

Authenticity and Provenance in Long-Term Digital Preservation: Analysis of the Scope of Content

Assoc. Prof. Dr. Regina Varnienė-Janssen

Research Center for the Organization of Libraries and Information,
Vilnius University, Faculty of Communication,
Email: regina.varniene-janssen@kf.vu.lt

Dr. Jūratė Kuprienė

independent researcher
Email: jurate.kupriene@gmail.com

Authenticity is a fundamental issue for the long-term preservation of digital objects; however, the validation of authenticity is a complex task – it requires the representation of provenance as a precondition for trust. For that matter, it is crucial to define the key conceptual elements that provide the foundation for such a complex framework. The main objective of this paper is to define how authenticity has to be managed in the digital preservation process: to identify the semantic units needed to support core preservation functions in order to ensure the trustworthiness of digital objects and ensure interoperability – i.e., the ability to exchange data between institutions. This paper presents results from the first stage of the research: the scope of authenticity and provenance content needed to support core preservation functions and ensure interoperability – the ability to exchange data between institutions and assess the added value of the Europeana and VEPIS regarding authenticity and provenance based on the adequacy of analyzed international models for authenticity and provenance.

Methodology: by applying the EU 2–3–6 concept of added value for electronic publishing and qualitative content analysis of research literature, we conceptualized the Content Creation Process, which refers to the structure and meaning of authenticity and provenance. This approach enabled us to identify the scope of the content of authenticity and provenance.

Keywords: authenticity, provenance, trust, digital environment, data quality, trustworthiness of digital objects, Lithuania's Information System of Virtual Electronic Heritage (VEPIS), Europeana.

1. Introduction

In order to guarantee the long-term understandability of preserved data, specialized approaches, models and technologies are needed. Long-term digital preservation is a set of processes, strategies and tools used to store and access digital data for long periods of time, during which technologies, formats, hardware, software and technical

communities are very likely to change. According to CASPAR,¹ long-term digital preservation is a process which transforms content: provides migration and emulation to adapt it to new constraints of rendering

¹ CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) Project. Lamb, D., Prandoni, C., Davidson, J., DCC (2009). Available at: <http://www.dcc.ac.uk/resources/briefing-papers/technology-watch-papers/caspar>

and playability, enhances content to preserve its intelligibility and reusability and ensures integrity and authenticity of the digital resources. Although digital objects are transformed within long-term preservation systems, their evidentiary capacity depends on their authenticity, reliability and accuracy established during their creation and maintained and preserved over time and across technological change. The key requirement for long-term preservation systems is their ability to manage authenticity and the history of record changes. Reliable re-use of digital materials is possible if materials are curated in such a way that their authenticity and integrity is retained. It is based on metadata, which, according to Tennis and Rogers, are “[...] machine- and human-readable assertions about information resources, enables intellectual, physical, and technical control over information resources” (Tennis, Rogers 2012).

This paper is of current importance and addresses a comparatively new topic, which has its roots in the early 1990s (Tennis, Rogers 2012), when the concepts of *authenticity* and *provenance in the digital environment* emerged. Defining and assessing *authenticity* is a complex task implying a number of theoretical and operational/technical activities. These include a clear definition of roles involved, a coherent development of recommendations and policies for building trusted repositories as well as the precise identification of each component of the custodial function. Therefore, it is crucial to define the key conceptual elements that provide the foundation for such a complex framework: we need to define how and on what basis authenticity has to be managed in digital preservation processes in order to ensure the trustworthiness of digital objects.

Authenticity is the fundamental issue of the long-term preservation of digital objects. This, in its turn, supports the verification of the provenance of digital resources despite technological and organizational changes of their context and helps to maximize the initial investment made in creating or collecting them. The relevance of *authenticity*, as a preliminary and central requirement, has been investigated by many international projects and researchers. Some of them focused on the long-term preservation of authentic digital records in the e-government environment, and others – on the scientific and cultural domains, while some devoted their efforts toward the identification of criteria and responsibilities in developing trusted digital repositories (Factor et. al. 2009; Giaretta 2011; InterPARES 3 Project, TEAM Canada 2016; PREMIS 2015; APARSEN 2012; Tennis, Rogers 2012; Guercio, Salza 2012; World Wide Web Consortium (W3C) and other projects, the results of which are analyzed in this paper). Before delving into the issues of authenticity and provenance in the digital environment, we shall discuss the concepts of authenticity and provenance.

According to the CASPAR Conceptual Model, authenticity “is the trustworthiness of the resource to be what it purports to be, untampered with the uncorrupted, based on the identity and integrity” (Giaretta 2011). The final results of this project state that “authenticity of digital objects must be supported by evidence provided by tracing history of its various migration and treatments, which occurred over time. Evidence is also needed to prove that the digital objects have been maintained using technologies and administrative procedures that either guarantee their continuing identity.”

According to MoReq, *authenticity* is the “concept, along with integrity, reliability and usability, one of the central characteristics of a record according to ISO 15489. An authentic record is one that can be proven to be what it purports to be (adapted from ISO 154891:2001, 7.2.2.) [...]” (DLM Forum Foundation 2011).

The Reference Model OAIS (ISO 14721) (CCSDS 2012, hereafter the OAIS) defines authenticity “as the degree to which a person (or system) may regard an object as what it is purported to be.” The OAIS states that the degree of the authenticity of a digital object is strongly related to criteria and procedures adopted to analyze and evaluate it, and authenticity is judged on the basis of evidence. Evidence is also needed to prove that digital objects have been maintained using technologies and administrative procedures that either guarantee their continuing identity and integrity or at least minimize risks of change from the time the objects were first set aside to the point at which they are subsequently accessed. Authenticity is regarded as being established by assessing the integrity and identity of the resource. The identity of a digital object refers to the whole of the characteristics of the digital object that uniquely identify it and distinguish it from any other object; in addition to its internal conceptual structure, it refers to its general context. The integrity of a resource refers to its wholeness. A resource has integrity when it is complete and uncorrupted in all its essential respects. The verification process should analyze and ascertain that the essential characteristics of an object are consistent with the inevitable changes brought about by technological obsolescence. For this reason, authenticity should be monitored continuously, so that every time when either a digital object

is somehow changed or a relationship modified, a verification of the permanence of that object’s relevant features could be guaranteed (CCSDS 2012; Factor et al. 2009; Giaretta 2011; InterPARES 3 Project, TEAM Canada 2016; W3C, PROV-O 2013; Moreau 2011 and other projects).

To validate the *authenticity* of a preserved digital object, provenance is needed. According to the World Wide Web Consortium, the “provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource. Provenance assertions are a form of contextual metadata and can themselves become important records with their own provenance” (W3C 2015). *Provenance* provides a 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. Provenance information is the information that documents the history of a digital object’s content information. This information tells the origin or source of the content information, any changes that may have taken place since it was originated, and who has provided its custody since it was originated. Provenance is often conflated with metadata. Metadata is used to represent properties of the objects, which are reflected in the models and ontologies for provenance. Because of a large diversity of models for provenance, the question of how metadata relates to provenance still remains important.

A digital preservation system is responsible for creating and preserving provenance information starting with ingesting a digital object; however, earlier provenance information should be provided by the producer. Procedures that must be followed to ensure

authenticity and to record provenance information are suggested by several projects. For example, according to the Authenticity Management Tool (ATM), developed by CASPAR project (Giaretta 2011), processes for ensuring maintenance and verification of the authenticity in terms of identity and integrity of the digital objects should be identified. It means that content and contextual information relevant to the authenticity during the entire preservation process should be provided by capturing and making understandable over time all the required information (Giaretta 2011). Going further along the description of concepts of authenticity and provenance, it is important to clarify how trust relates to provenance. *Trust* is a term with many definitions and uses; in many cases, establishing trust in an object or an entity involves analyzing its origins and authenticity. Provenance is related to trust because it is derived from provenance information and generally is a subjective judgment that depends on context and use. It can be argued that *provenance* is a platform for trust. The provenance of information is crucial to making determinations about whether information is trusted, how to integrate diverse information sources, and how to give credit to originators when reusing information (Buneman et al. 2008). As it has been mentioned above, the relevance of authenticity and provenance, as a preliminary and central requirement, has been investigated by many international projects. While the exact requirements for a system enabling provenance capture and retrieval depend on the scope of intended cases of use and the technology upon which it is based, it is important to make explicit the scope of the dimensions. The Requirements for Provenance on the Web provide the scope of the dimensions for modelling

the provenance that should be contained in provenance data (i.e., its content) (Groth et al. 2012). It is noteworthy that there is no exclusively unified model or schema that covers the entire scope of dimensions of authenticity and provenance and contains all the documentation that is needed for preserving e-information. The PREMIS focuses on encoding the preservation actions taking place before and during the ingestion of a digital object into an archival repository, while others, such as the PROV-O and OPM, focus on encoding the provenance history. The OAIS provides the outline that must be followed when developing a long-term digital archival repository as well as guidelines on what semantic information is needed for long-term preservation.

The main objective of this paper is to define how authenticity has to be managed during the digital preservation processes: to identify the semantic units needed to support core preservation functions in order to ensure the trustworthiness of digital objects and ensure interoperability – the possibility to exchange data between institutions. In order to achieve this objective, during the first stage of investigation, we focused on identifying the scope of authenticity and provenance content by drawing on the results of the qualitative analysis of various schemas and models for encoding authenticity and provenance information, developed by different international projects and consortiums, as well as on research literature.

The paper presents the results from the first stage of the research: the scope of authenticity and provenance content needed to support core preservation functions and ensure interoperability – the ability to exchange data between institutions and assessing the added value of the Europeana

and VEPIS regarding the authenticity and provenance based on the adequacy of the analyzed international models for authenticity and provenance.

The research is still being conducted, and further results on the remaining processes of the 2–3–6 concept of added value (Content Packaging, Market Making, Transport; Delivery Support and Services, and Interface and Systems) will be presented in future papers by the authors.

2. Methodology of the Research

The research presented in this article is much focused on the 2–3–6 concept for added value creation, which was developed by the European Commission and serves as the

ting 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 paper applies the 2–3–6 concept as a framework to identify developments in the creation of authenticity evidence and provenance information within the digital environment. 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 a qualitative analysis of scientific literature and the specifications. Considering that the Europeana is an aggregator on the international level², and that VEPIS³ is an

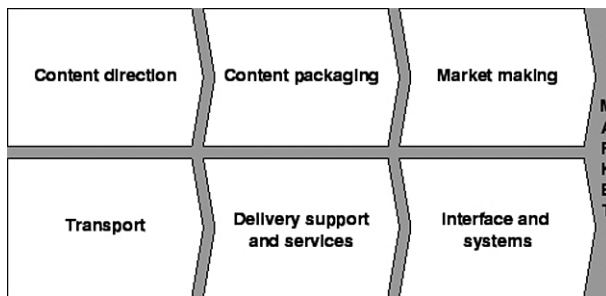


Fig. 1. The 2–3–6 concept of added value for electronic publishing, which was developed by the European Commission (Source: European Commission, DG XIII, Andersen Consulting 1996).

basis for creating added value of electronic publishing (Fig. 1) (European Commission, DG XIII, Andersen Consulting 1996).

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 a qualitative analysis of activities and behavior within a certain field and visualize it – e.g., to determine the impact of provenance on the trustfulness and accuracy of information. For this purpose, by relating value-crea-

² The European Digital Library (Europeana) has been established through the aggregation of heterogeneous content from multiple content providers, which needs to be delivered reliably and consistently using a commonly agreed metadata schema, typically by means of collaborative metadata mapping and delivery projects based on national or thematic aggregators. [...] system architecture, based on the implementation of a curation-aware, OAIS-compliant true digital repository capable of supporting comprehensive metadata ingestion, curation, preservation, transformation and harvesting services. A Curation-Oriented Thematic Aggregator . Available at: https://doi.org/10.1007/978-3-642-40501-3_13

³ <...>VEPIS as - OAIS-compliant digital reposi

aggregator on the national-level (Lithuanian) of cultural heritage⁴, it is feasible to determine the added value of these systems in regard to authenticity and provenance. Such a methodological approach correlates with the approach declared by the Australian Government Recordkeeping Metadata Standard (National Archives of Australia 2008), implying that the key categories of authenticity are *context*, *content of records* and *their management through time*. Such a methodological approach also correlates with the approach declared by the W3C Provenance Working Group, implying that the key categories of provenance are *Content*, *Management* and *Use* (W3C 2015).

By applying the EU 2–3–6 concept of added value for electronic publishing and the qualitative content analysis of research literature, we conceptualized the process of *Content Creation*, which refers to the structure and meaning of authenticity and provenance entities (dimensions). This approach enabled us to evaluate the scope of metadata in different application profiles and specifications for authenticity and provenance and to identify the added value of the Europeana and VEPIS regarding the completeness of the defined dimensions of the authenticity (trustworthiness) and provenance (history of creation, ownership, access and change) of preserved objects.

This paper presents the results from the first stage of our research: the scope of aut-

henticity and provenance content needed to support core preservation functions, ensure interoperability – the possibility to exchange data between institutions – and the added value model of the Europeana and VEPIS regarding the authenticity and provenance.

The broader study is still being conducted, and further results on the remaining processes of the 2–3–6 concept of added value (Content Packaging, Market Making, Transport; Delivery Support and Services, and Interface and Systems) will be presented in future papers by the authors.

3. The Model of Authenticity and Provenance Related to the Added Value of Content Creation Process in the VEPIS and Europeana

The first process of *Content Creation* refers to the structure and meaning of authenticity and provenance entities (dimensions). It is based on one function – *Identifying the Scope of Authenticity and Provenance Content*. In order to conceptualize this function, we analyzed the InterPARES general study report on the Application Profile for Authenticity Metadata, the Model Requirements for the Management of Electronic Records (MoReq2010), the PREMIS Dictionary, the CIDOC Conceptual Reference Model (CIDOC CRM), the CRM_{dig} ontology (the extension of CIDOC-CRM to support provenance metadata), ISO 23081, ISO 15489, the Open Provenance Model Ontology (PROV-O), the Open Provenance Model (OPM), the Requirements for Provenance on the Web by the W3C Provenance Working Group, the OAIS, the so-called provenance-aware application models as well as research literature on provenance and authenticity. We categorized the dimensions according to the

tory, outlines the application of CIDOC CRM and its extension CRM_{dig} and RDF<...> capable of data ingestion, supporting, curation, preservation, transformation and providing access services <...>. 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. ISBN 978-953-331-020-6.

kind of functionality they imply. The listed dimensions are not intended to themselves be part of a vocabulary for authenticity and provenance but as a framework for assessing how adequate are the analyzed systems of the Europeana and VEPIS for the models regarding authenticity and provenance. The key dimensions of authenticity and provenance have been identified and mapped as the following:

1. *Action/Event/Process/Aggregation/Entailment/Justification, Versioning;*
2. *Actor/Agent/Person/User;*
3. *Artefact/Entity/Object/Record/;*

4. *Class/Classification Code;*
5. *Date/Time/Time span;*
6. *Form/Type;*
7. *Identifier;*
8. *Original name/Title;*
9. *Place/Room;*
10. *Relation/Links;*
11. *Rights/Access Rights/Mandate;*
12. *Technology/Application;*
13. *Signature/Authentication.*

Since the terminology varies between models and frameworks, we have attempted to unify the definitions according to a functionality that they comply to.

Table 1. Dimension: Action/Event/Process/Aggregation/Entailment/Justification/Versioning

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
Action: 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 Europeana by EDM ⁴ edm:Event	Implemented within VEPIS by using CRM _{dig} <i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i>
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>		-----
Digital Machine Event , Class D7: 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 <i>CRM_{dig} version 3.2.1</i>	EDM has been aligned to CIDOC-CRM in its definition of an event-centric model <i>Isaac, A. (2013) Europeana Data Model Primer</i>	Implemented as D7 Digital Machine Event in line with CIDOC CRM Version 6.2.3, May 2018 <i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i>
Digitization Process , Class D2: the type of the process and techniques applied <i>Doer, Theodoridou. CRM_{dig}: A Generic Digital Model for Scientific Observation</i>		Implemented as D2 in line with CRM _{dig} , version 3.2.1 <i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i>

⁴ The Europeana Data Model (EDM) is a new proposal for structuring the data that the Europeana will be ingesting, managing and publishing. The Europeana Data Model is a major improvement on the Europeana Semantic Elements (ESE), the basic data model that Europeana began life with (Europeana Data Model Primer 14/07/2013). Available at: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Event: description of the outcome of a function that was performed previously and is retained to show the history of an entity</p> <p><i>MoReq Specification, version 1.1</i></p>	<p>Implemented within Europeana by EDM edm:Event</p>	<p>-----</p>
<p>Software Execution, Class D10: events by which a digital device runs a software program or series of computing operations on a digital object</p> <p>Provenance</p> <p><i>CRM_{dig} version 3.2.1</i></p>	<p>EDM has been aligned to CIDOC CRM in its definition of an event-centric model</p> <p><i>Isaac, A. (2013). Europeana Data Model Primer</i></p>	<p>Implemented as <i>D10</i> according to CRM_{dig} version 3.2.1</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Aggregation (entity): aggregations of records are accumulations of related record entities that, when combined, may exist at a level above that of a single record</p> <p><i>MoReq Specification Version 1.1, 2018, p.241</i></p> <p>Aggregation, Class ore:Aggregation. It is the pivotal object between the edm:ProvidedCHO and the edm:WebResource(s)</p> <p>Definition of the Europeana Data Model v5.2.8, 2017</p>	<p>Implemented according to EDM model ore:Aggregation</p>	<p>VEPIS is the national aggregator, cooperating with Europeana.</p> <p>Workflows presented and described in Section C18. OAI-PMH component of harvesting of</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Disposal hold: legal or other hold preventing the scheduled destruction of records</p> <p><i>MoReq Specification</i></p>		
<p>Entailment: explanations showing how facts were derived from other facts</p> <p><i>Groth et al. Requirements for Provenance on the Web</i></p>	<p>-----</p>	<p>Implemented as <i>Digital Machine Event and Measurement</i>, according to CIDOC CRM_{dig} thus allowing for correlation of the measured object with the device that</p>
<p>Justification: documentation recording why and how a particular decision is made</p> <p><i>Groth et al., "Requirements for Provenance on the Web"</i></p>	<p>-----</p>	<p>did the measurement</p>
<p>Versioning: 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>	<p>-----</p>	

Digital Events

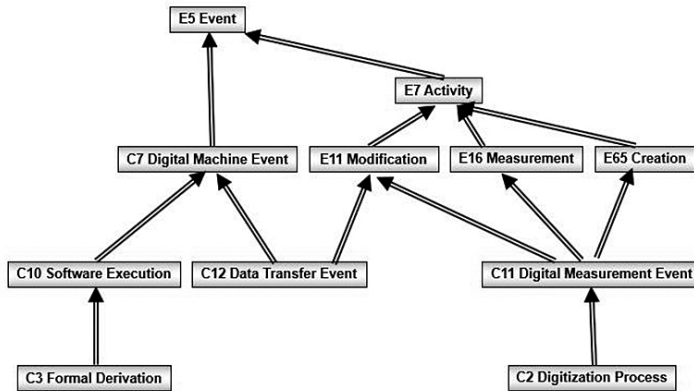


Fig. 2. An illustration of Digital Events according to the CIDOC CRM ontology (Doerr 2009).

The first dimension of the added value model is *Action/Event/Process/Aggregation/Entailment/Justification/Versioning* (see Table 1). According to the W3C Provenance Working Group, this dimension denotes the *activities* (or steps) that were carried out to generate the artefact in hand. According to the OPM, a *process* means *actions* or *series* performed on or caused by artefacts and resulting in new artefacts. According to the CRM_{dig}, a Class D7 Digital Machine Event is a D7 Digital Machine Event. This class comprises the 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's output of the partial events is regarded as the creation of an overall activity, as it is presented in Fig. 2.

Within the VEPIS, we implemented the terminology of CIDOC CRM and CRM_{dig} D2 Digitization Process, D10 Software Execution and D11 Digital Measurement, i.e., events that happened within VEPIS and actions measuring physical properties. According to the 36th Joint Meeting of CIDOC CRM SIG 2016, EDM is partly compliant with CIDOC CRM⁵: a further interpretation of EDM compliance with respect to ISO21127:2014 is underway⁶.

According to the MoReq Specification⁷,

⁵ The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

⁶ At the 36th Joint Meeting of the CIDOC CRM SIG and ISO/TC46/SC4/WG9 and the 29th FRBR – CIDOC CRM harmonization meeting, the SIG discussed the interpretation of EDM compliance with respect to ISO21127:2014. The SIG decided that a mapping of EDM to CRM should be published by SIG in the following formats: (a) X3ML, (b) text and (c) graph representation. These will be created by Athina and will be presented at the next SIG Heraklion, 1/8/2016. (<http://www.cidoc-crm.org/Issue/ID-292-is-edm-compliant-with-crm>).

⁷ The purpose of the MoReq2010 Project is to describe the minimum functionality required of a

the aggregations of records are the accumulations of related record entities that, when combined, may exist at a level above that of a single record. The aggregations of records may reflect relationships such as the shared characteristics or attributes or the existence of sequential relationships between related records, which include the System Identifier (M14.4.100), Created Timestamp (M14.4.9), Originated Date/Time (M14.4.61), First Used Timestamp (M14.4.32), Last Addition Timestamp (M14.4.48), Class Identifier (M14.4.4), Title (M14.4.104), Description (M14.4.16), Scope Notes (M14.4.97), Closed Timestamp (M14.4.5), Destroyed Timestamp (M14.4.17), Max Levels Of Aggregation (M14.4.52), Parent Aggregation Identifier (M14.4.63), and Aggregated Timestamp (M14.4.1) (MoReq Specification Version 1.1, p 241).

According to MoReq, the dimension **Disposal hold** is <...> a legal or other hold preventing the scheduled destruction of records which contains System Identifier (M14.4.100), Created Timestamp (M14.4.9), Originated Date/Time (M14.4.61) First Used Timestamp (M14.4.32), Held Record Identifier (M14.4.39), Held Aggregation Identifier (M14.4.37), Held Class Identifier (M14.4.38), Title (M14.4.104), Description (M14.4.16), Mandate (M14.4.51), Scope Notes (M14.4.97), Destroyed Timestamp (M14.4.17).”

Entailment represents explanations of 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, whereas some 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 (Groth et al. 2012).

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

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

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 a *Digital Machine Event* and *Measurement* (CIDOC CRM dig), thus allowing for a correlation of the measured object with the device that did the measurement.

The second dimension of content creation aggregates four definitions: **Actor / Agent/Person/User**, which are treated

MoReq2010® compliant records system, to define common processes, such as export and disposal, and to establish and standardize an underlying information model that includes entity types, data structures, meta-data element definitions and function definitions.

differently in various models and ontologies and by different projects (see Table 2).

According to the CIDOC CRM ontology, the dimension **Actor** is treated as Class E39 Actor and Class E39, “[...] people, either individually or in groups.”

According to the PREMIS, the dimension **Agent** “aggregates information about attributes or characteristics of agents (persons, organizations, or software) associated with rights management and preservation events in the life of a data object. Agent

Table 2. *Dimension: Actor/Agent/Person/User*

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Actor, Class E39: people, either individually or in groups</p> <p><i>CIDOC CRM</i></p>	<p>Implemented within Europeana edm:Agent edm:dataProvider</p>	<p>Implemented as <i>Actor</i> (Class E39) in line with CIDOC CRM</p>
<p>Agent (entity) aggregates information about attributes or characteristics of Agents (persons, organizations or software) associated with rights management and preservation events in the life of a data object. Agent information serves to identify an Agent unambiguously from all other Agent entities.</p> <p><i>PREMIS Data Dictionary for Preservation Metadata, version 3.0, 2015</i></p>	<p><i>EDM2E Model V 1.2, Specification</i></p>	<p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Agent, edm:Agent. This class comprises people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible. Equivalent to E39_Actor (CIDOC CRM)</p> <p><i>Definition of the Europeana Data Model v5.2.8, 2017</i></p>		
<p>Agent: contextual entity acting as a catalyst of processes enabling, facilitating, controlling or affecting its execution</p> <p><i>Moreau et al. The Open Provenance Model: Core specification (v1.1), p. 3</i></p>		
<p>Agent is treated as a contextual entity acting as a catalyst of processes, enabling, facilitating, controlling or affecting its execution and is presented in Class prov: Agent</p> <p><i>PROV-O: The PROV Ontology</i></p>		
<p>Agent: differently worded in various sectors. See text below</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>		

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Person Name, Class D21, comprises a proper noun name that identifies a person that acts as an entity. Subclass of E82 Actor Appellation</p> <p><i>CRM_{dig} version 3.2.1</i></p>	<p>EDM has been aligned to CIDOC-CRM in its definition of an event-centric model</p> <p><i>Isaac, A. (2013). Europeana Data Model Primer</i></p>	<p>Implemented as <i>Person Name</i> D21, a subclass of E82 Actor Appellation in line with CRM_{dig}, version 3.2.1</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Person: individuals or legally defined entities who are the subjects of rights and duties and are recognized by the juridical system or responsible for carrying out the action to which the object(s) relates or for the matter to which the object(s) pertains</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>		<p><i>Person</i>: implemented as instances of E21 Person of CIDOC CRM, whereas groups should be documented as instances of either E74 Group or its subclass E40 Legal Body</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Person: individuals or legally defined entities who are the subjects of rights</p> <p><i>PREMIS Data Dictionary for Preservation Metadata version 3.0, 2015</i></p>		<p>Implemented <i>Person</i> as instances of E21 Person of CIDOC CRM</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Person: a Person will be represented as the class foaf:Person (a subclass of edm:Agent) and will have a hasAddress property with the range of an Address class, which contains all details of the postal address</p> <p><i>Definition of the Europeana Data Model v5.2.8, 2017</i></p>	<p>Implemented in line with <i>Definition of the Europeana Data Model v5.2.8, 2017</i></p>	<p>Compliant with ESE</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

information serves to identify an Agent unambiguously from all other Agent entities” (PREMIS Editorial Committee 2015).

According to OPM, the dimension **Agent** could be understood as “[...] contextual entity acting as a catalyst of processes enabling, facilitating, controlling or affecting its execution” (Moreau 2011).

According to InterPARES⁸ (InterPARES

⁸ *International Research on Permanent Authentic Records in Electronic Systems (InterPARES)* has the task of developing knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form and of providing the basis for standards, policies, strategies and plans of action capable of ensuring the longevity of such material and the ability of its users to trust its authenticity as well.

2 Project Dictionary 2018), an *Agent* is differently treated in various sectors: “[...] one who is authorized to act for or in place of another. [Archives] n., An active communicating entity that can acquire a role; that is, an abstract representation of a function, service, or identity. [Sciences] n., Person or company licenced by the state to represent clients and negotiate their contracts for a standard agent’s fee. [Arts] n., People authorized to represent or act on behalf of someone else, particularly in business transactions. [Arts] n., An intermediary who performs various matters of business connected with the theatre. [Arts] n.”

In PROV-O, the *Agent* is treated as a contextual entity acting as a catalyst of processes enabling, facilitating, controlling or affecting its execution and presented in the class `prov:Agent` (W3C, PROV-O, 2013). This dimension could be described with the following properties: `prov:actedOnBehalfOf`, `prov:wasAssociatedWith` and `prov:wasAttributedTo`. It has subclasses `prov:Organization`, `prov:Person` and `prov:SoftwareAgent`.

According to CIDOC CRM and CRM_{dig} , the dimension *Person name* is Class D21, which comprises a proper noun name that identifies a person that acts as an entity. This class is a subclass of E82 Actor Appellation, people, either individually or in groups, who have the potential to perform intentional actions, for which someone may be held responsible.

According to InterPARES, the dimension *Person* is treated as “[...] individuals or legally defined entities who are the subject of rights and duties and are recognized by the juridical system or responsible for carrying out the action to which the object(s) relates or for the matter to which the object(s) pertains.”

According to PREMIS, a *Person* is treated as “[...] individuals or legally defined entities who are the subject of rights.”

The third dimension is *Artefact/Entity/Object/Record* (see Table 3).

According to the OPM, an *Artefact* means “[...] an immutable piece of state, which may have physical embodiment in a physical object or a digital representation in computer system” (Moreau et al. 2011).

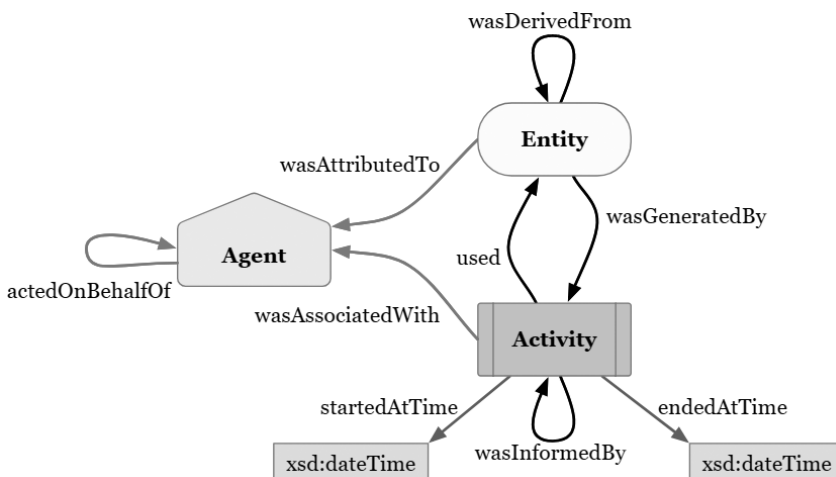


Fig. 3. Three starting point classes and properties that relate them (figure from W3C, PROV-O, 2013.).

Table 3. *Dimension: Artefact/Entity/ Object/Record*

Substantiating statements	Substantiating statements for Europeana	Substantiating statements for VEPIS
<p>Artefact: immutable piece of state, which may have a physical embodiment in a physical object or a digital representation in a computer system</p> <p><i>Moreau et al., The Open Provenance Model: Core specification (v1.1), p. 2</i></p>	<p>Implemented in line with the EDM ontology: edm:object, which has property edm:WebResource</p>	<p>-----</p>
<p>Digital object, Class D1: comprises identifiable immaterial items. Subclass of E73 Information Object, D9 Data Object, D14 Software D35 Area</p> <p><i>CRM_{dig} an extension of CIDOC-CRM 2016</i></p>	<p><i>EDM specification</i></p>	<p>Class D1 implemented within VEPIS</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Digital object: the class edm:object that is the URL of a suitable source image in the best resolution available on the website of the data provider from which small images could be generated for use on a portal. This will often be the same URL as given in edm:isShownBy</p> <p><i>Definition of the Europeana Data Model v5.2.8, 2017</i></p>		<p>Implemented in line with CIDOC CRM_{dig}, Class D1.</p> <p>Compliant with ESE within Europeana</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Object (digital object): part of one or more digital documents, and the metadata necessary to order, structure or manifest its content and form, requiring a given preservation action</p> <p><i>InterPARES dictionary, 2018</i></p>		<p>-----</p>
<p>Object entity: aggregates information about a digital object held by a preservation repository and describes those characteristics relevant to preservation management</p> <p><i>Premis Dictionary for Preservation Metadata, Version 3, p.11</i></p>		<p>-----</p>
<p>Entity: a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary</p> <p><i>PROV-O</i></p>		<p>-----</p>
<p>Digital record: a digital document that is treated and managed as a record (archives)</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>		<p>-----</p>
<p>Record, Class E14.2.12: includes elements of documentary form, annotations, context and medium. These are identified in the attributes, policies and procedures, and controls authenticity</p> <p><i>MoReq2010</i></p>		<p>-----</p>

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Conceptual Object: E55 Type > E28</p> <p><i>CIDOC CRM Version 6.2.3, May 2018</i></p>	<p>Partly compliant: further interpretation of EDM compliance with respect to ISO21127:2014 is underway</p>	<p>Conceptual Object E55 Type > E28. Implemented in line with CIDOC CRM, version 6.2.3, May 2018</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

According to CIDOC CRM_{dig}, the dimension **Digital Object** (Class D1), “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” (CRM_{dig}, 2016).

According to the PREMIS Dictionary for Preservation Metadata, this dimension is treated as an **Object entity** which “[...] aggregates information about a digital object held by a preservation repository and describes those characteristics relevant to preservation management. The only mandatory semantic units that apply to all categories of Object (Intellectual Entity, Representation, File, and Bitstream) are objectIdentifier and objectCategory” (PREMIS Editorial Committee 2015).

The EDM model for this dimension uses the term **Object**, which is a Class edm:object meaning “[...] the URL of a suitable source image in the best resolution available on the web site of the data provider from which small images could be generated for use in a portal. This will often be the same URL as given in edm:isShownBy”. The subclass edm:object is the property of edm:WebResource, which “leads to a thumbnail representing the digital object or, if there is no such thumbnail, the URL of the digital object in the best resolution available on the website of the data provider from which a thumbnail could be generated” (Dröge et al. 2015).

Another dimension often mentioned and described in various models and frameworks for authenticity and provenance in the content creation process is **Class/Classification Code** (see Table 4).

According to MoReq, this dimension treated as **Class** is one that denotes a “[...] business classification applied to records and aggregations of records Service Classification service System metadata, which includes the following metadata: System Identifier (M14.4.100); Created Timestamp (M14.4.9); Originated Date/Time (M14.4.61); First Used Time-

Table 4. *Dimension: Class/Classification code*

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Class: Business classification applied to records and aggregations of records</p> <p><i>MoReq Specification</i></p>	<p>Implemented within Europeana EDM and can be used as relation to the class edm:isRelatedTo</p>	<p>-----</p>
<p>Classification: systematic identification and/or arrangement of business activities and/or records into categories accordingly to logically structured conventions, methods, and procedural rules</p> <p>ISO 15489</p>		<p>-----</p>
<p>SignificantProperties: characteristics of a particular object subjectively determined to be important to maintain through preservation actions</p> <p><i>PREMIS Dictionary, 2008, v.2.0</i></p>		<p>-----</p>
<p>Classification code and registration number (A2.2.4)</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>		<p>List of coded data implemented within VEPIS</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

stamp (M14.4.32); Title (M14.4.104); Description (M14.4.16); Scope Notes (M14.4.97); Default Disposal Schedule Identifier (M14.4.11); Destroyed Timestamp (M14.4.17).”

According to ISO 15489, a dimension treated as **Classification** means the “[...] systematic identification and/or arrangement of business activities and/or records into categories accordingly to logically structured conventions, methods, and procedural rules.”

According to the *PREMIS Dictionary*, this dimension is **SignificantProperties**, which describes the characteristics of a particular object subjectively determined to be important to maintain through preservation actions.

According to the *InterPARES 2 Project Dictionary* (2018), this dimension is **Classification code and registration number** (A2.2.4).

The fifth dimension of the Content Creation Process is **Date/Time/Time span**, which is presented in Table 5.

InterPARES uses the definition **Date of object creation**, which means the date (and presumably time) of compilation and capture, the date of transmission from the originator, the date of receipt and capture and the transmission’s date and time.

According to the PREMIS, this dimension is treated as the class DateCreated-ByApplication, which registers the actual or approximate date and time when the object was created.

Table 5. *Dimension: Date/Time/Time span*

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Date of object creation: date (and presumably time) of compilation and capture, date of transmission from the originator, date of receipt and capture, and the transmission of date and time</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>	<p>Implemented within EUROPEANA EDM: edm:TimeSpan</p>	<p>-----</p>
<p>DateCreatedByApplication: the actual or approximate date and time the object was created</p> <p><i>PREMIS Data Dictionary</i></p>		<p>-----</p>
<p>TimeStamps. Timestamps must contain complete and accurate date and time data, including time zone information, which allows events to be ordered in the sequence in which they occurred</p> <p><i>MoReq</i></p>		<p>-----</p>
<p>Date, E50 Date, Subclass of: E49 Time Appellation. This class comprises specific forms of E49 Time Appellation. <i>CIDOC CRM Version 6.2.3 May 2018</i></p>	<p>EDM has been aligned to CIDOC CRM in its definition of an event-centric model</p> <p><i>Isaac, A. (2013). Europeana Data Model Primer</i></p>	<p>Implemented Class E50 according to CIDOC CRM</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10 Nr. V2.0. VEPIS specification</i></p>
<p>Time span, Class edm:TimeSpan, Subclass edm: NonInformationResource. Equivalent to: Time (ABC Harmony, E52 Time-Span (CIDOC CRM), dol:time-interval</p> <p><i>Definition of the Europeana Data Model v5.2.8, 2017</i></p>	<p>Implemented as Class edm:TimeSpan</p> <p>Equivalent to: Time (ABC Harmony, E52 Time-Span (CIDOC CRM), dol:time-interval.</p>	<p>Implemented as Time span, Class 52 in line with (CIDOC CRM)</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

According to the CIDOC CRM, this dimension includes some of the classes of E49 and E50. E49 comprises all forms of names or codes such as historical periods and dates, which are characteristically used to refer to a specific E52 Time-Span. The instances of the E49 Time Appellation may vary in their degree of precision and they may be relative

to other time frames – “Before Christ,” for example. Instances of the E52 Time-Span are often defined by reference to a cultural period or an event, e.g., “the duration of the Ming Dynasty.” The Class E 50 class comprises specific forms of the E49 Time Appellation.

Time Span is one of the core metadata elements in the Europeana data model. It is suggested to be mapped to *Time* or *Date* in other metadata schemas. It is one of key metadata elements in VEPIS as well.

The sixth dimension of the *Content Creation Process* is **Form/Type** (see Table 6).

According to MoReqg2010, the **Entity Type** means the “[...] definition of an entity, including a list of its system metadata and the functions that can be performed on it.” The subunit of the dimension is Description, M14.4.16.

According to the PREMIS Dictionary

Table 6. **Dimension: Form/Type**

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Entity Type: definition of an entity, including a list of its system metadata and the functions that can be performed on it. A subunit of the dimension is Description, M14.4.16</p> <p><i>MoReq2010.</i></p>	<p>Implemented within Europeana in line with EDM edm:type, edm:hasType dcterms:hasFormat</p>	-----
<p>Type: a designation used to identify the Object uniquely within the preservation repository system in which it is stored. The semantic units are objectIdentifierType and objectIdentifierValue</p> <p><i>PREMIS Dictionary for Preservation Metadata, Version 3.0</i></p>		-----
<p>Form: metadata that document the rules of representation that determine the appearance of an entity and convey its meaning: whether the object is a report, a letter, a contract etc.; its file format, wrapper, encoding etc.; its draft or version number; and its technology (digital representation, i.e., file format, encoding, wrapper etc.)</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>		-----
<p>Type, Class E55: comprises concepts denoted by terms from thesauri and controlled vocabularies used to characterize and classify instances of CRM classes</p> <p><i>CIDOC CRM, Version 6.2.3, May 2018</i></p>	<p>EDM has been aligned to CIDOC CRM in its definition of an event-centric model</p> <p><i>Isaac, A. (2013). European Data Model Primer</i></p>	<p>Implemented in line with CIDOC CRM</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Type, Property edm:type: the Europeana material type of the resource in the ESE namespaces</p> <p><i>Definition of the Europeana Data Model v5.2.8, 2017</i></p>	<p>Implemented in line with EDM edm:type</p>	<p>Implemented according to ESE</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

for Preservation Metadata, **Type** is a designation used to identify the Object uniquely within the preservation repository system in which it is stored. The semantic units are `objectIdentifierType` and `objectIdentifierValue`.

According to the InterPARES, **Form** means “metadata that document the rules of representation that determine the appearance of an entity and convey its meaning” and it can be defined as a “documentary form – that is, whether the document is a report, a letter, a contract, etc.”

According to the CIDOC CRM, **Type** is Class E55, which comprises concepts denoted by terms from thesauri and control-

led vocabularies used to characterize and classify instances of CRM classes.

According to the EDM, **Type** is Property `edm:type`, which denotes the Europeana material type of the resource (also exists in the ESE namespace). All digital objects in the Europeana must be classified as one of the five Europeana types (in upper case): TEXT, IMAGE, SOUND, VIDEO or 3D.

The seventh dimension is **Identifier** (see Table 7).

The dimension **Identifier**, which, according to the MoReq2010, is Identifier (M14.4.100), which is the entity (or metadata of) Event: “[...] description of the outcome of a function that was performed

Table 7. **Dimension: Identifier**

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Identifier (M14.4.100): the entity (or metadata of) <i>Event</i> “[...] Description of the outcome of a function that was performed previously and is retained to show the history of an entity System”</p> <p><i>MoReq2010, p. 256</i></p>	<p>It is implemented within Europeana EDM as metadata element to identify other dimensions Used as <code>dc:identifier</code></p> <p><i>Europeana data model, mapping guidelines v.2.4, 2017</i></p>	-----
<p>ObjectIdentifier: semantic units for Intellectual Entities, Representations, Files and Bitstreams [...]. They are <code>objectIdentifierType</code> and <code>objectIdentifierValue</code></p> <p><i>PREMIS Dictionary for Preservation Metadata, Version 3.0</i></p>		-----
<p>Identifier: a unique identifier assigned from the system, needed to retrieve and present the record</p> <p><i>ISO15489</i></p>		
<p>Identifier, Class E42, comprises strings or codes assigned to instances of <i>E1 CRM Entity</i> in order to identify them uniquely and permanently within the context of one or more organizations</p> <p><i>CIDOC CRM Version 6.2.3, May 2018</i></p>	<p>EDM has been aligned to CIDOC CRM in its definition of an event-centric model</p> <p><i>Isaac, A. (2013). Europeana Data Model Primer</i></p>	<p>Implemented in line with CIDOC CRM</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

previously and is retained to show the history of an entity System.”

In the PREMIS Dictionary for Preservation Metadata (Version 3.0), *Identifiers* are specified as “semantic units for Intellectual Entities, Representations, Files and Bitstreams.” They are objectIdentifierType and objectIdentifierValue.

According to the CIDOC CRM, the dimension *Identifier* is Class E42 comprising strings or codes assigned to instances of E1 CRM Entity in order to identify them uniquely and permanently within the context of one or more organizations. Such codes are often known as inventory numbers, registration codes etc. and are typically composed of alphanumeric sequences. The class E42 Identifier is not normally used for machine-generated identifiers used for automated processing unless these are also used by human agents, e.g., ISSN, ISBN and other codes.

The dimension *Original Name/Title* (see Table 8).

According to the MoReq2010, *Title* is defined as “the identifying name or title of the entity.” The PREMIS definition of this dimension is originalName and means “the name of the object as submitted to

or harvested by the repository, before any renaming by the repository.” The CIDOC CRM includes the dimension *E35 Title* as a “name given to the resource.” Since this dimension is a key for identifying the digital object, it is usually one of the core metadata elements in the applied metadata schemas, as in the Europeana and VEPIS.

The ninth dimension is *Place/Room* (see Table 9).

For the identification of provenance information in digital preservation, *Place/Room* is important, as it complies with the CIDOC CRM, the Europeana data model and the VEPIS specification.

The tenth dimension is *Relation/Links* (see Table 10).

The dimension *Relation/Links*, according to the InterPARES, is defined as *Relation* and means “the elements, that illuminate the connection of the object to other objects to which it relates, and its context, whether it is preserved or destroyed.” This dimension, according to the PREMIS, is defined as *Relation* as well and denotes the “Representation Information Network (dependency, provenance and documentation links).” According to the CIDOC CRM_{dig}, this dimension is defined as Annotation

Table 8. *Dimension: Original Name/Title*

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
Title: the identifying name or title of the entity	Implemented within Europeana in line with EDM as dc:title.	Implemented as E35 Title Compliant with ESE <i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i>
<i>MoReq2010 M14.4.104</i>		
originalName: the name of the object as submitted to or harvested by the repository, before any renaming by the repository		
<i>PREMIS</i> Title , Class E35: name given to the resource		
<i>CIDOC CRM</i>		

Table 9. **Dimension: Place/Room**

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>D23 Room, a 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>	-----	<p>Implemented within VEPIS by using CRM_{dig}</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Place, Subclass edm:Place: extent in space, in particular on the surface of the earth, in the pure sense of physics: independent from temporal phenomena and matter (equivalent to E53_Place (CIDOC CRM))</p> <p><i>DM2E Model V 1.2 Specification Europeana Data Model – Mapping Guidelines v2.4 06/10/2017</i></p>	<p>Implemented in line with EDM as the place of a physical publication or edm:WebResource of a web publication</p> <p><i>EDM documentation</i></p>	<p>Compliant with ESE</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

Table 10. **Dimension: Relation/ Links to the objects**

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Relation: elements that illuminate the connection of the object to other objects to which it relates, and its context, whether it is preserved or destroyed</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>	<p>Implemented within Europeana. In line with EDM Metadata elements: edm:isRelatedTo edm:isSimilarTo edm:osDerivativeOf</p>	-----
<p>Relation: Representation Information Network (dependency, provenance and documentation links)</p> <p><i>PREMIS</i></p>	<p><i>Europeana Data Model – Mapping Guidelines v2.4 06/10/2017</i></p>	-----
<p>Links to related objects document the same sequence of business activity or relating to the same person or case, if the object is part of a case file; details of embedded document links, including applications software and version under which the linked record was created</p> <p><i>ISO 15489</i></p>		-----

<i>Substantiating statements</i>	<i>Substantiating statements for Europeana</i>	<i>Substantiating statements for VEPIS</i>
<p>Annotation Object, Class D29, L43 annotates (is annotated by), L48 created annotation (was annotation create by)</p> <p><i>CIDOC CRM_{dig}</i></p>		<p>Implemented in line with CRM_{dig}</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>Relation, Subproperty edm:hasMet relates a resource with the objects or phenomena that have happened to or have happened together with the resource under consideration</p> <p><i>Europeana Data Model – Mapping Guidelines v2.4 06/10/2017</i></p>		<p>Compliant with ESE</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

Object, Class D29, L43 annotates (is annotated by), L48 created annotation (the annotation was created by).

According to the EDM, **Relation** is Subproperty edm:hasMet and “relates a resource with the objects or phenomena that have happened to or have happened together with the resource under consideration. The relation [...] allow linking between the associated classes and also has some more familiar descriptive properties, such as edm:dataProvider, that apply to the whole group. With the ability to express such relationships, providers should always try to ‘distribute’ their original descriptions onto objects that precisely match their holdings, i.e. choose the most appropriate level of granularity for the CHO” (Europeana, the Europeana data model – mapping, 2017).

The eleventh dimension is **Rights/Access Rights/Mandate** (see Table 11).

According to the PREMIS Data Dictionary for Preservation Metadata, (version 3.0) “[a]ccess is [...] assertion of one or more Rights or permissions pertaining to an Object and/or Agent.” The PREMIS Data Dictionary for Preservation Metadata

defines semantic units. Each semantic unit defined in the Data Dictionary is mapped to an entity that is organized within a simple data model. A semantic unit can therefore be understood as the property of an entity.

The model defines four entities important in regard to digital preservation activities: *Objects, Events, Agents* and *Rights* as it is presented in Fig. 4

According to the International Research on Permanent Authentic Records in Electronic Systems (InterPARES), the dimension **Access Rights** is treated as “[...] *Rights and access* metadata that identify any restrictions or privileges that apply to the object(s) (rights, access restriction code; access privileges code); indication of copyright or other intellectual rights” (InterPARES 3 Project 2009).

According to MoReq2010, the dimension **Access rights** is treated as a “[...] textual reference to a legal or other instrument that provides the authority for a disposal schedule or a disposal hold.”

The OAIS model treats this dimension for authenticity and provenance as *Access Rights Information*, which means “[...]”

Table 11. **Dimension: Rights/ Access Rights/ Rights Statement**

Substantiating statements	Substantiating statements for Europeana	Substantiating statements for VEPIS
<p>Rights Statement: an assertion of one or more Rights or permissions pertaining to an Object and/or Agent</p> <p><i>PREMIS Data Dictionary for Preservation Metadata version 3.0, 2015</i></p>	<p>Implemented within Europeana in line with the EDM model as the metadata element edm:rights</p>	<p>-----</p>
<p>Rights and access: metadata that identify any restrictions or privileges that apply to the object(s) (rights, access restriction code; access privileges code); indication of copyright or other intellectual rights</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>		<p>-----</p>
<p>Mandate: textual reference as a legal or other instrument that provides the authority for a disposal schedule or a disposal hold</p> <p><i>MoReq2010 M14.4.51 Mandate</i></p>		<p>-----</p>
<p>Access Rights Information: information that identifies the access restrictions pertaining to <i>Content Information</i>, including the legal framework, licensing terms, and access control</p> <p><i>Reference Model for an Open Archival Information System (OAIS) Recommended Practice CCSDS 650.0-M-2, 2012</i></p>	<p>Implemented in line with the OAIS reference model and ISO standard</p> <p><i>DM2E Model V 1.2 Specification Revision: Version 1.2</i></p>	<p>Implemented in line with the OAIS reference model and the ISO standard</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>
<p>E30 Right: comprises legal privileges concerning material and immaterial things or their derivatives</p> <p><i>CIDOC CRM Version 6.2.3 May 2018</i></p>	<p>EDM has been aligned to CIDOC-CRM in its definition of an event-centric model</p> <p><i>Isaac, A. (2013). Europeana Data Model Primer</i></p>	<p>Implemented as E30 Right of in line with CIDOC CRM</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10 Nr. V2.0. VEPIS specification</i></p>
<p>Rights, Class edm:WebResource: information about rights held in and over the resource</p> <p><i>Definition of the Europeana Data Model v5.2.8, 2017</i></p>	<p>Implemented in line with EDM as the URL of a resource describing licensing rights of the CHO.</p>	<p>Compliant with ESE</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification)</i></p>

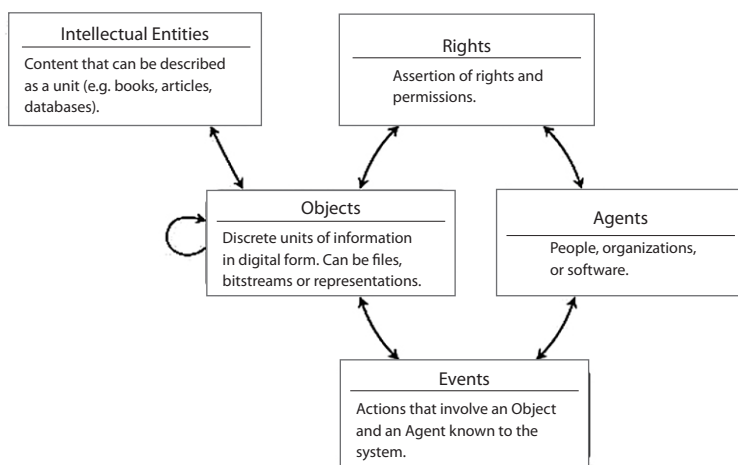


Fig. 4. A graphical illustration of the PREMIS Data Model (Source: Caplan 2009).

the information that identifies the access restrictions pertaining to the Content Information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related to both preservation (by the OAIS) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures” (CCSDS 2012).

According to the CIDOC CRM, the dimension *Access rights* is treated as a Class E30 Right. This class comprises legal privileges concerning material and immaterial things or their derivatives.

According to the EDM, the *Access rights* dimension is a metadata element `edm:rights`. The new developments around the Content Re-Use Framework provide new requirements for representing the rights statements for digital representations of cultural heritage objects (Europeana, Extending the Europeana, 2013).

This extension of EDM focuses on the creation of “complex” values for the existing property `edm:rights`. The indication

of an identifier of a rights statement (e.g., CC-BY) in a unique field (`edm:rights` on the EDM ore:Aggregation resource) no longer covers more complex requirements. This profile envisions that different access and re-use conditions can be provided for different views of a cultural object. It therefore allows the representation of individual views with specific rights statements.

The twelfth dimension is suggested as **Technology/Application** (see Table 12).

According to ISO 15489, this dimension is treated as “[...] application software and version under which the record was created or in which it was captured; business system from which the object was captured; standard with which the objects structure complies.”

According to the PREMIS, this dimension is treated as “information about the application that created the object (Application name, version, `dateCreatedByApplication`, `creatingApplicationExtension`).”

According to the InterPARES, “*technology* is metadata that identify the carrier(s) of the form and content of the record(s).”

Table 12. **Dimension: Technology/Application**

<i>Substantiating statements</i>	<i>Substantiating statements for EUROPEANA</i>	<i>Substantiating statements for VEPIS</i>
<p>Application software and version under which the record was created or in which it was captured; business system from which the object was captured; standard with which the objects structure complies</p> <p><i>ISO 15489</i></p>	-----	<p>The Administration DB: the name of organization carrying out the digitization; devices for digitization; archiving formats; metadata formats and standards, access formats etc</p>
<p>Information about the application that created the object (Application Name, Version), dateCreatedByApplication, creatingApplicationExtension)</p> <p><i>PREMIS Data Dictionary for Preservation Metadata version 3.0, 2015 -</i></p>	-----	-----
<p>Technology: metadata that identify the carrier(s) of the form and content of the record(s)</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>		
<p>Design or Procedure, Class E29 > P32 used general technique (was technique of) CIDOC CRMdig D14 Software CIDOC CRMdig D8 Digital device</p> <p><i>CIDOC CRM Version 6.2.3 May 2018</i></p>		<p>Implemented as <i>Digital Device for Digitization Processes</i></p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification</i></p>
<p>PROV-O: <i>prov:SoftwareAgent</i></p>	-----	-----

According to CIDOC CRM and CRM-dig, the dimension **Application/Technology** is a D14 Software Subclass of D1 Digital Object, which comprises software codes, computer programs, procedures and functions that are used to operate a system of digital objects. Within VEPIS, it is implemented as the Digital Device for Digitization Processes.

According to the PROV-O: The PROV Ontology, the dimension **Technology/Ap-**

plication could be presented as *PROV-O prov:SoftwareAgent*.

The last dimension identified in this research is **Signature/Authentication** (see Table 13). According to InterPARES, **Authentication** (including the digital signature, attestation etc.) is treated as an element that indicates the identity of the persons involved in the creation of the object.

According to the PREMIS, the dimension **Signature information** is treated as

Table 13. *Dimension: Signature/Authentication*

<i>Substantiating statements</i>	<i>Substantiating statements for the Europeana</i>	<i>Substantiating statements for the VEPIS</i>
<p>Authentication (including digital signature, attestation etc.): elements that indicate the identity of the persons involved in the creation of an object. The indication of the presence of a digital signature; corroboration – that is, an explicit mention of the means used to; attestation – that is, the validation of the object by those who took part in the issuing of it, and by witnesses to the action or to the “signing” of the object; subscription – that is, the name of the author or writer appearing at the bottom of the object/document; qualification of signature – that is, the mention of the title, capacity and/or address of the person or persons signing the object/document</p> <p><i>InterPARES 2 Project Dictionary, 2018</i></p>	<p>-----</p>	<p>Authentication information via LIBIS⁹ in accordance with the need</p> <p><i>VEPIS Techninis aprašymas. 2014-07-10, Nr. V2.0 (VEPIS Specification, 2014</i></p>
<p>Signature information: a container for PREMIS-defined and externally defined digital signature information used to authenticate the signer of an object and/or the information contained in the object</p> <p><i>PREMIS Data Dictionary</i></p>	<p>-----</p>	<p>-----</p>

“[...] a container for PREMIS-defined and externally-defined digital signature information, used to authenticate the signer of an object and/or the information contained in the object.”

Provenance information may be used for authentication purposes: for example, the creator of a digital object can be verified by representation information (provenance information) as it is realized in the OAIS model-based systems. Since the VEPIS and Europeana are OAIS-based systems, the authentication of digital objects is based on the OAIS-described processes.

Results from Investigating the Content Creation Process

The analysis of efforts of different expert groups and international projects specifying metadata dictionaries, subsets of the categories of authenticity and provenance and ontologies allowed us to conceptualize the *Content Creation Process*, which refers to the structure and meaning of authenticity and provenance records. We defined the scope of the content of authenticity and provenance, identifying the dimensions needed to support the core preservation functions and ensure interoperability – the ability to exchange data amongst institutions and assess the added value of the Europeana and VEPIS regarding the authenticity

⁹ LIBIS – the Lithuanian Integrated Library Information System.

and provenance based on the adequacy of analyzed metadata dictionaries, specifications or models. Our qualitative analysis showed that:

1. It is not possible to define one set of metadata that applies equally to all content types or organization types and covers the whole scope of the dimensions of authenticity and provenance needed for the long-term preservation of information. For example, the PREMIS focuses on encoding the preservation actions taking place before and during the ingestion of a digital object into an archival repository, while others, such as the PROV-O and OPM (for provenance), are focused on encoding the provenance history. The OAIS provides the outline that must be followed when developing a long-term digital archival repository as well as guidelines on what kind of semantic information is needed for long-term preservation. This, in turn, encourages the search for the opportunities of sharable preservation metadata specifications, which would ensure interoperability and the understanding of the digital object's metadata and its digital content.
2. When combining different metadata specifications or when embedding extension metadata, we often find that data models are mismatched or that semantic units overlap, e.g., *Process* within both the OPM and the requirements by the W3C Provenance Working Group, but their interpretation differs. Some definitions are semantically very close for both authenticity and provenance in those models, e. g., the *Agent* is understood as someone who is responsible for the events. On the other hand, some dimensions in the same specification

are treated differently according to the sector, e.g. *Agent* is differently treated in archives, sciences and art. Within various models, some of dimensions, for example, the *Process* within both the OPM and the Requirements for Provenance on the Web and the *Actor*, which are included in the three analyzed specifications and models (EDM, OPM, PROV-Ontology), bear the same names yet are to be interpreted differently. It is evident that the abundance of models and the terminology of authenticity and provenance is a major obstacle for the interoperability of systems and the common verification of authenticity and provenance. In order to support core preservation functions and ensure the ability to exchange data amongst the institutions of different sectors, most institutions take advantage of the sharable preservation of metadata specifications.

- 2.1. The granularity of the contemporary models varies. The ontology assumed by the OPM is minimal – it comprises only three classes and five associations between them. The CIDOC CRM ontology comprises 80 classes and 132 relations and possesses a rich structure of “intermediate” classes and relations, enabling queries at various levels of abstraction and granularity. For example, given the fact that the VEPIS is based on the CIDOC CRM, provenance information recorded according to the CRM_{dig} can be mapped onto an OPM-based view, but not the other way around.
- 2.2. The EDM provide three core classes, but according to EDM documentation, there is a mechanism to associate the related classes.

It has properties to allow linking between the associated classes and also has some more familiar descriptive properties, such as `edm:dataProvider`, that apply to the whole group. With the ability to express such relationships, providers should always try to “distribute” their original descriptions onto objects that precisely match their holdings, i.e., to choose the most appropriate level of granularity for the CHO.

3. The investigated systems (the Europeana and VEPIS) comply with the main requirements for long-term preservation systems, the authenticity of digital objects provided to the users and enable interoperability – the possibility to exchange data between systems, which is currently carried out in the following ways:

3.1. The Identity of the object in the Open Archival Information System (OAIS) Reference Model-based systems is strongly related to the PDI (Context, Provenance, Fixity, and Reference Information as defined in the 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 and metadata. As the VEPIS and Europeana are based on the

OAIS, it guarantees the authenticity of the object and metadata.

3.1. As the VEPIS is based on a CIDOC CRM_{dig} event-centric model, and the Europeana is based on an EDM event-centric model, it proves the implementation of the dimension *Action/Event/Process/Aggregation/Entailment/Justification, Versioning*, which creates functionality, thus providing metadata and context for the digitization process.

3.2. Provenance information within the VEPIS allows for referring to the versions of objects as they evolve, are modified or accessed over the time. In particular, it provides 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 and the CIDOC CRM dig models.

3.3. The Europeana, as an EDM-based system, has a mechanism for associating the related classes – `ore:Aggregation` class – and it is the pivotal object between the `edm:ProvidedCHO` and the `edm:WebResource(s)`, which provides the derivation chain and refers to its origin or source.

This, in turn, allows us to conclude that the Europeana and VEPIS support the function of *Identifying the Scope of Authenticity and Provenance Content* in the 2–3–6 of the model of added value chain through the dimensions regarding authenticity and provenance.

REFERENCES

- APARSEN (2012). D24.1 Report on Authenticity and Plan for Interoperable Authenticity Evaluation System. http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D24_1-01-2_5_incURN.pdf
- BUNEMAN, Peter, SANJEV Khanna and WANG-CHIEV Tan. *Why and Were: A Characterization of Data Provenance*. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1209&context=cis_papers.
- CAPLAN, P. (2009). Understanding PREMIS. <https://www.loc.gov/standards/premis/understanding-premis.pdf>
- CCSDS – Consultative Committee for Space Data Systems. (2012). Reference Model for an Open Archival Information System (OAIS). <https://web.archive.org/web/20131020200910/http://public.ccsds.org/publications/archive/650x0m2.pdf>
- CHARLES, V., Isaac, A. (2015). Enhancing the Europeana Data Model (EDM). https://pro.europeana.eu/files/Europeana_Professional/Publications/EDM_WhitePaper_17062015.pdf.
- 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, A., ENDERS, M. (2010). Digital Preservation Metadata Standards, In *ISO: Information Standards Quarterly*, 22:2 (2010): 5-13, http://www.niso.org/apps/group_public/download.php/4299/isqv22no2.pdf.
- DLM Forum Foundation (2011). Modular Requirements for Records Systems (MoReq2010). https://www.moreq.info/files/moreq2010_vol1_v1_1_en.pdf
- DOERR, M. (2009). An Ontological Approach to Digital Preservation Metadata. ICS-FORTH May 23, 2009. http://www.cidoc-crm.org/sites/default/files/carpar_for_Prague.ppt
- DOERR, M., Theodoridou, M. (2011). CRMdig: A Generic Digital Provenance Model for Scientific Observation, in 3rd USENIX Workshop on the Theory and Practice of Provenance (TaPP '11).
- DROGE, E., et al. (2015). DM2E, ICP-PSP-297274, Digitised Manuscripts to Europeana, Technical specification. https://dm2e.eu/files/DM2E_Model_V1.2.pdf
- European Commission, DG XIII, Andersen Consulting, (1996). Strategic Developments for the European Publishing Industry towards the Year 2000. <http://www.echo.lu/elpub2/en/exengl.pdf>
- Europeana (2013). Europeana Data Model Primer. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
- Europeana (2017). Definition of the Europeana Data Model v5.2.8 06/10/2017. At: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.8_102017.pdf
- Europeana (2017). Europeana Data Model – Mapping Guidelines v2.4. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf
- Europeana (2013). Extending the Europeana Licensing Framework. <https://pro.europeana.eu/post/extending-the-europeana-licensing-framework>
- 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. http://usenix.org/event/tapp09/tech/full_papers/factor/factor_html/
- GAVRILIS, D. DALLAS, C., ANGELIS, S. (2013). A Curation-Oriented Thematic Aggregator. In: Aalberg T., Papatheodorou C., Dobрева M., Tsakonas G., Farrugia C.J. (eds) Research and Advanced Technology for Digital Libraries. TPDL 2013. Lecture Notes in Computer Science, vol 8092, https://doi.org/10.1007/978-3-642-40501-3_13
- GIARETTA, D. (2011). Advanced Digital Preservation. Springer Heidelberg Dordrecht London New York e-ISBN 978-3-642-16809-3. DOI 10.1007/978-3-642-16809-3
- GROTH, P. et al. (2012). Requirements for Provenance on the Web. In International Journal of Digital Curation, Vol. 7, No. 1, <https://www.isi.edu/~gil/papers/groth-et-al-ijdc12.pdf>
- GUERCIO, M., SALZA, (2012). Managing Authenticity through the Digital Resource Lifecycle. In Italian Research Conference on Digital Libraries IRCDL 2012: Digital Libraries and Archives pp. 249–260, 2013. https://link.springer.com/chapter/10.1007/978-3-642-35834-0_25
- InterPARES 3 Project (2009). InterPARES 2 Project Chain of Preservation (COP) Model Metadata. http://www.interpares.org/ip2/display_file.cfm?doc=ip2_cop-model_metadata_v1.0.pdf
- InterPARES 3 Project, TEAM Canada (2016). General Study 15 – Application Profile For Authenticity Metadata: General study Report. V.2.3. http://www.interpares.org/ip3/display_file.cfm?doc=ip3_canada_gs15_final_report.pdf

International Organization for Standardization (2016). ISO 15489-1:2016. Information and documentation – Records management. <https://www.iso.org/standard/62542.html>

LAMB, D., PRANDONI, C., DACIDSON, J., DCC (2009). <http://www.dcc.ac.uk/resources/briefing-papers/technology-watch-papers/caspar>

MOREAU, L. et al. (2011). The Open Provenance Model Core Specification (v1.1). In *Future Generation Computer Systems*, Vol. 27 (2), Issue 6. P. 743–756. <https://doi.org/10.1016/j.future.2010.07.005>

National Archives of Australia (2008). Australian Government Recordkeeping Metadata Standard version 2.0. <http://www.naa.gov.au/information-management/managing-information-and-records/describing/metadata/AGRKMS/index.aspx>

PREMIS Editorial Committee (2015). PREMIS Data Dictionary for Preservation Metadata. Version 3.0. <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

ROGERS, C. A literature review of authenticity of records in digital systems from ‘machine-readable’

to records in the cloud (2016). <http://oaji.net/articles/2017/3932-1484337719.pdf>

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. <http://dcpapers.dublincore.org/pubs/article/view/3662>

The InterPARES 2 Project Dictionary (2018). http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf&CFID=15051643&CFTOKEN=77329071

VEPIS Techninis aprašymas (Specifikacija) (2014). Nr.V2.0.

W3C (2013). Provenance Working Group, https://www.w3.org/2011/prov/wiki/Main_Page

W3C (2013). PROV-O: The PROV Ontology, W3C Recommendation 30 April 2013. <https://www.w3.org/TR/2013/REC-prov-o-20130430/#description-starting-point-terms>

W3C Provenance Incubator Group (2015). Overview of Provenance on the Web. <https://www.w3.org/2005/Incubator/prov/wiki/images/0/02/Provenance-XG-Overview.pdf>

AUTENTIŠKUMAS IR PROVENIENCIJA ILGALAIKIO SKAITMENINIO IŠSAUGOJIMO KONTEKSTE: TURINIO APRĖPTIES ANALIZĖ

Regina Varnienė-Janssen, Jūratė Kuprienė

S a n t r a u k a

Autentiškumas yra pamatinis siekis skaitmeninių išteklių ilgalaikio išsaugojimo srityje; kita vertus, autentiškumo tikrinimas – kompleksinė užduotis, kuriai įgyvendinti reikalinga proveniencija, kaip patikimumo prielaida. Atsižvelgiant į tai, labai svarbu apibrėžti pagrindinius šio proceso elementus, sudarančius šios sudėtingos koncepcijos pagrindą. *Pagrindinis straipsnio tikslas* – nustatyti autentiškumo valdymo procesą skaitmeninio išsaugojimo kontekste: identifikuoti semantinius elementus, užtikrinančius ilgalaikio išsaugojimo funkcijas ir skaitmeninių objektų patikimumą bei funkcinių sistemų suderinamumą, t. y. galimybę institucijoms keistis duomenimis. Straipsnyje pateikiami pirmojo tyrimo etapo rezultatai: autentiškumo ir proveniencijos turinio aprėptis, užtikrinanti pagrindines ilgalaikio išsaugojimo funkcijas ir funkcinių sistemų suderinamumą. Straipsnyje, taikant Europos Sąjungos 2–3–6 elektroninės leidybos priedamosios vertės kūrimo

konceptinį požiūrį, analizuojami autentiškumo ir proveniencijos modeliai ir standartai, taip pat vertinama priedamoji EUROPEANOS ir VEPIS vertė autentiškumo ir proveniencijos požiūriu, kuri yra grindžiama atitiktimi analizuotiems standartams ir modeliams, naudojamiems šiose sistemose.

Metodologija. Taikant Europos Sąjungos 2–3–6 elektroninės leidybos priedamosios vertės koncepciją ir remiantis kokybine mokslinės literatūros analize, tyrime konceptualizuojamas *turinio kūrimo* procesas ilgalaikio išsaugojimo informacinėse sistemose, susijęs su autentiškumo ir proveniencijos struktūros ir prasmės nustatymu. Taikant šį metodą, buvo nustatyta autentiškumo ir proveniencijos turinio aprėptis.

Pagrindiniai žodžiai: autentiškumas, proveniencija, patikimumas, skaitmeninė aplinka, duomenų kokybė, skaitmeninių objektų patikimumas, turinio aprėptis, Virtualaus elektroninio paveldo informacinė sistema (VEPIS), Europėana.

Įteikta 2018 m. rugsėjo 10 d.