

CARARE 2.0: a metadata schema for 3D Cultural Objects

Andrea D'Andrea

Centro Interdipartimentale di Servizi di Archeologia
Università degli Studi di Napoli "L'Orientale"
Naples, Italy
dandrea@unior.it

Kate Fernie

MDR Partners (Consulting) Ltd
London - United Kingdom
kate.fernie@mdrpartners.com

Abstract— One of the features of the digital data is that they cannot be understood without information about their meaning and the ways they have been created. 3D ICONS, a project funded by the European Commission, has brought together partners expert in 3D digitization of the archaeological and architectural heritage to contribute content to Europeana. The project has had as one of its main objectives the quality control of 3D data and establishing a metadata schema to support the provenance and paradata required for quality assurance of 3D models. This paper describes provenance in the CRM_{dig} schema, the *paradata principles* of the London Charter and how provenance and paradata could be relevant for the new strategy of Europeana. The schema, which builds on earlier work in the CARARE project, aims to foster the adoption of a clearer approach to describing the features of cultural object, the techniques and the methodologies chosen for the digitization and the motivation behind the creation of a digital object. Complete knowledge of the digital resource allows for more efficient reuse and increased usability of the resources on-line.

Keywords— 3D Objects, Metadata, 3D-ICONS, Europeana, EDM, CARARE, Provenance, Paradata

I. Introduction

Large-scale production of digital data, given the variety of processes involved and the complexity of relationships amongst digital objects, requires an innovative knowledge management system. Scientific data cannot be interpreted and reused correctly without information about the meaning of the data, the way in which it has been created and the conditions. In the field of the 3D all these aspects are particularly relevant given the range of technologies, instruments and methodologies available for data-capture and data-processing, and the diversity of motivations and rationales behind any 3D reconstruction.

Metadata can store information about the life-cycle of a 3D object from the initial generation phase to later uses, storage and the possible reuses. Metadata allows field professionals to keep track of the instrument settings (calibration, tolerances, errors), the physical object (status, conditions) and the possibilities to improve or reprocess the model.

3D ICONS, a project funded by the European Commission's ICT Policy Support programme, has brought

together partners expert in 3D digitization of the archaeological and architectural heritage to contribute content to Europeana [1]. A main objective of the 3D ICONS project has been to establish a pipeline for the capture, processing and publication of 3D models, which incorporates quality control of the digital data being produced. This paper reports on the solutions that the project has adopted its first year concerning metadata and the way the London Charter and other quality recommendations have been implemented [2]. 3D-ICONS has prepared an update to the metadata schema prepared by the CARARE project [3] to support the provenance and paradata required for quality assurance of 3D models [4]. The project also intended to contribute to the maintenance of the CARARE schema by monitoring developments in the core module of Europeana Data Model (EDM) [5].

As far as provenance is concerned, the paper describes the CRM_{dig} schema adopted and customized by the 3D-COFORM project [6] [7]. It also addresses the *paradata principles* contained in the London Charter and how paradata in the metadata schema are relevant for Europeana. In the Section 3 the paper introduces EDM and the CARARE schema, describing the current status of these two metadata systems and the mapping between them. Finally, both the new elements introduced by 3D ICONS to the CARARE 2.0 schema and the mapping to EDM are described.

II. Provenance and Paradata

A. Provenance

According to the W3C Provenance Incubator Group "Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource" [8]. In the Cultural Heritage domain scientific data are often based on direct observation or direct measurement of a physical object using devices to create a digital object. In the digital acquisition and processing phases it is particularly important to record the technologies and methodologies used, and the rationale behind the creation of the digital object. To make all the technological and scientific choices made in creating a model explicit it is necessary to record information linked to the origins and derivations of digital resources. The CRM_{dig} was implemented

as an extension of the CIDOC-CRM ontology with this specific objective in mind.

The scope of the CRM_{dig} is to describe all the processes starting at the level of human activities or actions, which in turn initiate "machine events" on devices and computers, and form a connected graph through the data, people and things involved in multiple events in various roles. The relevant context of these actions comprise descriptions of objects, people, places and times, which in turn may be related to other things.

Recently the European Project 3D-COFORM addressed the large-scale production of 3D objects for scientific and cultural purposes using the CRM_{dig}, which was tested, revisited and implemented in RDFS [9].

The CRM_{dig} model is particularly appropriate in describing a typical workflow, from acquisition to processing, synthesis, presentation and, finally, reuse. Creating in this way a complex semantic network of relationships. The model is designed to provide reliable registration of the capture device instruments, the parameters used in data-acquisition (geometry, light sources, obstacles, sources of noise/reflections, etc.) and in the subsequent processing phase (registration, meshing, texturing, decimation, simplification, etc.). CRM_{dig} also allows for a clear description of the organization of data-acquisition and the system adopted for the alignment of the shots (targets, TLS, GPS).

After considering the objectives of 3D-ICONS, the CRM_{dig} model was chosen because it allows a simple and clear description of the processes carried out to digitize and render a 3D model.

B. Paradata

While Provenance is a record of the technical processes involved in creating digital objects, Paradata can be described as information about the human processes of understanding and interpreting. Paradata include descriptions of the evidence used to interpret an artefact and its reconstruction and descriptions of the methodological premises within a research project. It is intended to communicate the process through which one or more artefacts were processed and interpreted. Paradata makes the methodological premises and the research targets behind digitization explicit; it enables for example alternative hypothesis and factual evidence supporting the reasoning behind the implementation of a 3D model to be expressed [10].

In the framework of computer-based visualisation numerous scientific initiatives have underlined the importance of ensuring clear distinctions between evidence and hypothesis, and between different levels of probability. A set of principles reflecting the distinctive properties of computer-based visualisation technologies and methods, must be made explicit like any other information concerning the cultural object. The London Charter, an international initiative dedicated to these issues provides guidelines and principles for the use of computer-based visualisation methods in relation to intellectual integrity, reliability, documentation, sustainability and access. The Charter aims to promote technical rigour in

digital heritage visualisation, and ensure that computer-based visualisation processes and outcomes can be evaluated properly by users.

The London Charter does not implement a metadata schema or prescribe specific methods; it establishes broad principles for the use of computer-based visualisation upon which the intellectual integrity of such methods and outcomes depend.

In the framework of 3D-ICONS, principle 4.6 is particularly relevant. This covers the Documentation Process (Paradata): *"Documentation of the evaluative, analytical, deductive, interpretative and creative decisions made in the course of computer-based visualisation should be disseminated in such a way that the relationship between research sources, implicit knowledge, explicit reasoning, and visualisation-based outcomes can be understood"*.

The CIDOC-CRM and the CRM_{dig} allow alternative hypothesis, factual evidences, motivations and rationale at the basis of the implementation of 3D content to be modelled. In section IV we will show how provenance and paradata elements have been introduced into the CARARE 2.0 metadata schema according to the principles of the documentation process contained in the London Charter.

III. EDM and CARARE schemas

A. Europeana Data Model (EDM)

Europeana is a portal for millions of items from Europe's galleries, archives, libraries and museums [11]. Recently it introduced the EDM along with a new Data Exchange agreement to make a qualitative change in the way Europeana deals with metadata provided by institutions and to enable Linked Open Data. Of interest here is the clear distinction made in EDM between the technical creation of a digital resource submitted by a provider, the physical object this resource is about, and the digital representations of this object accessible over the web [12].

EDM facilitates the approach of the Semantic Web and it can be used as an anchor to which various finer-grained models can be attached, making them at least partly interoperable. In this sense, EDM is an attempt to overcome the respective information perspectives of the various communities constituting Europeana, such as museums, archives, audio-visual collections and libraries. EDM is not built on any particular community but rather adopts an open, cross-domain framework that can accommodate particular community standards.

As EDM is an integration medium for collecting, connecting and enriching the descriptions provided by Europeana content providers. It can accommodate any element (class or property) found in a content provider's description and forms an open set that can be easily extended.

EDM is based on three core classes that result from the package of data provided to Europeana:

- the cultural heritage object provided, for example a painting, movie, music score, book (edm:ProvidedCHO);
- one or more accessible digital representations of this object, some of which will be used as previews (edm:WebResource);
- an aggregation to represent the result of this provider's activity (ore:Aggregation).

The first two classes allow the distinction between the physical objects, expected to be the focus of users' interest, and their digital representations visualized online to be captured. The third, following the ORE approach, demonstrates that the provided object, together with the digital representations provided by data provider can be regarded as one logical whole.

Among the possible models for descriptive metadata, EDM distinguishes “*object-centric*” and “*event-centric*” approaches. EDM allows either approach.

The *ObjectCentric* model focuses on the provided object: information are expressed in the form of statements that link the described object and its features; they can be simple strings or more complex resources denoting entities from the real world. Most metadata practices making use of the Dublin Core metadata set [DC] can be seen as an application of such an approach. To support enrichment, EDM includes a number of classes devoted to the representation of “contextual” entities:

- edm:Agent: for representing persons or organizations;
- edm:Event: for events;
- edm:Place: for spatial locations;
- edm:TimeSpan: for time periods or dates;
- skos:Concept: for knowledge organization systems such as thesauri or classification schemes.

In the *Event-centric* approach the focus is on characterizing the various events in which objects have been involved. This approach underlies models such CIDOC-CRM and may suit the data of some Europeana providers. The idea is that this approach will lead to finding richer networks of entities and meanings - by representing the events that constitute an object's history - than with the object-centric approach.

It is not the aim of EDM to capture the full complexity of a model like CIDOC-CRM. Nor is its scope to express the full diversity of all object-centric approaches. Rather, EDM provides a small set of concepts, entities, attributes, properties and classes to which more specialized constructs can be attached.

The *EDM mapping guidelines* [13] gives guidance for providers that want to map their data to EDM. It contains definitions of the properties, information about the data types that can be used as values and the obligation level of each property. The report provides also examples of original data, converted to EDM and diagrams showing the distribution of

the properties amongst the classes. *The EDM object templates* [14] are also important giving information on the implementation of EDM in Europeana, which elements are currently implemented, the data type, attributes and the obligation (mandatory, recommended, optional).

B. CARARE Schema

CARARE [3] was a three-year project which involved the network of heritage organisations, archaeological museums, research institutions and specialist digital archives in making their digital content available to Europeana. The project established an aggregation service for archaeology and architecture resources and enabled the integration of 3D and Virtual Reality content in Europeana.

As information about the archaeological and architecture heritage domain belongs to a broad and diverse set of heritage organizations across Europe, described according to different standards and management procedures, one of CARARE's main objectives was to ensure interoperability between the native metadata held by heritage organizations and the metadata used by Europeana. For the integration of digital resources representing such heterogeneous assets, one of the most relevant task of CARARE was to establish a metadata schema able to map the existing original metadata into a common output schema.

Conceptually the CARARE record is focused on a heritage asset and its relations to digital resources, activities and to collection information. An object in the CARARE schema consists of the Heritage Asset Identification (HA) wrapped together with the related Digital Resources (DR), Activities (A) and Collection information (C). The cardinality of themes and elements has been specified to enable the harvesting of the real data actually present in CARARE content providers' datasets and to meet Europeana's requirements.

C. CARARE Mapping

The CARARE mapping [15] [16] uses EDM classes to describe different types of information about:

- Heritage assets, such as monuments, buildings or other real world objects, identified by a set of particular characteristics that refer to their identity, location, related events, etc. Information carried by a heritage asset includes: textual metadata (such as title, etc.), thumbnails and other digital objects.
- Real world cultural objects with their digital representations which provide other sources of information about the heritage asset (historic drawings and photographs, publications, archive materials etc.).
- Born-digital resources related to these objects, such as 3D models.

EDM requires a unique identifier for each class; for this reason CARARE needed to introduce new identifiers resource based on the local identifier provided by the project partners. In EDM, each ProvidedCHO gives rise to an aggregation class, whose role is to bundle a ProvidedCHO with digital representations of the provided objects made available on the

web (WebResources). CARARE created unique identifiers for each ore:Aggregation. CARARE created unique identifiers for each ore:Aggregation. These aggregation identifiers are web-enabled, in the sense that they redirect to a landing page that CARARE created for each object. The landing page provides a unique identifier that can be used in the data, but also visually aggregates the heritage asset with its representations, collection information and information about related activities

IV. CARARE 2.0

One of the main goals of 3D-ICONS was to develop a metadata schema able of capturing all the semantics present in the digitations process (*provenance*) and in understanding and interpretation of data objects (*paradata*) and to build on the results of the CARARE project [17]. Thus work was completed in the first year of the project on updating the CARARE schema by adding classes or entities and properties to make the original schema compliant to the 3D-ICONS requirements.

Provenance and paradata were innovative for the CARARE schema, but it has possible to add this information without substantially changing the original schema or its mapping to EDM. In CARARE the Activity theme is designed to hold metadata about an event. In describing the different digitization processes it is important to distinguish different types of events (figure 1 illustrates the data acquisition process and figure 2 illustrates post processing activities). For this

reason, 3D ICONS proposed extending the definition of the Event theme as follows:

- **EventType** – general classification of the type of event or activity which took place, e.g. survey, archaeological excavation, digitization, rebuilding. Use of a controlled vocabulary is recommended.
- **Methods** – the methods used in this specific activity, e.g. open area excavation, sample survey, augering, boring, stratigraphic, restoration, conservation, re-pointing, photogrammetric survey etc. Use of a controlled vocabulary is recommended.

According to the principles of the London Charter information should be provided to define aims and objectives of 3D data-capture. Two new elements were proposed for the Activity theme to capture this information:

- **Had General Purpose** (source = CIDOC CRM) – this is a free text description of the general goal or purpose of an Activity. For example this could include practicing, preparing, monitoring, researching, designing, testing etc.
- **Had Specific Purpose** (source = CIDOC CRM) - a free text note describing the specific goal or purpose of this activity. For example, carrying out 3D data acquisition, restoration of a part of a building, completing a survey, constructing a building, etc.

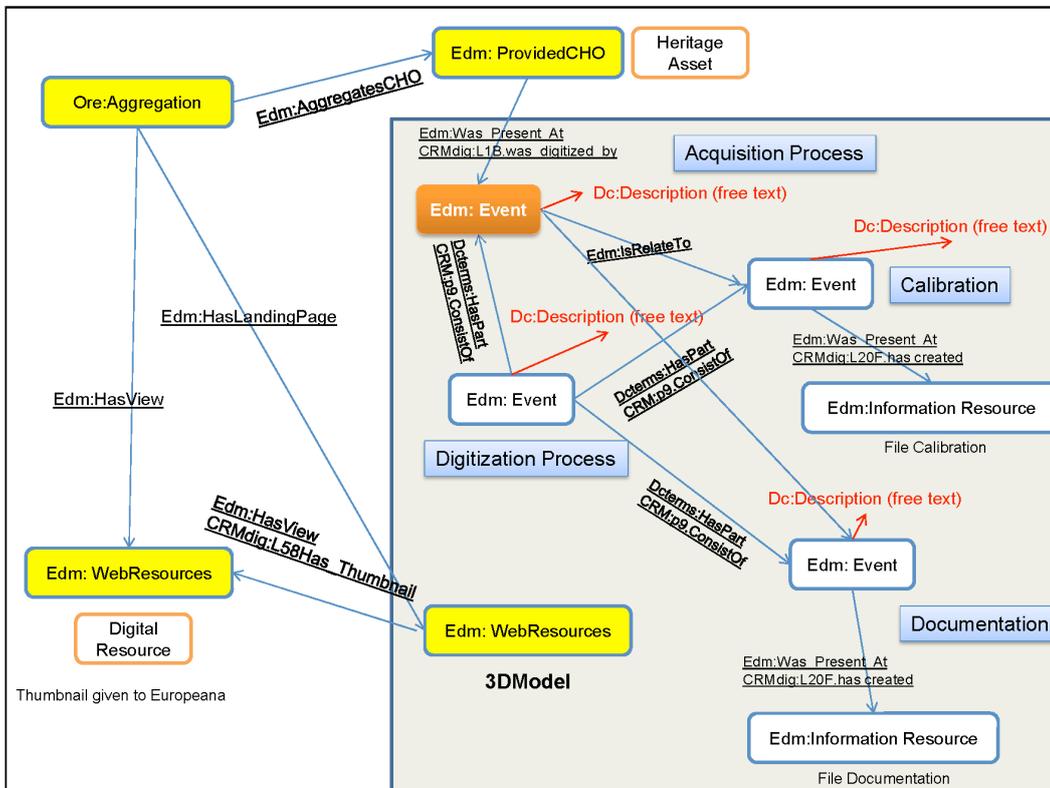


Fig. 1. Schema for the workflow of the data-acquisition process. On the left side the main EDM elements. In the grey rectangle the events chain introduced in the CARARE Schema 2.01

The CARARE V1.0 schema included an element, “Material and Techniques Used”, to register information about the digital machines and software used for the digitization process. This was specified a simple string. As no distinction was made between machines and software (concerning different processes) there was potential for misunderstandings. Ideally each technique and/or material would be described in separate field.

Consequently two separate elements were recommended for inclusion in the CARARE v2.0 schema:

- **Techniques** – the techniques used in this specific activity. Use of a controlled vocabulary is recommended.
- **Materials** – the materials used during the event method. Use of a controlled vocabulary is recommended.

To define the digitization process used to create the 3D final model the CARARE 2.0 schema adds some relations to explain the relationships between the Heritage Asset, Digital Resource and the Activity related to digitization processes (fig. 3). These reuse relations from CIDOC-CRM and in particular from CRM_{dig}.

The relations are:

- **Was digitized by** – this is the relation between a Heritage Asset and an Activity in which it was digitized. (It is a specialization of Was Present At). Give the id number of the target record or a URI.
- **Has Created** (source = CRM_{dig}) this is the relation between an Activity and a digital resource or digital file that it created; includes raw data files, processed data files and final models published online. Give the id of the target record, the file-name or a URI.
- **Consists of** (source = CRM_{dig}) this is a repeating group of elements which allows the specific activity (or activities) that took place during the overall Event to be described.
- **Created Derivative** – this is the relation which defines the reuse of a Digital Resource to create derivatives, e.g. during the different processing phases of digitization. It is a specialization of Is Derivative Of. Give the id number of the target record or a URI.

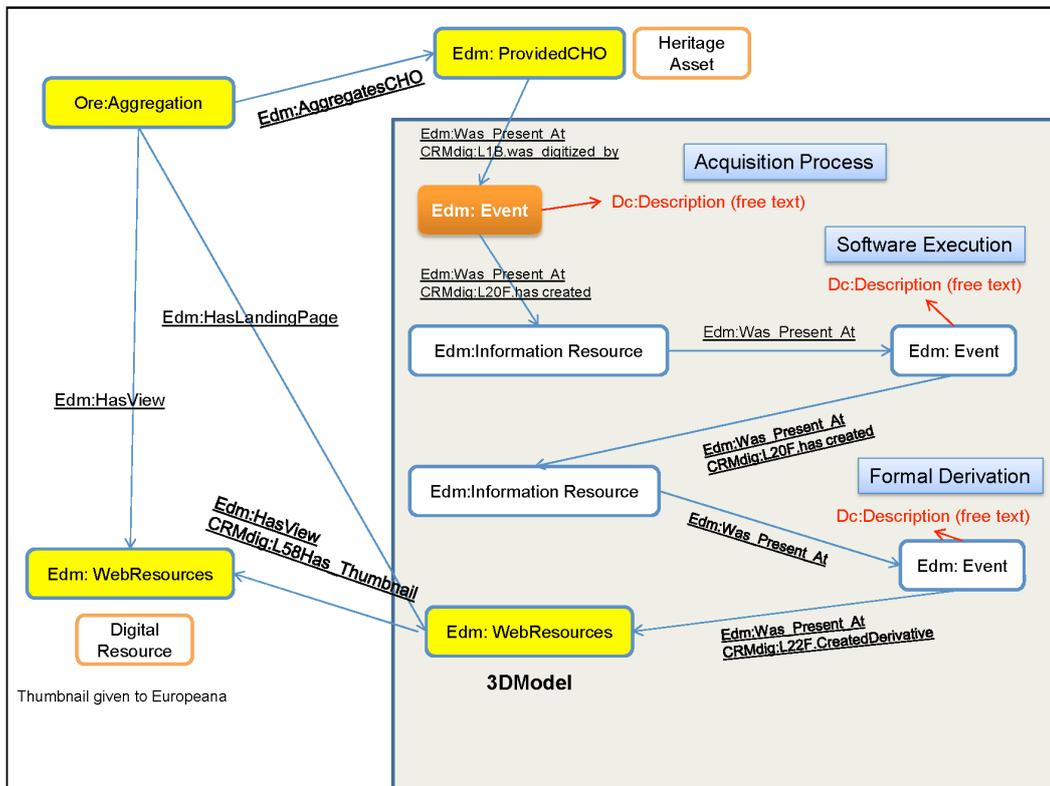


Fig. 2. Schema for the workflow of the post-processing. On the left side the main EDM elements. In the grey rectangle the events chain introduced in the CARARE Schema 2.01

CARARE 1.0 was designed to be used for metadata harvesting and to provide an intermediate schema between native metadata and EDM. For this reason another task of the 3D-ICONS project in establishing the CARARE 2.0 was to build on the experience of the CARARE project in providing

over 2 million records and more than 40 distinct collections to Europeana and the latest developments in the implementation of EDM by Europeana. This resulted in a simplification and a closer alignment between the CARARE 2.0 schema and EDM; the CARARE 2.0 schema can be described as an application

profile and extension of EDM designed to include the requirements of the archaeological and architectural heritage domain and of 3D digitisation processes.

The last OWL version of EDM has been aligned to CIDOC-CRM Core Classes and some properties of CIDOC-

CRM have been reused in EDM allowing a more simple integration of CRM_{dig} into EDM. In this way all the new entities and properties have been correctly mapped on the corresponding EDM elements.

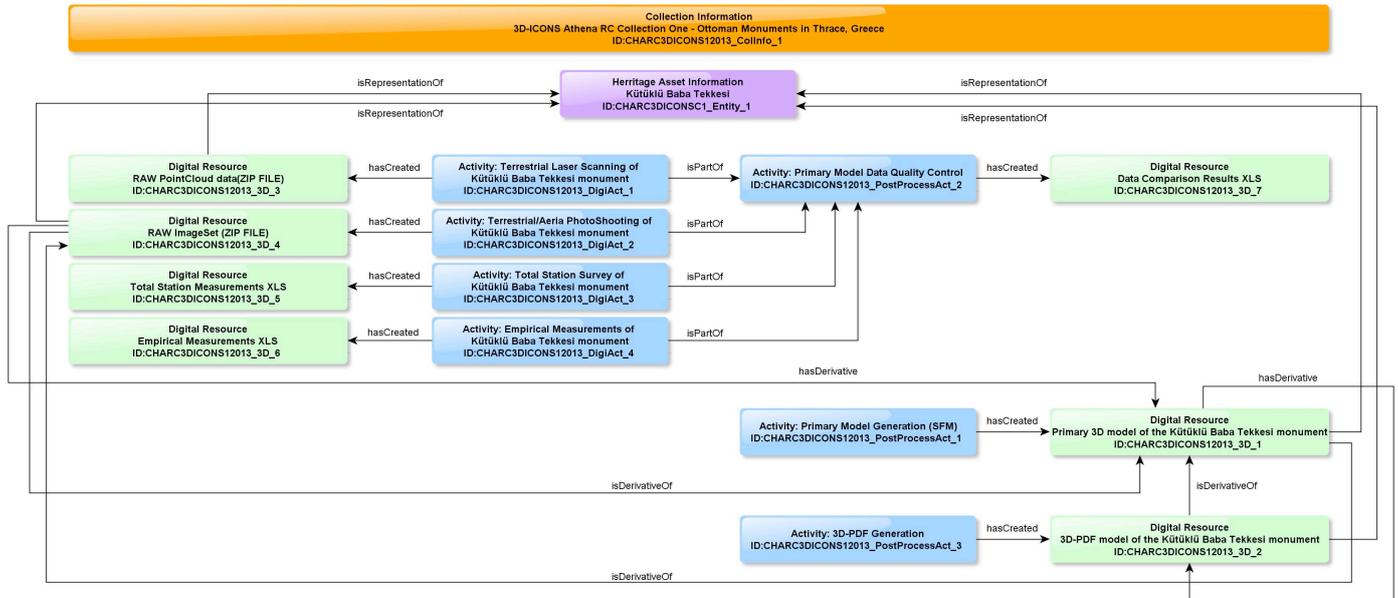


Fig. 3. CARARE 2.01 Schema. The pipeline: from 3D data-capture to the final 3D Model

v. Conclusions

The CARARE v2.0 schema is described in a report, submitted at the end of the first year of the 3D-ICONS project, [13]. The updated metadata schema is a good compromise between the need to store a complete record of the digitization processes, provenance and paradata, and an easy way to standardized the metadata for 3D models. The update to the CARARE schema, covering paradata and provenance concepts, assures a simple mapping with EDM.

We hope the CARARE 2.0 schema will encourage Europeana providers to adopt a clearer approach to describing the features of the cultural object, the techniques and the methodologies chosen for its digitization and the motivation behind the creation of the digital object. Complete knowledge of the digital resource allows for a more efficient reuse of the archive and increase the usability of the resources available on-line. Furthermore it will be easier to compare models, their complexity, any eventual innovation in their creation and their reliability.

The 3D models of which will be made available from the 3D ICONS project will show best practices and be based on guidelines useful for new projects and 3D data collections. The metadata provided to Europeana will be published under the terms of the Creative Commons CC0 1.0 Universal Public Domain Dedication and can therefore be re-used by third parties without any restrictions [18].

Europeana recently launched the portal data.europeana.eu to make metadata openly available to public and private sectors alike so they can use it to develop innovative applications and create new web services and portals,. So far only a few collections have been published according to the Linked Open Data (LOD) protocol. The data are represented in EDM and the described resources are addressable and de-referencable by their URIs. Links between Europeana resources and other resources in the Linked Data Web enable the discovery of semantically related resources.

In the future all Europeana metadata will be published as Linked Open Data thus enabling data providers to benefit from Europeana's efforts to link them to semantically related resources on the Web [19].

Acknowledgment

3D-ICONS is funded by the European Commission under the *Competitiveness and Innovation Framework Programme* (CIP), contract no. 297194. However, this paper reflects only the authors view and the European Community is not liable for any use that may be made of the information contained herein. The authors wish to thank Anestis Koutsoudis (CETI - Greece) for the figure 3 and for the implementation of some tests about the schema.

References

- [1] <http://www.3dicons-project.eu>. (accessed September 15, 2013)
- [2] London Charter for the Use of 3D Visualisation in the Research and Communication of Cultural Heritage", was circulated and was adopted

- as the first official draft on June 2006 <http://www.londoncharter.org/>. (accessed September 15, 2013)
- [3] <http://www.carare.eu>. (accessed September 15, 2013)
- [4] <http://3dicons-project.eu/eng/Resources/Documentation/CARARE-2.0-schema>. (accessed September 15, 2013)
- [5] <http://pro.europeana.eu/edm-documentation>. (accessed September 15, 2013)
- [6] M. Doerr, M. Theodoridou, "CRMdig: A generic digital provenance model for scientific observation", Proceedings of TaPP 2011: 3rd USENIX Workshop on the Theory and Practice of Provenance, Heraklion, Greece (June 2011). http://static.usenix.org/events/tapp11/tech/final_files/Doerr.pdf. (accessed September 15, 2013)
- [7] <http://www.3d-coform.eu>. (accessed September 15, 2013)
- [8] W3C Provenance Incubator Group. http://www.w3.org/2005/Incubator/prov/wiki/What_Is_Provenance. (accessed September 15, 2013)
- [9] http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs. (accessed September 15, 2013)
- [10] H. Denard, "A New Introduction to the London Charter" in A. Bentkowska-Kafel, D. Baker & H. Denard (eds.) *Paradata and Transparency in Virtual Heritage Digital Research in the Arts and Humanities Series* (Ashgate, 2012) 57-71. http://www.londoncharter.org/fileadmin/templates/main/docs/ch6_denard.pdf. (accessed September 15, 2013)
- [11] <http://www.europeana.eu/portal/aboutus.html> (accessed September 15, 2013)
- [12] EDM Data Model elements, Version 5.2.3 (24/02/2012) Europeana v 1.0 <http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22>. (accessed September 15, 2013)
- [13] EDM Mapping Guidelines v.1.0.1 (24/02/2012) <http://pro.europeana.eu/documents/900548/ea68f42d-32f6-4900-91e9-ef18006d652e>. (accessed September 15, 2013)
- [14] EDM Object templates: <http://europeanalabs.eu/wiki/EDMObjectTemplatesProviders>. (accessed September 15, 2013)
- [15] Charles V., Isaac A., Fernie K., Dallas C., Gavrilis D. and S. Angelis, "Achieving interoperability between the CARARE schema for monuments and sites and the Europeana Data Model". Proc. Int'l Conf. on Dublin Core and Metadata Applications, 2013, forthcoming.
- [16] <http://pro.europeana.eu/documents/900548/8303c3d2-10b8-4f36-9db5-6a3dd06bd400> (accessed September 15, 2013)
- [17] <http://3dicons-project.eu/eng/Media/Files/D6.1-Report-on-Metadata-Thesauri>. (accessed September 15, 2013)
- [18] <http://pro.europeana.eu/web/guest/data-exchange-agreement>. (accessed September 15, 2013)
- [19] Isaac A., Clayphan R., Haslhofer B., "Europeana: Moving to Linked Open Data", ISQ Information Standards Quarterly Spring/Summer 2012, volume 24, issues no. 2/3, http://www.niso.org/apps/group_public/download.php/9407/IP_Isaac-et-al_Europeana_isqv24no2-3.pdf. (accessed September 15, 2013)