



Informacijos mokslai  
2020, vol. 88, pp. 66–82

ISSN 1392-0561 eISSN 1392-1487  
DOI: <https://doi.org/10.15388/Im.2020.88.32>

# Ontologies and Technologies for Integrating and Accessing Digital Cultural Heritage: Lithuanian Approach

**Regina Varnienė-Janssen**

Faculty of Communication, Vilnius University  
E-mail: [regina.varniene-janssen@kf.vu.lt](mailto:regina.varniene-janssen@kf.vu.lt)

**Albertas Šermokas**

Faculty of Mathematics and Informatics, Vilnius University  
E-mail: [albertas.sermokas@mif.vu.lt](mailto:albertas.sermokas@mif.vu.lt)

**Summary.** Web technologies are the key for the implementing and ensuring the full range of user needs in the digital age. On the other hand, the issue of unified representation of digital content from diverse memory institutions in order to ensure semantic integrity still remains a matter of urgency. Semantic interoperability of information and data is essential in an integrated system. In this paper, we analyze and describe an ontology-based metadata interoperability approach and how this approach could be applied for memory institution data from diverse sources which do not support ontologies. In particular, we describe the use of the CIDOC CRM ontology as a mediating schema within Lithuania's Information System of the Virtual Electronic Heritage (hereinafter "VEPIS"). The paper introduces the role of the CIDOC CRM based Thesaurus of Personal Names, Geographical Names and Historical Chronology (hereinafter "BAVIC"), which operates as a core ontology within VEPIS by allowing to understand things and relationships between things as well as identify the time and space of things. The paper also focuses on trust of the cultural information on the Web. Users make trust judgments based on provenance that may or may not be explicitly offered to them. In particular, we describe how provenance is managed within digital preservation and access processes within VEPIS and define whether this management meets the W3C Provenance Incubator Group's Requirements for Provenance on the Web. The paper is based on the results of the research initiated in 2018–2019 at the Faculty of Communication and the Faculty of Mathematics and Informatics of Vilnius University by authors of this paper.

**Keywords:** Ontology based integration, metadata interoperability, CIDOC CRM, provenance, Requirements for Provenance on the Web

## Ontologijos ir technologiniai sprendimai skaitmeninio kultūros paveldo integravimui ir prieigai: Lietuvos patirtis

**Santrauka.** Saityno technologijos sudaro galimybę tenkinti įvairiapusiškas informacines skaitmeninės eros vartotojų reikmes. Kita vertus, iki šiol aktuali problema išlieka atminties institucijų į saityną teikiamo skaitmeninio turinio semantinis integralumas. Informacijos ir duomenų turinio semantinis suderinamumas ypač aktualus integruotoms sistemoms. Straipsnyje apibūdinama ontologijomis grindžiamų metaduomenų koncepcija. Straipsnyje aprašomas

**Received:** 29/10/2011. **Accepted:** 03/02/2020

Copyright © 2019 Regina Varnienė-Janssen, Albertas Šermokas. Published by Vilnius University Press. This is an Open Access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

CIDOC/CRM ontologijos kaip tarpininkavimo schemas vaidmuo VEPIS sistemoje. Straipsnis taip pat supažindina su Asmenvardžių, vietovardžių ir istorinės chronologijos tezauru (BAVIC), VEPIS atliekančiu pamatinės ontologijos vaidmenį (leidžia suprasti esybes ir jų santykius, jų santykį su laiku ir erdve). Kita straipsnyje analizuojama problema yra susijusi su kultūros informacijos turinio patikimumu saityne. Vartotojai apie informacijos ir duomenų patikimumą sprendžia remdamiesi proveniencija, kuri gali arba negali būti jiems tiesiogiai pateikiama. Straipsnyje analizuojama, kaip proveniencija yra valdoma VEPIS skaitmeninto turinio ilgalaikio išsaugojimo ir jos sklaidos procesų metu, ir kartu nustatoma, ar šie procesai atitinka proveniencijos saityne *W3C Provenance Incubator Group* reikalavimus. Straipsnyje remiamasi Vilniaus universiteto Komunikacijos fakulteto ir Matematikos fakulteto 2018–2019 m. straipsnio autorių inicijuoto tyrimo rezultatais.

**Pagrindiniai žodžiai:** ontologija pagrįstas suderinamumas, metaduomenų suderinamumas, CIDOC CRM, proveniencija, reikalavimai proveniencijai saityne (Requirements for Provenance on the Web).

---

## Introduction

Web technologies are the key for the implementing and ensuring the full range of user needs in the digital age which were identified in the Recommendations<sup>1</sup> of the World Wide Web Consortium (W3C) as *Web for All*, *Web on Everything*, *Web for Rich Interaction*, *Web of Data and Services*, and *Web of Trust*. This paper analyses two aspects of the information on the Web: *Web for Rich Interaction* and *Web of Trust*.

Various cultural domains provide different sets of digital cultural content: descriptions (metadata) of fine art, archaeological and cultural heritage, archival documents, literature, natural history, geology and collections of digital objects. Currently, memory institutions seek to integrate digital information and data resources because of the well-known advantages of this integration. Such integration raises the problem of interoperability of data and information. In this paper we describe an ontology-based metadata interoperability approach. In particular, we describe the use of the CIDOC CRM ontology as a mediating schema within VEPIS. However, it is not a mere technological issue: we have to achieve the semantic interoperability as well. The Semantic Web, which is based on Semantic Web technologies (RDF, OWL, SKOS, SPARQL, etc.) provides a language that expresses both data and rules for reasoning about data. By using Web languages, such as RDF and OWL, it is possible to create semantically rich data models. Drawing on the results from the research by Diego Calvanese, et al. (1998), Michael Doerr, et al. (2002), (2003), David Giaretta (2011), Gordon Dunsire and Mirna Willer (2018), Oreste Signore (2007) and Crofts, et al. (2001), we could assume that a core ontology is a key structural block essential for enabling the integration of information from diverse sources. However, the problem is that the VEPIS sources do not use any common ontology that could serve as a common ground for semantic interoperability. Within VEPIS, the role of such an ontology is performed by BAVIC, which is based on CIDOC CRM and diverse sources in such a way that relations between various entities are main carriers of semantic information and provide semantic interoperability of VEPIS objects.

With the increasing amount of data available on the Web, users need to have reliable means to obtain their provenance to decide whether they can trust information they ac-

---

<sup>1</sup> Web3C Mission. Available at: <https://www.w3.org/Consortium/mission>.

cess. The paper focuses on provenance in the digital environment, which provides critical foundation for *assessing authenticity, enabling trust and allowing reproducibility*. Users make trust judgments based on provenance that may or may not be explicitly offered to them. The paper analyses how provenance is managed within digital preservation and access processes within VEPIS and defines whether it meets the W3C Provenance Incubator Group's Requirements for Provenance on the Web. The paper presents the results from the research initiated at Vilnius University in 2018–2019.

*Methodology of the research.* We use qualitative analysis of research papers and draw on the potential of using methods of information technologies for integrating heterogenic collections of the digital culture within integrated systems and their presentation on the Semantic Web. Another methodological tool is the *Requirements for Provenance on Web*, which define *Content Category* referring to what types of information would need to be represented in a provenance record, *Management Category* referring to mechanisms that make provenance available and accessible within a system and *Use Category* implying that provenance records may need to accommodate a variety of uses as well as diverse users. The application of the Requirements for Provenance on the Web for a particular integrated system (VEPIS) contributes to the novelty of the research.

## 1. Ontology for integration of data and information

### 1.1. From metadata to an object-oriented approach

According to Oreste Signore<sup>2</sup>, data integration at a metadata level “<...> cannot exploit the full richness of possible associations among different information items. The association mechanism remains in mind of the user”. Since a metadata vocabulary schema does not organise entities into a hierarchy, Oreste Signore argues that an ontology based metadata relationship is a very efficient way of integrating information. The value of ontologies, in particular CIDOC CRM, for integrating information from various domains and different data schemas is presented in papers by various authors (Crofts, Doerr & Gill, 2003; Kakali, et al. 2007). According to Constanca Kakali et al. (2007), metadata are used to describe resources in terms of elements and facilitate discovery and access to information. Ontologies define entities on an abstract level with the insertion to conceptualize a domain interest. Ontologies do not provide specific elements for the description of the resource.

With the involvement of ten memory institutions within VEPIS in 2010, the number of metadata formats from source systems increased as well: at present MARC (UNIMARC, MARC 21), ESE, EAD, CDWA Lite and DC are applied as it illustrated in Fig 1. The architecture of VEPIS relies on the approach that it is more feasible to have mappings from many metadata schemas to a single a core ontology, (CIDOC CRM) than to apply mappings between numerous schemas.

<sup>2</sup> Signore, O. (2007). The Semantic Web and Cultural Heritage: Ontologies and Technologies Help in Accessing Museum Information. Available at: <https://www.semanticscholar.org/paper/The-Semantic-Web-and-Cultural-Heritage-%3A-Ontologies-Signore/ab4f0ee826bbea097f4d7aa4d5244b67c0caea6>.

Table 1 presents a fragment of the mapping of some new subfields of Field 325 *Reproduction Note* of UNIMARC/B (2017) to CIDOC CRM and CRMdig. In the context of digital provenance, of particular importance are the following subfields of Field 325 *Reproduction Note*: 325\$a, 325\$b, 325\$c, 325\$d, 325\$e, 325\$f, 325\$g, 325\$h, 325\$i, 325\$n, 325\$u, 325\$v, 325\$x, 325\$y and 325\$z. Although at the development phase of VEPIS in 2010-2012, some subfields of 325 *Reproduction Note* were implemented to provide important information on digital objects regarding provenance, the data were modelled on the basis of the MARC 21 because their equivalents in UNIMARC at that time did not exist in the official mapping schema. Therefore the authors of this paper consider that it is appropriate to move to the formal structure of the UNIMARC/B fields and update the mapping of UNIMARC/B to CIDOC CRM and CRMdig in line with the updates published on the website of IFLA.<sup>3</sup> Since the resources are tightly connected, the integration of the new updates in the UNIMARC /A presented on the IFLA website<sup>4</sup> will ensure the presentation of more related data about the entities: one can move from a digital copy to other copies of the same publication, to other editions, to other works of the same author, to works about the author and so on. If the data is verified and supplemented with external sources and different author appellations are correctly associated with the relevant author, it is much easier for a user to be satisfied with the data.

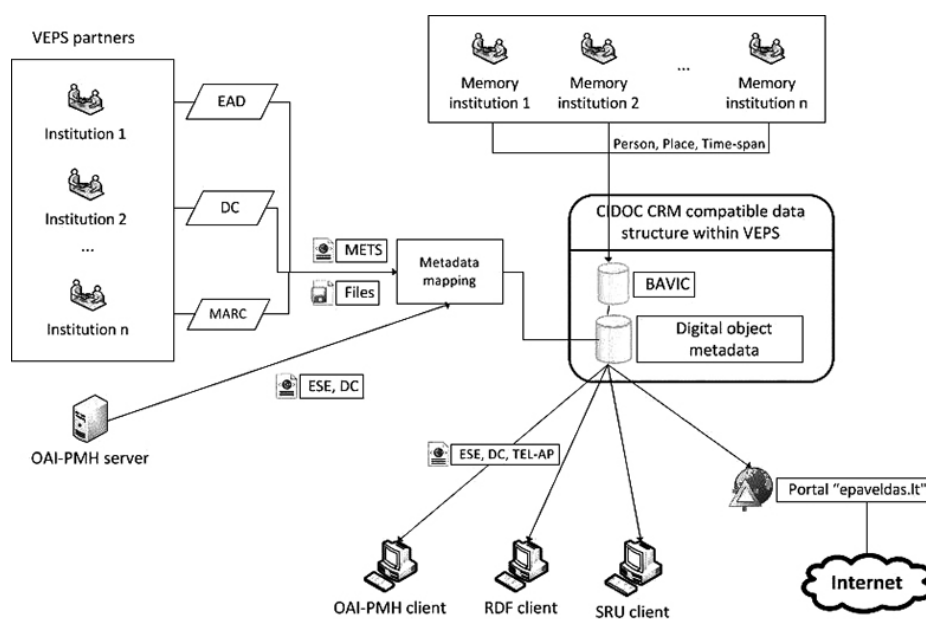


Fig. 1. Workflows of descriptive metadata within VEPIS

<sup>3</sup> UNIMARC bibliographic, 3<sup>rd</sup> edition( with updates) . IFLA. Available at: <https://www.ifla.org/publications/unimarc-bibliographic--3rd-edition--updates-2012-and-updates-2016>.

<sup>4</sup> UNIMARC Authorities, 3<sup>rd</sup> edition ( with updates). Available at : IFLA <https://www.ifla.org/publications/unimarc-authorities--3rd-edition--updates>.

For ingest data to VEPIS central database and creating XML document instances METS<sup>5</sup> is used. It expresses the structure of digital library objects, associated descriptive and administrative metadata and the names and locations of files that comprise the digital object. The description of the object in METS serves as a linking element between various parts of the document and its versions. Regarding the mapping methodology, all schemas used by various partners are integrated by converting descriptive metadata to UNIMARC/B and mapping to CIDOC CRM, which functions as a universal schema allowing aggregation of digital objects and information related to them.<sup>6</sup> All this enables data exchange and data integration. It allows bringing together disparate data sources and combining them into a single stream of data. The event-centric core model CIDOC CRM, which contains 86 classes and 137 properties in RDF and OWL, is a major step from entity-relation to object-oriented approach. The use of CIDOC CRM instead of a set of metadata elements opens up more advanced search and resource discovery possibilities.

Table 1. A fragment of the mapping of UNIMARC/B (2017) to CIDOC CRM prepared by the authors

UNIMARC/B field		CIDOC CRM class domain	CRM property	CRM range class
<b>325</b>	<b>Reproduction note</b>			
<b>325\$a</b>	<b>Text of unstructured note</b> Used only for the complete text of an unstructured note	D1 <i>Digital Object</i> (subclass of E73), (instance = the publication exemplified by the item being described)	P3 <i>Has Note</i>	E62 <i>String</i> (value = "content of 320\$a")
<b>325\$b</b>	<b>Type of reproduction</b> The mode of reproduction (e.g. digitization)	D2 <i>Digitization Process</i>	L1 <i>Digitized (Was_Digitized_by)</i> : E18 <i>Physical Thing</i>	E18 <i>Physical Thing</i>
<b>325\$c</b>	Place where the reproduction is published or distributed	E7 <i>Activity</i>	L 29 <i>Has Responsible Organization (is Responsible Organization for) Digitization</i>	E40 <i>Legal Body</i> (value = "content of subfield 210\$a" of the record that would be established to describe the reproduction")
<b>325\$d</b>	The name of the agency that makes the reproduction available	E7 <i>Activity</i>	L29 <i>Has Responsible Organization (is Responsible Organization for) Distribution</i>	E40 <i>Legal Body</i> (value = "content of subfield 210\$c" The name of the agency that makes the reproduction available")

<sup>5</sup> Metadata Encoding and Transmission Standard (METS). Official Web site: Available at: <https://www.loc.gov/standards/mets/>

<sup>6</sup> CIDOC CRM. Available at: (<http://www.cidoc-crm.org/versions-of-the-cidoc-crm> (<http://www.cidoc-crm.org/>)).

## 1.2. The BAVIC Thesaurus for integrating data about Agents, Places and Time

There have been done many attempts on unified representation of names, dates, and places, which have different meaning in a multicultural distributed environment by creating the uniform authority files, e.g. VIAF<sup>7</sup>, and standard vocabularies, e.g. TNG<sup>8</sup>, ULAN<sup>9</sup>, etc. Integration is often attempted at metadata level: MARC (UNIMARC, MARC 21) or DC formats. According to Martin Doerr, Jane Hunter and Carl Lagoze (2003), many metadata vocabularies are largely resource-centric, inadequately expressing entities such as people, places, ideas, and etc. For example, UNIMARC/A provides a set of properties that are associated with a primary resource, the “library object”. The values of some of these properties, e.g. of “Personal name” in Field 200 in UNIMARC/A are entities themselves having poor interoperability across other systems. Martin Doerr, Jane Hunter and Carl Lagoze (2003)<sup>10</sup> state that “In contrast, a core ontology provides underlying formal model tools that integrate sources and perform a variety of functions”. In order to provide value added services in VEPIS achieve better integration of diverse cultural sources, i.e. offer abstracted information and knowledge rather than returning documents (in the manner of most current Web search engines), the CIDOC CRM based BAVIC<sup>11</sup> has been developed. BAVIC is based on the CIDOC CRM data model and is compatible with the VEPIS data structure. It brings together authority records created for diverse cultural domains. In this stage of the development, BAVIC was built automatically by making use of possibilities of information technologies to relate and identify information about personal names, geographical names and chronological data from diverse cultural sectors.

The figure illustrates description of a manuscript, which was created by person *Stepon Batory* (E82.1) in the role of *Author* (P14.1) and which has type *King of Poland and Grand Duke of Lithuania* (P2). CRMdig is used for describing all stages of the creation of the content of digital cultural heritage. It is based on events that relate physical objects, digital objects, actors, times and places. The figure also illustrates the events within VEPIS: class

<sup>7</sup> VIAF (The Virtual International Authority File). Joint project with the Library of Congress, the Deutsche Nationalbibliothek, and the Bibliothèque nationale de France, in cooperation with an expanding number of other national libraries and other agencies, VIAF explores virtually combining the name authority files of participating institutions into a single name authority service. As of the winter of 2011, there are 21 authority files of personal, corporate, and conference names from 18 organizations participating in VIAF. Available at: <https://www.oclc.org/en/viaf.html>.

<sup>8</sup> TNG is intended to aid cataloging, research, and discovery of art historical, archaeological, and other scholarly information. However, its unique thesaural structure and emphasis on historical places make it useful for other disciplines in the broader Linked Open Data cloud. Available at: <https://www.getty.edu/research/tools/vocabularies/tgn/>

<sup>9</sup> The ULAN is a structured vocabulary containing names and other information about artists, patrons, firms, museums, and others related to the production and collection of art and architecture. Names in ULAN may include given names, pseudonyms, variant spellings, names in multiple languages, and names that have changed over time (e.g., married names). Among these names, one is flagged as the *preferred name*. Available at: <https://www.getty.edu/research/tools/vocabularies/ulan/about.html>.

<sup>10</sup> Doerr, M., et al. (2003) Towards a Core Ontology for Information Integration. Research Gate. Available at: : <https://www.researchgate.net/publication/31914729>.

<sup>11</sup> Authors: Rimvydas Laužikas and Vyngintas Vaitkevičius. Available at: <http://www.kf.vu.lt/en/structure/institutes/department-of-museology>.



D2 *Digitization Process* comprises events that result in the creation instances of D9 *Data Object* that represent appearance and form of instance of D9 *Data Object* that represents appearance or form of an instance of E84 *Information Carrier* (manuscript). Class D9 comprises instances of D1 *Digital Object*.

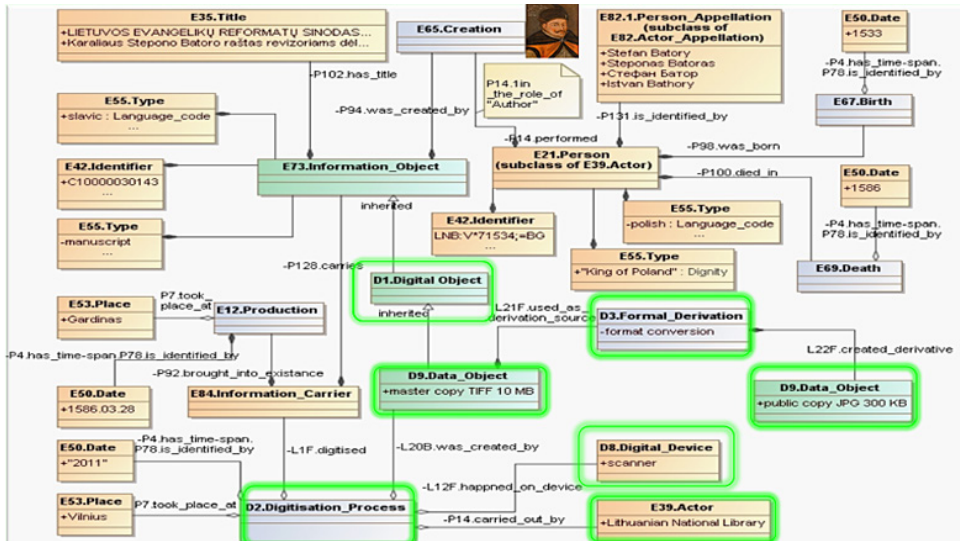


Fig. 2. An example of metadata integration from diverse systems and representation according to CIDOC CRM within VEPIS (specification of VEPIS)

VEPIS integrates two metadata categories: authority (BAVIC) and descriptive (metadata of digital objects linked with BAVIC), thus ensuring that semantic queries and provenance metadata refer to the versions of objects as they evolve and are modified or accessed over time. In particular it provides for a representation of how one version (or parts thereof) was derived from another version. VEPIS also provides the derivation chain (“ByWhom”), which documents the history of the content information and refers to its origin or source, to any changes that may have taken place since it was originated and to those who has had custody of it since it was originated as we can see in the figure. CRMdig captures and models requirements regarding the provenance of digital objects.

For example, Fig. 3. user wishes to get the creator of the book *Aitvaras* (“The Kite”). In terms of UNIMARC, the user searches for records for UNIMARC/B 200 title = *Aitvaras*. The query is then propagated to an ontology mediator where it is transformed by using a set of mapping rules from UNIMARC/A to CIDOC into equivalent query terms of CIDOC CRM paths such as E33 (Linguistic Object-) - P102 has title(is title of) = *Aitvaras*- P 94 (was created by) -E65 (Creation Event) E39 (Actor) – P131 (is identified by/ identifies) -E82<sup>12</sup> (Actor Appellation) = Judita Vaičiūnaitė.

<sup>12</sup> E82 Actor Appellation Deprecated use E41 Appellation instead. Available at: <http://www.cidoc-crm.org/sites/default/files/2019-03-26-CIDOC%20CRM%20b.pdf>.

## 2. Provenance as a basis for authenticity and trust of cultural data on the Web

### 2.1. Methodology of the research of provenance data within VEPIS

Michael Factor, et al. (2009) states that defining and assessing authenticity are complex tasks, including a clear definition of roles involved, coherent development of recommendations and policies for building trusted repositories, and precise identification of each component of the custodial function. Therefore the relevance of *authenticity and provenance* as a preliminary and central requirement for long term preservation digital content has been investigated by many international projects and researchers: Michael Factor, et al., (2009); David Giarretta, et al. (2011); J. T. Tennis (2012); Guercio, Salza, (2012); Grodon Dunsire and Mirna.Willer (2018), the InterPARES Project, World Wide Web Consortium (W3C) as well as other authors and projects.

The analysis of the research papers on provenance shows that provenance provides critical foundation for assessing *authenticity*, enabling *trust* and allowing *reproducibility*. It is essential for decision makers to make trust judgments about the information they use over the Semantic Web. According to the OAIS Reference Model<sup>13</sup>, provenance provides information about the events that occur during the lifecycle of digital objects (related to a license holder, registration and copyright). It guarantees the authenticity of the object because the customer is informed and has a certain “knowledge base” about the digital object, which may change over time. According to Factor et al. (2009), *trust* is a term with many definitions and uses, but in many cases establishing trust in an object or an entity involves analyzing its origin and authenticity. Trust is related to provenance because it is derived from provenance information and typically is a subjective judgment that depends on the context and use. It can be argued that provenance is a platform for trust. According to the *Requirements for Provenance on Web*, provenance encompasses the initial sources of information used as well as any entity and process involved in producing a result. That is why provenance data representation and management is important at any segment of the life cycle of a digital object. We analysed the provenance information within VEPIS in line with the methodology of the Requirements for Provenance on the Web, which defines *Content Category* referring to what types of information would need to be represented in a provenance record, *Management Category* referring to mechanisms that make provenance available and accessible within a system and *Use Category* implying that provenance records may need to accommodate a variety of uses as well as diverse users.

---

<sup>13</sup> Consultative Committee for Space Data Systems. (2012). Reference Model for an Open Archival Information System (OAIS). Available at: <https://web.archive.org/web/20131020200910/http://public.ccsds.org/publications/archive/650x0m2.pdf>.



## 2.2. Provenance as a basis for authenticity of the cultural content within VEPIS: results of the research

Since VEPIS is based on the OAIS<sup>14</sup> reference model, PDI within the Archival Information Package provides information about events that occur during the lifecycle of digital objects (related to a license holder, registration and copyright). It guarantees the authenticity of the object. One of the most important insights embedded in the OAIS Reference Model is that the “Content Information” to be preserved by an archive is composed not only of a “set of bit sequences” (the “data object”) but is also associated with sufficient preservation description information. However, Moreau (2009) draws attention to the fact that this provenance approach is inherent in closed systems and states that in the context of the Web, a broader approach is required by which chunks of provenance representation can be brought together to describe the provenance of information flowing across multiple systems<sup>15</sup>.

As it has been mentioned, in order to provide value added services in VEPIS and achieve better integration of sources from diverse cultural sectors, i.e. offer abstracted information and knowledge rather than returning documents (in the manner of most current Web search engines), the CIDOC CRM based BAVIC<sup>16</sup> was developed. The research by authors of this paper was intended to ascertain whether the provenance information at VEPIS ensures the authenticity of the objects at a sufficient level and how the information of provenance is represented on the Semantic Web.

### 2.2.1. Content of provenance data within VEPIS

According to the Requirements for Provenance on the Web, the first Category of provenance data is *Content*, which refers to the structure and meaning of provenance records. Since VEPIS is based on CIDOC CRM, (ISO 21127) and CRMdig, *Content Category* within this System is also based on this ontology.

Table A illustrates how *Content Category* of the Requirements is realized within VEPIS.

Table A <i>Content Category</i>		
Dimension	Description	Substantiating statements for VEPIS
Groth, et al., Requirements for Provenance on the Web		

<sup>14</sup> According to the specification, the infrastructure of VEPIS is based on the OAIS model, thus ensuring full management of data, their provenance and the workflows as well as the control of these processes within this information system.

<sup>15</sup> Moreau, L. The Foundations Provenance on the Web, p.2. Available at: <https://eprints.soton.ac.uk/268176/1/psurvey.pdf>.

<sup>16</sup> Authors: Rimvydas Laužikas and Vyngintas Vaitkevičius, Available at <http://www.kf.vu.lt/en/structure/institutes/departments-of-museology>.

<b>Object</b>	The artefact that a provenance statement is about	CIDOC CRM: E73 <i>Information Object</i> ; CRMdig: D1 <i>Digital Object</i> (subclass of E73)
<b>Attribution</b>	The sources or entities that contributed to the creation the artefact in question	CIDOC CRM: E21 <i>Person</i> , E74 <i>Group</i> ; CRMdig: D21 <i>Person Name</i>
<b>Versioning</b>	Records of changes to or between artefacts over time and what entities and processes were associated with those changes	CIDOC CRM: E1 <i>CRM Entity</i> ; CRMdig: D7 <i>Digital Machine Event</i> (subclass of D11 <i>Digital Measurement Event</i> ), data from UNIMARC 300 field (“Notes”)
<b>Justification</b>	Documentation recording why and how a particular decision is made	CIDOC CRM: E84 <i>Information Carrier</i> ; CRMdig: D11 <i>Digital Measurement Event</i> (subclass of D7 <i>Digital Machine Event</i> ) and E16 <i>Measurement</i> (superclass of D2 <i>Digitization Process</i> )
<b>Entailment</b>	Explanations showing how facts were derived from other facts	CRMdig: D12 <i>Data Transfer Event</i> (subclass of D7 <i>Digital Machine Event</i> )

According to the Requirements for Provenance on the Web, the first dimension of *Content Category* is *Object*: *Content Category* is *Object*: any statements about provenance and a possibility to refer to it. *Object* on the Web is a resource, essentially E73 *Information Object* (CIDOC CRM) and D1 *Digital Object* CRMdig), a subclass of E73, which can be identified with an URI (PURL in VEPIS). The second dimension is *Attribution*. In VEPIS, it is people, organizations, and other identifiable groups that contributed to the creation of the digital artefact: E21 *Person*, E74 *Group* (CIDOC CRM) and D21 *Person Name* (CRMdig.). These attributes are directly related with entities of BAVIC (personal names, geographical names and chronological data) and their metadata that describe these objects and extend the information about the metadata of VEPIS objects and play an important role in the search for objects on the Semantic Web.

According to the Requirements for Provenance on the Web, a special status should be attributed to the dimension *Versioning* in a provenance representation. It can be often difficult to understand whether a resource has changed its version because the representations of resources may differ but the underlying resource should be constant.

*Justification* is another dimension of *Content Category*. According to the W3C Provenance Working Group, it is the justification of decisions, which means why and how a particular decision is made. The purpose of justification is to allow those decisions to be discussed and understood. Some provenance information may be directly asserted by the relevant sources of some data or actors in a process, while other information may be derived from that which was asserted.

The dimension *Entailment* of *Content Category* represents explanations that show how facts were derived from other facts. Some provenance information may be directly

asserted by relevant sources of some data or actors in a process, while other information may be derived from that which was asserted. A standard way for implementing *Versioning*, *Justification* and *Entailment* within VEPIS is the realization of the following components: *Format Conversion*, *Data Verification* and *Logging Events* (Varnienė-Janssen, R. and Šermokas A., 2018).

### 2.2.2. Management of provenance data within VEPIS

According to the Requirements for Provenance on the Web, the second Category of provenance data is *Management Category* referring to mechanisms that make provenance available and accessible within a system. Table B illustrates how *Management Category* is realized within VEPIS. Several components and standards are applied: *Publication and Access*, the portal *epaveldas.lt*, *CIDOC CRM*, *RDF*, which comply with each of the dimensions of *Management Category*.

Table B. <i>Management Category</i>		
Dimension <i>Groth et al., Requirements for Provenance on the Web</i>	Description	Substantiating statements for VEPIS
<b>Publication</b>	Making provenance available on the Web	<i>Publication</i> within VEPIS is realized by the component <i>Publication and Access</i>
<b>Access</b>	The ability to find the provenance for a particular artefact	<i>Access</i> is realized via the portal <a href="http://www.epaveldas.lt">http://www.epaveldas.lt</a> and automatic data import using the OAI-PHM protocol
<b>Dissemination</b>	Defining how provenance should be distributed and controlled	<i>Dissemination</i> : BAVIC and metadata of digital objects are based on the CIDOC CRM and CRMdig and are in the RDF form in line with the XML schema, thus ensuring provenance related query services
Scale	Dealing with large amounts of provenance	<i>Scale</i> within VEPIS has been only partially realized

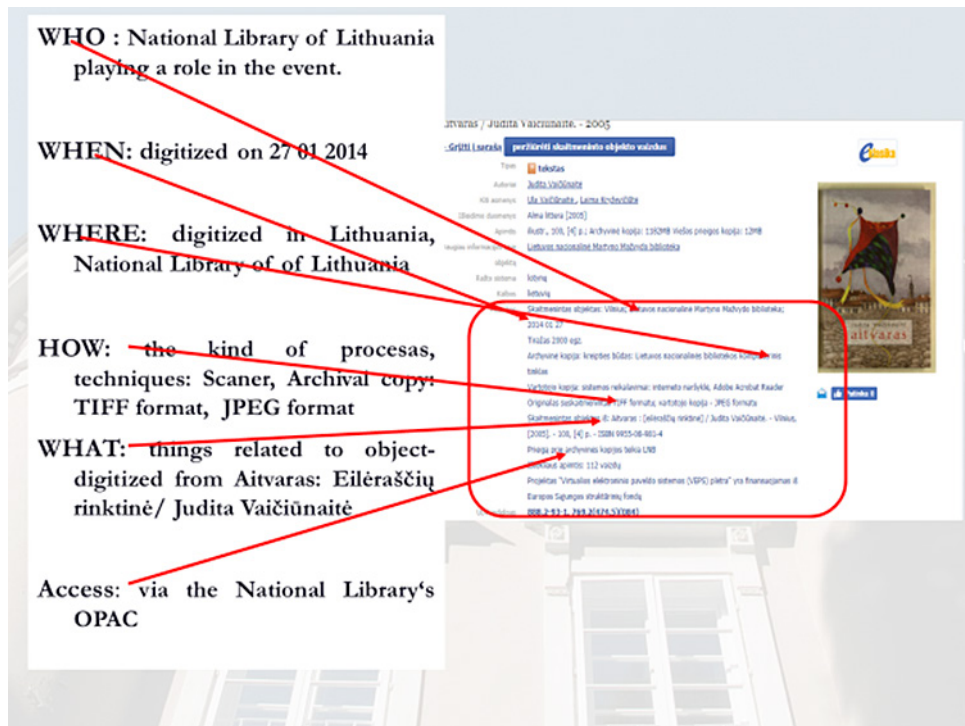
*Publication* within VEPIS is realized by the component *Publication and Access*. The portal's interface has all the accessibility features according to the recommendations of the European Union's WAI (Web Accessibility Initiative) and is intuitive, understandable and easy for users and is realized via portal <http://www.epaveldas.lt>. Another way to access is automatic data import via the OAI-PHM protocol. The search of the provenance information is based on CRMdig.

*Dissemination*: BAVIC and metadata of digital objects are based on CIDOC CRM and CRMdig and are in the RDF form in line with the XML schema, thus ensuring provenance-related query services by providing data about the creator of the object, the earlier versions

of the item, events that changed the custody of the item, input that influenced the result, the master version of the object and the scanner/resolution of the digital object (see Fig. 3).

*Scale* within VEPIS has been only partially realized. BAVIC ensures formulation of queries and organizing search results and permits obtaining information about the object from all the VEPIS partners independent of media types within VEPIS. However, it does not guarantee access to information about the investigation of the object that has been carried out or its results across numerous repositories.

As the example of the realization of Management of provenance data is presented in the Fig.3



### Fig. 3. Queries within VEPIS

Figure 3 illustrates queries within VEPIS, which are realized as follows: *Get the creator of the object* – *Get the earlier versions of the item* – *Get the events that changed the custody of the item* – *Get the master version of the object* – *Get the scanner/resolution of the digital object* – *Get access to the object*. On the other hand, the thesaurus BAVIC, which serves as a framework for semantic search, has to be extended by semantic relationship between the entities in order to improve searching on the Semantic Web for provenance information from heterogeneous data repositories. This issue will be solved during later stages after evaluating the results of the integration of BAVIC and VEPIS data.

### 2.2.3. Use of provenance data within VEPIS

According to the Requirements for Provenance on the Web, the third *Use Category* of provenance data implying that provenance records may need to accommodate a variety of uses as well as diverse users. An important consideration is how to make provenance information understandable for its users as well as provide appropriate presentation and visualization, compare artefacts according to their origin, imperfections, trust and interoperability.

<b>Table C. Use Category</b>		
<b>Dimension</b> Groth, et al., Requirements for Provenance on the Web	<b>Description</b>	<b>Substantiating statements for VEPIS</b>
<b>Understanding</b>	How to enable the end user consumption of provenance	Realized within VEPIS by the component <i>Publication and Access</i>
<b>Interoperability</b>	Combining provenance produced by multiple different systems	We could refer to interoperability only in the sense that VEPIS aggregates data from diverse systems and all descriptive information is converted into UNIMARC including provenance data (however, it is not interoperable as regards search)
<b>Comparison</b>	Comparing artefacts through their provenance.	Not implemented within VEPIS
<b>Accountability</b>	Using provenance to assign credit or blame.	Not implemented within VEPIS
<b>Trust</b>	Using provenance to make trust judgments	Specific components: <i>Component of Metadata Verification</i> , which ensures control of metadata and <i>Component of Logging Events</i> , which tracks the import of digitized objects
<b>Imperfections</b>	Dealing with imperfections in provenance records	Specific components: <i>Component of Metadata Verification</i> , which ensures control of metadata and <i>Component of Logging Events</i> , which tracks the import of digitized objects
<b>Debugging</b>	Using provenance to detect failures or bugs	Components implemented within VEPIS: <i>Component of Metadata Verification</i> and <i>Component of Logging Events</i>

There are several components for understanding the provenance and validating the authenticity of a preserved data object and within VEPIS: *Component of Metadata Verification*, which ensures control of metadata loaded into VEPIS in line with the requirements for quality, comprehensiveness and excellence of data, and *Component of Logging Events*, which tracks the import of digitized objects from VEPIS data providers and systems supporting the OAI-PMH protocol and verifies whether information about digitized objects

satisfy / do not satisfy the requirements for quality, comprehensiveness and excellence of data. However, we have to admit that we could refer to interoperability only in the sense that VEPIS aggregates data from diverse systems and all descriptive information is converted into UNIMARC including provenance data (but it is not interoperable as regards search).

According to Moreau (2009), a powerful argument for provenance is that it can help make systems transparent so that it becomes possible to determine whether a particular use of information is appropriate under a set of rules. Such capability helps make systems and information accountable. Our analysis showed that the dimensions *Comparison* and *Accountability* were not implemented within VEPIS. For to this reason, VEPIS does not support the possibility to compare artefacts through their provenance and assign credit or blame. *Debugging* is realized within VEPIS by *Component of Metadata Verification* and *Component of Logging Events*.

## Summarizing conclusions

1. Semantic interoperability of metadata and data within the cultural domain is one of main issues within integrated systems. In our attempt to accomplish the goal of the research, we analysed the role of CIDOC CRM as a mediating tool for integrating metadata represented in different schemas from various cultural domains of Lithuania. The authors of this paper consider that it is appropriate to update the mapping of UNIMARC/B and UNIMARC/A to CIDOC CRM and CRMdig within VEPIS in line with the updates published on the website of IFLA.
2. The creation of the BAVIC Thesaurus encompassing personal names, geographical names and chronological data from diverse cultural domains by applying methods of information technologies answered its purpose as it allowed providing more semantic links and better interoperability of VEPIS objects and using these links for searching and presentation of data. Furthermore, we propose that the BAVIC Thesaurus data structure be further developed by extending semantic relationships between entities and improving representation and management of entities of authority records on the national level and drawing on the information from international thesauri of similar nature.
3. The qualitative analysis of the *Requirements for Provenance on the Web* and the specification of VEPIS and its services, allowed us to conclude that VEPIS, which is based on CIDOC CRM, CRMdig, RDF the OAIS Reference Model as well as on *Component of Metadata Verification*, *Component of Logging Events* and *Component of Publication and Access*, meets the main Requirements for Provenance on the Web as it supports the following functionality:
  - ✓ Provides support for three major categories of provenance: the *content* of provenance information, the *management* of provenance as it exists on the Web, and the *use* of provenance.
  - ✓ Provides metadata and context of the digitization process referring to the master version and derivation chain. All this creates trustworthy provenance information and provides access to it by using open protocols.



- ✓ The portal [www.epaveldas.lt](http://www.epaveldas.lt) allows querying the most relevant facts and retrieving complete descriptions encoded in this model by generic CIDOC CRM terms without the need to refer to its specific properties. The user has the possibility to identify the creator of the object, earlier versions of the item, the events that changed the custody of the item as well as to find out how results were derived (what input influenced the result), identify the master version of the object and the scanner / resolution of the digital object and information about access of the resource. We can conclude that VEPIS satisfies the requirement defined by the W3C Provenance Incubator Group that provenance on the web should include information about the creation and publication of Web resources and information about access of those resources as well as activities related to their discussion, linking and reuse.

### ***Future developments***

This research has identified key directions for the development of VEPIS regarding the provenance in order to ensure the representation and exploiting provenance information on the Web.

*Interoperability.* At present VEPIS aggregates data from diverse systems and all descriptive information is converted into UNIMARC including provenance data and mapped to CIDOC CRM; however, it is not interoperable as regards search. In order to meet the Requirements for Provenance on the Web, the BAVIC Thesaurus, which serves as a framework for semantic search, has to be extended by entities and their semantic relationship in order to improve searching on the Semantic Web for provenance information from heterogeneous data repositories.

*Accountability.* In order to meet the Requirements for Provenance on the Web regarding accountability, new services and functions need to be established so that the possibility to compare artefacts through their provenance and assignment of credit or blame could be exploited.

*Representation of metadata.* The extensiveness of the metadata has a profound impact on the reliability of information. It is, therefore, very important to harmonize descriptive metadata regarding the provenance information of VEPIS objects.

*Representation of data.* In order to achieve utmost conformance of VEPIS to the Requirements for Provenance on the Web, it is essential to warrant the coordination of activities of all institutions related to VEPIS, ensure extensiveness of metadata and their conformance to uniform requirements and supplement the database of BAVIC with authority files and provide monitoring of these data.

### **References**

1. Definition of the CIDOC Conceptual Reference Model Volume A. Version 6.2.5, 2018 Available at: <http://www.cidoc-crm.org/sites/default/files/2019-03-26-CIDOC%20CRM%20b.pdf>.
2. DOERR, M. (2009). *An Ontological Approach to Digital Preservation Metadata*. ICS-FORTH May 23, 2009. Available at: [http://www.cidoc-crm.org/sites/default/files/carpar\\_for\\_Prague.ppt](http://www.cidoc-crm.org/sites/default/files/carpar_for_Prague.ppt).

3. DOERR, M., HUNTER, J. and LAGOZE, C. (2003). Towards a Core Ontology for Information Integration, *Journal of Digital Information*, Volume 4, Issue 1, Article N 169, 2003-04-09, (April 2003), Available at: <https://www.researchgate.net/publication/31914729>.
4. DUNSIRE, G. WILLER, M. (2018). Authority versus authenticity the shift from labels to identities. *Authority, Provenance, Authenticity, Evidence: Selected papers from the Conference and School Authority, Provenance, Authenticity, Evidence*. Zadar, Croatia, October 2016. / Edited by Mirna Willer, Anne J. Gilliland and Marijana Tomić . Zadar: Sveučilište u Zadaru, 2018, p.87-115.
5. CALVANESE, D., De GIACOMO, G., LENZERINI, M., NARDI, D. and ROSATI, R. (1998). Description Logic Framework for Information Integration . Available at: <https://www.dis.uniroma1.it/~degiacono/papers/1998/CDLNR98kr.pdf>
6. GIARETTA, D. (2011). *Advanced Digital Preservation*. Springer Heidelberg Dordrecht London New York e-ISBN 978-3-642-16809-3.
7. GIARETTA, D., MATTHEWS, B., BICARREGUI, J., LAMERT, S., GUERCIO, M., MICHETTI, G. and SAWYER, D. (2009). Significant Properties, Authenticity, Provenance, Representation Information and OAIS David./ UC Office of the President iPRES 2009: the Sixth International Conference on Preservation of Digital Objects. Available at: <https://escholarship.org/content/qt0wf3j9cw/qt0wf3j9cw.pdf>.
8. GROTH, P. et al. (2012). Requirements for Provenance on the Web. In *International Journal of Digital Curation*, Vol. 7, No. 1. Available at: <https://www.isi.edu/~gil/papers/groth-et-al-ijdc12.pdf>.
9. GUERCO, M., SALZA, (2013). Managing Authenticity through the Digital Resource Lifecycle. *Italian Research Conference on Digital Libraries IRCDL 2012: Digital Libraries and Archives* pp 249-260. Available at: [https://link.springer.com/chapter/10.1007/978-3-642-35834-0\\_25](https://link.springer.com/chapter/10.1007/978-3-642-35834-0_25).
10. KAKALIC, L., LOURDI, I., STASINOPOULOU, T., BOUNTOURI, L. and PATHEODOROU, C., DOERR, M. (2007). Integrating Dublin Core metadata for cultural heritage collections using ontologies. *Proc. International Conference on Dublin Core and Metadata Application*. Available at: <https://dcpapers.dublincore.org/pubs/article/view/871>.
11. FACTOR, M., HENIS, E., NAOR, D., RABINOVICI-COHEN, S., RESHEF, P., Ronen, S. (2009). Authenticity and provenance in long term Digital preservation: modeling and implementation in preservation aware storage. Available at: [http://usenix.org/event/tapp09/tech/full\\_papers/factor/factor\\_html/](http://usenix.org/event/tapp09/tech/full_papers/factor/factor_html/).
12. MOREAU, L. (2009). *The Foundations for Provenance on the Web*. November 3, 2009 Available at: <https://eprints.soton.ac.uk/268176/1/psurvey.pdf>.
13. MOREAU, Luc, The Foundations for Provenance on the Web Foundations and Trends, *Web Science*, 2:2-3 (2010): 99-241, <http://dx.doi.org/10.1561/18000000010>, 6.
14. Requirements for Provenance on the Web. Available at: [https://www.w3.org/2005/Incubator/prov/wiki/User\\_Requirements](https://www.w3.org/2005/Incubator/prov/wiki/User_Requirements).
15. SIGNORE, O. Ontology Driven Access to Museum Information. Available at: <http://www.w3c.it/papers/cidoc2005.pdf>.
16. TENNIS, J. T., ROGERS, C. (2012). Authenticity Metadata and the IPAM: Progress toward the InterPARES Application Profile. In *Proc. Int'l Conf. on Dublin Core and Metadata Applications*. The Kuching Proceedings. Available at: <http://dcpapers.dublincore.org/pubs/article/view/3662>.
17. VARNIENĖ-JANSSEN, R., ŠERMOKAS, A. Provenance in the Context of Digital Cultural Heritage Content: The Lithuanian Approach. *Authority, Provenance, Authenticity, Evidence: Selected papers from the Conference and School Authority, Provenance, Authenticity, Evidence*. Zadar, Croatia,

October 2016. / Edited by Mirna Willer, Anne J. Gilliland and Marijana Tomić. Zadar: Sveučilište u Zadaru, 2018, p. 213-257. ISBN 978-953-331-220-0.

18. VARNIENĖ-JANSSEN, R., KUPRIENĖ, J. Authenticity and Provenance in Long-Term Digital Preservation: Analysis of the Scope of Content. *Informacijos mokslai*, 2018, t. 82, p. 131-160. ISSN 1392-0561. eISSN 1392-1487. Available at: <http://www.zurnalai.vu.lt/informacijos-mokslai/article/view/12291>.

19. VARNIENĖ-JANSSEN, Regina; JUŠKYS, Jonas. Strategic, Methodological and Technical Solutions for the Creation of Seamless Content of the Digital Cultural Heritage: Lithuanian Approach. *Summer School in the Study of Historical Manuscripts: Proceedings / Referees*: Istvan Kecsmeti, PhD; Laila Vejzovic, MLS; Tinka Katic, PhD. Zadar: Sveučilište u Zadru, 2013, p. 349–369. ISBN 978-953-331-020-6. Available at: [http://www.unizd.hr/Portals/41/elektronicka\\_izdanja/summer2904\\_tisak.pdf](http://www.unizd.hr/Portals/41/elektronicka_izdanja/summer2904_tisak.pdf).

20. W3C Provenance Incubator Group, Semantic Web Activity World Wide Web Consortium, *Overview of Provenance on the Web*. Available at: <https://www.w3.org/2005/Incubator/prov/wiki/images/0/02/Provenance-XG-Overview.pdf>.