

PROVENANCE IN THE CONTEXT OF DIGITAL CULTURAL HERITAGE CONTENT *THE LITHUANIAN APPROACH*

Regina Varnienė-Janssen

*Faculty of Communication, Vilnius
University, Vilnius, Lithuania*

Albertas Šermokas

*Faculty of Mathematics and
Informatics, Vilnius University,
Vilnius, Lithuania*

KEYWORDS:

*provenance, digital
environment, data
quality, trustfulness of
information, Lithuania's
Information System of the
Virtual Electronic Heritage
(VEPIS)*

ABSTRACT

The main objective of this article is to highlight issues of provenance in the digital environment as a fundamental concept of data quality. The article seeks to theoretically substantiate the notion that the provenance of information is crucial for determining the trustworthiness of information. By applying the EU 2–3–6 concept of added value for electronic publishing and qualitative content analysis of research literature, we conceptualized the following processes of provenance: Content Creation, Content Packaging, Market Making, Delivery Support and Services, and Interface and Systems. This approach also enabled us to identify the functions of provenance as well as their dimensions, and define provenance-related added value in Lithuania's Information System of the Virtual Electronic Heritage (VEPIS).

Introduction

Before delving into issues of provenance within the digital environment, we will discuss the provenance of information in traditional media. The Oxford Wordfinder dictionary states that provenance (from Latin “*provenire*, “come forth”, “arise”) is “the place of origin or history of a work of art, etc.”¹ The most frequent indications of the book’s provenance are exlibris, stickers, inscriptions and stamps. Identification of provenance shows the history and development of a library and its cultural connections; it also helps reconstruct older collections of books. As an example we can take the section of the website of the National Library of Lithuania “Provenances,” and analyse how the provenance of documents preserved at the Library is treated. We searched the website for Mykolas Kleopas Oginskis (1765–1833), one of the most sophisticated members of the Oginsky family and a citizen of the Federal Monarchy of the Kingdom of Poland and Grand Duchy of Lithuania, the centenary of whom we celebrated in 2015. He was a man of European outlook, a patriot of his homeland, politician, diplomat and composer (undoubtedly everyone knows his famous “Polonaise”). At the age of 20, he became a member of the Parliament (Seimas) and diplomat. We can see how this person and other persons related to him are treated within the Provenance Catalogue of the National Library. The Catalogue provides a list of owners of documents and types of provenance indicators (?), the most important of which are inscriptions, stamps, stickers and exlibris. We retrieved a list of provenances of personal documents. Since the list did not include Mykolas Kleopas Oginskis’s name, we searched for documents owned by the second wife of his son Irenėjus Kleopas Oginskis, Olga Kalinowska-Ogińska, (1808–1863), who was a lady-in-waiting at the palace of Czar Alexander II. She was well-known in Saint Petersburg for her excellent musical education and taste. The picture shows her inscription in a document of the Oginsky family preserved at the National Library (Figure 1). Provenances of documents provide significant historical information and evidence of historical and cultural traditions as well as the history of publishing.

1 *Reader’s Digest Oxford Complete Wordfinder* (Pleasantville, N.Y.: Reader’s Digest Association, 1993), 1227.



FIGURE 1 The seal of Olga Oginskienė (Kalinowska-Oginska, 1820–1899) in the book “Histoire de L’Angleterre”

Everything that was said above about provenance within traditional media is known by everyone who is interested in this field. However, within the digital environment, this issue gets more complicated. The World Wide Web Consortium (W3C) defines provenance as a record that describes people, institutions, entities and activities involved in producing, influencing or delivering a data element or thing. Therefore, it is no accident that the investigation of provenance in the digital environment has formed a major focus since 2005, when the definition of provenance was formulated. That definition states that, “Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource. Provenance provides a critical foundation for assessing authenticity, enabling trust, and allowing reproducibility.”² Requirements for provenance on the Web (Web3) were subsequently defined and ontologies, models and standards identifying starting points for provenance representations as well as elements of the provenance architecture created. Provenance is essential for tracking attribution and credit in curated databases. It is essential for decision makers to make trust judgments about the information they use over the

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

Semantic Web. Therefore, an increasing number of scientific publications regarding provenance have appeared. As far back as 2010, Luc Moreau highlighted the importance of provenance by stating that “A total of 425 papers have been identified. The first publication dates back from 1986 and describes an auditing technique to assist analysts in understanding and validating data results. [...] about half the papers were published in the last two years.”³

The main objective of this paper is to highlight issues of provenance in the digital environment as a fundamental concept of data quality drawing on the results of the qualitative analysis of research literature and documentation of Lithuania’s Information System of the Virtual Electronic Heritage (hereafter “VEPIS”, *Virtuali elektroninio paveldo informacinė sistema*), as well as to evaluate the added value of VEPIS regarding provenance.

Methodology of the research

The research presented in this paper is primarily focused on the 2–3–6 concept for added value creation developed by the European Commission and serving as the basis for the 2–3–6 concept for creating added value in electronic publishing (Figure 2).⁴ It is a concept of value chains. Value activities are defined as processes that create value for the customer. By applying this concept, it is possible to do qualitative analyses of activities and behaviour within a certain field and visualize them, e.g., the impact of provenance on the trustworthiness and accuracy of information. For this purpose, by relating value-creating activities, the role of every agent is determined. The analysis of the dynamic development of the resulting combinations shows the potential of every model, standard and system. This article applies the 2–3–6 Concept as a framework to identify developments in the creation of provenance information within the digital envi-

3 Luc Moreau, “The Foundations for Provenance on the Web Foundations and Trends,” *Web Science*, 2:2-3 (2010): 99-241, doi: 10.1561/1800000010, <https://eprints.soton.ac.uk/271691/1/survey.pdf>.

4 *Strategic Developments for the European Publishing Industry Towards the Year 2000*. Created by European Commission, DG XIII, Andersen Consulting, 1996, <http://www.echo.lu/elpub2/en/exengl.pdf>.

ronment. This research is based on the presumption that this structure reflects the current structure (the *status quo*) of creating digital provenance, which is the main criterion for data quality. We start with the assignment of identified core processes and their activities to relevant players such as models, working groups and implemented systems, and proceed to qualitative analysis of scientific literature and the VEPIS Specification as well as its operational functions. VEPIS is the largest national system in Lithuania using the OAIS Reference Model for the preservation of museum artefacts, archival documentation and library documents, modern and rare books manuscripts, and so forth. Taking into account the functionality of VEPIS on the national and international levels, it is feasible to determine provenance-related added value of this system.

Having briefly introduced the methodology of the research, we will proceed to the results of the research itself. Drawing on this established model for the provenance-related added value of VEPIS I, we might conclude that the added value of VEPIS is generated by employing the following six processes: *Content Creation, Content Packaging, Market Making, Transport, Delivery Support and Services, and Interface and Systems*. Such a methodological approach correlates with the approach declared by the W3C Provenance Working Group, implying that the key categories of provenance are *Content, Management and Use*. Therefore, the research is based on the presumption that the above-described methodology is suitable for investigating provenance within the digital environment and analysing services and products of VEPIS from the point of view of added value.

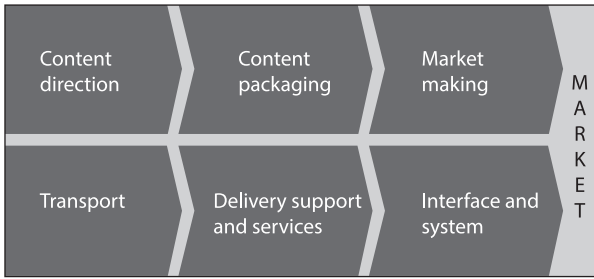


FIGURE 2 The 2-3-6 Concept for added value for electronic publishing⁵

5 Strategic Developments for the European Publishing Industry Towards the Year 2000, op. cit.

Research results: The model of provenance related added value of VEPIS

Content Creation Process

The process of *Content Creation* refers to the structure and meaning of provenance records. It is based on one function – *Identifying the Scope of Provenance Content*. In order to conceptualize this function, we analysed the Open Provenance Model (OPM), Requirements for Provenance on the Web by the W3C Provenance Working Group, the ontology of CRM-dig (the extension of CIDOC-CRM to support provenance metadata), OAIS, RDF and the so-called provenance-aware applications models as well as research literature on provenance.

TABLE 1 The model of provenance-related added value of VEPIS: *Content Creation Process*

CONTENT CREATION PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Identifying the Scope of Provenance Content</i>	<i>Object</i>	The artefact that a provenance statement is about	
	<i>D1 Digital Object</i>	Digital object comprises identifiable immaterial items that can be represented as sets of bit sequences such as data sets, e-texts, etc. <i>CRM_{dig}, version 3.2.1</i>	Implemented within VEPIS by using CRM _{dig}
	<i>Attribution</i>	The source or entities that contributed to creating the artefact in question <i>Groth et al., “Requirements for Provenance on the Web”</i>	

The following dimensions were established: *Object*, *Digital Object*, *Attribution*, *Artefact*, *Agent*, *Person Name*, *Place (Room)*, *Digital Machine Event*, *Process*, *Digitization Process*, *Software Execution*, *Digital Measurement*, *Versioning*, *Justification* and *Entailment*.

The first dimension is *Object*, which states that we need to establish first what the artefact is that a provenance statement is about. The Requirements for Provenance on the Web by the W3C Provenance Working Group state that “On the Web this will be a Web resource, essentially anything that can be identified with a URI such as Web documents, datasets, assertions or services. Sometimes provenance information refers to a particular portion or aspect of the artefact. It is especially challenging to keep track of provenance during the object’s lifetime as it migrates among different systems. For example, objects may be organized in collections, selected subgroups or portions of some modified objects or the document may be disseminated on the Web and its copies further modified.”⁶

The second dimension is *D1 Digital Object*, which, according to CIDOC CRMdig, “comprises identifiable immaterial items that can be represented as sets of bit sequences such as data sets, e-texts, images, audio or video items, software, etc. and are documented as single units. Any aggregation of instances of *D1 Digital Object* and treated as single unit is also regarded an instance of *D1 Digital Object*. This means that, for instance, the content of a DVD, an XML file on it or an element of this file are regarded as distinct instances of *D1 Digital Object* mutually related by the *P106 is composed of (forms part of) the property*. *D1 Digital Object* does not depend on a specific physical carrier and it may exist on one or more carriers simultaneously.”⁷

The third dimension is *Attribution*; according to the Requirements for Provenance on the by the Web W3C Provenance Working Group, it refers to the sources (i.e., typically any Web resource or document that has an associated URI, Web sites or data) or entities (i.e., people, organizations

6 Paul Groth, Yolanda Gill, James Cheney and Simon Miles, “Requirements for Provenance on the Web,” *The International Journal of Digital Curation*, 7:1 (2012): 39-56, doi: 10.2218/ijdc. V7i1.213, <http://www.ijdc.net/index.php/ijdc/article/viewFile/203/272>.

7 CRMdig: *Model for provenance metadata*. Version 3.2.1, April, 2016, http://www.ics.forth.gr/isl/CRMext/CRMdig/docs/CRMdig_v3.2.1.pdf.

and other identifiable groups) that contributed to the creation of the artefact in question).⁸

TABLE 2 The model of provenance-related added value of VEPIS: *Content Creation Process* (cont.)

CONTENT CREATION PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Identifying the Scope of Provenance Content</i>	<i>Artifact</i>	Immutable piece of state, which may have a physical embodiment in a physical object, or a digital representation in a computer system. <i>Moreau et al., "The Open Provenance Model: Core specification (v1.1)"p. 2</i>	
	<i>Agent</i>	Contextual entity acting as a catalyst of processes enabling, facilitating, controlling or affecting its execution. <i>Moreau et al., "The Open Provenance Model: Core specification (v1.1), p.3"</i>	

The fourth dimension is *Artifact*, which, according to the Open Provenance Model (hereafter OPM),⁹ is an immutable piece of state, which may have physical embodiment in a physical object, or a digital representation in computer system.

The fifth dimension *Agent*, according to OPM, represents a contextual entity acting as a catalyst of processes enabling, facilitating, controlling or affecting its execution.

8 Groth et al., "Requirements for Provenance on the Web".

9 Luc Moreau et al., "The Open Provenance Model Core Specification (v1.1)," *Future Generation Computer Systems*, 27 (2011): 743-756, <https://pdfs.semanticscholar.org/4181/c0d8284d8737f61232365747690297e31c82.pdf>.

TABLE 3 The model of provenance related added value of VEPIS: *Content Creation Process* (cont.)

CONTENT CREATION PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
Identifying the Scope of Provenance Content	D21 Person Name	<p>Subclass of E82 Actor Appellation The class <i>E82 Actor Appellation</i> comprises the proper noun name that identifies a person who acts as an entity.</p> <p><i>CRM_{dig}, version 3.2.1</i></p>	Implemented within VEPIS by using CRM _{dig}
	D23 Room (Place)	<p>Subclass of E53 Place. This class comprises a small scale space that contains manipulable objects and returns the bodily experiences of how people assimilate image schemata.</p> <p><i>CRM_{dig}, version 3.2.1</i></p>	Implemented within VEPIS by using CRM _{dig}

The sixth dimension, according to CRMdig, is *D21 Person Name*. It is a subclass of *E82 Actor Appellation*. This class comprises a proper noun name that identifies a person who acts as an entity. Properties: *L51 has first name* and *L52 has last name* (CRMdig).

The seventh dimension, according to CRMdig, is *Room*, which in turn may be related to other things. It is a subclass of *D23 Room* of *E53 Place*. This class comprises small-scale space that contains manipulable objects and retains the bodily experiences of how people assimilate image schemata (CRMdig).

TABLE 4 The model of provenance related added value of VEPIS: *Content Creation Process* (cont.)

CONTENT CREATION PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Identifying the Scope of Provenance Content</i>	<i>D7 Digital Machine Event</i>	<p>Events that happen on physical digital devices following a human activity that intentionally caused its immediate or delayed initiation and results in the creation of a new instance of <i>D1 Digital Object</i> on behalf of the human actors</p> <p><i>CRM_{dig}, version 3.2.1</i></p>	Implemented within VEPIS by using CRM _{dig}

The eighth dimension, according to CRMdig, is *D7 Digital Machine Event*. This class comprises events that happen on physical digital devices following a human activity that intentionally caused its immediate or delayed initiation and results in the creation of a new instance of *D1 Digital Object* on behalf of the human actor (agent). The input of a *D7 Digital Machine Event* may take the form of parameter settings and/or data to be processed. Some *D7 Digital Machine Events* may form part of a wider *E65 Creation Event*. In this case, all the machine output of the partial events is regarded as creation of the overall activity.

TABLE 5 The model of provenance related added value of VEPIS: *Content Creation Process* (cont.)

CONTENT CREATION PROCESS			
	Dimension	Substantiating statements	Substantiating statements for VEPIS
Functions <i>Identifying the Scope of Provenance Content</i>	Process	Activities (or steps) that were carried out to generate the artefact at hand <i>Groth et al., "Requirements for Provenance on the Web"</i>	Implemented within VEPIS by using CRM _{dig}
	Process	Action or series of actions performed on or caused by artefacts and resulting in new artefacts <i>Moreau et.al., "The Open Provenance Model: Core specification (v1.1)"</i>	
	D2 Digitization Process	The type of the process and techniques applied <i>Doer, Theodoridou, "CRM_{dig}: A Generic Digital Model for Scientific Observation"</i>	Implemented within VEPIS by using CRM _{dig}
	D10 Software Execution	Events by which a digital device runs a software program or series of computing operations on a digital object <i>CRM_{dig} version 3.2.1</i>	Implemented within VEPIS by using CRM _{dig}
	D11 Digital Measurement Event	Actions measuring physical properties using a digital device <i>CRM_{dig} version 3.2.1</i>	Implemented within VEPIS by using CRM _{dig}

Process, according to the W3C Provenance Working Group, means *activities* (or steps) that were carried out to generate the artefact at hand. According to OPM, the *process* means *actions* or *series* performed on or caused by artefacts and resulting in new artefacts. According to CRMdig, the class Digitization Process (D2) comprises events that result in the creation of instances of *D9 Data Object* that represent the appearance and/or form of an instance of *E18 Physical Thing* such as paper documents, paintings etc. Within VEPIS, we implemented the terminology of CRMdig *D2 Digitization Process*, *D10 Software Execution* and *D11 Digital Measurement*, i.e. events that happened within VEPIS and actions measuring physical properties.

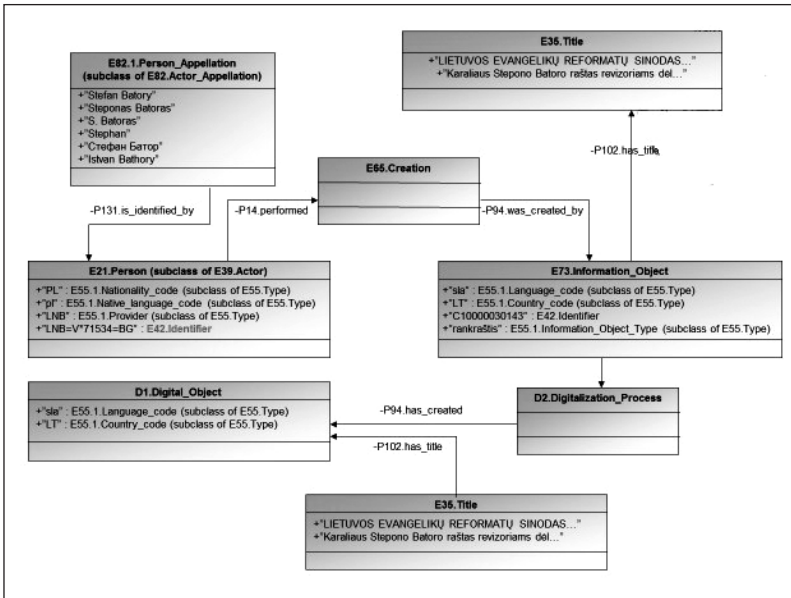


FIGURE 3 The digitization process of the manuscript by Steponas Batoras, the Duke of the Grand Duchy of Lithuania and King of Poland

The Figure 3 illustrates the digitization process and its description in CRMdig classes within VEPIS. The Person Steponas Batoras, the Duke of the Grand Duchy of Lithuania, is identified by Class *E21* as being of Polish nationality; he performed creation (Class *E65 was created*). The In-

formation Object is the manuscript in the Ruthenic language, which has the title “Lietuvos Evangelikų reformatų sinodas“. This information object was digitized during *D2 Digitization Process* and created as *D1 Digital Object*, which has the same title “Karaliaus Stepono Batoro raštas revizoriams dėl [...]”.

TABLE 6 The model of provenance related added value of VEPIS: *Content Creation Process* (cont.)

CONTENT CREATION PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Identifying the Scope of Provenance Content</i>	<i>Versioning</i>	<p>Records of changes to or between artefacts over time and what entities and processes were associated with these changes</p> <p><i>Groth et al., “Requirements for Provenance on the Web”</i></p>	Implemented within VEPIS by using CRM _{dig}
	<i>Justification</i>	<p>Documentation recording why and how a particular decision is made</p> <p><i>Groth et al., “Requirements for Provenance on the Web”</i></p>	Implemented within VEPIS by using CRM _{dig}

The dimension *Versioning*, according to W3C Provenance Working Group, has to be understood as records of changes to or between artefacts over time, and what entities and processes were associated with those changes.

Justification is another dimension of the content of provenance creation. According to 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.

TABLE 7 The model of provenance related added value of VEPIS: *Content Creation Process* (cont.)

CONTENT CREATION PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Identifying the Scope of Provenance Content</i>	<i>Entailment</i>	<p>Explanations showing how facts were derived from other facts</p> <p><i>Groth et al., “Requirements for Provenance on the Web”</i></p>	Implemented within VEPIS by using CRM _{dig}

Entailment is the dimension of the process *Content Creation* and 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. In general, one fact may entail another: it is important in the case of provenance data that is inherently describing the past, for which the majority of facts cannot now be known.

Versioning, Justification and *Entailment* within VEPIS as an OASIS and service-oriented architecture, is based on messages which invoke services and which may themselves act as clients for other services. *Versioning, Justification* and *Entailment* encompass the execution of a computer program to which we can explicitly point, a physical act to which we can only refer, and some action performed by a person that can only be partially represented. *Versioning, Justification* and *Entailment* also serve as a standard way within VEPIS to find provenance for a given resource. Within VEPIS, *Versioning, Justification* and *Entailment* are modelled as subclasses of both *Digital Machine Event* and *Measurement*, thus allowing for correlation of the measured object with the device that did the measurement.

Luc Moreau states that “many provenance approaches support a notion of time, so that users can refer to executions or data products according

to the time at which they took place or were produced.”¹⁰ Within VEPIS the place where the object was created (Vilnius), organization (National Library of Lithuania or other institution) and time (when the digital object was produced) are all presented in the field for notes of UNIMARC format.

Results from investigating the Content Creation Process

Having summarized the conceptualization of the first process, we could state that at present there is no uniform terminology of the content of provenance, though some of the terms are duplicates and some are identical. Within various models, some of dimensions bear the same names (e.g., *Process* within both OPM and requirements by the W3C Provenance Working Group) but their interpretation differs. Within VEPIS the CRMdig terminology and the dimensions *Digital Object*, *Person Name*, *Place (Room)*, *Digital Machine Event*, *Digitization Process*, *Software Execution* and *Digital measurement* were implemented, thus allowing for the identification of the context of the digitization process and derivation chain. We could also state that it correlates with the dimensions of *Versioning*, *Justification* and *Entailment* identified by the W3C Provenance Working Group in the Requirements for Provenance on the Web.

Given the fact that VEPIS is based on CRMdig, within which the concept *Event* allows tracing the history of an object, it is possible to adequately describe the conditions under which the new version of the object was created. We conclude that the *Content Creation Process* generates provenance-related added value within VEPIS.

The three dimensions of the scope of the provenance content – *Versioning*, *Justification* and *Entailment* – are closely connected with the identity of the resource. The identity of a resource is intended to have a very wide meaning, not only its unique designation and/or identification. Identity refers to *the whole* of the characteristics of a resource that uniquely identifies it and distinguishes it from any other resource. In addition to its

10 Moreau, “The Foundations for Provenance on the Web Foundations and Trends”, 26.

internal conceptual structure, it refers to its general context (e.g., legal and technological). Given that VEPIS is based on the OAIS Reference Model which strongly relates *Context Provenance*, *Fixity*, and *Reference Information* within the system, it could be stated that VEPIS provides for the identification of preserved objects.

Another important aspect of provenance information is *verification of authenticity* in terms of the identity and integrity of the digital objects by providing content and contextual information during the whole preservation process. Since VEPIS is based on the OAIS model, the Preservation Description Information (hereafter “PDI”) is specially focused on by describing the past and present states of the content information. PDI is made up of four types of information, which must be presented in an Archival Information Package. One of the types of information is provenance, which describes the source of content information. Provenance information also can be considered “event-based” metadata describing the object as a dynamic entity. This information is to be collected, maintained and interpreted together as much as possible as a set of relationships defining the resource itself. Since VEPIS is based on the OAIS model and contains this type of information, the Archival Information Packages, which are designed following the Metadata Encoding and Metadata Transmission Standard (METS), encode descriptive, administrative and structural metadata in XML. VEPIS provides metadata about the digitization process and object in context, since it is both the object itself and the relationships that provide complete meaning to it. We could conclude that VEPIS documents the complete history of the resource.

Another standard way for realizing *Versioning*, *Justification* and *Entailment* of the object within VEPIS are components which are based on CRM dig. One of them is *Component of Format Conversion*, which (1) ensures the compatibility of descriptive metadata of digitized objects from diverse institutions by transforming supplied metadata following the CIDOC-CRM data model and by using semantic software applications allowing syntactical interoperability to be established between terms within controlled vocabularies of different institutions belonging

to the same problematic area; (2) links terms from different controlled vocabularies, (3) establishes syntactical links between terms to be used in searching and discovering objects and associated terms from different control vocabularies; (4) identifies semantic groups of objects: principal classes of CIDOC CRM including personal names, geographical names, time periods and subjects; and (5) automatically saves descriptive metadata of the work into digital copies adjusted for presenting online. The dimensions *Versioning*, *Justification* and *Entailment* within VEPIS are also realized by the component *Data Verification*. This ensures control of data loaded into VEPIS in line with the requirements for quality, comprehensiveness and excellence of data. The dimensions *Versioning*, *Justification* and *Entailment* within VEPIS are also realized by the component *Logging IS Events*. This enables presenting information from the log of the manager of the VEPIS data, i.e., (1) tracking the import of digitized objects from VEPIS data providers and systems supporting the OAI-PMH protocol, (2) presenting information about digitized objects which satisfy/do not satisfy the requirements for quality, comprehensiveness and excellence of data.

Content Packaging Process

TABLE 8 The model of provenance-related added value of VEPIS: *Content Packaging Process*

CONTENT PACKAGING PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
Implementation of Provenance Data Models	Open Provenance Model	<p>Defines a core set of rules that identify the valid inferences that can be made on provenance representation</p> <p><i>Moreau et.al., "The Open Provenance Model: Core specification (v1.1)"</i></p>	

CONTENT PACKAGING PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
Implementation of Provenance Data Models	W3C Provenance Working Group Model	Modelling what should be contained in the provenance data (its content) will have different concerns from developing software to capture and maintain the data (its management), and different concerns again from those providing solutions to solve specific user problems (its use) <i>Groth et al., "Requirements for Provenance on the Web"</i>	
	CRMdig	CRMdig is the key to provenance metadata. It describes and integrates the digital provenance with the physical object that has been measured. CRMdig is an ontology as well as the RDF schema to encode metadata about the steps and methods of production ("provenance") of digitization products and synthetic digital representations Doer, Theodoridou, "CRMdig: A Generic Digital Model for Scientific Observation"	Implemented within VEPIS
	Implemented within VEPIS		
	OAIS Reference Model	Provides a framework for understanding archival concepts needed for long-term preservation of digital information and its access <i>ISO 14721:2012 Space data and information transfer systems – Open archival information system (OAIS) – Reference Model</i>	

Going further along the added value chain of provenance, we come to the process *Content Packaging*, which refers to structuring and modelling provenance information. This process is based

on two functions: *Implementation of Provenance Data Models* and *Adaptation of Provenance Metadata Standards*. The first function is realized by the following dimensions: OPM, CRMdig, OAIS, RDF and the so-called *Provenance-Aware Applications Models*.

The first dimension of the function *Implementation of Provenance Data Models* is OPM, which: (1) allows provenance information to be exchanged between systems by means of a compatibility layer based on a shared provenance model, (2) allows developers to build and share tools that operate on such a provenance model, (3) defines provenance in a precise, technology-agnostic manner, (4) supports digital representation of provenance for any “thing”, whether produced by computer systems or not, (5) allows multiple levels of description to coexist, and (6) defines a core set of rules that identify the valid inferences that can be made on provenance representation. According to the specification of the OPM, it allows for the capture of the causal dependencies between artefacts, processes and agents. There are some definitions of dependencies (ontology): causal relationship, artefact used by a process, artefacts generated by processes, processes triggered by processes, artefact derived from artefact, and processes controlled by Agent and Role.¹¹

The second dimension of the function is *W3C Working Group Model*, which models what should be contained in the provenance data (its content); “it will have different concerns from the developing software to capture and maintain the data (its management) and again different concerns from those providing solutions to solve specific user problems (its use).”¹² It provides description of the categories and their dimensions. The Working Group also developed the dependencies (ontology) (PROV-O). Three classes provide the basis for PROV-O ontology: *Entity* (a physical, digital, conceptual or other kind of thing with some fixed aspects; entities, which may be real or imaginary; *Activity* (something that occurs over a

11 Moreau et al., “The Open Provenance Model Core Specification (v1.1)”.

12 Groth et al., “Requirements for Provenance on the Web”.

period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using or generating entities); and *Agent* (something that bears some form of responsibility for an activity taking place, for the existence of an entity or for another agent's activity). The PROV-O ontology provides a set of classes, properties and restrictions that can be used to represent and interchange provenance information generated in various systems and under various contexts. It may also be customized to create new classes and properties to model provenance information for different applications and domains.

The third dimension is of the function *Implementation of Provenance Data Models* is *CRMdig*. This model describes and integrates digital provenance with the physical object that has been measured. *CRMdig* is both an ontology and the RDF schema to encode metadata about the steps and methods of production (“provenance”) of digitization products and synthetic digital representations. It is also the key to provenance metadata. The description of processes starts at the level of human activities or actions, which in turn, among others, “initiate machine events” on devices and computers and forms a connected graph through the data involved in multiple events in roles as input and output.¹³

The fourth dimension of the function *Implementation of Provenance Data Models* is *OAIS ISO 14721:2012*, defines the reference model for an open archival information system (OAIS). It meets a set of responsibilities defined in this standard; it allows an OAIS archive to be distinguished from other uses of the term “archive”. It also expands consensus on the elements and processes for long-term preservation and access of digital information, promotes a larger market which vendors can support and guides the identification and production of OAIS-related standards.¹⁴ In the context of this paper, it is important that OAIS suggests that PDI should contain provenance information documenting the history of the data object.

13 *CRMdig: Model for Provenance Metadata*.

14 *ISO 14721:2012 Space data and information transfer systems – Open archival information system (OAIS) – Reference Model*, <https://www.iso.org/standard/57284.html>.

TABLE 9 The model of provenance-related added value of VEPIS: *Content Packaging Process* (cont.)

Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
Implementation of Provenance Data Models	RDF	<p>Resource Description Framework (RDF) is a family of W3C originally designed as a metadata model</p> <p><i>RDF Schema 1.1. W3C Recommendation 25 February 2014</i></p>	Implemented within VEPIS
	Provenance-Aware Applications	<p>Provenance systems that produce a description of their execution</p> <p><i>Groth, Paul et al. An Architecture for Provenance Systems¹⁵</i></p> <hr/> <p>Deterministic model: the location of any piece can be uniquely described by path. This model uses a variation of existing edge-labeled tree models for semistructured data.</p> <p><i>Buneman, et al. Why and Where: A Characterization of Provenance¹⁶</i></p>	

The fifth dimension of the function *Implementation of Provenance Data Models* is *Resource Description Framework (RDF)*, which was originally designed as a metadata model. It has come to be used as a general method for the conceptual description or modelling of information that is implemented in the Web using a variety of syntax notations and data serialization formats. It is also used in knowledge management applications.

15 Groth, Paul et al. *An Architecture for Provenance Systems*, https://www.researchgate.net/publication/39994555_An_Architecture_for_Provenance_Systems.
 16 Buneman, Peter, Sanjeev Khanna and Wang-Chiew Tan. *Why and Where: A Characterization of Data Provenance*, https://repository.upenn.edu/cgi/viewcontent.cgi?article=1209&context=cis_papers.

The sixth dimension of the function *Implementation of Provenance Data Models* is *Provenance-Aware Application Model*. Systems based on this model produce a description of their execution and store this information in the provenance store. The provenance store also provides querying facilities to enable services to retrieve provenance of data items. The development of the architecture has been strongly influenced by the service-oriented architectural style, according to which services or actors interact with each other by exchanging messages.

TABLE 10 The model of provenance related added value of VEPIS: *Content Packaging Process* (cont.)

CONTENT PACKAGING PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Adaptation of Provenance Metadata Standards</i>	<i>Descriptive Metadata Standards</i>	<p>MARC and MODS (PREMIS): describe the intellectual entity through properties such as author and title and support discovery and delivery of digital content. It may also provide a history context by, for example, specifying which print material was the original source for digital derivation (source provenance)</p> <p>Dappert and Enders. Digital Preservation Metadata Standards</p>	<p>Implemented within VEPIS: EAD for archives; UNIMARC and MARC21, for libraries; and CDWA Lite and ESE, for museums</p>

CONTENT PACKAGING PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Adaptation of Provenance Metadata Standards</i>	Structural, Technical and Administrative Metadata Standards	<p>METS: METS standard, which is universally regarded as the most convenient XML schema reflecting the hierarchical structure of digital libraries, the related descriptive administrative metadata as well as names and locations of files comprising the digital object.</p> <p><i>Dappert and Enders. Digital Preservation Metadata Standards</i></p>	Implemented within VEPIS

The second function of the process *Content Packaging* is *Adaptation of Provenance Metadata Standards*. The W3 Provenance Working Group concluded that provenance could be viewed as metadata, but not all metadata is provenance. Provenance provides a substrate for deriving different trust metrics and provenance records can be used to verify and authenticate among other uses. Three dimensions were established: *Descriptive Metadata* (MARC and MODS), *Structural Metadata*, *Technical Metadata* and *Administrative Metadata* (METS).

According to Angela Dappert and Markus Enders,¹⁷ the main descriptive provenance metadata are *MARC* and *MODS*, which describe the in-

17 Angela Dappert and Markus Enders, "Digital Preservation Metadata Standards," *ISO: Information Standards Quarterly*, 22:2 (2010): 5-13, http://www.niso.org/apps/group_public/download.php/4299/isqv22no2.pdf.

tellectual entity through properties such as author and title and support discovery and delivery of digital content. This metadata may also provide a history context by, for example, specifying which print material was the original source for digital derivation.

Descriptive Metadata. In order to ensure long-term preservation of resources, it is essential to create separate bibliographic records for the original and its digitized object as it allows data to be included about physical and technical attributes of the digitized object, data about the rights and terms of use, and coded data necessary for long-term preservation and determining the publication from which the copy was made.

Within VEPIS the following standards are applied: EAD for archives; UNIMARC and MARC21, for libraries; and CDWA Lite and ESE, for museums.

UNIMARC¹⁸ (the main metadata format of VEPIS) contains the following new fields for provenance: (1) 703 PERSONAL NAME – PROVENANCE OR OWNERSHIP (new), which includes the name of a person regarding any type of provenance and/or ownership (previous ownership or custodianship, author of signature, informal dedication annotations, etc.) that apply uniquely to the item; (2) 713 CORPORATE BODY NAME – PROVENANCE OR OWNERSHIP (new), which contains the name of a corporate body regarding any type of provenance and/or ownership (previous ownership or custodianship, author of signature, informal dedication, annotations, etc.) that apply uniquely to the item; and (3) 723 FAMILY NAME – PROVENANCE OR OWNERSHIP (new) which contains the name of a family regarding any type of provenance and/or ownership (previous ownership or custodianship) that applies uniquely to the item. The new fields provide for reflecting the provenance of documents in traditional form; however, reflection of digital provenance within descriptive information will remain the function of notes fields as long as appropriate solutions will be made.

For archiving digitized objects, VEPIS applies the METS standard, which is universally regarded as the most convenient XML schema re-

18 UNIMARC *Bibliographic*, 3rd edition (with updates), 2017, <https://www.ifla.org/publications/unimarc-bibliographic--3rd-edition--updates-2012-and-updates-2016?og=33>.

flecting the hierarchical structure of digital libraries, the related descriptive administrative metadata as well as names and locations of files comprising the digital object. It also allows integral sets of digital objects to be built, within which qualitative control is performed, corrupt images are returned for re-digitization and objects are linked to the OCRed full-text files or bibliographic objects. The METS format is also applied for exporting objects with their attributed metadata to the central database of VEPIS. The METS package used to export metadata comprises detailed object-related information and references to external objects (digital objects or sound files). The system applies the following sections of METS: “Identification data of the object”, “Descriptive metadata”, “Administrative metadata”, “File section” and “Structural map”.

Conclusions from investigating the Content Packaging Process

The analyses of some these models showed that granularity of contemporary models varies, e.g., the ontology assumed by OPM is minimal. It comprises only 3 classes: (*Artefact*, *Process* and *Agent*) and five associations between them: (*Used as an artifact*, *wasGeneratedBy a process*, *WasControlledBy an agent*, *wasTriggeredBy a process* and *WasDerivedFrom an artefact*).

According to Theodoridou et al. “provenance information recorded according to CRMdig can be mapped to an OPM-based view, but not the other way round“, namely, “the ontology assumed by OPM does not explicitly model the concept *Event*; it influences the representation adequacy because events allow tracing the history of an object and enable the integration of several information that concern the object. Without the notation of event and of physical objects that are carriers (devices), it is impossible to adequately describe the conditions under which the process was done.”¹⁹ Given the fact that VEPIS is based on the CIDOC CRM ontology comprising 80 classes and 132 relations and possessing a rich structure of “intermediate” classes and relations enabling queries at var-

19 Maria Theodoridou et al., “Modeling and querying provenance by extending CIDOC CRM,” *Distributed and Parallel Databases*, 27:2 (2010): 169-210, doi 10.1007/s10619-009-7059-2.

ious levels of abstraction and granularity; and that it is a generic model for recording the “what happened” on the human scale, we conclude that the process *Content Packaging* generates provenance-related added value within VEPIS.

The function *Adaptation of Provenance Metadata Standards* within VEPIS is realized by the Component *Maintenance of Digitized Objects*, which enables:

- preparation of METS packages consisting of descriptive structural metadata and administrative metadata and for exporting to the database of VEPIS;
- supplementing metadata of digitized objects provided by VEPIS data providers with technical data generated by the system and linking them with the object within the portal www.epaveldas.lt.

Market Making Process

TABLE 11 The model of provenance-related added value of VEPIS: *Market Making Process*

MARKET MAKING PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Management</i>	<i>Publication</i>	<p>Making provenance available on the Web</p> <p><i>Groth et al., “Requirements for Provenance on the Web”</i></p>	Published on www.epaveldas.lt
	<i>Access</i>	<p>The capability to find provenance for a particular artefact</p> <p><i>Groth et al., “Requirements for Provenance on the Web”</i></p>	Capability to find provenance for a particular artefact via www.epaveldas.lt

MARKET MAKING PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Management</i>	<i>Access</i>	<p>Provenance related query services: CRM_{dig} allow querying of the most relevant facts and retrieving full descriptions encoded in this model by generic CIDOC CRM terms</p> <p><i>CRM_{dig} version 3.2.1</i></p>	<p>Semantic querying implemented within VEPIS: BAVIC and metadata of digital objects in RDF form in line with the XML schema</p>
		<p>Querying Provenance information can be made accessible in some manner, and there must be mechanisms to find provenance for a given resource</p> <p><i>W3C, Provenance Working Group</i></p> <p><i>W3C, Provenance RDF Named Graph</i></p>	<p>Implemented within VEPIS by using CRM_{dig}</p>
	<i>Dissemination</i>	<p>Defining how provenance should be distributed and controlled</p> <p><i>Groth et al., "Requirements for Provenance on the Web"</i></p>	<p>Realized within VEPIS by the component <i>Publication and access</i>. The portal has all the accessibility features in line with the recommendations of the EU's <i>Web Content Accessibility Guidelines 2.0</i></p>
	<i>Scale</i>	<p>Dealing with large amounts of provenance</p> <p><i>Groth et al., "Requirements for Provenance on the Web"</i></p>	

The process *Market Making* refers to portal services, research tools and distribution. It was conceptualized on the basis of the *Requirements for Provenance on Web* formulated by the W3C Provenance Working Group. We established one function within this process, *Management*, which, according to the model, defines how provenance should be distributed. The function is realized by the following dimensions: *Publication, Access, Dissemination, Scale*.

The first dimension, *Publication*, according to the W3 Provenance Working Group, means that provenance information must be made available on the Web. Related issues include how provenance is exposed, discovered and distributed. The publisher of provenance information should be associated with reference records.

Publication of provenance information within VEPIS is realized by the component *Publication and Access*, which ensures that:

- the portal’s interface has all the accessibility features according to the recommendations of the European Union’s WAI (Web Accessibility Initiative) (*Web Content Accessibility Guidelines 2.0*);
- the portal’s interface is intuitive, understandable and easy for users with an appropriate level of computer literacy (ECDL or higher);
- the interface is controlled by means of a mouse and keyboard or touchscreen;
- user interfaces satisfy the requirements of today’s ergonomics as defined by the standard “LST EN ISO 9241-110:2006 Ergonomics of human-system interaction – Part 110: Dialogue principles (ISO 9241-110:2006)“.

The second dimension is *Access* which, according to the *W3C Provenance Working Group*, is the capacity to find provenance information for a particular artefact. The authors of the *CRMdig Model for provenance metadata* claim that *CRMdig* ensures provenance-related query services retrieving full descriptions encoded in this model by generic CIDOC CRM terms. The *Provenance WG Wiki* states that there must be mechanisms to find provenance for a given resource to make provenance information

accessible.²⁰

The third dimension according to the *W3C Provenance Working Group* is *Dissemination*, which defines how provenance should be distributed and controlled.

The example in Figure 4 below explicitly shows provenance information: the book by Judita Vaičiūnaitė has been digitized, its archival copy is preserved within the computer network of the National Library of Lithuania and it is accessible to the user via a Web browser.

Aitvaras / Judita Vaičiūnaitė. - 2005

Grįžti į sąrašą **peržiūrėti skaitmeninto objekto vaizdus**

Tipas **tekstas**

Autoriai [Judita Vaičiūnaitė](#)

Kiti asmenys [Ula Vaičiūnaitė](#), [Laima Kryževičiūtė](#)

Išleidimo duomenys **Alma littera [2005]**

Apimtis **ilustr., 100, [4] p.; Archyvinė kopija: 1182MB Viešos prieigos kopija: 12MB**

Daugiau informacijos apie **Nacionalinė biblioteka**

objektą

Rašto sistema **lotynų**

Kalbos **lietuvių**

Pastabos **Skaitmenintas objektas: Vilnius; Lietuvos nacionalinė Martyno Mažvydo biblioteka; 2014 01 27**

Tiražas 2000 egz.

Archyvinė kopija: kreipties būdas: Lietuvos nacionalinės bibliotekos kompiuterinis tinklas

Vartotojo kopija: sistemos reikalavimai: interneto naršykle, Adobe Acrobat Reader
Originalas suskaitmenintas TIFF formatu; vartotojo kopija - JPEG formatu
Skaitmenintas objektas iš: Aitvaras : [eileiraščių rinktinė] / Judita Vaičiūnaitė. - Vilnius, [2005]. - 100, [4] p. - ISBN 9955-08-981-4
Prieiga prie archyvinės kopijos teikia LNB

Išteklaus apimtis: 112 vaizdų

Projektas "Virtualios elektroninio paveldo sistemos (VEPS) plėtra" yra finansuojamas iš Europos Sąjungos struktūrinių fondų

UDK indeksas **888.2-93-1, 769.2(474.5)(084)**

Atsisiųsti el. knygą pasirinktu formatu:

Patinka 0

FIGURE 4 Provenance information available at www.epaveldas.lt

Within VEPIS the dimension *Dissemination* is realized by the component *Publication and Access*, which defines access policies. Figure 4 illustrates that provenance information within VEPIS is presented in the Lithuanian language and provides provenance information (presented within field 300 of UNIMARC) containing:

- system requirements for the user: an Internet browser. Adobe Ac-

20 W3C, *ProvenanceRDFNamedGraph*, <https://www.w3.org/2011/prov/wiki/ProvenanceRDFNamedGraph>.

robat Reader, the digital original (TIFF) and the format for the user (JPEG) as well as the mode of access to the archival copy (the computer network of the National Library of Lithuania);

- the digital copy from the source of the digital object: *Aitvaras : [eilėraščių rinktinė] / Judita Vaičiūnaitė. – Vilnius, [2005]. – 100, [4] p. – ISBN 9955-08-981-4;*
- the project under which the digitization was done (financed by the EU Structural Funds);
- accessibility data (the access to the archival copy is provided by the National Library).

The *Fourth* dimension, according to the *W3C Provenance Working Group*, is *Scale* and deals with the amount of provenance. The scale of provenance information is a major concern as the size of provenance records may far exceed the scale of the artefacts themselves. According to the *W3C Requirements*, “this poses significant data management and querying challenges, which databases and other systems are only beginning to address. Trade-offs must be made regarding the granularity of the provenance records kept and the actual amount of detail needed by users of provenance.”²¹

When discussing the dimension *Scale* within the added value model of VEPIS, we must state that this problem has been only partially solved within VEPIS. The CIDOC CRMdig-based component *Information Presentation* within VEPIS allows for the implementation of the Common Thesaurus of Personal Names Geographical Names and Historical Chronology (BAVIC), functioning as an aid for formulating queries and organizing search results. These tools permit the receipt of information about the object from all the VEPIS partners independent of media types within VEPIS; however, they do not guarantee access to information about studies of the object that have been carried out, or their results across many published repositories. Taking into account the dimension *Scale of Provenance*, we could state that there is no interaction of the research and culture sectors in Lithuania. The Lituanica research institutions and infra-

21 Groth et al., “Requirements for Provenance on the Web”.

structures are supervised by the Ministry of Education, whereas memory institutions are administered by the Ministry of Culture. For this reason, heritage objects (movable and immovable) are beyond the context of their research investigation: information about the digitization of heritage objects is in the domain of the Ministry of Culture and scholarly research data regarding such objects, in the domain of research institutions. Therefore, information systems of research data and those of cultural heritage are not related: there are no integrated search platforms within remote databases and no coordination of international partnerships. Such a divide is inconvenient for the user seeking to get needed comprehensive information about completed investigations and their results.

The process *Transport* refers to the technical issues of publishing provenance on the Web and of transferring data between geographically-distributed components of system. It is based on the WAN (wide area network) and LAN (local area network) infrastructure without specific elements for provenance data; therefore, it is not being analyzed in detail in this publication.

Delivery Support and Services

TABLE 12 The model of provenance related added value of VEPIS: *Delivery Support and Services*

DELIVERY SUPPORT AND SERVICES			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Providing Access and Exchange the Data</i>	<i>Protocol Compatibility</i>	<p>Z39.50: Most popular synchronous protocol based on client/server architecture and MARC format structures</p> <p><i>Dappert and Enders. Digital Preservation Metadata Standards</i></p>	<p>Implemented within VEPIS: Z39.50, i.e. UNIMARC/A corresponds to the format UNIMARC/B on the level of content designators</p>

DELIVERY SUPPORT AND SERVICES			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
Providing Access and Exchange the Data	Protocol Compatibility	<p>SRU (Search/Retrieve via URL): a standard synchronized search protocol for Internet search queries utilizing CQL (Common Query Language), which is a syntax for representing queries</p> <p><i>Morgan, An Introduction to the Search/Retrieve URL Service (SRU)</i></p>	In VEPIS: implemented within Providing Access to the Network component
		<p>OAI-PMH: a protocol for metadata harvesting, which enhances the description of resources on the Web</p> <p><i>Sharma et al. Exploring OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting²²</i></p>	Implemented within The Component of the OAI-PMH Harvester
		<p>RDF, which is a low-barrier mechanism for repository interoperability</p> <p><i>RDF Schema 1.1. W3C Recommendation 25 February 2014</i></p>	Implemented within the Component of RDF

The process *Delivery Support and Services* is based on the following two functions: *Providing Access and Exchanging Data* and *Application of Permanent Identifiers*. The first function is realized by protocols and standard models for data interchange on the Web used for Internet searching and

22 Shruti Sharma, J.P. Gupta and A.K. Sharma. "Exploring OAI-PMH: Open Archives Initiative Protocol For Metadata Harvesting," International Journal of Advanced Research in Computer Science, 1:2 (2010), <http://www.ijarcs.info/index.php/Ijarcs/issue/view/2>.

for data exchange (Z39, SRU, OAI-PMH and RDF, which are low-barrier mechanisms for repository interoperability). The function *Providing Access and Exchanging Data* of the process *Delivery Support and Services* within VEPIS is realized by the following components:

- 1) *Providing Access to the Network*, which enables:
 - automated transfer of digital copies indicated in the metadata into the VEPIS database by using the OAI-PMH v2.0 protocol;
 - submitting the received data to the VEPIS component *Data Submission*;
 - providing a graphic interface to the management of repository functions;
 - delivering metadata of cultural heritage objects and digitized content to the European virtual space;
 - conversion of preserved metadata into formats supported by the OAI-PMH repository (oai_telap, oai_dc, oai_ese).

- 2) *OAI-PMH Harvester*, which enables:
 - managing integration into VEPIS of the content from other OAI-compatible systems for preserving digital content;
 - automatic loading of digital object metadata into the VEPIS database;
 - actions of the OAI-PMH harvester (defined by the OAI-PMH protocol): identifying submitted metadata formats, records and descriptions of records;
 - submitting records by using OAI-PMH.

- 3) *RDF*, which enables:
 - presentation of descriptive metadata of BAVIC and digitized objects in the RDF form (Resource Description Framework, <http://www.w3.org/RDF>) and CIDOC CRM-compatible XML schema)
 - presentation of elements of descriptive metadata of diverse institutions, as well as BAVIC records and semantic links between their elements, in the RDF form in line with principles of the Semantic Web (<http://www.w3.org/2001/sw/>).

TABLE 13 The model of provenance related added value of VEPIS: *Delivery Support and Services* (cont.)

DELIVERY SUPPORT AND SERVICES			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Application of Persistent Identifiers</i>	URN	Standard, permanent and unique identifier for digital resources on the Internet	Not implemented
	DOI	Unique identifier of digital objects	Implemented
	URL (PURL)	Used temporarily until URN is implemented	Implemented: PURL

The second function, *Application of Persistent Identifiers*, is implemented by introducing the following three dimensions: URN, DOI and PURL. URN (Universal Resource Name) is a standard, persistent and unique identifier for digital resources on the Web. URNs and URLs are forms of Uniform Resource Identifiers (URIs). URNs use ISBN and ISSN as NIDs. VEPIS has URNs based on the National Bibliography Numbers (NRNs) with NBNs as NID.

DOI (Digital Object Identifier) is a system of unique identifiers based on the Handle System and allowing the allocation of a unique digital identifier to commercial digital publications. Since the DOI structure is similar to those operating within ISBN (publisher, content, etc.), it is used within metadata in VEPIS.

Within VEPIS, URI is realized as the permanent URL (PURL). VEPIS, which operates in line with the METS standard, contains the description of provenance in line with CRMdig expressed in the RDF/XML form.

All functions from the process *Delivery Support and Services* are used in the software of the process *Interface and System*.

Interface and System

TABLE 14 The model of provenance-related added value of VEPIS: *Interface and System*

INTERFACE AND SYSTEMS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Access via the Public Online Catalogue (OPAC)</i>	<i>Access</i>	How to enable the end user to consume provenance	Application for Web pages; access at www.epveldas.lt
<i>Access and Automatic Data Import via OAI-PHM</i>	<i>Access</i>	Automatic data import via the OAI-PHM protocol	Application for access and automatic data import via OAI-PHM

The process *Interface and Systems* refers to Internet services for user data. It is based on the following two functions: access via the public online catalogue and application for access and automatic data import via the OAI-PHM protocol.

Within VEPIS, open technologies are applied for consuming provenance. User interfaces are realized on the basis of Web-based principles via a standard Web. The search of the provenance information is based on CRMdig allowing to:

- identify the creator of the object,
- identify earlier versions of the item,
- identify the events that changed the custody of the item,
- identify the master version of the object,
- identify the scanner/resolution of the digital object.

The portal's application has all the accessibility features in line with the recommendations of the European Union's WAI (Web Accessibility Initiative) (Web Content Accessibility Guidelines 2.0).

The portal's application is intuitive, understandable and easy to use for users with an appropriate level of computer literacy (ECDL or higher).

Interface and System Process

TABLE 15 The model of provenance-related added value of VEPIS: *Interface and System Process*

INTERFACE AND SYSTEMS PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
Use	<i>Understanding</i>	How to enable the end user to consume provenance <i>Groth et al., “Requirements for Provenance on the Web”</i>	Queries regarding provenance in VEPIS are based on the paths of CRM _{dig} ClassName – PropertyName – TargetClassName <i>Specification of VEPIS</i>
	<i>Interoperability</i>	Integrating provenance produced by multiple systems <i>Groth et al., “Requirements for Provenance on the Web”</i>	VEPIS integrates provenance information produced by 9 partners, Europeana and TEL

The process *Interface and Systems* refers to the functionality of the portal’s application. It is based on one function, *Use*. According to the W3C Provenance Working Group, “provenance records may be expected to serve a variety of purposes.”²³ The W3C Provenance Working Group identified seven dimensions of the *Use* category, which we will use within our model: *Understandability*, *Interoperability*, *Comparison*, *Accountability*, *Trust*, *Imperfection* and *Debugging*.

The first dimension of the function *Use* is *Understandability*, which

²³ Groth et al., “Requirements for Provenance on the Web”.

means how to make provenance information understandable to the consumer and how appropriate presentation and visualization of provenance information has to be created. This dimension is explicitly related to *Publication* and *Dissemination*, both dimensions of the *Market making* process; they, as it was mentioned before, are realized within VEPIS by the component *Publication and Access*, which guarantees that:

- the portal's application has all the accessibility features according to the recommendations of the European Union's WAI (Web Accessibility Initiative) (*Web Content Accessibility Guidelines 2.0*);
- the portal's application is intuitive, understandable and easy for users with an appropriate level of computer literacy (ECDL or higher);
- the interface of the application is controlled with the help of a mouse and keyboard or touchscreen;
- user interfaces of the application satisfy the requirements of today's ergonomics as defined by the standard "LST EN ISO 9241-110:2006 Ergonomics of human-system interaction – Part 110: Dialogue principles (ISO 9241-110:2006)".

The second function is *Interoperability*, which, according to the W3C Provenance Working Group, means that provenance information may be obtained from heterogeneous systems and various representations. This dimension is interoperable with the above-discussed dimension *Scale* of the process *Market making*. As regards the *Intereoperability* dimension, we should state that this issue has been only partially solved within VEPIS as it was mentioned above. 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).

Another important dimension of the function is *Comparison of Artefacts Based on Their Origins*. According to the W3 Provenance Working Group, two artefacts may seem very different while their provenance may indicate significant commonalities. Conversely, two artefacts may seem alike, while their provenance may reveal significant differences. VEPIS integrates var-

ious abundant data about objects from libraries, museums and archives which may be of diverse origin, but the data about them may be the same.

A very important dimension of the function *Use* is *Trust*. According to the W3C Provenance Working Group, *Trust* is often based on attribution information; it checks reputation of the entities involved in provenance. Another dimension of *Use* connected with the quality of the information is *Imperfection*, which, according to the W3C Provenance Working group, means “using provenance information may imply handling imperfections. Provenance information may be incomplete in that some information may be missing or incorrect, if there are errors. Provenance information may also be provided with some uncertainty or be of a probabilistic nature”.²⁴

And the last dimension within *Use* is *Debugging*. According to the W3C Provenance Working Group, it means using provenance to detect failures or bugs. This Group also states that without a record of provenance, it would be impossible to determine whether such a “bug” in the processes had indeed occurred.

TABLE 16 The model of provenance related added value of VEPIS: *Interface and System Process* (cont.)

INTERFACE AND SYSTEMS PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
<i>Use</i>	<i>Comparison</i>	Comparison of artefacts based on their origins. Two artefacts may be widely different while their provenance may indicate significant commonalities <i>Groth et al., “Requirements for Provenance on the Web”</i>	Not implemented within VEPIS
	<i>Accountability</i>	Using provenance to assign credit blame <i>Groth et al., “Requirements for Provenance on the Web”</i>	Not implemented within VEPIS

24 Ibid., 53.

INTERFACE AND SYSTEMS PROCESS			
Functions	Dimension	Substantiating statements	Substantiating statements for VEPIS
Use	<i>Accountability</i>	<p>“information accountability means the use of information should be transparent so it is possible to determine whether a particular use is appropriate under a given set of rules, and that the system enables individuals and institutions to be held accountable for misuse.”</p> <p>Weitzner et al., “Information accountability”²⁵</p>	
	<i>Trust</i>	<p>Using provenance to assign credit blame</p> <p>Groth et al., “Requirements for Provenance on the Web”</p>	Specific components implemented within VEPIS
	<i>Imperfection</i>	<p>Dealing with imperfections in provenance records</p> <p>Groth et al., “Requirements for Provenance on the Web”</p>	Specific components implemented within VEPIS
	<i>Debugging</i>	<p>Using provenance to detect failures or bugs</p> <p>Groth et al., “Requirements for Provenance on the Web”</p>	

There are several components for validating the authenticity of a preserved data object 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

25 Daniel J. Weitzner et al., “Information accountability,” *Communications of the ACM*, 51:6 (2008): 81-87, doi: <http://doi.acm.org/10.1145/1349026.1349043>.

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. Therefore, we may conclude that within VEPIS, provenance documents the history of creation, ownership, accesses and changes that have occurred over time for a particular data object.

Conclusions

Qualitative analysis of scientific literature as well as the specification of VEPIS and its services allowed for the establishment of a model for provenance-related added value in VEPIS. Drawing on this model, we could conclude that VEPIS, which is based on OAIS, CRMdig and RDF, supports the following functionality:

1. Providing 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;
2. Provenance information within VEPIS allows referring to the versions of objects as they evolve, are modified or accessed over the time. In particular it provides for a representation of how one version (or parts thereof) was derived from another version due to the components of VEPIS modelled in line with the OAIS reference model. The identity of the object is strongly related to PDI (Context, Provenance, Fixity, and Reference Information as defined in OAIS) and helps users to understand the environment of the resource. PDI within the Archival Information Package provides events that occur during the lifecycle of digital objects (license holder, registration and copyright). It guarantees the authenticity of the object. VEPIS as a CRMdig-based system also provides the derivation chain (“ByWhom”), which documents the history of the content information and refers to its origin or source, any

changes that may have taken place since it was originated and who has had custody of it since it was originated.

3. There is a standard way within VEPIS to find provenance information for a given resource: a standard WEB application within VEPIS based on the OPAC.
4. VEPIS uses a standard way to present several basic provenance entities: (1) PURL, for referring to an object (resource), (2) a person/entity to which the object is attributed (3) a processing step carried out by a person/entity in creating a new object.
5. Queries within VEPIS regarding provenance are based on paths of CRMdig. We can identify the following query requirements: the creator of the object, earlier versions of the item and the events that changed the custody of the item as well as allowing users to find out how any result was derived (what input influenced the result), identifies the master version of the object and the scanner/resolution of the digital object.
6. We must acknowledge some imperfections of VEPIS, the most fundamental of which is the fact that there is no standard way to obtain provenance regarding the object from heterogeneous systems and different representations. Taking into account that information systems of research data and those of cultural heritage are not related, there are no integrated search platforms within remote databases and no coordination of international partnerships. For this reason, the information from provenance records created by different systems is not integrated either. Therefore, we could conclude that in the future, when the research and cultural domains do become interoperable, the provenance of a given artefact will acquire the potential to be specified by multiple systems and be integrated.

REFERENCES

- BUNEMAN, PETER, SANJEEV KHANNA and WANG-CHIEW TAN. *Why and Were: A Characterization of Data Provenance*, https://repository.upenn.edu/cgi/viewcontent.cgi?article=1209&context=cis_papers.
- CRMdig: Model for provenance metadata. Version 3.2.1. April, 2016, http://www.ics.forth.gr/isl/CRMext/CRMdig/docs/CRMdig_v3.2.1.pdf.
- DAPPERT, ANGELA and MARKUS ENDERS. “Digital Preservation Metadata Standards,” *ISO: Information Standards Quarterly*, 22:2 (2010): 5-13, http://www.niso.org/apps/group_public/download.php/4299/is-qv22no2.pdf.
- DOERR, MARTIN and MARIA THEODORIDOU. “CRMdig: A Generic Digital Provenance Model for Scientific Observation,” in *3rd USENIX Workshop on the Theory and Practice of Provenance (TaPP ‘11)*, June 20–21, 2011, Heraklion, Crete, Greece. Available at: https://www.usenix.org/legacy/event/tapp11/tech/final_files/Doerr.pdf
- GROTH, PAUL, YOLANDA GIL, JAMES CHENEY and SIMON MILES. “Requirements for Provenance on the Web,” *The International Journal of Digital Curation*, 7:1 (2012): 39-56, doi: 10.2218/ijdc. V7i1.213 <http://www.ijdc.net/index.php/ijdc/article/viewFile/203/272>.
- GROTH, PAUL et al. *An Architecture for Provenance Systems*, https://www.researchgate.net/publication/39994555_An_Architecture_for_Provenance_Systems.
- ISO 14721:2012 *Space data and information transfer systems – Open archival information system (OAIS) – Reference Model*, <https://www.iso.org/standard/57284.html>.
- MOREAU, LUC. “The Foundations for Provenance on the Web Foundations and Trends,” *Web Science*, 2:2-3 (2010): 99-241, doi: 10.1561/1800000010, <https://eprints.soton.ac.uk/271691/1/survey.pdf>.
- MOREAU, LUC et al. “The Open Provenance Model Core Specification (v1.1),” *Future Generation Computer Systems*, 27 (2011): 743-756, <https://pdfs.semanticscholar.org/4181/c0d8284d8737f61232365747690>

- 297e31c82.pdf.
- MORGAN, ERIC LEASE. "An Introduction to the Search/Retrieve URL Service (SRU)," *ARIADNE Web Magazine for Information Professionals*, <http://www.ariadne.ac.uk/issue40/morgan/?ref=Fuckonly.com>.
- RDF Schema 1.1*. W3C Recommendation 25 February 2014, <https://www.w3.org/TR/rdf-schema/>.
- Reader's Digest Oxford Complete Wordfinder*. Pleasantville, N.Y.: Reader's Digest Association, 1993.
- SHARMA, SHRUTI, J.P. GUPTA and A.K. SHARMA. "Exploring OAI-PMH: Open Archives Initiative Protocol For Metadata Harvesting," *International Journal of Advanced Research in Computer Science*, 1:2 (2010), <http://www.ijarcs.info/index.php/Ijarcs/issue/view/2>.
- Strategic Developments for the European Publishing Industry Towards the Year 2000*. Created by European Commission, DG XIII, Andersen Consulting, 1996, <http://www.echo.lu/elpub2/en/exengl.pdf>.
- THEODORIDOU, MARIA et al. "Modeling and querying provenance by extending CIDOC CRM," *Distributed and Parallel Databases*, 27:2 (2010): 169-210, doi 10.1007/s10619-009-7059-2.
- UNIMARC Bibliographic*, 3rd edition (with updates), 2017, <https://www.ifla.org/publications/unimarc-bibliographic--3rd-edition--updates-2012-and-updates-2016?og=33>.
- WEITZNER, DANIEL J. et al. "Information accountability," *Communications of the ACM*, 51:6 (2008): 81-87, doi: <http://doi.acm.org/10.1145/1349026.1349043>.
- W3C Provenance Incubator Group, Semantic Web Activity World Wide Web Consortium. *Overview of Provenance on the Web*, <https://www.w3.org/2005/Incubator/prov/wiki/images/0/02/Provenance-XG-Overview.pdf>.
- W3C. *ProvenanceRDFNamedGraph*, <https://www.w3.org/2011/prov/wiki/ProvenanceRDFNamedGraph>.
- W3C. *Provenance Working Group*, https://www.w3.org/2011/prov/wiki/Main_Page.

RELATED LITERATURE

BUNEMAN, PETER and SUSAN B. DAVIDSON. *Data provenance: the foundation of data quality*. September 1, 2010, <http://www.sei.cmu.edu/measurement/research/upload/Davidson.pdf>.

GORDON DUNSIRE. "The Web of data for archival materials," in *Records, archives and memory: Selected papers from the conference and school on records, archives and memory studies, University of Zadar*, edited by Mirna Willer, Anne J. Gilliland and Marijana Tomić, 293-308. Zadar: Sveučilište u Zadru, 2015.

DUNSIRE, GORDON and MIRNA WILLER. "Standard library metadata models and structures for Semantic Web," *Library Hi Tech News*, 28:3 (2011): 1-12.

BRIAN LAVOIE. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide*. 2nd edition, DPC Technology Watch Report 14-02 October 2014, doi: <http://dx.doi.org/10.7207/twr14-02>.

MCDONOUGH, JEROME P. "METS: standardized encoding for digital library objects," *International Journal of Digital Libraries*, 6:2 (2006): 148-158, <https://link.springer.com/article/10.1007/s00799-005-0132-1>.

PROV-DM: The PROV Data Model. W3C Recommendation of 30 April 2013, <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>.

PROVENIJENCIJA U KONTEKSTU KULTURNE BAŠTINE DIGITALNIH SADRŽAJA LITAVSKI PRISTUP

KLJUČNE RIJEČI:

*provenijencija, digitalno
okruženje, kvaliteta
podataka, pouzdanost
informacija, Litavski
informacijski sustav
virtualne elektroničke
baštine (VEPIS)*

SAŽETAK

Glavni je cilj ovog članka naglasiti pitanja provenijencije u digitalnom okruženju kao temeljnog koncepta kvalitete podataka. Člankom se želi teoretski potkrijepiti pojam da je provenijencija informacija ključna za određivanje njihove vjerodostojnosti. Primjenom koncepta dodane vrijednosti 2-3-6 EU za elektroničko izdavaštvo i kvalitativne analize sadržaja istraživačke literature, koncipirali smo sljedeće procese provenijencije: Kreiranje sadržaja, Pakiranje sadržaja, Formiranje tržišta, Podrška i usluge isporuke te Sučelje i sustavi. Ovakav pristup omogućio nam je i prepoznavanje funkcija provenijencije, kao i njihovih dimenzija te određivanje dodane vrijednosti povezane s provenijencijom u Litavskom informacijskom sustavu virtualne elektroničke baštine (Lithuania's Information System of the Virtual Electronic Heritage, VEPIS).