



**Vilnius
universitetas**

CIDOC CRM Based Provenance Metadata within VEPIS: for Interoperability of Cultural and Research Data

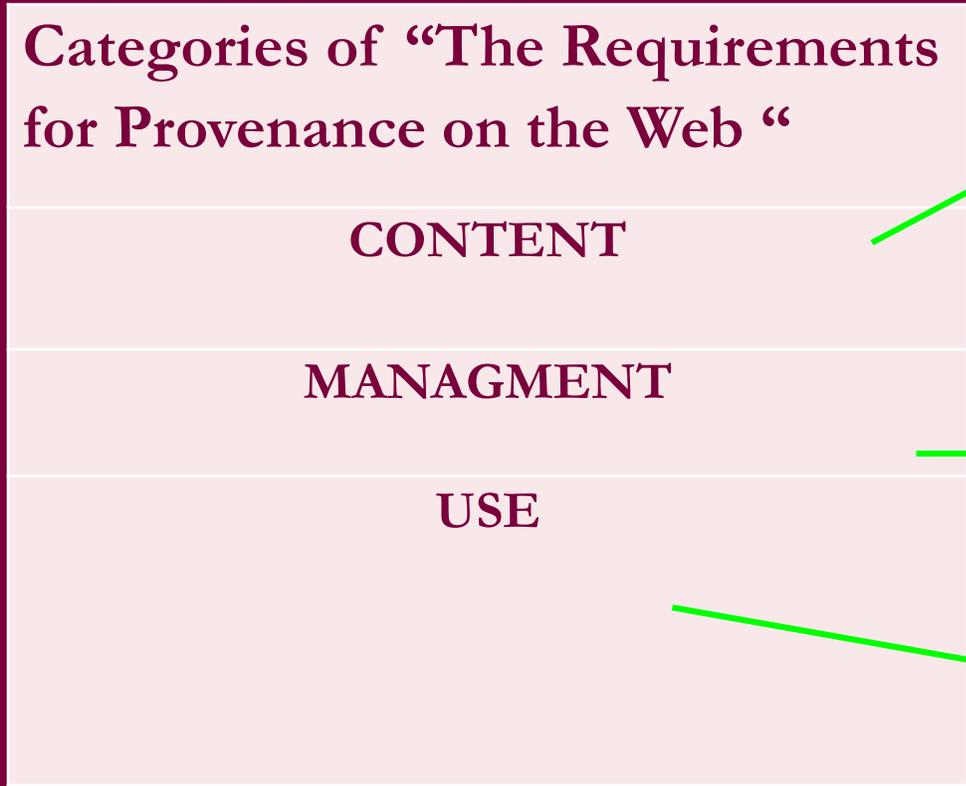
**Assoc. Prof Regina Varnienė-Janssen
Affiliate Prof. Albertas Šermokas**

RODBH 2019, Leipzig

Objective of the Research

The report analyses provenance based on the CIDOC CRM Reference Model and CRMdig as a fundamental concept for ensuring data quality and integration of cultural and research data.

In 2017-2018, for analysing the compliance of provenance metadata with the W3C Incubator Group's Requirements for Provenance on the Web, (https://www.w3.org/2005/Incubator/prov/wiki/User_Requirements) the researchers at the Vilnius University's Faculties of Communication and Mathematics and Informatics choose biggest Lithuania's system - Virtual Information System of Electronic Heritage (hereinafter "VEPIS", *Virtuali elektroninio paveldo informacinė sistema*).



What is the meaningful scope of the provenance content and what types of information would should be represented in a provenance record

How this provenance information has to be managed in order to make provenance available and accessible within a system and

How provenance records accommodate a variety of uses and diverse users/consumers

Fig. 1. Methodology of The Research



According to the Requirements for Provenance on the Web „provenance is a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing.

In particular provenance is crucial in deciding whether information is to be trusted, how it should be integrated with other diverse information sources, and how to give credit to its originators when reusing it.

In an open and inclusive environment such as the Web, where users find information that is often contradictory or questionable, provenance can help those users to make trust judgements.“

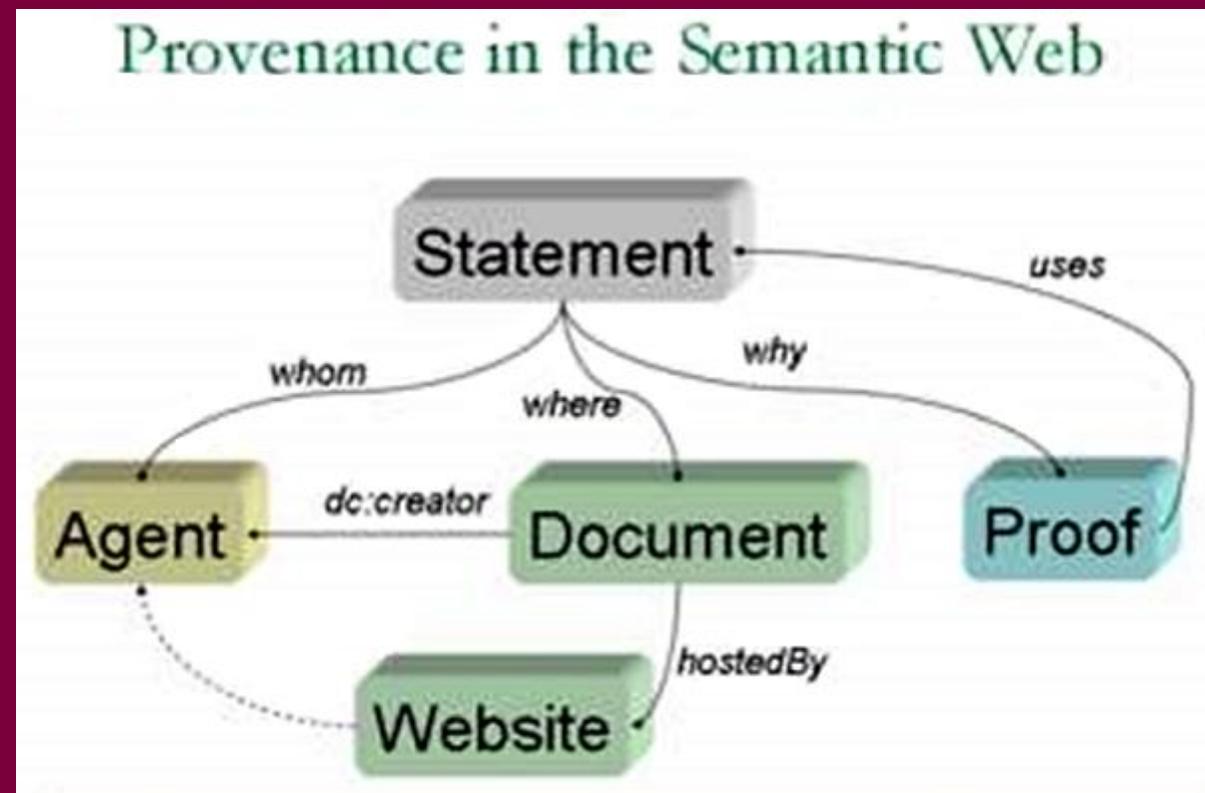
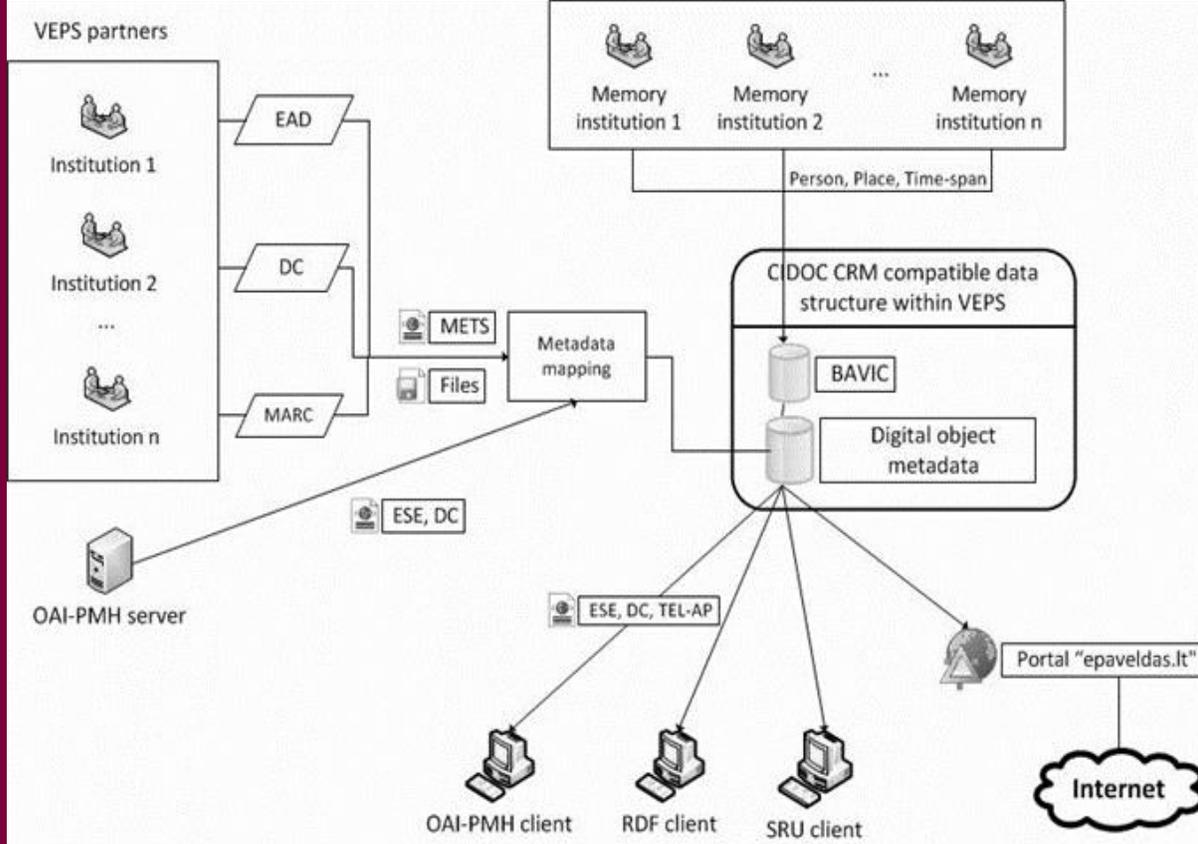


Fig. 2. Recommendation REC-prov-dm-20130430, W3C

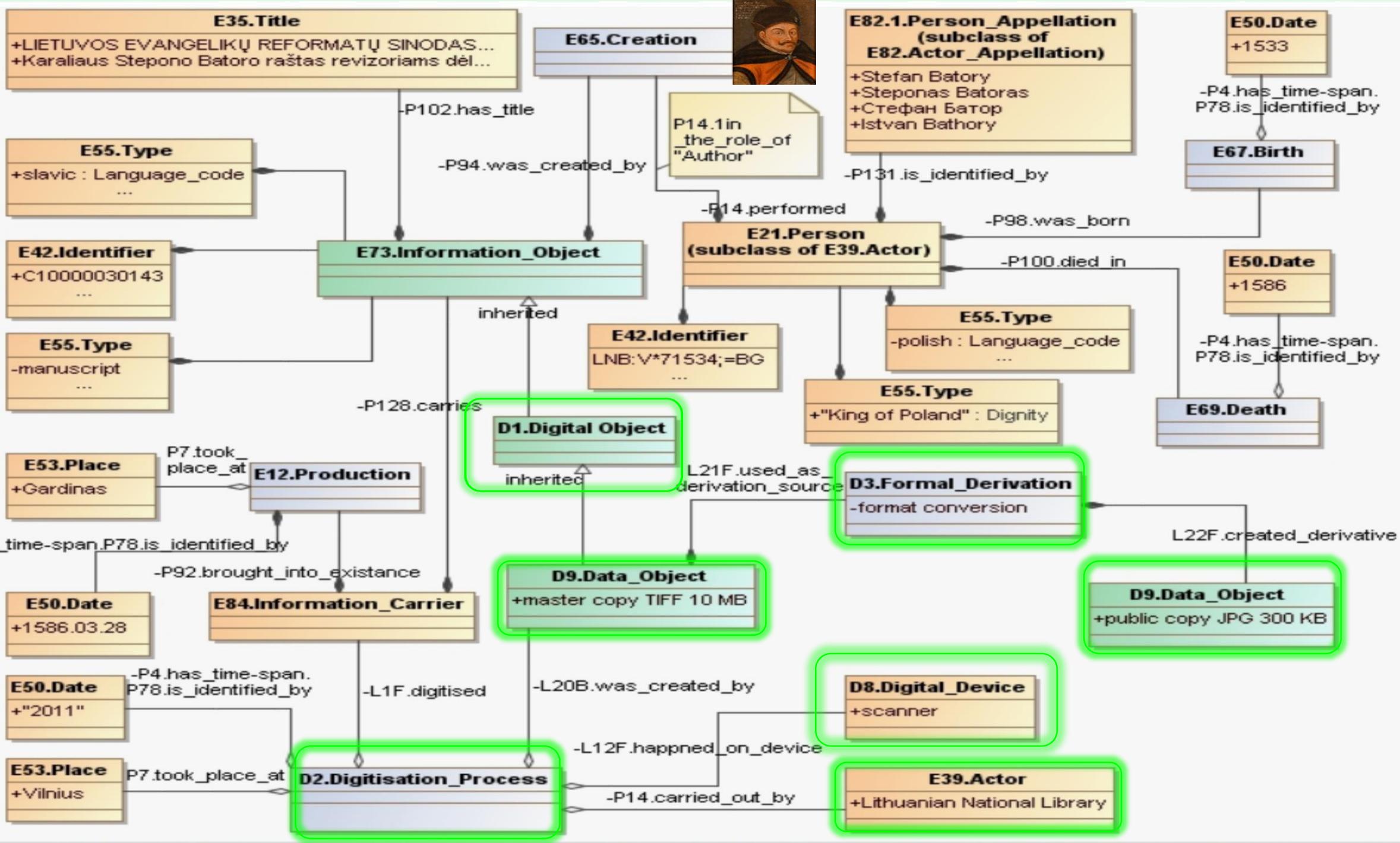


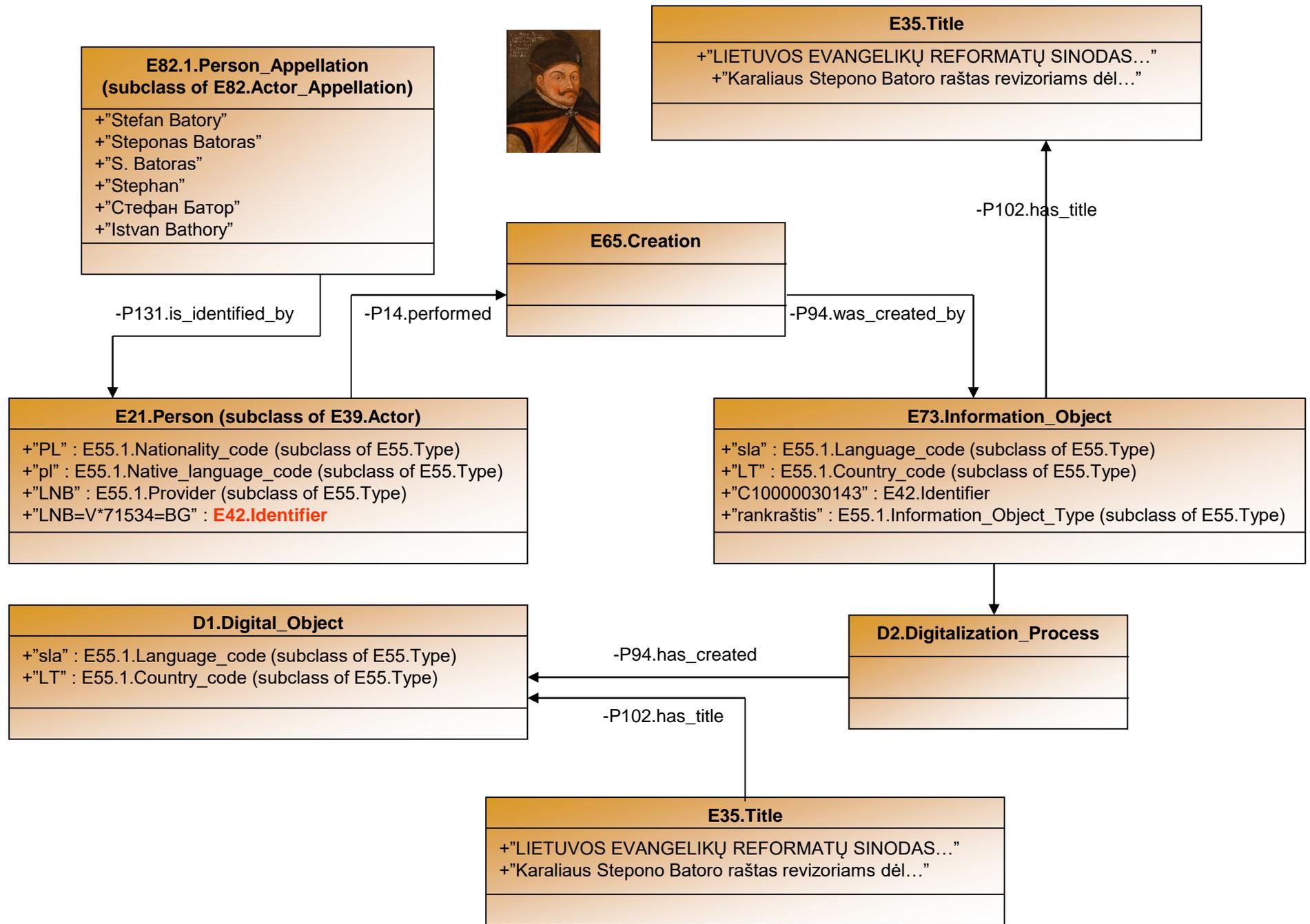
If the metadata are consistent and the relationships are well described, chains or data paths are created that correlate, **PEOPLE, EVENTS, OBJECTS, PLACES, TIMES** which are related with the object.

Such consistency is ensured by the CIDOC CRM ontology, which has been realized with the help of the Thesaurus of Personal Names, geographical Names and Historical Chronology (BAVIC) serving as a formal language. It:

- ✓ Identifies the general information content within data of diverse formats by linking them with CIDOC CRM, thus ensuring their integration;
- ✓ Retains the content and meaning of documents during the automatic transferring of data from local to the VEPIS data structures by applying an algorithm;
- ✓ Establishes a semantic quering of digital objects using the knowledge base of Thesaurus BAVIC and Semantic technologies;
- ✓ Ensures data exchange and their delivery export from VEPIS, e.g Europeana;

Fig. 3 . Metadata flows within VEPIS





| Dimensions | Description | Substantiating statements for VEPIS |
|----------------------|---|---|
| Object | The artefact, which is described by provenance data (identified by URI) | CIDOC CRM: E73 Information Object ; CRMdig: D1 Digital Object (subclass of E73) |
| Attribution | Sources and entities which were involved in creating the artefact | CIDOC CRM: E21 Person, E74 Group ; CRMdig: D21 Person Name |
| Process | Actions (or stages) by which the artefact was created | CIDOC CRM: E65 Creation (subclass of E7 Activity); E66 Formation (subclass of E7 Activity); CRMdig: D2 Digitization Process, D7 Digital Machine Event |
| Versioning | Recording artefacts and changes of their relationships and related entities and processes | CIDOC CRM: E1 CRM Entity ; CRMdig: D7 Digital Machine Event (subclass of D11 Digital Measurement Event), data from UNIMARC 300 field (Notes) |
| Justification | Documentation recording reasons and process of taking solutions | CIDOC CRM: E84 Information Carrier ; CRM dig: D11 Digital Measurement Event (subclass of D7 Digital Machine Event) and E16 Measurement (superclass of D2 Digitization Process) |
| Entailment | Explanations how certain facts have been derived from other facts | CRM dig: D12 Data Transfer Event (subclass of D7 Digital Machine Event) |



Some Remarks Concerning the Provenance Metadata Consistency

Vilniaus
universitetas

| UNIMAR field | | CIDOC CRM class domain | CRM property | CRM range class |
|--------------|--|---|--|---|
| 325 | Reproduction note | | | |
| 325 \$a | Text of unstructured note | E73 Information Object (instance = the publication exemplified by the item being described); D1 Digital Object (subclass of E73.) | P3 has note | E62 String (value = "content of 320\$a") |
| 325 \$b | Type of reproduction | D1 Digital object, subclass of E73. | P3.1 has type | E62 String (value = "content of reproduction note"); 325 \$b "note of the type" |
| 325 \$d | Agency responsible for the reproduction | E7 Activity | L 29 has responsible organization (is responsible organization for) | E40 Legal Body (value = 325 \$d "The name of the agency that makes the reproduction available") |
| 325 \$e | Date of publication for the reproduction | D7 Digital Machine Event; D2 Digitization Process | L31 has starting date-time (was starting date-time of) L32 has ending date-time (was ending date-time of) | D7 Digital Machine Event (value= content 325 \$e Date of Publication for the Reproduction |



Realization of the Management Category Within VEPIS

Vilniaus
universitetas

| Dimensions | Description | Substantiating statements for VEPIS |
|----------------------|--|--|
| Publication | Making provenance available on the Web | Publication within VEPIS is realized by the component <i>Publication and Access</i> . The portal's interface has all the accessibility features according to the recommendations of the European Union's WAI (Web Accessibility Initiative) and is intuitive, understandable and easy for users. |
| Access | The ability to find the provenance for a particular artifact | Access is realized via portal http://www.epaveldas.lt and automatic data import via the OAI-PMH protocol. User interface is realized on the basis of Web-based principles via a standard Web. The search of the provenance information is based on CRMdig |
| Dissemination | Defining how provenance should be distributed and controlled | Dissemination : BAVIC Thesaurus and metadata of digital objects are based on the the CIDOC CRM and CRMdig and are in the RDF form in line with the XML schema, thus ensuring provenance-related query services:providing data about the creator of the object; the earlier versions of the item; the events that changed the custody of the item; input that influenced the result; the master version of the object; and the scanner/resolution of the digital object. (See Fig 6). |
| Scale | Dealing with large amounts of provenance | Scale within VEPIS has been only partially realized: BAVIC ensures formulation of queries and organizing search results and permits obtaining information about the object from all the VEPIS partners independent of media types within VEPIS; however, it does not guarantee access to information about investigation of the object that have been carried out or their results across many published repositories. |



Realization of the Use Category Within VEPIS

Vilniaus
universitetas

| Dimensions | Description | Substantiating statements for VEPIS |
|-------------------------|---|--|
| Understanding | How to enable the end user consumption of provenance. | Realized within VEPIS by the component <i>Publication and Access</i> . The portal's interface has all the accessibility features according to the recommendations of the European Union's WAI (Web Accessibility Initiative) and is intuitive, understandable and easy for users |
| Interoperability | Combining provenance produced by multiple systems | We could refer to interoperability only in the sense that VEPIS aggregates data from diverse systems and all descriptive information is converted into UNIMARC including provenance data (however, it is not interoperable as regards search). |
| Comparison | Comparing artefacts by their provenance | Not implemented within VEPIS |
| Accountability | Using provenance to assign credit or blame | Not implemented within VEPIS |
| Trust | Using provenance to make trust | Specific components: <i>Component of Metadata Verification</i> , which ensures control of metadata and <i>Component of Logging Events</i> , which tracks the import of digitized objects |
| Imperfections | Dealing with imperfections in provenance records | Specific components: <i>Component of Metadata Verification</i> , which ensures control of metadata and <i>Component of Logging Events</i> , which tracks the import of digitized objects |
| Debugging | Using provenance to detect failures or bugs | Specific components implemented within VEPIS : <i>Component of Metadata Verification, Component of Logging Events</i> |

Summarizing Conclusions

Qualitative analysis of research literature as well as the specification of VEPIS and its services allowed us to conclude that VEPIS, which is based on CIDOC CRM and CRM dig and RDF, meets the W3C Incubator Group's Requirements for Provenance on the Web and 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;

LIETUVOS EVANGELIKŲ REFORMATŲ SINODAS. XI : Lietuvos ir Baltarusijos evangelikų reformatų vietovės. 9 : Dainiai. 240 : [Karaliaus Stepono Batoro raštas revizoriams dėl Dainių kaimo (Raseinių pav.) žemės apmatavimo ir perdavimo valdyti Daratai Šemetienei (Szemetowa) pagal apsikeitimo teisę]. - 1586.III.28

peržiūrėti skaitmeninto objekto vaizdus

| | |
|-----------------------------------|--|
| Tipas | tekstas |
| Kiti asmenys | Stefan Batory |
| Išleidimo duomenys | s.n. [1586] |
| Apimtis | 2 lap., 1 dok.; Archyvinė kopija: 42MB Viešos prieigos kopija: 1MB |
| Daugiau informacijos apie objektą | Lietuvos nacionalinė Martyno Mažvydo biblioteka |
| Rašto sistema | lotynų |
| Kalbos | rusų, senoji (slavų) |
| Pastabos | Rankraštinis dokumento originalas, su antsp. Defekt. |
| | Skaitmenintas objektas: Vilnius; Lietuvos nacionalinė Martyno Mažvydo biblioteka; 2019 |
| | 03.19 |
| | Archyvinė kopija: kreipties būdas: Lietuvos nacionalinės bibliotekos kompiuterinis tinklas |
| | Vartotojo kopija: sistemos reikalavimai: interneto naršyklė |
| | Originalas suskaitmenintas TIFF formatu; vartotojo kopija - JPEG formatu |
| | Skaitmenintas objektas iš: LIETUVOS EVANGELIKŲ REFORMATŲ SINODAS. 240 : [Karaliaus Stepono Batoro raštas revizoriams dėl Dainių kaimo (Raseinių pav.) žemės apmatavimo ir perdavimo valdyti Daratai Šemetienei (Szemetowa) pagal apsikeitimo teisę]. - 1586.III.28. - 2 lap., 1 dok. |
| | Prieiga prie archyvinės kopijos teikia LNB |



Summarizing Conclusions

- ✓ **WHO:** playing the role in the Event - the National Library of Lithuania has digitized a manuscript
- ✓ **WHERE:** the place of the EVENT - at the National Library of Lithuania;
- ✓ **WHEN:** the time of the EVENT - 19 March 2019
- ✓ **HOW:** the kind of processes - digitization
- ✓ **HOW:** provenance records are available for users - a Web browser
- ✓ **HOW:** access to archival copy - access at the National Library of Lithuania (www.epavldas.lt)

Fig 6. <http://www.epavldas.lt>

https://www.epaveldas.lt/object/recordDescription/LNB/C1B0002930092

558f443908ae47a3... 39. Lietuvos Respub... Search results... ed

Aitvaras / Judita Vaičiūnaitė. - 2005

[view digital images of the object](#)

Type text

Authors [Judita Vaičiūnaitė](#)

Other people [Ula Vaičiūnaitė](#), [Laima Kryževičiūtė](#)

IPublishing data Alma littera [2005]

Dimension data iliustr., 100, [4] p.; Archyvinė kopija: 1182MB Viešos prieigos kopija: 12MB

Object provider [Lietuvos nacionalinė Martyno Mažvydo biblioteka](#)

Print system lotynų

Languages lietuvių

Note

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 naršyklė, Adobe Acrobat Reader

Originalas suskaitmenintas TIFF formatu; vartotojo kopija - JPEG formatu

Skaitmenintas objektas iš: Aitvaras : [eilėraščių 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 index **888.2-93-1, 769.2(474.5)(084)**

Download e-book in format:

[EPUB](#) [MobiPocket](#) [PDF](#) [HTML](#) [TXT](#)



Summarizing Conclusions

There is a standard way within VEPIS to find provenance information for a given resource - on the portal <http://www.epaveldas.lt>

- (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/institution in creating a new object.

Fig. <http://www.epaveldas.lt>

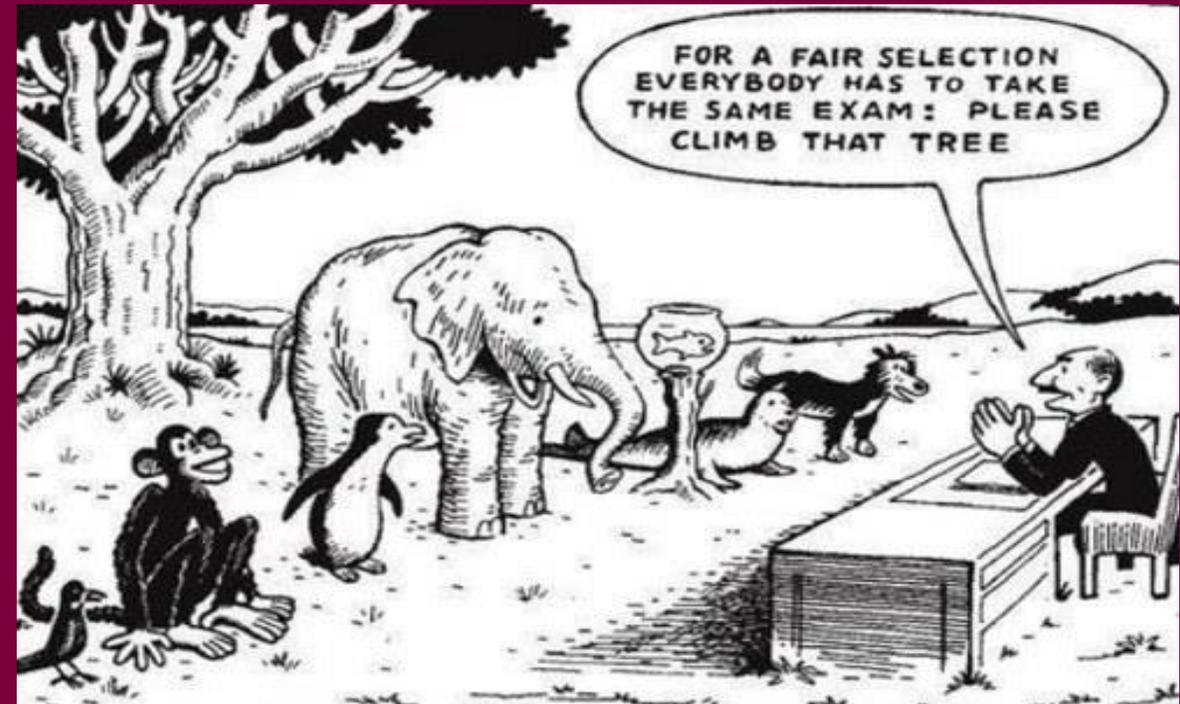
Imperfections and Next Steps of the Research

Vilniaus
universitetas

We must acknowledge that Lithuanian information systems of the research and cultural domain are not related yet, there are no integrated search platforms, information from provenance records created by diverse systems is not integrated either.

In this context, VEPIS also has some imperfections, the fundamental of which is the fact that there is no standard way to obtain provenance regarding the object from heterogeneous systems and diverse representations.

In order to make information systems of research and cultural domains become interoperable, we will continue modeling a standard way to obtain provenance regarding the object from heterogeneous systems and diverse domains.



theodysseyonline.com