenės plėtros 2011–2019 metų programa : patvirtinta Lietuvos Respublikos Vyriausybės 2011 m. kovo 16 d. nutarimu Nr. 301. // Valstybės žinios 19, 33(2011), 19-25. Available also at: [http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc\\_l?p\\_id=394457&p\\_query=&p\\_tr2=](http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc_l?p_id=394457&p_query=&p_tr2=) [cited: 2011-10-20].
- Lietuvos kultūros paveldo skaitmeninimo koncepcija : patvirtinta Lietuvos Respublikos Vyriausybės 2005 m. rugpjūčio 25 d. nutarimu Nr. 933. // Valstybės žinios 30, 105(2005), 5-6. Available also at: [http://www3.lrs.lt/pls/inter2/dokpaieska.showdoc\\_l?p\\_id=260975](http://www3.lrs.lt/pls/inter2/dokpaieska.showdoc_l?p_id=260975) [cited: 2011-10-20].
- Lietuvos kultūros paveldo skaitmeninimo, skaitmeninio turinio saugojimo ir prieigos strategija : patvirtinta Lietuvos Respublikos Vyriausybės 2009 m. gegužės 20 d. nutarimu Nr. 493. // Valstybės žinios 4, 66(2009), 19-29. Available also at: [http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc\\_l?p\\_id=389331](http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc_l?p_id=389331) [cited: 2011-10-20].
- Metadata Encoding and Transmission Standard (METS). Official Web site [cited: 2011-10-20]. Available at: <http://www.loc.gov/standards/mets/>

## Further reading

- Building a heterogeneous network of digital libraries on the Semantic Web / Sebastian Ryszard Kruk, Mariusz Cygan, Piotr Piotrowski, Krystian Samp, Adam Westerski, Stefan Decker [cited: 2011-10-20]. Available at: <http://semdl.info/presentations/8>
- Doerr, Martin. The CIDOC conceptual reference module : an ontological approach to semantic interoperability of metadata. // AI magazine 24, 3(2003), 75-92 [cited: 2011-10-20]. Available at: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1720/1618>
- Doerr, Martin. DELOS task 5.5 – ontology-driven interoperability [cited: 2011-10-20]. Available at: [http://www.delos.info/files/pdf/events/paris06/T5.5\\_Ontology-Driven\\_Interop\\_Doerr.pdf](http://www.delos.info/files/pdf/events/paris06/T5.5_Ontology-Driven_Interop_Doerr.pdf)
- Doerr, Martin. Harmonized EDM-CRM-FRBRoo-CRMdig [cited: 2011-10-20]. Available at: [http://www.tpd12011.org/images/stories/tpdl/EDM-DC-ORE-CRM-FRBR-CRMdig\\_Integration.pdf](http://www.tpd12011.org/images/stories/tpdl/EDM-DC-ORE-CRM-FRBR-CRMdig_Integration.pdf)
- Doerr, Martin. Modelling intellectual processes : the object-orient[ed] FRBR model. Berlin, Germany, September 25, 2011 [cited: 2011-10-20]. Available at: [http://www.tpd12011.org/images/stories/tpdl/EDM\\_for\\_TPDL.pdf](http://www.tpd12011.org/images/stories/tpdl/EDM_for_TPDL.pdf)
- Doerr, Martin; Athina Kritsotaki. Documenting events in metadata : the 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2006) [cited: 2011-10-20]. Available at: <http://www.cidoc-crm.org/docs/fin-paper.pdf>

- Doerr, Martin; Athina Kritsotaki; Steven Stead. Sample of a period thesaurus [cited: 2011-10-20]. Available at: <http://www.cidoc-crm.org/docs/period/period.html>
- Doerr, Martin; Athina Kritsotaki; Stephen Stead. Which period is it? : a methodology to create thesauri of historical periods [cited: 2011-10-20]. Available at: [http://www.cidoc-crm.org/docs/ParousCAA\\_4.ppt](http://www.cidoc-crm.org/docs/ParousCAA_4.ppt)
- Doerr, Martin; Maria Theodoridou. CRMDig : a generic digital provenance model for scientific observation [cited: 2011-10-20]. Available at: [http://www.usenix.org/event/tapp11/tech/final\\_files/Doerr.pdf](http://www.usenix.org/event/tapp11/tech/final_files/Doerr.pdf)
- Doerr, Martin; Stephen Stead. Harmonized models for the digital world: CIDOC CRM, FRBROO, CRMDig, EDM. Berlin, Germany, September 25, 2011 [cited: 2011-10-20]. Available at: [http://www.tpd12011.org/images/stories/tpdl/crm\\_for\\_TPD1.pdf](http://www.tpd12011.org/images/stories/tpdl/crm_for_TPD1.pdf)
- Erlangen CRM / OWL 2011-04-04 based on CIDOC CRM 5.0.2 January 2010 [cited: 2011-10-20]. Available at: <http://erlangen-crm.org/previous-versions>
- Joiner, Gregory; Douglas Reid. Implementing digital provenance on the World Wide Web using Semantic Web technology [cited: 2011-10-20]. Available at: [http://semtech2011.semanticweb.com/uploads/handouts/THURS\\_1110\\_Joiner\\_3982.pdf](http://semtech2011.semanticweb.com/uploads/handouts/THURS_1110_Joiner_3982.pdf)
- Le Boeuf, Patrick. Using an ontology-driven system to integrate museum information and library information : paper presented on the occasion of the Symposium on Digital Semantic Content Across Cultures, Paris, the Louvre, 4-5 May 2006 [cited: 2011-10-20]. Available at: [http://www.seco.tkk.fi/events/2006/2006-05-04-websemantique/presentations/articles/LeBoeuf-DigSemAcrossCultures\\_200605\\_LeBoeuf.pdf](http://www.seco.tkk.fi/events/2006/2006-05-04-websemantique/presentations/articles/LeBoeuf-DigSemAcrossCultures_200605_LeBoeuf.pdf); [http://www.cidoc-crm.org/docs/DigSemAcrossCultures\\_200605\\_LeBoeuf.ppt](http://www.cidoc-crm.org/docs/DigSemAcrossCultures_200605_LeBoeuf.ppt)
- LIDO – Lightweight Information Describing Objects. Version 1.0, November 2010 / Erin Coburn, Richard Light, Gordon McKenna, Regine Stein, Axel Vitzthum [cited: 2011-10-20]. Available at: <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>
- Lourdi, Irene; Christos Papatheodorou; Martin Doerr. Semantic integration of collection description : combining CIDOC/CRM and Dublin Core Collections Application Profile. // D-Lib magazine 15, 7/8 (July/August 2009) [cited: 2011-10-20]. Available at: <http://www.dlib.org/dlib/july09/papatheodorou/07papatheodorou.html>
- Modeling and querying provenance by extending CIDOC CRM / Maria Theodoridou, Yannis Tzitzikas, Martin Doerr, Yannis Marketakis, Valantis Melessanakis. // Distributed and parallel databases 27, 2 (2010), 169–210 [cited: 2011-10-20]. Available at: <http://www.springerlink.com/content/53225114q136g68v/full-text.pdf>
- The role of ontologies in semantic digital libraries / Sebastian Ryszard Kruk, Bernhard Haslhofer, Piotr Piotrowski, Adam Westerski, Tomasz Woroniecki [cited: 2011-10-20]. Available at: <http://www.ukoln.ac.uk/nkos/nkos2006/papers/paper-013.pdf>
- Using an ontology for interoperability and browsing of museum, library and archive information / Patrick Le Boeuf, Patrick Sinclair, Kirk Martinez, Paul Lewis, Geneviève Aitken and Christian Lahanier [cited: 2011-10-20]. Available at: [http://eprints.ecs.soton.ac.uk/14790/1/D4\\_Le\\_Boeuf.pdf](http://eprints.ecs.soton.ac.uk/14790/1/D4_Le_Boeuf.pdf)

### **Biographical sketch**

**Dr. Regina Varnienė-Janssen** has been Director of the National Centre for Digitisation and Virtual Electronic Heritage System at the Martynas Mažvydas National Library of Lithuania since January 2011. She is the author of over 80 scientific and methodological publications on bibliographic control, cataloguing, bibliographic and publishing standards, design of information systems, development of the Lithuanian Integrated Library Information System, Virtual Electronic Heritage System, in the Lithuanian and international professional journals.

She actively participated in the activities of various national and international organisations. She was chair of the Technical Committee 47 “Information and documentation” of the Lithuanian Standards Board (1998–2007), a member of the IFLA Permanent UNIMARC Committee (1999–2005) and the Governing Board of the ISSN International Centre (2000–2007). At present she is a member of the Digitisation Board at the Ministry of Culture of the Republic of Lithuania, the Standing Committee of the IFLA Bibliography Section (since 2003), the Scientific Board of the Martynas Mažvydas National Library of Lithuania (since 1992) and the Governing Board of the Lithuanian Integrated Library Information System (LIBIS) (since 1996). She is the manager of the national project “Development of the Virtual Electronic Heritage System (VEPS)” financed from the European Structural Funds. The Project started at the beginning of 2010, and it will be finally implemented in the midsummer of 2012.

She is assistant professor at the Institute of Library and Information Sciences of Vilnius University’s Faculty of Communication. She teaches courses on Information Processing and Retrieval, and Management of Cultural Heritage Projects.

**Jonas Juškys** is a project manager at ASSECO Ltd, Vilnius, Lithuania. He has been involved in the following key projects: “Creation of the Lithuanian Integrated Library Information System (LIBIS)”, “Creation of the Information System of the National Library of Belarus”, and “Development of the Virtual Electronic Heritage System (VEPS)”.

## **STRATEŠKA, METODOLOŠKA I TEHNIČKA RJEŠENJA ZA STVARANJE POVEZANOG SADRŽAJA DIGITALNE KULTURNE BAŠTINE LITVANSKI PRISTUP**

### **Sažetak**

Razmatra se oblikovanje konceptualnog pristupa za stvaranje povezanog sadržaja digitalne kulturne baštine u Litvi. Prvi se dio rada usredotočuje na novu paradigmu u litvanskom zakonodavstvu o kulturnoj baštini i na financijska pitanja. Drugi dio opisuje što je naučeno iz inicijative Nacionalne knjižnice Litve Martynas Mažvydas i još devet baštinskih ustanova vezanih za stvaranje povezanog sadržaja kulturne baštine. U glavnim crtama prikazuje se primjena formalne referentne ontologije CIDOC CRM i njezinog dodatka CRM<sub>dig</sub> za različite formate podataka, a posebno za podršku implementacije algoritma za automatsku transformaciju podataka iz lokalnih u nacionalne podatkovne strukture (odnosno Virtualni elektronički baštinski sustav) bez gubljenja podataka. Raspravlja se i o semantičkom ispitivanju digitalnih objekata, njihovom podrijetlu i metapodacima primjenom BAVIC-a (Zajedničkog tezaurusa osobnih imena, geografskih naziva i povijesne kronologije Litve) unutar Virtualnog elektroničkog baštinskog sustava.

**Ključne riječi:** zakonodavstvo o kulturnoj baštini, CIDOC CRM, CRM<sub>dig</sub>, RDF, metapodaci, digitalni objekti, modeliranje podrijetla, BAVIC (Zajednički tezaurus osobnih imena, geografskih naziva i povijesne kronologije Litve)