3D Digitisation of Icons of European Architectural and Archaeological Heritage

D6.1: Report on Metadata and Thesaurii

Authors:
Andrea D'Andrea (CISA)
Kate Fernie (MDR)

Revision History

Rev.	Date	Author	Org.	Description
1.0	30/01/2013	A. D'Andrea	CISA	First draft for review and discussion
1.0	30/01/2013	K.Fernie	MDR	First draft for review and discussion
1.0	05/02/2013	C. Chamzas	CETI	Review and discussion
1.0	05/02/2013	A. Koutsoudis	CETI	Review and discussion
1.0	05/02/2013	D. Gavrilis	NTUA	Review and discussion
1.0	05/02/2013	D. Tsiafaki	CETI	Review and discussion
1.0	05/02/2013	N. Sofou	NTUA	Review and discussion
1.0	05/02/2013	F. Niccolucci	CISA	Review and discussion
1.0	05/02/2013	P. Ronzino	STARC	Review and discussion
1.0	05/02/2013	G. Ferrandino	CISA	Review and discussion
1.0	05/02/2013	C. Vastenhoud	KMKG	Review and discussion
1.1	05/03/2013	A.D'Andrea	CISA	Second draft for review and discussion
1.1	05/03/2013	K. Fernie	MDR	Second draft for review and discussion

Revision: [Final]

Authors:
Andrea D'Andrea (CISA)
Kate Fernie (MDR)

TABLE OF CONTENTS

1. Executive Summary

This deliverable reports on the solutions adopted by the project as far as metadata is concerned, how the London Charter and other quality recommendations have been implemented, the preferred thesauri and authority lists. The report has been prepared by CISA with the support of other partners involved (CETI, KMKG, MAP and MDR).

One of 3D-ICONS main objectives was to prepare an extension to the CARARE schema to support provenance, transformation and paradata required for quality assurance of 3D models. In addition, the project intends to contribute to the maintenance of the CARARE schema by monitoring developments in the core module of Europeana metadata schema (EDM). The report also offers a survey on the thesauri and authority files used by the partners and finally it contains some recommendations on the future possible scenarios in the LOD paradigm.

In the introduction (**Chapter 2**) the report focuses on the description of the starting point of the project in terms of definition of metadata for 3D objects. This introduces the CARARE metadata schema and the new EDM schema highlighting the state of art of the two schemas and showing some recent case-studies based on the integration of the two metadata schemas (CARARE-EDM). As far as provenance is concerned, this will be introduced through the CRMdig schema recently adopted and customized by the 3D-COFORM project. Finally, the report addresses the *paradata principles* contained in the London Charter and how the input of paradata in the metadata schema could be relevant for the new strategy of Europeana.

Chapters 3 and 4 exposes the definition of the CARARE+ schema based on the objectives of the 3D-ICONS and the results achieved in the mapping of CARARE on the EDM schema. Particularly it outlines how to describe provenance and paradata implementing the event model and correlating different events or activities that the object has taken part in. Some graphical schemas or template will help partners to encode their data on the basis of the CARARE+ schema and the EDM mapping schema.

Chapter 5 presents a short survey of the thesauri used in some CH institutions providing a brief introduction about the institutions responsible for their publication and the management.

Finally, in **Chapter 6,** some possible future scenarios suggested by the recent release of the technical requirements for the implementation of LOD in the framework of a new Europeana data pilot project will be discussed.

In the **appendix** an updated proposal for version 2.0 of the CARARE metadata schema is presented following the experience gained as a result of aggregating metadata for Europeana from the CARARE content providers and taking into account the requirements of the 3D ICONS project.

2. Introduction

2.1 CARARE

CARARE[1] is a Best Practice Network funded under the ICT PSP 2009 programme of the European Commission. It is designed to involve the network of heritage organisations, archaeological museums, research institutions and specialist digital archives:

- in making the digital content that they hold available to Europeana;
- establishing a sustainable aggregation service for archaeology and architecture resources;
- enabling the integration of 3D and Virtual Reality content in Europeana.

CARARE is a three-year project addressed to make digital content, including unique archaeological monuments, architecturally important buildings, historic town centres, industrial monuments and landscapes, available to Europeana's users. As archaeological and architecture heritage domain belonging to a broad and diverse set of heritage organizations across Europe, was described according different standards and management procedures, one of CARARE's main objectives has been to ensure interoperability between the native metadata held by heritage organisations 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 create a metadata schema able to map the existing original metadata into a common output schema. The project has established a metadata schema used to mediate between the original data and the standards and schemas defined by Europeana.

The CARARE metadata schema[2] is a harvesting schema intended for delivering metadata to the CARARE service environment about an organisation's online collections, monument inventory database and digital objects. The strength of the schema lies with its ability to support the full range of descriptive information about monuments, building, landscape areas and their representations. It does not support activities such as monument management and protection.

The CARARE metadata schema is an application profile, based on MIDAS Heritage and POLIS DTD and built on existing standards and best practice from a number of different countries in Europe. The CARARE schema's focus is on the detailed description of monuments, events in which the monument has been involved and resources which represent and provide sources of information about the monument. The CARARE schema includes the Digital Resource set which covers the information needed for the digital resources being made accessible to the CARARE and Europeana service environment.

Conceptually the CARARE record is focussed on a heritage asset and its relations to digital resources, activities and to collection information.

The top level elements in a CARARE record are:

CARARE Wrap – The CARARE schema wrap, it wraps one or many CARARE records.

CARARE – The CARARE start element. It wraps the Heritage Asset with the other information resources (Collection information, Digital Resource and Activity):

---

[1] CARARE is funded by the European Commission's ICT Policy Support Programme – contract n° 250445. The project started on 1st February 2010. www.carare.eu.
[2] The paragraph is based on Report on "*Carare Metadata schema outline v.1.1*" (November 2011), http://www.carare.eu/eng/content/download/4277/35150/file/CARARE%20metadata%20schema%20outline%20v1.1.pdf

Heritage asset Identification (HA) – the descriptive information and metadata about the monument, historic building, archaeological landscape area, shipwreck, artefact, etc. The ability to create relations between heritage asset records allows the relationships between individual monuments that form parts of a larger complex to be expressed.

Digital resource (DR) – these are digital representations and sources of information (images, texts, videos, audio, 3D models) about the heritage asset being provided to the service environment (e.g. Europeana). They are often digital representations of monuments or of parts of monuments.

Activity (A) - these are events that the heritage asset has taken part in, such as Creation, Field investigation, Research and analysis Historical events, Change in use, etc.

Collection (C) – this is a collection level description of the data being provided to the service environment.

CARARE Core schema
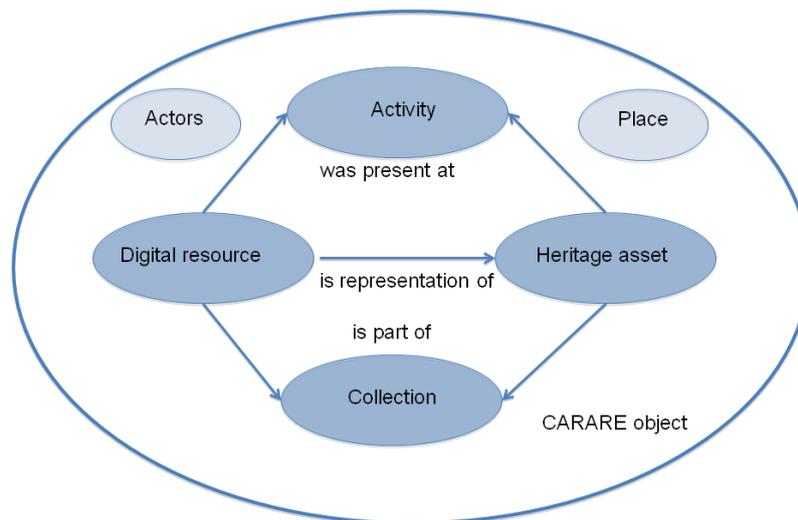
In the CARARE schema each of these top elements are global wrappers which are contained within a CARARE wrapper. An object in the CARARE schema consists of the Heritage Asset Identification (HA) wrapped together with the related Digital Resources (DR), Activities (A) and Collection information (C). The 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. Since Digital Resources are required for Europeana, the schema specifies that it is mandatory to include at least one Digital Resource in a CARARE record, while it is highly recommended that each CARARE record should include a single Heritage Asset. For each of the 4 top-level themes the CARARE schema defines a rich set of descriptive information and also administrative metadata.

As all of these sources had to be aggregated and delivered into the common format, CARARE has been mapped on the schema which Europeana uses to describe its content, the Europeana Data Model (EDM). The digital content submitted by the project has been described using the EDM 5.2[3]. As EDM is a conceptual model the CARARE elements have been mapped directly to EDM Classes or to particular EDM paths. On the

---

[3]   Deliverable   D2.2.3_3.4   "*EDM-CARARE   metadata   mapping*"   (January   2011), http://www.carare.eu/eng/Media/Files/D2.2.3_3.4-EDM-CARARE-metadata-mapping

basis of this approach a monument has been considered an instance of the **Edm:ProvidedCHO** class and the features that refers to its nature have been grouped under the this EDM entity.

Use of the main EDM classes
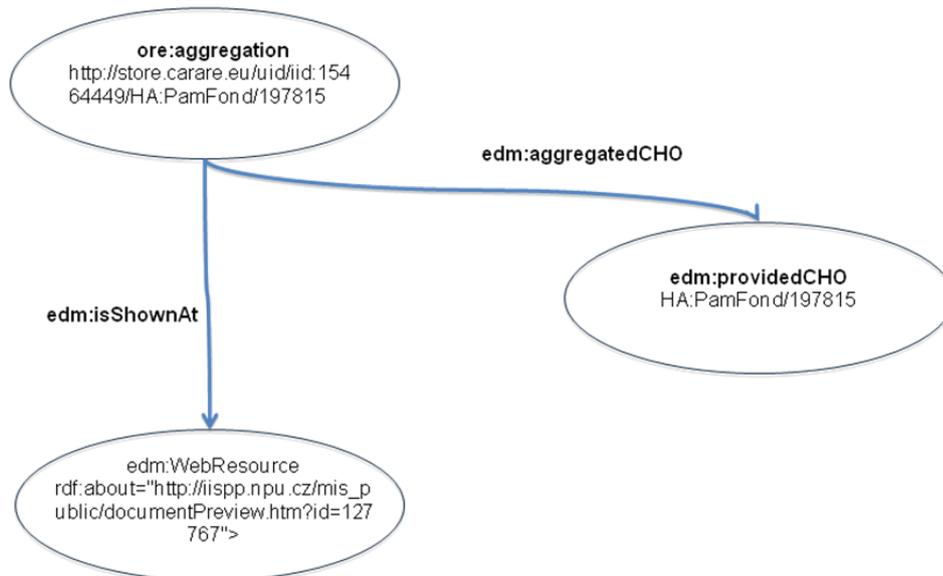
2.2 The Europeana Data Model (EDM)

The Europeana Data Model (EDM)[4], is a qualitative change in the way Europeana deals with the metadata gathered from data providers and aggregators. It is aimed at solving some of the issues observed with the ESE, by providing extra expressivity and flexibility.

In particular, it makes a distinction between the intellectual and technical creation that is submitted by a provider (a bundle of resources about an digital object created by the provider), the object this structure is about, and the digital representations of this object, which can be accessed over the web.

Also, EDM adheres to the modelling principles that underpin the approach of the Web of Data (*Semantic Web*). In that approach, there is no such thing as a fixed schema that dictates just one way to represent data. A common model like EDM can be seen instead as an anchor to which various finer-grained models can be attached, making them at least partly interoperable at the semantic level, while the data retain their original expressivity and richness. It does not require changes in the local approaches, although any changes in local practice that increase the cross-domain usefulness of the data is encouraged, such as the use of publicly accessible vocabularies (for persons, places, subjects etc.). In this sense, EDM is an attempt to transcend 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 standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate particular community standards.

As EDM[5] is aimed at being an integration medium for collecting, connecting and enriching the descriptions provided by Europeana content providers, it can include any element (i.e., class or property) found in a content provider's description. Giving an account of all these elements is clearly an impossible task, since they form an open set, i.e. a set that can be extended as new providers join the Europeana information space.

There is however a well-identified set of elements that EDM uses in order to carry out its task. These elements can be divided into two main categories:

1. The elements re-used from other namespaces, and
2. The elements introduced by EDM.

EDM re-uses from the following namespaces:

- The Resource Description Framework (RDF) and the RDF Schema (RDFS) namespaces (http://www.w3.org/TR/rdf-concepts/)
- The OAI Object Reuse and Exchange (ORE) namespace (http://www.openarchives.org/ore)
- The Simple Knowledge Organization System (SKOS) namespace (http://www.w3.org/TR/skos-reference/)
- The Dublin Core namespaces for elements (http://purl.org/dc/elements/1.1/, abbreviated as DC), terms (http://purl.org/dc/terms/, abbreviated as DCTERMS) and types (http://purl.org/dc/dcmitype/, abbreviated as DCMITYPE)

This motivation derived for the general goal of Europeana to exploit the richness of all available data in order to support the richest possible functionality, justifies three fundamental design principles:

---

[4]   "*Europeana   Data   Model   Primer*",   (26/10/2011)   Europeana   v1.0, http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5
[5] "*Definition of the Europeana 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

- D1. EDM allows data integration in an open environment: it is impossible to anticipate all data contributed
- D2. EDM allows for rich functionality, possibly via extensions
- D3. EDM should re-use existing (standard) models as much as possible

A number of requirements and principles have strongly influenced the design of EDM as it stands now while trying to figure out the motivation for some modelling choices in EDM. Noticeable requirements are:

- R1. distinction between "provided objects" (painting, book, movie, archaeology site,
- archival file, etc.) and their digital representations
- R2. distinction between objects and metadata records describing an object
- R3. multiple records for the same object should be allowed, containing potentially
- contradictory statements about this object
- R4. support for objects that are composed of other objects
- R5. compatibility with different abstraction levels of description (e.g. if a provider wishes
- to submit descriptions that follow the distinctions introduced in FRBR Group 1 [FRBR])
- R6. EDM provides a standard metadata format that can be specialized
- R7. support for contextual resources, including concepts from controlled vocabularies.

EDM has three core classes of resources that will result from the package of data provided to Europeana:

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

The first two allow capturing the distinction between "works", which are expected to be the focus of users' interest, and their digital representations, which are the elements manipulated in information systems like Europeana. The third, following the ORE approach, demonstrates that the provided object, together with the digital representations from one Europeana data provider can be regarded as one logical whole.
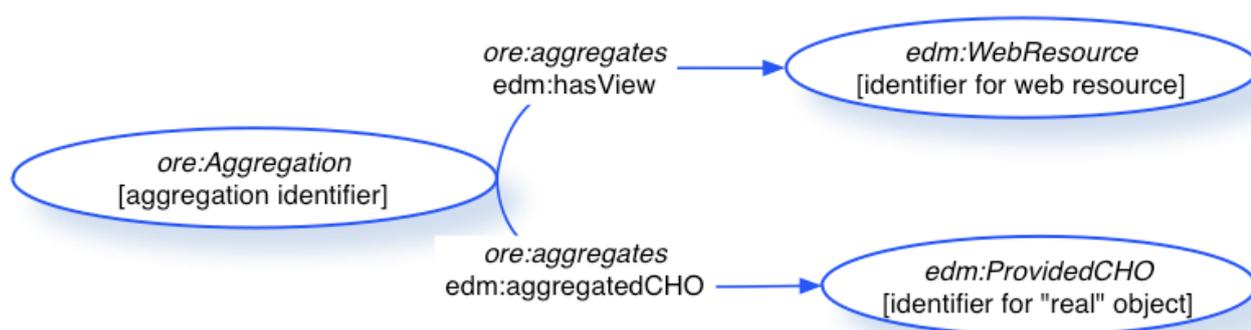
Visualization of the three core EDM classes for data providers

Aggregations enable capturing a description of the "digital environment" of an object submitted to Europeana, and attaching descriptive information to the various resources that take part in this environment. This mechanism remains however agnostic with regard to *which* descriptive data that should be provided. EDM therefore includes a set of "descriptive" and "contextual" properties that capture the different features of a resource, as well as relate it to the other entities in its context.

Among the possible approaches for descriptive metadata, one can distinguish *"object-centric"* and *"event-centric"* approaches. EDM provides constructs that allow representing metadata to follow either approach.

There are also classes in EDM that allow for capturing rich data. This section deals with these in order of complexity starting with the object centric approach, then looking at enriching this data with contextual classes and finally looking at the more complex event-centric approach.

The **ObjectCentric** model focuses on the object described: information comes in the form of statements that provide a direct linking between the described object and its features—be they 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.
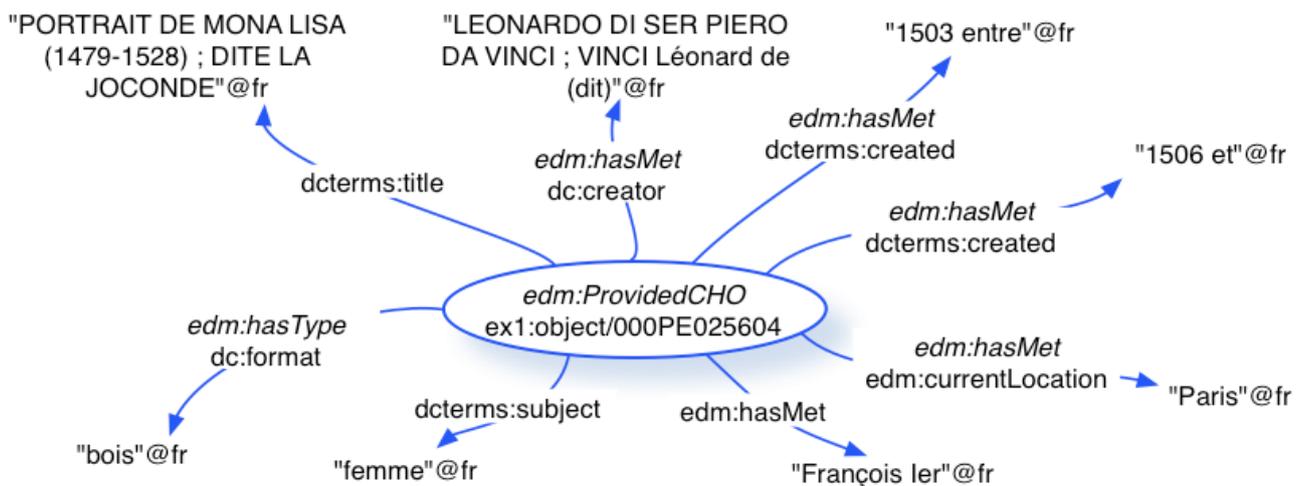
An object-centric description

To support the modelling of such semantic enrichment and to support further enrichment, EDM features a number of classes devoted to the representation of "**contextual**" entities:

- edm:Agent, to be used for representing persons or organizations
- edm:Event, for events
- edm:Place, for spatial entities
- edm:TimeSpan, for time periods or dates
- skos:Concept, for all entities from knowledge organization systems like thesauri,
- classification schemes (including some place gazetteers or person authority files)…
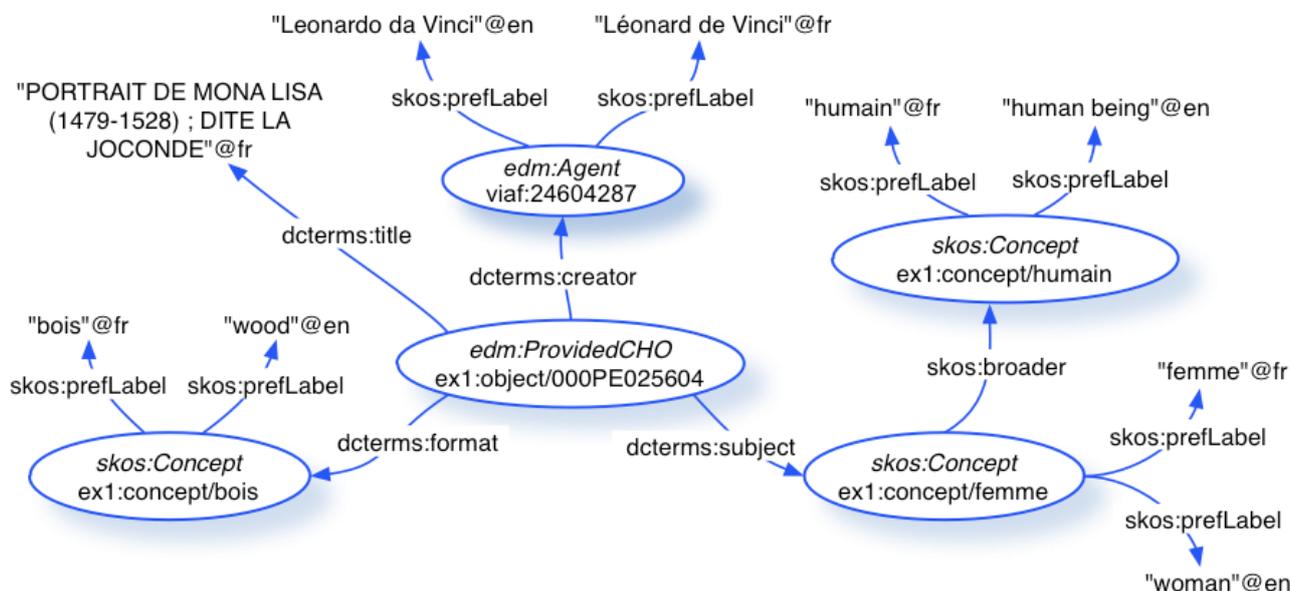
An object-centric description enriched using contextual entity

Event-centric approaches consider that descriptions of objects should focus on characterizing the various events in which objects have been involved. The idea is that it will lead to establishing richer networks of entities—by representing the events that constitute an object's history—than with the object-centric approach. This approach underlies models such CIDOC-CRM[6] and may suit the data of some (but of course not all) Europeana providers.

It is not the aim of EDM to capture the full complexity of a model like CIDOC-CRM. Nor can it capture the full diversity of all object-centric models. Rather, EDM provides a small set of properties and classes to which more specialized constructs can be "attached".

The *"EDM mapping guideline"*[7] report gives guidance for providers wanting 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. It also has an example of original data, the same converted to EDM and diagrams showing the distribution of the properties amongst the classes. "*The EDM object templates*"[8] shows which properties apply to which class and states the data type and the obligation of the values.

---

[6] http://cidoc.ics.forth.gr/
[7] "*Europeana Data Model. Mapping Guidelines v.1.0.1*" (24/02/2012), http://pro.europeana.eu/documents/900548/ea68f42d-32f6-4900-91e9-ef18006d652e
[8] http://europeanalabs.eu/wiki/EDMObjectTemplatesProviders.
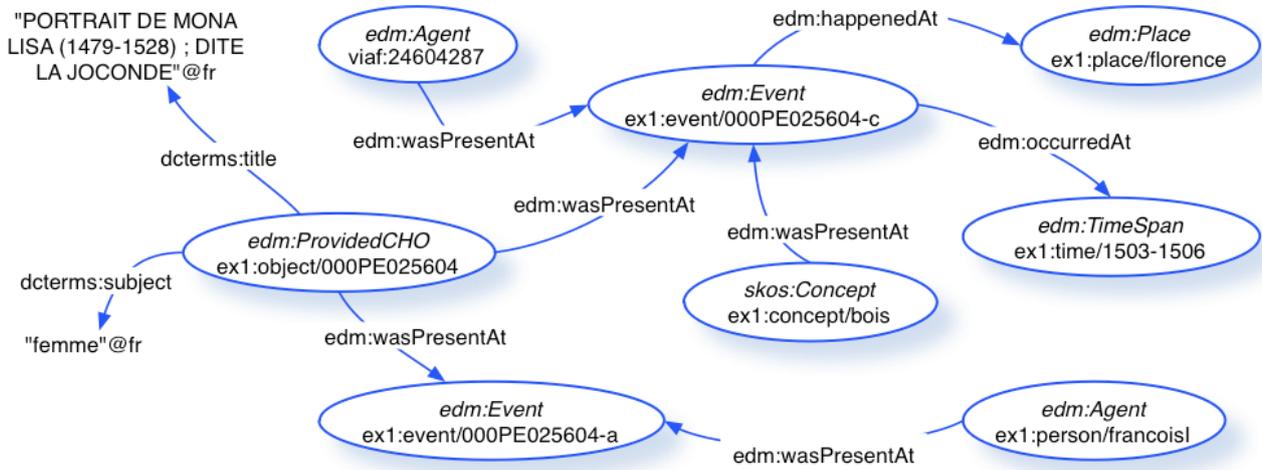
An event-centric description

2.3 CARARE-EDM[9]

As already described in section 2.1, CARARE first established a domain specific metadata schema based on existing standards from the archaeology and architecture domain and then further work was carried out to create a mapping from the CARARE schema to EDM. Using EDM, CARARE modelled a network of connections between the heritage assets themselves and other resources that are related to the assets. These resources describe either specific real objects with their digital representations or "born-digital" objects. In addition, it used the EDM contextual classes to model entities such as places and concepts.

CARARE's initial mapping to EDM focused on the three core **edm** classes implemented by Europeana in 2011. Some of the elements of the core classes were not included in the initial implementation by Europeana but have since been added, e.g. in **edm:WebResource**. Europeana has also implemented the contextual classes during 2011-12; **edm:Place** class was added to the CARARE to EDM mapping and transformation in summer 2012.

2.3.1. Use of the main EDM classes and Identifier creation

The CARARE mapping 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

- digital representations which provide other drawings and sources of information about the heritage asset (historic photographs, publications, archive materials etc.)

- Born-digital resources related to these objects, such as 3D models.

EDM requires a unique identifier for each class. Since unique identifiers could not be guaranteed in the original data, CARARE creates new identifiers or each 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 **WebResources** (digital representations of the provided objects made available on the Web). CARARE creates unique identifiers for each **ore:Aggregation resource**.

These aggregation identifiers are web-enabled, in the sense that they redirect to a landing page that CARARE creates 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.

---

[9] The paragraph is based on "*EDM Case Study: CARARE and EDM*"
http://pro.europeana.eu/documents/900548/8303c3d2-10b8-4f36-9db5-6a3dd06bd400

CARARE object
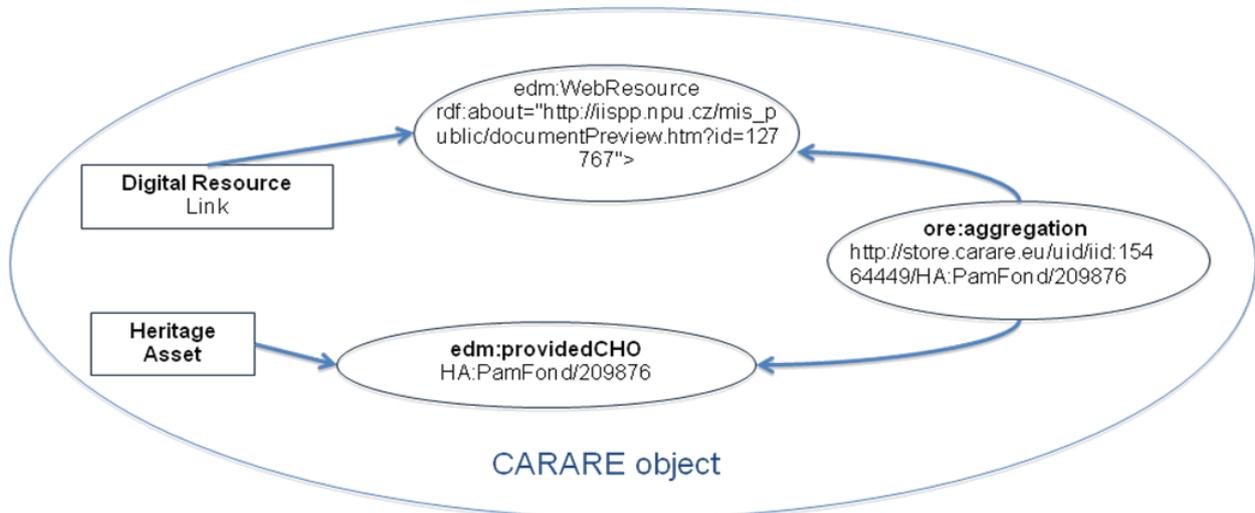
2.3.2. Representing archaeological objects and the cultural objects related to them

In terms of EDM representation, the different information sources presented on the landing page of an archaeological asset give rise to different **ProvidedCHOs**. Each print, map, or book about an archaeological place counts as a separate object provided to Europeana. These new identifiers assigned to each object make it possible to create an explicit link between the heritage asset and a related object. For instance, the relation between a heritage asset and a document describing it is expressed using the property **Edm:isRepresentationOf** with the URIs of the corresponding resources.
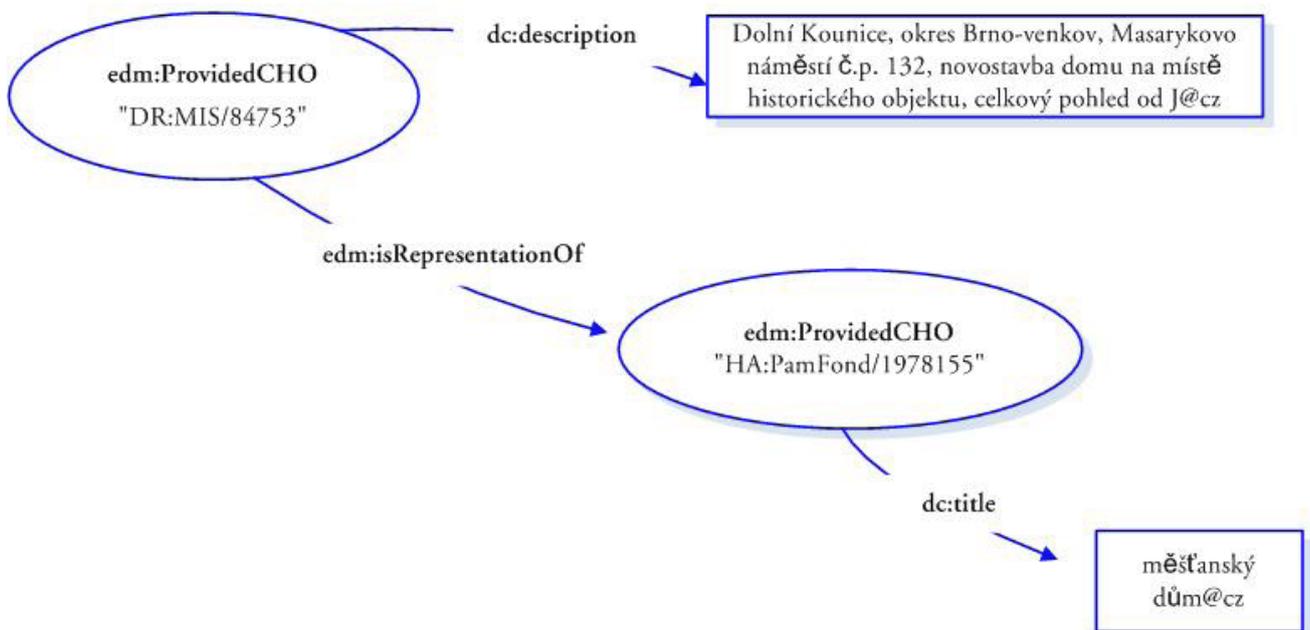
2.3.3. Data conversion decisions

The datasets aggregated by CARARE present a lot of differences in terms of data granularity. When converting this data to EDM, CARARE has to make decisions that are specific to each dataset. Three scenarios have been defined:

- Scenario 1: a heritage asset such as a building is related to a series of cultural objects such as historic drawings, books and photographs in a CARARE object. The heritage asset and each of the cultural object give rise to **providedCHOs** if we want to provide metadata for each object.

- Scenario 2: a book, photograph or map relates to more than one heritage asset and are contained in more than one CARARE object. Each heritage asset creates a **providedCHO** and each cultural object creates a **providedCHO**. Any duplicate **providedCHO** created as a result of the book or picture being referenced by more than one heritage asset are removed.

- Scenario 3: a heritage asset is represented by a series of photographs which are published online with simple descriptive metadata. A **providedCHO** is created for the Heritage asset and this is aggregated together with **edm:WebResources** created for each of the photographs.

2.3.4. Richer Web Resources

For each cultural heritage object (a monument, a building or other real world object), CARARE provides relations to multiple digital resources. EDM allows these different resources to be connected to each other, and also permits detailed description of these resources. One of the requirements of EDM is the separation of information related to a cultural heritage object from the information describing the digital representation of this object. This distinction is particularly important when dealing with rights metadata: an object and its digital representations might have different, or even contradictory, rights statements which determine the conditions for re-use of the content. In addition to the rights information provided in the **ore:Aggregation**, CARARE provides rights information specific to these digital resources.

2.3.5. Representation of Place entities

The spatial dimension is an important concept in the archaeology and architecture heritage domain. EDM allows the representation and the description of entities such as places, by a specific class: **edm:Place**. CARARE uses the **edm:Place** class to describe information related to a specific place separately from the **ProvidedCHO**. The relation between the **ProvidedCHO** and its related lace is created via the property **dcterms:spatial**. Each place has spatial coordinates and a label, sometimes provided in different languages. The inclusion of spatial coordinates in the data enables archaeology and architecture data sets to be included in geo-portals alongside other datasets used in planning, development control, tourism and other map-based services. In the Europeana context, the spatial information is used to enrich other datasets and to provide new features such s map browsing.

2.3.6. Representation of Agent entities

CARARE has completed a mapping to the **edm:Agent** class, but this has not yet been implemented in the CARARE to EDM transformation. In the current implementation of **edm:Agents** are related to **providedCHOs** through **dc:contributor**; **dc:creator** or **dc:publisher**.

2.3.7. Representation of Concept entities

In the current implementation of EDM, Concepts may be linked to **providedCHOs** through **dc:subject**. CARARE has completed a mapping to the **skos:Concept** class, although this has not yet been implemented in the CARARE to EDM transformation. This is mostly because the project has not identified any commonly used standardized vocabularies.

Representation of Place entities

Representation of Agent entities

Representation of Concept entities
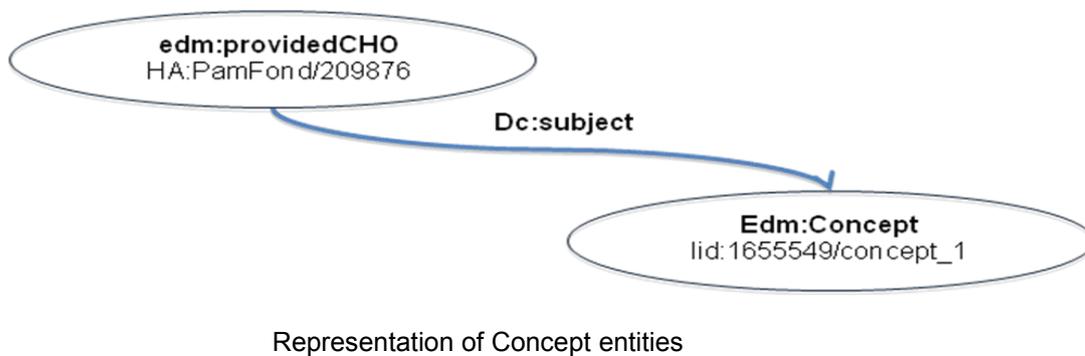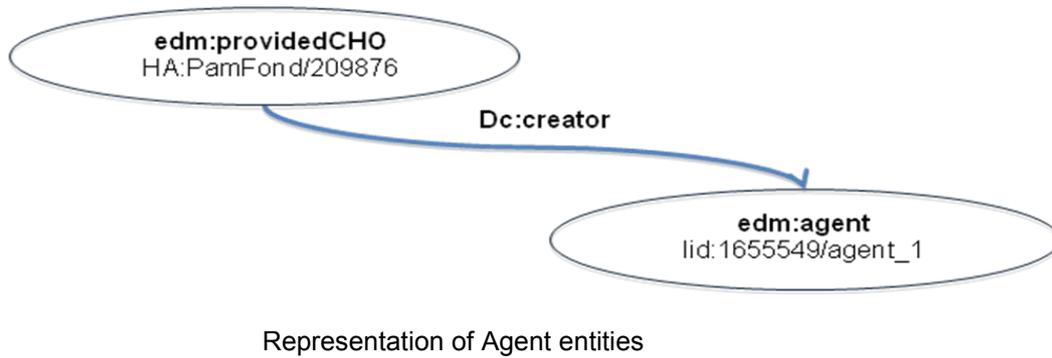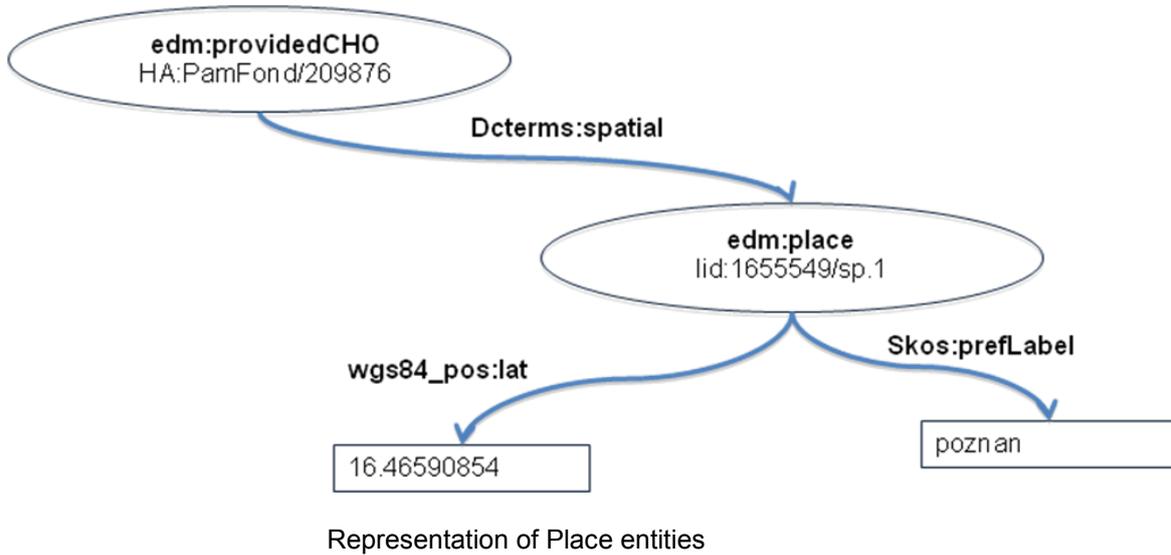
The repository of CARARE[10] provided by the partners. It is possible to retrieve records encoded according the CARARE schema or in EDM. The next figure shows the on-line tool, MINT, to search and visualized data.

The next figures shows partially the record (HA UAEMAT-17 from CETI) mapped on EDM schema.

```
1    <rdf:RDF>
2      <edm:ProvidedCHO rdf:about="HA:UAEMAT-17">
3        <dc:date>1925-01-01</dc:date>
4        <dc:date xml:lang="en">Inter-War Period (1922-1940)</dc:date>
5        <dc:date xml:lang="el">Περίοδος Μεσοπολέμου (1922-1940)</dc:date>
6        <dc:description xml:lang="en">The tobacco warehouse of Hermann Spierer is
         located in the district of Santa Barbara. It was built by the architect
         Konrad
                                    ...........
32       <dc:format>text</dc:format>
33       <dc:identifier>17</dc:identifier>
34       <dc:language xml:lang="en">en</dc:language>
35       <dc:language xml:lang="el">el</dc:language>
36       <dc:publisher xml:lang="en">Cultural and Educational Technology Institute -
         Research Centre Athena</dc:publisher>
37       <dc:source xml:lang="en">Cultural and Educational Technology Institute -
         Research Centre Athena</dc:source>
38       <dc:subject xml:lang="en">Warehouse</dc:subject>
39       <dc:subject xml:lang="el">Αποθήκη Εμπορευμάτων</dc:subject>
40       <dc:subject xml:lang="en">architecture</dc:subject>
```

---

[10] http://store.carare.eu.

```
60   <edm:WebResource rdf:about="http://www.ceti.gr/carare/CETI_SBD_17_EN.pdf">
61     <dc:rights>Cultural and Educational Technology Institute - Research Centre
       Athena, Ινστιτούτο Πολιτιστικής και Εκπαιδευτικής Τεχνολογίας – Ερευνητικό
       Κέντρο Αθηνά</dc:rights>
62   </edm:WebResource>
63   <edm:Place rdf:about="iid:1598812/SP.1">
64     <wgs84_pos:lat>24.141987</wgs84_pos:lat>
65     <wgs84_pos:long>41.150610</wgs84_pos:long>
66     <skos:prefLabel xml:lang="en">Municipality of Drama</skos:prefLabel>
67     <skos:note>Tobacco Warehouse of Ioannis Anastasiadis, 6, Ayia Barbara's
       Street, Drama, 66100, Greece</skos:note>
68   </edm:Place>
69   <ore:Aggregation
     rdf:about="http://store.carare.eu/uid/iid:1598812/HA:UAEMAT-17">
70     <edm:aggregatedCHO rdf:resource="HA:UAEMAT-17"></edm:aggregatedCHO>
71     <edm:hasView
       rdf:resource="http://www.ceti.gr/carare/CETI_SBD_17_EN.pdf"></edm:hasView>
72     <edm:dataProvider>Cultural and Educational Technology Institute - Research
       Centre Athena</edm:dataProvider>
73     <edm:provider>CARARE</edm:provider>
```

The record is not compatible with the RDF schema requirements missing at the top of the namespace declaration.The mapping of CARARE datasets was implemented by the MINT tool. The next figure shows the procedure as far as mapping is concerned: a mapping template can be set up and afterwards user can chose to transform other datasets encoded on the basis of the same template. Through a simple XSLT file is easy to map CARARE data to EDM schema. Currently NTUA is testing a version 2 of MINT[11].

---

[11] http://oreo.image.ntua.gr:8080/mint2/Home.action

2.4 CRMdig[12]

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*"[13]. In the Cultural Heritage domain, scientific data is based on results from observation, in particular on measurements by devices creating digital output. The systematic large-scale production of digital scientific objects, the diversity of the processes involved and the complexity of describing historical relationships among them, requires the need for an innovative knowledge management system. In order to capture and handle all the semantic information linked to the monitoring, management and documentation of the origins and derivation of digital resources, CRMdig has recently been implemented as an extension of the event-model CIDOC-CRM[14] ontology.

We have chosen CIDOC-CRM as the core conceptual schema because it describes the material past as it is observed or documented by archaeologists, historians and museum experts of all disciplines. It is an event-centric core model, implemented, among other forms, in RDFS and OWL. Instances of the CIDOC-CRM model can be merged to huge meaningful networks of knowledge about historical facts and contextual relationships. The use of the CRM and extensions of it enables an easy integration of provenance data with descriptions of the observed reality and integrated reasoning.

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

Recently in the context of the European Project 3D-COFORM[15], addressed to a large-scale production of 3D objects for scientific and cultural use, the CRMdig schema has been tested, revisited and implemented in RDFS[16]. The model has been extended by specializations and used in a Repository Infrastructure designed as a workspace for massive 3D model production, built on a SESAME triple store for the metadata and a distributed data object repository.

The basic classes of CRMdig comprise a hierarchy of event classes and a hierarchy of digital things. The Digital Machine Event is a very generic notion essential for e-science. One considers that any such event happens on behalf of a human Actor responsible for it, and that its products belong to this Actor. It is further specialized into Digital Measurement Event and Formal Derivation.

CRMdig is able to model the physical circumstances of scientific observation resulting in digital data. The ontology is particularly appropriate to describe typical workflows (acquisition, processing, synthesis, presentation) creating a complex semantic network of relationships and to support complex queries which can be resolved by following deep data paths of direct or inferred relationships in the semantic network. Depending on the quality of the required reasoning, more specializations may be introduced. It contains the constructs at the level of OPM and other models, and even more, since it is integrated in ISO21127, which allows for connecting the Provenance view with other parts of reality.

---

[12] The paragraph is based on 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://www.usenix.org/events/tapp11.

[13] W3C Provenance Incubator Group. http://www.w3.org/ 2005/Incubator/prov/wiki/.

[14] http://www.cidoc-crm.org. CIDOC-CRM is a standard ISO 21127:2006.

[15] 3D-COFORM is an European Project funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 231809. http://www.3d-coform.eu.

[16] CRMdig 2.5. http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs.

By considering the objectives of 3D-ICONS we chose CRMdig because it allows a simple and clear description of the processes carried out to digitize and render a 3D model. Other initiatives, mainly DCMI's Provenance Task Group[17], use the provenance element to describe the metadata creation process.
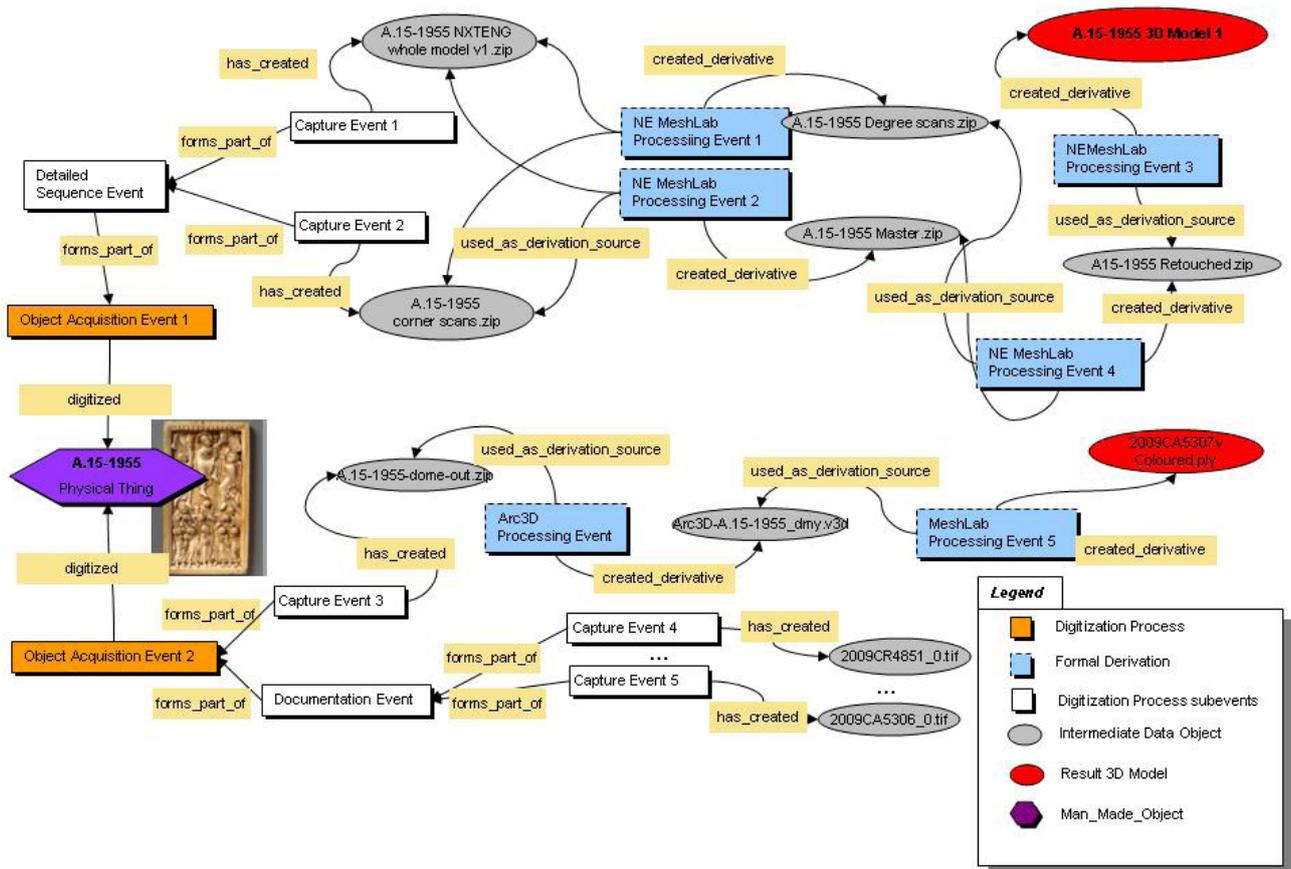
CRDdig Schema. The digitations and processing of a cultural object shown in Fig. 4 (see note 10)

---

[17] http://dublincore.org/groups/provenance/

2.5 Paradata[18]

Numerous scientific initiatives have underlined the importance of ensuring both that computer-based visualisation methods are applied with scholarly rigour, and that the outcomes of research that include computer-based visualisation should accurately convey to users the status of the knowledge that they represent, such as distinctions between evidence and hypothesis, and between different levels of probability.
A set of principles is therefore needed to ensure that digital heritage visualisation is, and is seen to be, at least as intellectually and technically rigorous as longer established cultural heritage research and communication methods. Such principles, reflecting the distinctive properties of computer-based visualisation technologies and methods, must to be made explicit like any other information concerning the cultural object. This approach allows the description of the features of the physical object and its digital replicas also as far as the digitization process is concerned.

While the Provenance deals with "*a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource",* Paradata have been defined as information about human processes of understanding and interpretation of data objects. Examples of paradata include descriptions stored within a structured dataset of how evidence was used to interpret an artefact, or a comment on methodological premises within a research publication. It is closely related, but somewhat different in emphasis, to "contextual metadata", which tend to communicate interpretations of an artefact or collection, rather than the process through which one or more artefacts were processed or interpreted.

The London Charter[19] is an international initiative seeking to establish principles for the use of computer-based visualisation methods and outcomes in the research and communication of cultural heritage. The Charter defines principles for the use of computer-based visualisation methods in relation to intellectual integrity, reliability, documentation, sustainability and access, particularly in order to:

- Promote intellectual and technical rigour in digital heritage visualisation,

- Ensure that computer-based visualisation processes and outcomes can be properly understood and evaluated by users.

In doing so, the Charter aims to enhance the rigour with which computer-based visualisation methods and outcomes are used and evaluated in heritage contexts, thereby promoting understanding and recognition of such methods and outcomes. The Charter does not seek to prescribe specific aims or methods, but rather establishes those broad principles for the use, in research and communication of cultural heritage, of computer-based visualisation upon which the intellectual integrity of such methods and outcomes depend.

The London Charter is based on 6 principles:

- **Implementation**
  The principles of the London Charter are valid wherever computer-based visualisation is applied to the research or dissemination of cultural heritage.

- **Aims and Methods**

---

[18] The paragraph is based on 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
[19] 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/.

A computer-based visualisation method should normally be used only when it is the most appropriate available method for that purpose.

- **Research Sources**
  In order to ensure the intellectual integrity of computer-based visualisation methods and outcomes, relevant research sources should be identified and evaluated in a structured and documented way.

- **Documentation**
  Sufficient information should be documented and disseminated to allow computer-based visualisation methods and outcomes to be understood and evaluated in relation to the contexts and purposes for which they are deployed.

- **Sustainability**
  Strategies should be planned and implemented to ensure the long-term sustainability of cultural heritage-related computer-based visualisation outcomes and documentation, in order to avoid loss of this growing part of human intellectual, social, economic and cultural heritage.

- **Access**
  The creation and dissemination of computer-based visualisation should be planned in such a way as to ensure that maximum possible benefits are achieved for the study, understanding, interpretation, preservation and management of cultural heritage.

For the objective of 3D-ICONS, particularly relevant is the principle 4.6 on the Documentation Process (Paradata). The Charter establishes that *"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"*.

3. The CARARE Schema Version 2.0

The scope of 3D-ICONS project is twofold:

- Making available 3D contents for Europeana;

- Developing a metadata schema able to capture all the semantic present in the digitations process (*provenance*) and in understanding and interpretation of data objects (*paradata*).

While as far as the data-acquisition, data-capture and publication of 3D cultural object is concerned, the existing technologies allow the former scope to be easily adderrssed, the latter, being innovative also for Europeana, has to be addressed with particular attention.

3D-ICONS has identified a need to capture metadata relating to the digitization processes used and the provenance of 3D models to make it clear to end-users how and why a particular reconstruction was produced.

One of the tasks of 3D-ICONS was therefore to update the CARARE schema by adding classes or entities and properties that would have made the original schema compliant with the 3D-ICONS requirements. Besides, as CARARE works like an intermediate schema between existing European standard and EDM, another step was to check the compatibility between CARARE and last advances in EDM. In the introduction both issues have been addressed and finally the metadata required has been specified as an extension to the CIDOC-CRM called CRMdig.

Thanks to the recent developments of integration between CARARE and EDM[20] and to the publication of object templates of EDM[21], the updating of CARARE results in simplification. The last OWL version of EDM[22] has been aligned to CIDOC-CRM Core Classes and some properties of CIDOC-CRM has reused in EDM allowing a more simple integration of CRMdig into EDM.

The areas of interests of provenance and paradata are innovative for the CARARE schema, but it is possible to add this information without changing substantially the original schema and the mapping to EDM.

In CARARE, the Information Resources addressed to hold metadata about an event is an **Activity**. In this description of different digitization processes it is important to distinguish the different types of events. For this reason we propose the extension of the definition of the Event Type attribute for Activity. In Event Type is possible to record the general purpose and specific purpose of the event as well.

The change of definition proposed for this theme are:

- **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.

---

[20] See note 9

[21] http://europeanalabs.eu/wiki/EDMObjectTemplatesProviders.

[22] http://europeanalabs.eu/svn/europeana/trunk/corelib/corelib-solr-definitions/src/main/resources/eu/rdf/EDM-v524-120820.owl (December 2012)

Other information, that should be provided according to the principles of the London Charter, has the scope to define aims and objectives of the 3D data-capture. The change proposed for this theme consists of two new attributes for the theme Activity:

- **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.

To register digital machines and software used for the digitization process and formal derivation we use in CARARE the existing element "Material and Techniques Used" included in the theme Activity. The CARARE schema considers this element as a simple string and recommends the use of a controlled vocabulary. As in CARARE there is no distinction between digital machines and software (concerning different processes) we propose a particular mapping to EDM to avoid possible misunderstandings (c.f. section 4). Each technique and/or material has to be compulsory described in separate field: for example "material and techiniques=laserscanner Faro Cam2"; and "material and techiniques=Meshlab". If we register "material and techiniques=laserscanner Faro Cam2; Meshlab" it is not possible to have a correct mapping with EDM. So we have to assign a specific term for each activity (digitization or processing) that the object has taken part in. Consequently, the new definition of the two elements are:

- **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.

In order to define the digitization process adopted to create the 3D final model we propose to add some relation addressed to explain the relations between the Heritage Asset, Digital Resource and the Activity related to the digitization processes. The properties reuse those from CIDOC-CRM and in particular from CRMdig.

The relations proposed 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 = CRMdig) 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 = CRMdig) 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.
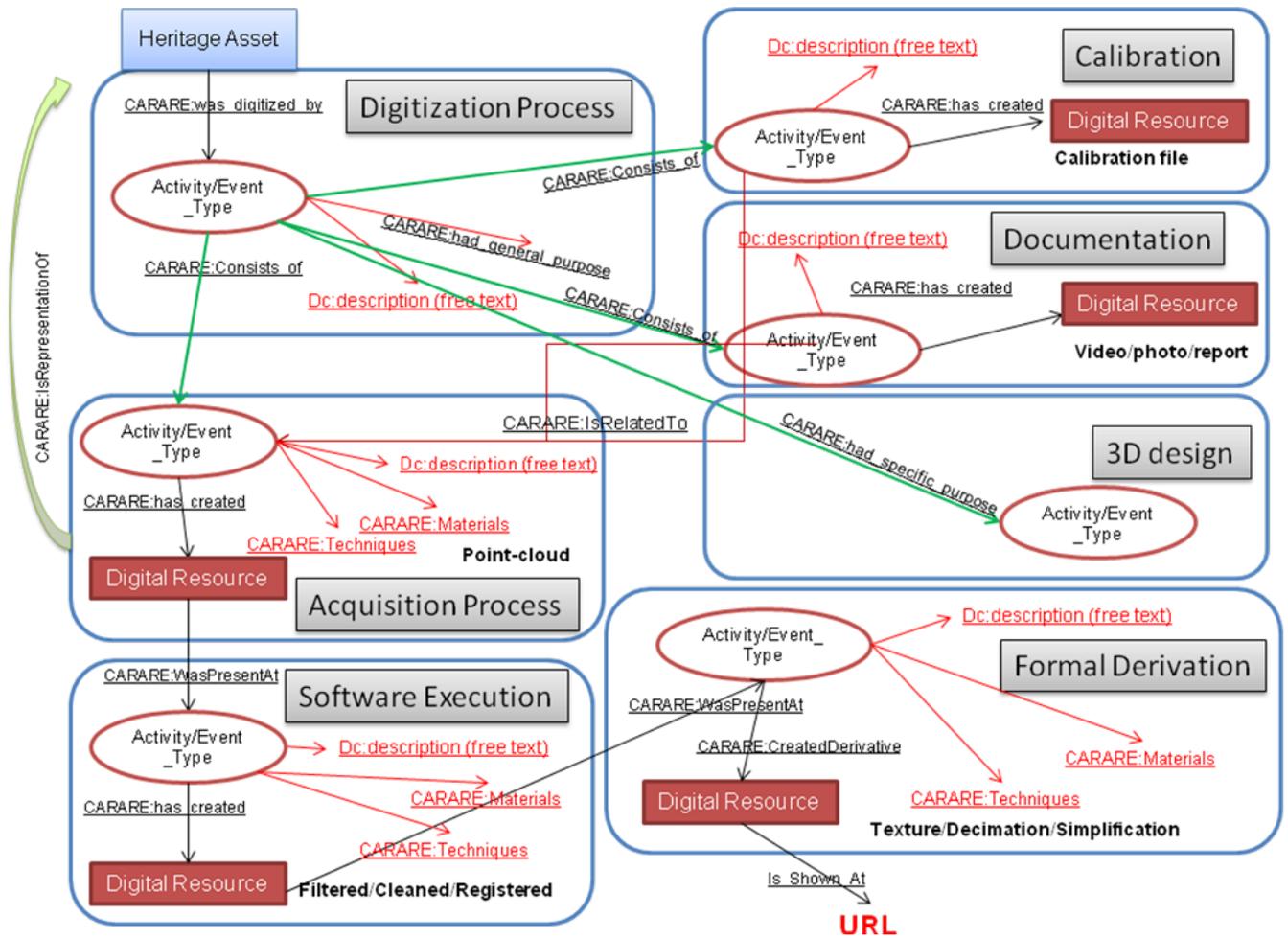
The 3D digitization workflow represented in the CARARE 2.0 Schema

4. CARARE Templates and Mapping to EDM

In order to use EDM schema, it has been necessary to create an update of the mapping from CARARE to EDM. For this scope the relation introduced of the new version of CARARE have been mapped on EDM schema.

- **Was_digitized_by** is a specialization of edm:wasPresentAt and associates the class ProvidedCHO with Event:

CARARE	HeritageAsset	was_digitized_by	Activity
EDM	edm:ProvidedCHO	was_digitized_by	edm:Event

- **Has_created** is a specialization of edm:wasPresentAt and associates an Event with WebResource:

CARARE	Activity	has_created	DigitalResource
EDM	edm:Event	has_created	edm:WebResource

- **Consists_of** is a specialization of dcterms:hasPart and defines a relation between two Events:

CARARE	Activity	consists_of	Activity
EDM	edm:Event	consists_of	edm:Event

- **Had_general_purpose** is a specialization of dc:description

CARARE	Activity	had_general_purpose	literal
EDM	edm:Event	had_general_purpose	Literal / reference

- **Had_specific_purpose** is specialization of edm:isRelatedTo and links two Events

CARARE	Activity	had_specific_purpose	Activity
EDM	edm:Event	had_specific_purpose	edm:Event

- **Created_derivative** associates an Event with a Web Resource

CARARE	Activity	created_derivative	DigitalResource
EDM	edm:Event	created_derivative	edm:WebResource

In order to capture and record the devices and the software utilized for the digitisation process it needs extend the proposed schema using other specific elements from CRMdig. In particular we propose other two relations both as specialization of Edm:was_Present_At:

- **Used_software_or_firmware** - associates an Event with an Information Resource[23]

CARARE	Activity	• Materials	Literal
EDM	edm:Event	• Used_software_or_firmware	Edm:InformationResource

- **Happened_on_device** - associates an Event with a Non Information Resource

CARARE	Activity	• Materials	Literal
EDM	edm:Event	• Happened_on_device	Edm:NonInformationResource

Furthermore to define digital machines and software we can use the following concepts (source CRMdig) mapped on the corresponding Edm:entities:

DigitalDevice = Edm:NonInformationResource (= CARARE:Materials)

Software = Edm:InformationResource (= CARARE:Materials)

In order to map correctly CARARE:Materials we propose the following schema:

If the value in CARARE for Activity/Event_type is = DigitizationProcess then the value for "materials" in CARARE corresponds to an Edm:NonInformationResource = DigitalDevice and the relation linking the event to the NonInformationResource is edm:WasPresentAt = Happened_on_device

If the value of Activity/Event_type is = SoftwareExecution or Formal Derivation then the value of materials is an Edm:InformationResource = Software and the relation linking the event to the edm:InformationResource is edm:WasPresentAt = Used_software_or_firmware

The proposed relations are presented in formal way.

Element name	**L1B.was_digitized_by**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#L1B.was_digitized_by
Label	Was Digitized By
Definition	This property associates an resource of ProvidedCHO with an Event related to digitization process.
Subproperty of	edm:wasPresentAt
Domain	edm:ProvidedCHO
Range	edm:Event
Obligation&Occurrence	
Example	The painting was digitized by digitization process

---

[23] For *Information resource* and *non information resources* elements, that are not still implemented in the Object_Template, see the *EDM Data Model elements, Version 5.2.3 (24/02/2012) Europeana v 1.0*

Element name	**L20F.has_created**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#L20F.has_created
Label	Has Created
Definition	This property describes the creation of an information resources by means of an acquisition, calibration, documentation process and software execution.
Subproperty of	edm:wasPresentAt
Domain	edm:Event
Range	edm:InformationResource
Obligation&Occurrence	
Example	Digital data obtained by different measurement machines.

Element name	**P9F.consists_of**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#P9F.consists_of
Label	Consists Of
Definition	This property captures the relation between two events and it specifies the decomposition of a resource of event in subsidiary events. Each sub-event can be related to another one with the EDM relation edm:isRelatedTo.
Subproperty of	dcterms:hasPart
Domain	edm:Event
Range	edm:Event
Obligation&Occurrence	
Example	The digitization process consists of an acquisition, documentation and calibration event.

Element name	**P21F.had_general_purpose**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#P21F.had_general_purpose
Label	Had General Purpose
Definition	This property describes an intentional relation between an resource of Event and its goal or purpose. This property allows to describe Paradata in a general meaning.
Subproperty of	dc:description
Domain	edm:Event
Range	
Obligation&Occurrence	
Example	

Element name	**P20F.had_specific_purpose**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#P20F.had_specific_purpose
Label	Had Specific Purpose
Definition	This property captures the relation between an resource of an Event and an resource of different Event related to activities or other actions, taken in preparation for the digitization

	process.
Subproperty of	edm:isRelatedTo
Domain	edm:Event
Range	edm:Event
Obligation&Occurrence	
Example	

Element name	**L22F.created_derivative**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#L22F.created_derivative
Label	Created Derivative
Definition	This property describes the derivation of a digital object in another one through decimation, simplification or other similar processing.
Subproperty of	edm:isDerivativeOf
Domain	edm:Event
Range	edm:WebResource
Obligation&Occurrence	
Example	A of an image produces the same image but with different formal aspects.

Element name	**L23F.used_software_or_firmware**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#L23F.used_software_or_firmware
Label	Used Software Or firmware
Definition	This property describes software used during the digitization process
Subproperty of	edm:wasPresentAt
Domain	edm:Event
Range	edm:InformationResource
Obligation&Occurrence	
Example	

Element name	**L12F.happened_on_device**
Namespace	CIDOC-CRMDig
URI	http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs#L12F.happened_on_device
Label	Happened On Device
Definition	This property describes the relation between a digitization event and material items such as computer, scanner, camera etc.used to process
Subproperty of	edm:wasPresentAt
Domain	edm:Event
Range	edm:NonInformationResource
Obligation&Occurrence	
Example	

4.1 Minimal requirements

To assure a homogeneous classification of the heritage assets, activities and digital resources and to guarantee an high level of compatibility with the contextual scheme of EDM, the following elements have to be considered obligatory.

HERITAGE ASSET
It corresponds to a metadata for an **Edm:ProvidedCHO: value = Dc:Identifier**

DublicCore	Carare	EDM	
dc:identifier	car:id	dc:identifier	Museum inventory or reference = Edm:ProvidedCHO
dc:title	car:title	dc:title	Literal
dc:subject	car:subject	dc:subject	Literal or URI Thesauros
dc:description	car:description	dc:description	Free text (DBPedia, Wikipedia)
dc:date	car:date	dc:date	Literal
dc:rights	car:rights copyright	dc:rights	Literal or URI (VIAF, FOAF)
dcterms:temporal	car:temporal display date	dcterms:temporal	Literal
dcterms:spatial	car:spatial	dcterms:spatial	Literal or URI (Geonames)
dc:language	car:language	dc:language	Literal

Optional

dc:relation	Reference to another HA
car:collection	Literal or reference
dcterms:extent (size of the object)	Literal
dcterms:references	Literal
wgs84:long	Longitude coordinates
wgs84:lat	Latitude coordinates
wgs84:alt	Elevation value
dcterms:extent (radius)	Literal (only for the radius of the spherical container for the HA)

ACTIVITY
To be repeated for all events that the HeritageAsset has taken part in.
For each HeritageAsset we can have multiple activities. It corresponds more or less to an Edm:Event

DublicCore	Carare	EDM	
dc:Identifier		dc:identifier	Counter
	car:eventType	edm:HasType	Literal or reference to a concept (URI)
	car:spatial	edm:OccurredAt	Value => Edm:Place: Literal or URI (Geonames, etc.)
	car:temporal	edm:HappenedAt	Value => Edm:TimeSpan: Literal
	car:actor	edm:HasMet	Value => Edm:Agent: literal or URI (VIAF, FOAF, etc.)
car:Materials	Edm:NonInformationResource (digital devices) **Optional** if the activity is for "*software execution*" or "*formal derivation*")	Literal or references. URI for event_type = data acquisition, calibration, documentation (Digitization process)
car:Materials	Edm:InformationResource (Software)	Literal or references. URI for

	Optional only if the activity is for "*documentation*"	event_type = software execution, formal derivation (data processing)
car:isPartOf	dcterms:isPartOf	Reference to an event: Software execution and multiple formal derivations are part of **Data processing**
car:hasPart	dcterms:hasPart	Reference to an event: **Digitization process** has different parts (or events) acquisition process, calibration and documentation
car:description	dc:description	free text
dc:description	car:hadGeneralPurpose	Free text
	car:hadSpecificPurpose	edm:isRelatedTo = SKOS:concept
	car:hasCreated	edm:InformationResource = nome.file

Optional (to register the positioning of the digital device in relation to the physical object)

wgs84:long	Longitude coordinates
wgs84:lat	Latitude coordinates
wgs84:alt	Elevation value

WEB_RESOURCE
It corresponds to the final 3Dmodel delivered to Europeana. It is an Edm:IsShownBy (the model) and Edm:IsShownAt (the model + information).

DublicCore	Carare	EDM	
dc:identifier	car:id	Edm:WebResource	reference
dc:rights	car:rights	dc:rights	Literal
		edm:rights	Literal

ORE AGGREGATION
It corresponds to a reference or URI of the HA **(**Edm:ProvidedCHO)

DublicCore	Carare	EDM	
		edm:AggregatedCHO	Reference or URI of the HA (Edm:ProvidedCHO)
		edm:Provider	References URI. The name of the organisation that sends data to Europeana
		edm:DataProvider	References URI The name of the organisation who supplies data to Europeana indirectly via an aggregator
		edm:IsShownAt	This is a URL that will be active in Europeana Interface. It will lead users to the digital object displayed on the provider's web site in its full information context.
		edm:IsShownBy	This is a URL that will be active in Europeana Interface. It will lead users to the digital object displayed
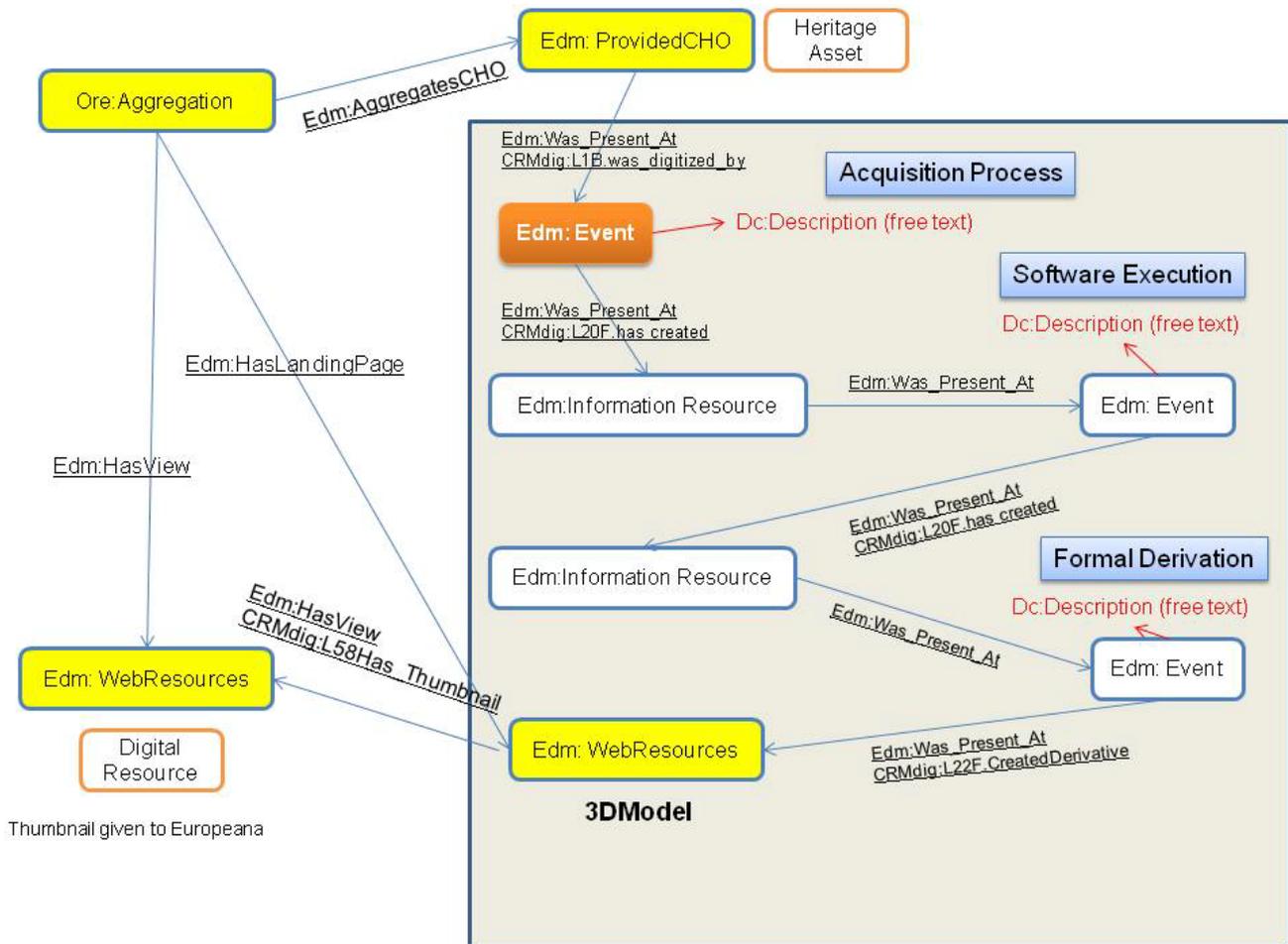
EDM Schema: From data-acquisition process to the final 3D model on the web
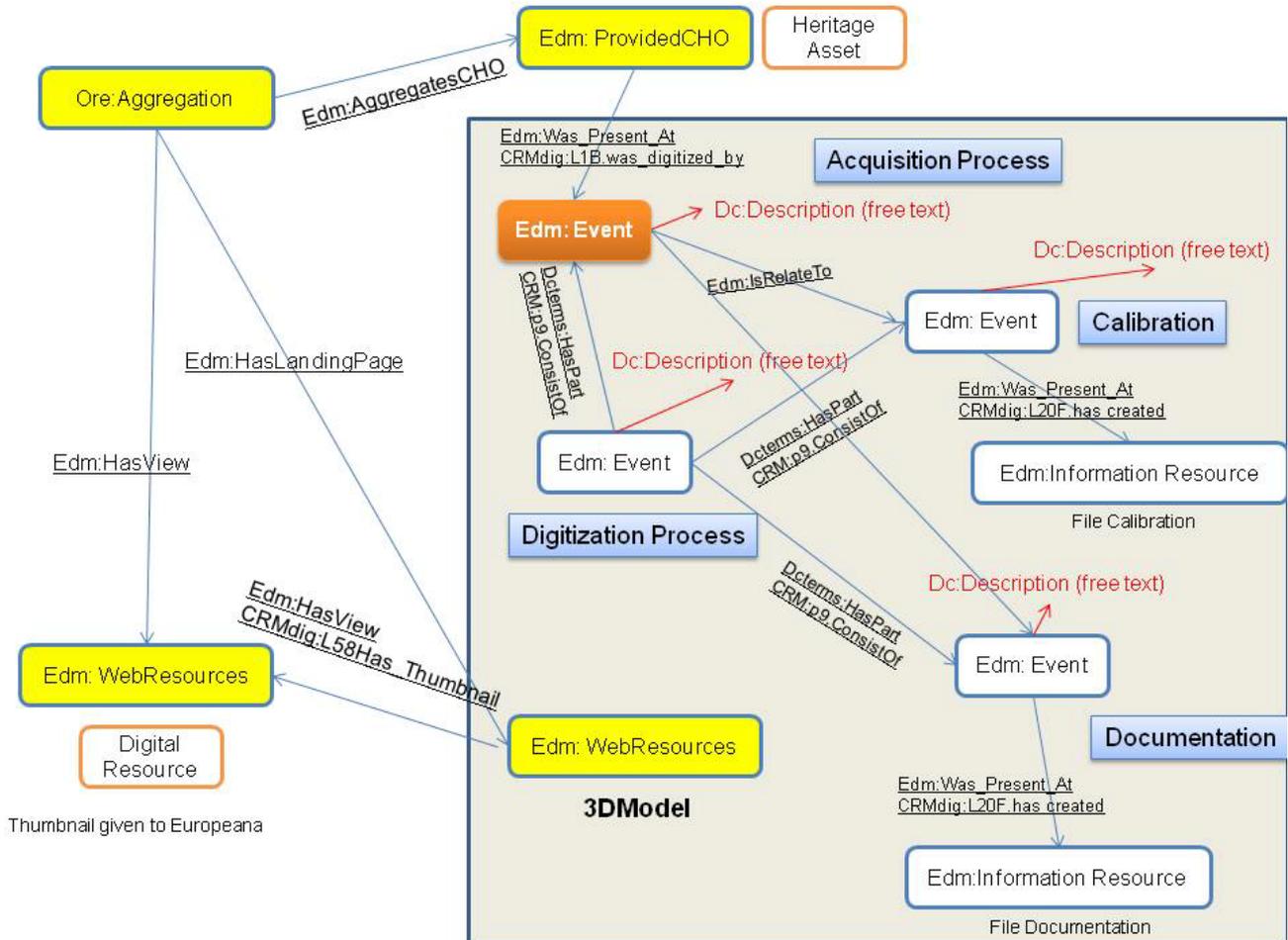
EDM Schema: Modelling of the digitization process
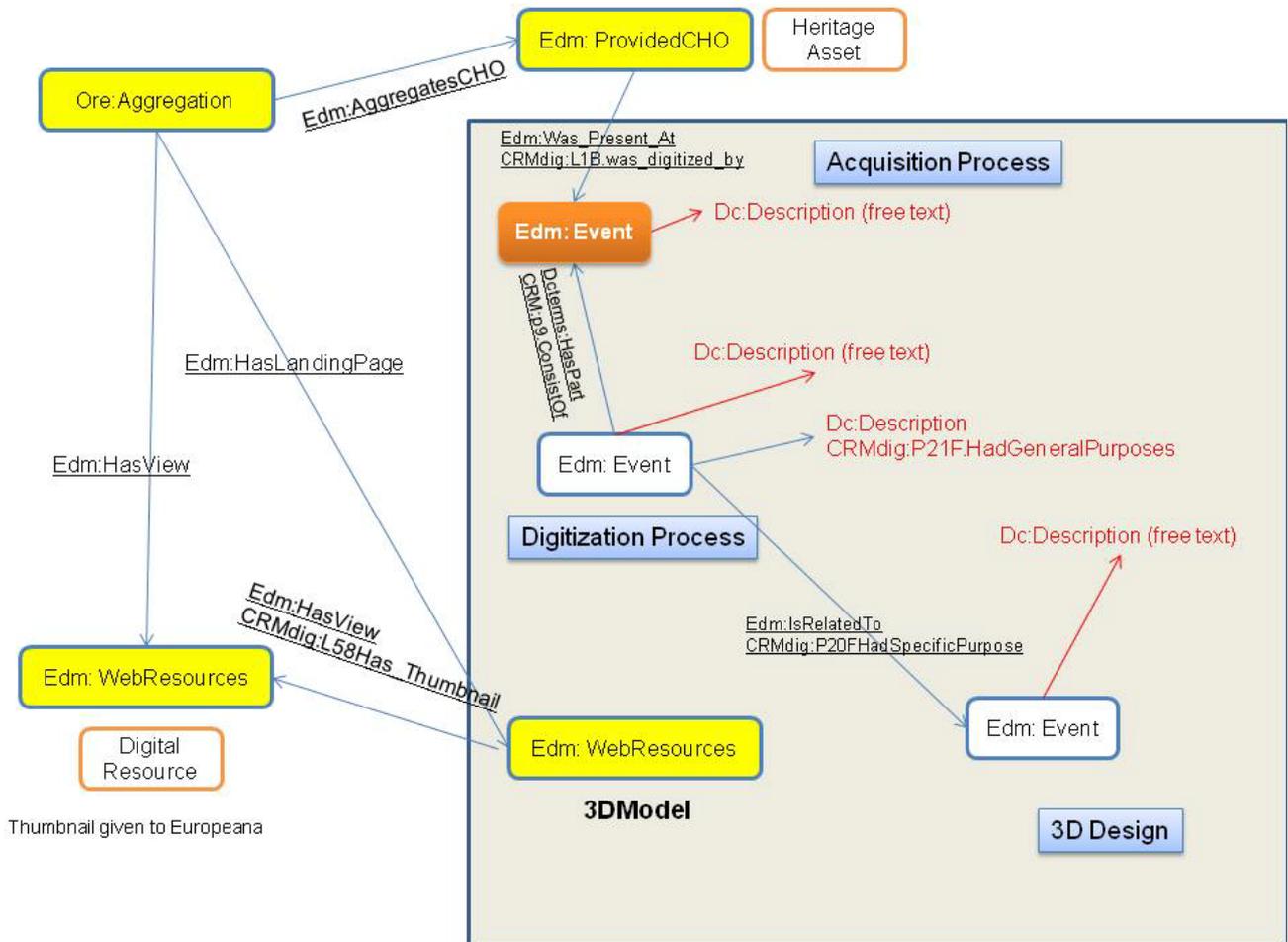
EDM Schema: Modelling of the paradata

EDM Schema: Modelling of the Digital Machine linked to Digitization Process

EDM Schema: Modelling of the use of Software in the Software Execution or Formal Derivation events.

5. Thesauri and Authority Lists[24]

Thesauri play an important role in information storage and retrieval systems. To assist different user communities in metadata compilation and information retrieval, libraries, museums and other CH institutions have created controlled vocabularies to describe and manage their assets. There are thesauri specific to fields, disciplines, institutions, and even collections. Many of these thesauri have been standardized using SKOS. We present a short survey of those used in some European CH communities giving a brief summary of the institutions responsible for the publication and the management.

5.1. The ICCD/CulturalItalia Portal

CulturaItalia (CI)[25] is a portal promoted and managed by the Ministry of Cultural Heritage and Activities (Italy). The CI repository hosts the digital resources from several Museums, Public and Private Cultural Institutions and part of the data from the Istituto Centrale per il Catalogo e la Documentazione (ICCD) catalogue. It contains 2 million records. The digital resources related to the national CH can be consulted through the "Metadata Index", which is an ordered set with a tree structure. The resources of the index are classified according to the thesaurus PICO 4.2 that is a Dublin Core qualified profile[26]. The metadata gathered in the CI repository have been mapped to the PICO metadata schema and recently PICO has been mapped on EDM. The thesaurus is a controlled vocabulary designed to manage and organize heterogeneous information coming from different cataloguing systems. The thesaurus is composed of keywords hierarchically structured according to the resource categories. It is *open* and can be further developed, in order to ensure a better organization of information. Moreover, it is codified in SKOS. The ICCD Thesaurus of Architecture has been recently developed but it is available only in a printed version. The framework of the thesaurus, hierarchically organized, combines a set of conventional vocabularies, referring to various types of buildings, of great value to the knowledge of Cultural Heritage and design practice. The thesaurus refers to the most popular residential architectural types and their most usual architectural and structural components, and is an open vocabulary.

5.2. The Getty Research Institute

The Getty Provenance Index Databases[27] provide instant access to data from primary sources relevant to the history of collecting. The index contains nearly one million records of western European works of art from the late 16th to the early 20th century, including:

- Archival Inventories: it contains 5,200 inventories and more than 260,000 individual records about artworks existing in a specific collection at the time the inventory was made.
- Sales Catalogues: it contains 8,400 catalogues and more than 740,000 individual records of works of art for sale and can be used to trace the sales and ownership history for a work of art.
- Dealer Stock Books: maintained by galleries and art dealers, the database contains 15 stock books and more than 43,700 individual records.
- Payments to Artists: it contains 1,000 recorded payments to artists made in Rome between 1576

---

[24] This paragraph is based mainly on: *Inventories and multilingualism issues: Multilingualism and Thesaurus*" edited by the MINERVA Plus Working Group 3, http://www.minervaeurope.org/structure/workinggroups/inventor/multilingua/documents/Multiligualismv1_printed.pdf (printed version); and P. Ronzino , N. Amico , F. Niccolucci, "*Federating specialized digital libraries*", Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage, October 18-21, 2011, Prato, Italy.

[25] http://www.culturaitalia.it/

[26] http://www.culturaitalia.it/pico/thesaurus/4.2/thesaurus_4.2.0.skos.xml

[27] http://www.getty.edu/research/tools/provenance/

and 1711.
  • Public Collections: descriptions and provenances of paintings held by public institutions.

The Getty Provenance Index Databases is based on the Art & Architecture Thesaurus (AAT)[28] that is a structured vocabulary containing around 125,000 terms about architecture, fine arts, decorative arts, archival materials and material culture. It ranges from Antiquity to the present and covers all kinds of art objects and architectural elements and their descriptions. The thesaurus may be consulted in two ways: hierarchically and the alphabetically. The layout of the hierarchical display is based on concepts arranged by facets from general to specific, or from abstract concepts to concrete artefacts. At the top of the AAT hierarchy is positioned the root and below this level the seven facets are located. Within the facets there are more levels organized in 36 hierarchies containing 2.949 guide terms and 27.992 concepts. The vocabulary, originally in XML, has been converted to SKOS.

5.3. English Heritage

PastScape[29] is an online resource from English Heritage's National Monuments Record (NMR)[30], providing information about many of England's ancient and historical sites, buildings and monuments. The information within PastScape is taken directly from the NMR's national historic environment database that contains nearly 400,000 records on the archaeology and buildings of England and its territorial waters. The records are referred to the Thesaurus of Monument Types[31]. Images of England is another English Heritage database containing a photographic record of England's 370,000 listed buildings[32]. All the lists of heritage assets are being unified into the National List for English Heritage[33].

English Heritage is continuously developing new thesauri to provide structure and guidelines for the standardization of terms to be used when creating new records of the past. Among the available thesauri, there is the Thesaurus of Monument Types, which is a standard produced by the Royal Commission on the Historical Monuments of England (RCHME) in 1995. The purpose of this thesaurus is to standardise the terms used to describe archaeological sites or standing buildings in the UK. The terms may be accessed hierarchically or alphabetically. This thesaurus is now maintained by English Heritage, and is available on-line. It has been converted to SKOS[34]. Other thesauri include:

  • Buildings material thesaurus: construction materials for monuments, relating to the built and buried Heritage.
  • Components thesaurus: elements of a monument relating to the built or buried Heritage.
  • Evidence thesaurus: terminology covering the existing physical remains of a monument, or the means by which a monument has been identified where no physical remains exist.

---

[28] http://www.getty.edu/research/tools/vocabularies/aat/

[29] http://www.pastscape.org.uk/

[30] http://www.english-heritage.org.uk/professional/archives-and-collections/nmr/

[31] http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=1

[32] http://www.imagesofengland.org.uk/

[33] http://www.english-heritage.org.uk/professional/protection/process/national-heritage-list-for-england/

[34] http://hypermedia.research.glam.ac.uk/resources/terminology/

5.4. The European Heritage Network (HEREIN)[35]

The database on heritage policies in Europe (HEREIN) provides a multilingual overview of the heritage policies pursued in European countries. The new version of the system HEREIN 3 focuses on an integrated approach between Archaeological Heritage and Landscape Heritage, which still in many countries fall under different policies. The multilingual thesaurus attached to the HEREIN project and database intends to offer a terminological standard for national policies dealing with architectural and archaeological heritage. This tool is intended to help the user of the website when surfing through the various on-line national reports, covering more than 500 terms in sixteen languages. The terms of the thesaurus are gathered into nine classes which represent nine top terms around which concept are organized. Each class forms a structured grouping of descriptors and non-descriptors around a top term.

5.5. Mérimée[36]

Mérimée is a database on monumental French heritage. It was created in 1978 and put online in 1995 by the Ministry of Culture of France, Direction for Architecture and Heritage. The Mérimée database contains the information from the service of historical monuments and *l'Inventaire général du patrimoine culturel*, addressing religious architecture, domestic, agricultural, educational, military and industrial. The base Mérimée includes about 200.000 records with 147.000 for inventory and 42.000 for historic buildings. The Mérimée thesaurus was developed by the Direction de *l'Architecture et de Patrimoine, Ministère de la Culture et de la Communication*, in parallel with the homonymous repository It is maintained by the *Inventaire général du patrimoine culturel, des Monuments historiques*, and by the *Médiathèque de l'Architecture et du Patrimoine*. It ranges from the prehistory until 1970. Each term of the thesaurus is classified according to functional categories related to religious use, funeral, industrial and so on. It also includes all necessary references, definitions, summaries, and usage notes. For each term there is the correspondence with the American and English equivalents used in the thesaurus of ATT and the RCHME. Part of the descriptors have also been translated into the Italian ICCD thesaurus.

5.6. Royal Museums of Art and History of Bruxelles

Since 2008 the Museum has created multiple thesauri with the scope to classify archaeology, art history, decorative arts, musical instruments objects from all continents. Currently there are four thesauri for object name, materials, techniques and geographical reference. They were created by a manager and they were imported manually. All of them can be consulted on line. The sources are Art and Architecture Thesaurus (Getty), British Museum Object Name Thesaurus, PACTOLS (Frantiq), AmMovE-thesaurus (Province East-Flanders), AICIM (Fédération Wallonie-Bruxelles) and specialised literature. The thesaurus can be exported in different formats (Excel/Access/Word/XML). The terms are in Dutch, French and English. The object name and the geographical references thesauri are compliant with the ISO standard 5964.

---

[35] http://www.coe.int/t/dg4/cultureheritage/heritage/herein/default_en.asp
[36] http://www.culture.gouv.fr/culture/inventai/patrimoine/

6. Linked Open Data (LOD)[37]

In the framework of the recent Europeana Linked Open Data Pilot (data.europeana.eu[38]) Europeana is making data openly available to the public and private sectors alike so they can use it to develop innovative applications and to create new web services and portals. LOD is a data publishing technique that uses common web technologies to connect related data and make it accessible on the Web. Linked Open Data implies that reuse restrictions have been removed from the metadata. Moving to such a model may mean that in future the portal is seen as the reference application of Europeana but that its main function is that of a rich data service that allows third parties to take the data freely and reuse it to create new knowledge and applications.

The data is represented in EDM and the described resources are addressable and dereferenceable by their URIs. Links between Europeana resources and other resources in the Linked Data Web enable the discovery of semantically related resources. Europeana Staff developed an approach that allows Europeana data providers to opt for their data to become Linked Data and converts their metadata to EDM, benefiting from Europeana efforts to link them to semantically related resources on the Web. With that approach and thanks to a first Linked Data version of Europeana, the resulting datasets have been published on the Web. All Europeana datasets can be explored and queried through an experimental SPARQL endpoint provided by Ontotext at europeana.ontotext.com.

Currently, Europeana does not harvest metadata in the EDM format. For this reason it is necessary to convert legacy ESE data into EDM. This implies creating resources for the main EDM classes, and distributing ESE metadata fields over these various resources The resulting data does not realize the full potential offered by EDM. Source data in Europeana is of varying degrees of richness and is all mapped to ESE, which is based on simple text string values. While achieving interoperability, this often entails losing some of the richness of the more detailed formats. In particular, it means any provider that has used contextual resources (authority files, thesauri, etc.) will have lost those relationships. Additionally, data.europeana.eu includes semantic connections to external (linked data) sources.

Europeana serve links to other linked data services, already maintained by Europeana providers. Currently this only includes the Swedish cultural heritage aggregator (SOCH). But the vast majority of external links comes from *semantic enrichment* realized at the Europeana Office, connecting Europeana items to places (as provided by *GeoNames*), concepts (from the *GEMET* thesaurus), persons (from *DBpedia*) and time periods (from an *adhoc* time period vocabulary).

Support for Open Data innovation is at the root of Europeana's new Data Exchange Agreement, the contract that libraries, museums, and archives agree to when their metadata goes into Europeana. The Data Exchange Agreement has been signed by all the national libraries, by leading national museums and by many of the content providers for entire countries, such as Sweden's National Heritage Board. The new Data Exchange Agreement dedicates the metadata to the Public Domain and came into effect from 1 July 2012, after which all metadata in Europeana are available as Open Data.

The following explains in more detail the data that can be found for every class of resources served by data.europeana.eu:

---

[37] The paragraph is based on: the web pages at
Haslhofer B., Isaac A*., "data.europeana.eu - The Europeana Linked Open Data Pilot"*, DCMI Global Meetings & Conferences,DC-2011, The Hague, http://dcevents.dublincore.org/IntConf/dc-2011/paper/view/55/14; and 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-etal_Europeana_isqv24no2-3.pdf.
[38] http://pro.europeana.eu/linked-open-data

Item (Provided Cultural Heritage Object): Item resources represent objects (painting, book, etc.) for which institutions provide representations to be accessed through Europeana. ProvidedCHO is the main entry points in data.europeana.eu. A Provided CHO is the hub of the network of relevant resources. When applicable, the Europeana URIs for these objects also link, via owl:sameAs statements, to other linked data resources about the same object. In this Linked Data pilot, no descriptive metadata (creator, subject, etc.) is directly attached to item URIs. It is instead attached to the *proxies* that represent a view of the object, from a specific institution's perspective (either a Europeana provider or Europeana itself). Europeana in future could change this approach and duplicate all the descriptive metadata at the level of the item URI. Such an option is costly in terms of data verbosity, but it would enable easier access to metadata, for data consumers less concerned about provenance.

Provider's proxy: These resources (are used as subjects of descriptive statements (creator, subject, date of creation, etc.) for the item, which are contributed by a Europeana provider. In the OAI-ORE model, proxies enable the separation of different views for a same resource, in the context of different aggregations. This allows us to distinguish the original metadata for the object from the metadata that is created by Europeana, an important requirement for us. Descriptive properties that apply to these proxies mostly come from Dublin Core. Proxies are connected to the item they represent a facet of, using the ore:proxyFor property. They are attached to the aggregation that contextualises them, using ore:proxyIn.

Provider's aggregation: These resources provide data related to a Europeana provider's gathering of digitised representations and descriptive metadata for an item. They are related to digital resources about the item, be they files directly representing it (edm:object and edm:isShownBy) or web pages showing the object in context (edm:isShownAt). They may also provide controlled rights information applying to these resources (edm:rights). Other statements provided in the same ESE record as the descriptive metadata for the item – but that do not always clearly apply to it – are also attached to aggregations. Finally, provenance data is given in statements using edm:provider (the direct provider to Europeana in the data aggregation chain) or edm:dataProvider (the cultural institution that curates the object). The aggregation is connected to the item resource using the edm:aggregatedCHO property.

Europeana's proxy: The second type of proxies served at data.europeana.eu. Europeana proxies provide access to the metadata created by Europeana for a given item, distinct from the metadata provided by the provider. Here one can find edm:year statements, indicating a normalised date associated with the object. It also serve millions of generic edm:hasMet enrichments, created by Europeana from a range of ESE descriptive fields. These statements connect a Europeana proxy to places from GeoNames, concepts from the GEMET thesaurus, persons from DBpedia and periods from an adhoc time vocabulary. Finally, a proxy is connected to the item it represents a facet of, using the ore:proxyFor property, as well as to the aggregation that contextualises it, using ore:proxyIn.

Europeana's aggregation: a Europeana aggregation bundles together the result of all data creation and aggregation efforts for a given item. It aggregates the provider's aggregation (using ore:aggregates), which in turn will connect to the provider's proxy. Next to the provider aggregation, one can find the digitised resources Europeana.eu serves for the item, i.e., an object page (edm:landingPage) and a thumbnail (using a combination of edm:hasView and foaf:thumbnail). The Europeana proxy is also connected to this aggregation, as mentioned above.

Resource map: OAI-ORE Resource maps are constructs for indicating meta-level statements about the creation and publication of ORE data (ORE aggregations and their aggregated resources). Europeana is exploring their use as a contextualisation mechanism for the Europeana aggregation. Maps are connected to an item they are about using foaf:primaryTopic, and to its corresponding Europeana aggregation using ore:describes. They sum up the provenance of data using dc:creator and dc:contributor statements. Crucially, they also indicate, in a machine-readable way, that the (RDF) data served at data.europeana.eu is provided under the CC0 open license.

The following RDF namespace abbreviations are currently used in data.europeana.eu:

rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
owl:	http://www.w3.org/2002/07/owl#
dc:	http://purl.org/dc/elements/1.1
dcterms:	http://purl.org/dc/terms/
foaf:	http://xmlns.com/foaf/0.1/
edm:	http://www.europeana.eu/schemas/edm/
ese:	http://www.europeana.eu"
xhtml;	http://www.w3.org/1999/xhtml/vocab#"
cc:	http://creativecommons.org/ns#"
ore:	http://www.openarchives.org/ore/terms/
str:	http://exslt.org/strings

File dumps for 20 million objects from Europeana providers are downloadable in RDF at http://data.europeana.eu/download/2.0/. As far as 3D LOD Data are concerning, currently there are only 3 datasets containing in total 23 records provided by ECLAP (10), 3D-COFORM (11) and CYPRUS LIBRARY (2). The schema used for the publication of 3D content is not different from that adopted for the other typologies of objects.

The conversion of metadata in RDF and the following transformation in LOD is automatically managed by Europeana. Currently particular task is not required by the data providers. The transition from Europeana URIs to dereferencable HTTP URIs for EDM aggregations and proxies is the major challenge in the conversion process. The main Europeana production system and the Europeana Linked Open Data Prototype are still two distinct systems so a bridge is needed between the identification mechanisms in place.
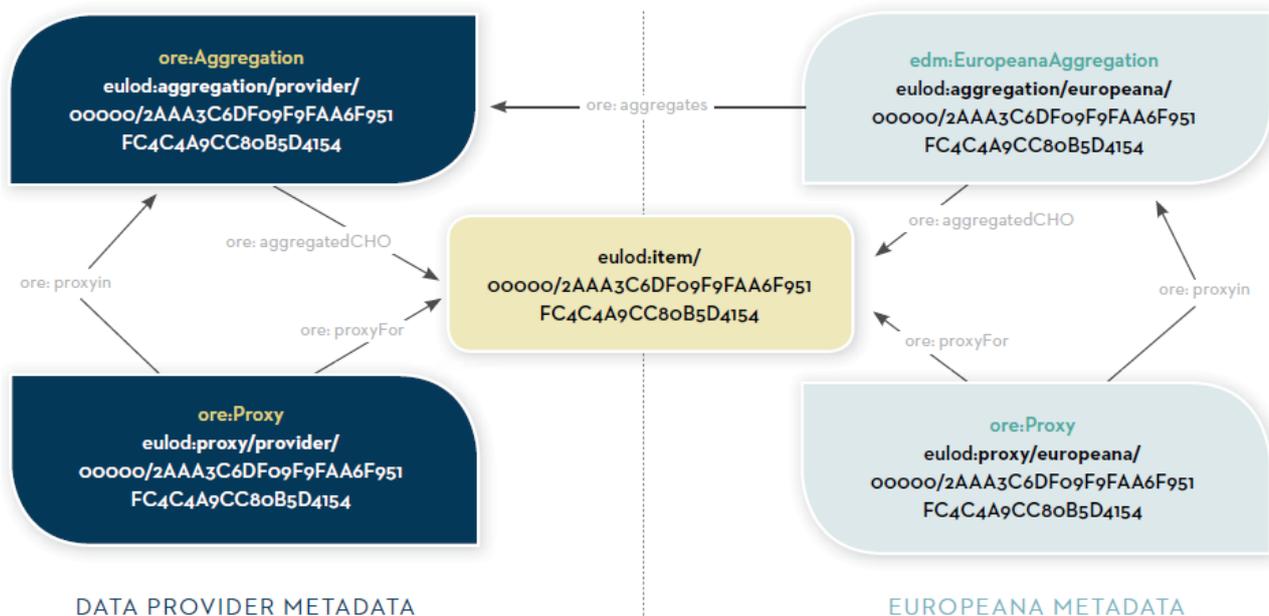
Basic structure of EDM networked resources

7. References (all web-sites have been visited on 30/01/2013)

3D-COFORM = http://www.3d-coform.eu.

CARARE = www.carare.eu.

CARARE REPOSITORY = http://store.carare.eu.

CARARE Metadata schema outline v.1.1" (November 2011)
http://www.carare.eu/eng/content/download/4277/35150/file/CARARE%20metadata%20schema%20outline%20v1.1.pdf

CIDOC-CRM = http://www.cidoc-crm.org.

CIDOC-CRM RDFS 5.0.4 = http://www.cidoc-crm.org/rdfs/cidoc_crm_v5.0.4_official_release.rdfs

CRMdig = 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://www.usenix.org/events/tapp11.

CRMdig RDFS 2.5 = http://www.ics.forth.gr/isl/rdfs/3D-COFORM_CRMdig.rdfs.

CULTURA ITALIA = http://www.culturaitalia.it/

CULTURA ITALIA PICO 4.2.0. = http://www.culturaitalia.it/pico/thesaurus/4.2/thesaurus_4.2.0.skos.xmlù

DCMI Provenance Task Group = http://dublincore.org/groups/provenance/

EDM-CARARE metadata mapping" (January 2011)
http://www.carare.eu/eng/Media/Files/D2.2.3_3.4-EDM-CARARE-metadata-mapping

EDM Case Study: CARARE and EDM
http://pro.europeana.eu/documents/900548/8303c3d2-10b8-4f36-9db5-6a3dd06bd400

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

EDM Linked Open Data = http://pro.europeana.eu/linked-open-data

> Haslhofer B., Isaac A*., "data.europeana.eu - The Europeana Linked Open Data Pilot*", DCMI Global Meetings & Conferences,DC-2011, The Hague, http://dcevents.dublincore.org/IntConf/dc-2011/paper/view/55/14;

> 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-etal_Europeana_isqv24no2-3.pdf.

EDM Mapping Guidelines v.1.0.1 (24/02/2012)
http://pro.europeana.eu/documents/900548/ea68f42d-32f6-4900-91e9-ef18006d652e

EDM Object Templates = http://europeanalabs.eu/wiki/EDMObjectTemplatesProviders.

EDM Primer (26/10/2011) Europeana v1.0

http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5

EDM RDFS 5.24-120820 = http://europeanalabs.eu/svn/europeana/trunk/corelib/corelib-solr-definitions/src/main/resources/eu/rdf/EDM-v524-120820.owl

ENGLISH HERITAGE THESAURUS = http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=1

ENGLISH HERITAGE THESAURUS ver SKOS =
http://hypermedia.research.glam.ac.uk/resources/terminology/

GETTY PROVENANCE INDEX DATABASES = http://www.getty.edu/research/tools/provenance/

GETTY ATT = http://www.getty.edu/research/tools/vocabularies/aat/

HEREIN = http://www.coe.int/t/dg4/cultureheritage/heritage/herein/default_en.asp

IMAGES OF ENGLAND = http://www.imagesofengland.org.uk/

LONDON CHARTER = http://www.londoncharter.org/.

NATIONAL MONUMENTS RECORD = http://www.english-heritage.org.uk/professional/archives-and-collections/nmr/

MÉRIMÉE = http://www.culture.gouv.fr/culture/inventai/patrimoine/

MULTILINGUALISM AND THESAURUS, edited by the MINERVA Plus Working Group 3
http://www.minervaeurope.org/structure/workinggroups/inventor/multilingua/documents/Multiligualismv1_printed.pdf (printed version);

PARADATA = 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

PASTSCAPE = http://www.pastscape.org.uk/

THE NATIONAL HERITAGE LIST FOR ENGLAND =
http://www.english-heritage.org.uk/professional/protection/process/national-heritage-list-for-england/

THESAURI = P. Ronzino , N. Amico , F. Niccolucci, "*Federating specialized digital libraries"*, Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage, October 18-21, 2011, Prato, Italy.

W3C PROVENANCE = http://www.w3.org/ 2005/Incubator/prov/wiki/.

8. Appendix

Version 2.0 of the CARARE Metadata Schema

The CARARE metadata schema, v.2.0

Kate Fernie, Dimitris Gavrilis and Stavros Angelis

With contributions from Andrea D'Andrea, Franc Zakrajsek, Paola Ronzino, Anestis Koutsoudis and Christodoulos Chamzas

Contents

Revision History

Rev.	Date	Author	Org.	Description
1.0		Christos Papatheodorou, Phil Carlisle, Christian Ertmann-Christiansen, Kate Fernie	CARARE	CARARE metadata schema Version1.0
2.0	21/2/13	K Fernie et al	MDR	Final draft of version 2.0 of the CARARE metadata schema

1. Introduction

This is a proposal for version 2.0 of the CARARE metadata schema following the experience gained as a result of aggregating metadata for Europeana from the 24 CARARE content providers and taking into account the requirements of the 3D ICONS project.

Version 1.0 of the CARARE schema built on existing standards (MIDAS Heritage, CIDOC CRM, LIDO etc) and was the work of the members of the CARARE metadata working group, the DCU metadata team and the English Heritage Data Standards Unit including: Christos Papatheodorou, Phil Carlisle, Christian Ertmann-Christiansen, Kate Fernie, Maria Emilia Masci, Oliver Mamo, Börje Justrell, Sven Ole Clemens, Vassilis Tzouvaras, Dimitris Gavrilis, Stavros Angelis, Constantia Kakali, Giannis Tsakonas, Panos Constantopoulos, Costis Dallas, Sólborg Una Pálsdóttir, Effie Patsatzi, Lena Inger Larsen, Daniel Pletinckx, Nasos Drosopoulos, Vykintas Vaitkevičius, Rimvydas Laužikas, Phil Carlisle, Gillian Grayson and Stephen Stead.

This version of the schema takes into account the experience gained from mapping more than 40 datasets from 20 different countries to version 1.0 of the CARARE metadata schema during the CARARE project, and of transforming CARARE metadata to EDM for supply to Europeana. This experience suggested areas where the schema might be simplified. The main changes proposed in version 2.0 are as follows:

- The scope of the Heritage Asset has been broadened to include printed materials, archives and born-digital objects relating to the archaeological and architectural heritage.
- Digital Resource has been simplified to focus on the type, format and location of the online resource.
- Heritage Asset becomes mandatory; there must be at least one in each CARARE object. It remains mandatory to include at least one digital resource in each CARARE object.
- The record information has been simplified.
- The rights statements have been simplified and metadata rights clarified.
- The references section of the Heritage Asset has been simplified.
- Provenance has been added to Heritage Asset.
- Spatial information has been updated.
- Elements for types of relations from heritage assets, digital resources and activities have been specified for clarity.

In addition to these changes, to meet the needs of the 3D ICONS project to enable information to be captured about the provenance of 3D objects. The following changes are included:

- Activities have been extended to allow for the recording of sub-ordinate events that take place during a larger campaign.
- New elements for types of relations have been added based on CRM-DIG.

2. Outline of the CARARE schema

It is important to note that this is a harvesting schema intended for delivering metadata to the CARARE service environment of an organisation's online collections, monument inventory database and digital objects. It does not support activities such as monument management and protection. The strength of the schema lies with its ability to support the full range of information about monuments, building, landscape areas, artefacts and born-digital objects, related events and their representations.

The schema is an application profile based on MIDAS Heritage and the CIDOC CRM. MIDAS Heritage is a detailed standard intended for the full documentation of all aspects of heritage management not all of which are relevant to the CARARE service environment. The CARARE schema's focus is on the detailed description of heritage assets, events in which the asset has been involved and digital resources which are available online and their provenance. It follows the structure of MIDAS Heritage enhanced by the expressiveness of LIDO and EDM.

2.1 Wrappers

Conceptually the CARARE record is focussed on a heritage asset and its relations to digital resources, activities and to collection information. The top-level elements in a CARARE record are:

CARARE Wrap – The CARARE schema wrap, it wraps one or many CARARE records.

CARARE – The CARARE start element. It wraps the Heritage Asset with the other information resources (Collection information, Digital Resource and Activity):

    Heritage asset – holds the metadata for a monument, building or cultural object including printed materials and born-digital objects, including descriptive and administrative metadata. One.

    Digital resource – holds the metadata about a digital resource including its online location. One or more.

    Activity – holds the metadata about an event or activity. Zero or more.

    Collection information – holds the collection-level description. Zero or one

Heritage assets are "first-class" citizens in the CARARE schema and it is mandatory for each CARARE record to include one heritage asset and at least one digital resource – in this way the schema provides for the description of cultural objects including historical images whose exact location is no longer certain and born-digital cultural objects.

We recommend the inclusion of collection information in CARARE records to provide the context for the collection.

The schema also allows for the option of providing information about Activities.

3. CARARE object record

3.1   CARARE

This is the start element for the CARARE object record. It contains a unique identifier for the record and wraps the four main themes:
- Collection information
- Heritage asset identification
- Digital resource
- Activity

3.2   Collection information

The following elements provide a collection level description of the resources being harvested

- ID – a unique identifier for the collection.
- Title – the title of the collection. The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred.
- Source – organisation that is the source of the collection
- Contacts – for the collection
- Rights – associated with the collection as a whole
- Language – of the metadata. Specifies the default language of the records in the collection; deviations in particular records are specified in the record metadata, and deviations in particular elements are specified using the xml:lang attribute where allowed. Specified (like the xml:lang attribute) using ISO 639-1:2002, i.e. standard two letter language codes (en, fr, etc.).
- Statement – free text. The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred.
- Creation – information about how the resources being harvested were collected includes:
    o Created on – when the collection was created
    o Query - The query used to extract the data.
- Keywords – for the collection. The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. Use of a controlled vocabulary such as Getty Arts and Architecture Thesaurus is recommended, and the vocabulary used may be indicated using an attribute.
- Coverage – of the collection
    o Temporal – general temporal coverage of the collection
    o Spatial – general spatial coverage of the collection, e.g. the country covered.
- DC:Relation

3.3   Heritage Asset Identification Set

The CARARE information set for heritage assets is based on the MIDAS Heritage standard, however the elements are compatible with the POLIS DTD and the CIDOC Core Data Index for Archaeological Sites.

The scope of this information set includes archaeological monuments, historic buildings, industrial monuments, archaeological landscape areas, shipwreck, artefacts and ecofacts, as well as books, printed materials, printed maps, drawings, photographs, films and born-digital 3D models which relate to the archaeological and architectural heritage. The ability to create relations between heritage asset records allows the relationships between individual monuments that form parts of a larger complex to be expressed, for example the Parthenon, Propylaea and the Erechtheum are part of the Acropolis of Athens.

Please note that global types are defined in section 11. The heritage asset information set includes:

Record information (source = MIDAS) (global) – The ID, the language of the metadata, creation information and other metadata describing the record. The ID element of this information block holds the ID assigned by the content provider and is used to add a unique ID for the CARARE object. .Mandatory, see section 8 below.

Appellation (global) – This is the name or title of the heritage asset and the identifier (ID) and may be repeated if, for example, a monument is known by more than one name or has more than one ID number, If the heritage asset has a name in an alternate language the XML:lang attribute should be used. The preferred/alternate attribute may be used to indicate which name is preferred.

Description (source = MIDAS) (global) – of the features of the archaeological monument, historic building, industrial monument, archaeological landscape area, shipwreck, artefact or ecofact, or the subject of the book, printed material, printed map, drawing, photograph, films and born-digital 3D models which relate to the archaeological and architectural heritage. This element may be repeated, for example in cases where there is a both a short summary and a full description.

Actors (source = MIDAS) (global) – the actors involved with this monument, for example famous inhabitants, the architect etc. May be repeated.

Designations (source = MIDAS);
This is information about any designations for a monument or building which provide it with protection in law. There may be more than one designation.
- Protection type – the type of designation or protection.
- Grade – the grade or level of protection.
- Date from – the calendar date from which the protection came into force.
- Date to – the calendar date until which the monument is protected
- Display date – this is a text element to allow for a descriptive date, e.g. '1950s'.

Conditions (source = MIDAS);
This is information about the condition of a monument or building. The element is repeatable.
- Condition – the observed condition (e.g. good, fair, bad, poor, part destroyed, under restoration.)

- o Condition Assessment - A detailed assessment of the condition of a Heritage Asset and any treatment required and an estimation of the percentage of the monument affected.
- o Condition Date – the date when the assessment of condition was made. Free text to allow for a descriptive date, e.g. '1950s'.
- o Was Present At – relation to an associated event/activity at which the assessment was made

Provenance (source = DCMI Terms) – A free-text statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.

Characters (source = MIDAS);
This is a set of index information to describe the character of the monument
- o **Heritage asset type** (source = MIDAS) – classification of the monument, building, landscape feature, artefact or ecofact primarily with respect to its function or use, e.g. house. Use of a controlled vocabulary is recommended, and the vocabulary used may be indicated using an attribute. No particular common vocabulary is recommended for use.
    - Term
    - Namespace – this is the name/location of the controlled vocabulary from which the term is taken.
- o **Temporal** (source = MIDAS) (Global) (see section 11)
- o **Materials** (source = MIDAS) – the basic materials of which a heritage asset is composed, e.g. brick, stone, tile, paper etc. Use of a controlled vocabulary is recommended, and the vocabulary used may be indicated using an attribute. No particular common vocabulary is recommended for use.
- o **Inscriptions** (source = MIDAS) – text inscribed on a monument or building, if any. The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred. The type of inscription may be indicated using an attribute. Use of a controlled vocabulary to indicate the type of inscription is recommended, and the vocabulary used may be indicated using an attribute. No particular common vocabulary is recommended for use.
- o **Dimensions** (source = MIDAS, LIDO)
    - **Extent** – this is to note the part of the heritage asset being measured, e.g. Base.
    - **Measurement type** – e.g. height, length, width, depth, shape (e.g. oval)
    - **Units** – e.g. metres, centimetres
    - **Scale**
    - **Value** - e.g. 150. The valuetype attribute allows the accuracy of the measurement to be indicated (exact, approximate)
- o **Craft** (source = MIDAS) – this is a set of information to describe shipwrecks if any
    - Placeofregistration – the name of the place where the ship was registered
    - Nationality – of the ship
    - Constructionmethod – the method by which the ship was constructed, e.g. clinker,
    - Propulsion – the means of propulsion of the ship, e.g. steam, sail etc.
    - Lastjourneydetails – Details about the last journey of the ship

- Departure – Port of departure
- Destination – Port of destination
- Cargo – the cargo the ship was carrying, may be repeated.
- Mannerofloss – a description of how the ship was lost, e.g. 'ran aground'.
- Dateofloss – the date when the ship was lost. Free text.

Spatial (source = MIDAS) (Global) (see section 11)
This is information about the place at which the heritage asset is located, included named places, postal address, the map coordinates and geometry of the heritage asset.

Repository location (source = LIDO) – identification of the institution with custody of the artefact and possibly the current location.

Publication statement (Global) (see section 11)

Rights (source = DC) – a statement of any rights associated with the heritage asset.

References – these are sources of information about the heritage asset in publications and archival sources (for example, bibliographic references etc.). Do not include the digital objects (image, text, video, audio, 3D model, etc.) which your organisation is making accessible to Europeana – these should be described as Digital Resources, not References. Source = MIDAS + DCMI Terms. The information includes:

- **Appellation** – the ID and name given to the information source.
- **Actors** (source = MIDAS) (Global) **–** (creator, author, contributor, editor, etc.)
- **Type** (archive, file, record, book, chapter, article etc.) Use of a controlled vocabulary is recommended, and the vocabulary used may be indicated using an attribute. No particular common vocabulary is recommended for use.
- **Rights** (source = MIDAS)
- **Publication statement** (Global) (see section 11)
- **Note** (source = MIDAS) (Global)
- **Link (Source = LIDO)** – this is the URL where users can find the reference online

Has Representation – the relation between a heritage asset and a digital resource in which it is represented. Give the id number of the target record or a URI.

Is Successor Of – the relation between a heritage asset and an earlier one, e.g. a church is successor of an earlier church on the same site. Give the id number of the target record or a URI.

Was Present At – this is the relation between a heritage asset and an Activity that it was present at, e.g. a castle was present at a siege, a cross was present at a digitization event. Give the id number of the target record or a URI.

Has Part – used for heritage assets that incorporate other assets, e.g. a landscape monument has parts (is made up of) smaller monuments. Give the id number of the target record or a URI.

Is Part Of – this is the relation between a heritage asset and the larger heritage asset of which it is part, e.g. a gatehouse is part of a fortification. Give the id number of the target record or a URI.

Is Replica Of – this is the relation between a replica and the original heritage asset, for example a scale model and the original, or a 3D reconstruction and the original building. Give the id number of the target record or a URI.

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.

DC:Relation – this is for general relations from the Heritage Asset. Give the id number of the target record or a URI.

3.4   Digital Resource

These are digital resources (image, text, video, audio, 3D model) that are being made accessible to the service environment (e.g. CARARE, Europeana). Use this to describe those digital objects which are available online for end-users. Source = Europeana Data Model + LIDO + MIDAS + DCMI Terms.

The information set includes:

Record information (source = MIDAS) (Global) - The ID, language, creation information and other metadata describing the record. The ID element of this information block holds the ID assigned by the content provider, cf. section 8.

Appellation – the ID and the name given to the information source (see section 11).

Description (source = MIDAS) (Global) – The description of the resource, e.g. describe the view of the monument.

Note (source = MIDAS) (Global)

Actors – The actors involved in the creation of a digital resource

Type (source = DCMI Terms) – The nature or genre of the resource. Use of the DCMI controlled vocabulary is recommended, and the vocabulary used may be indicated using an attribute.

Format (source = DCMI Terms) – the file format of the resource. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types (MIME).

Format Details (source = DCMI Terms) – Additional information about the file or its production that could be of use in selecting an appropriate viewer for the resource, such as specific codecs used.

Extent (source = DCMI Terms) – the size or extent of the resource, including the unit of measurement.

Publication statement (see section 11)

Created (source = DCMI Terms) – this is the date when the digital resource was created

Language (source = DC) – use for the language of the resource, e.g. the language sub-titles or a voice-over in a movie or a Virtual Reality model of a monument. Specified (like the xml:lang attribute) using ISO 639-1:2002, i.e. standard two letter language codes (en, fr, etc.). Mandatory for text-type resources.

Link (source = LIDO) – the URL of the resource. A reference to the digital object on the content provider's web site in the best available resolution/quality (i.e. a link to the resource as a text, image, sound, or video file, **not** to the webpage that contains it).

Object (source = Europeana) – A URL to a thumbnail. The data given here will allow the automatic generation of a thumbnail by Europeana for its functionality.

IsShownAt (source = ESE v3.3) – A URL to the digital object on the content provider's website in its full information context (i.e. a link to the webpage that contains the digital object and contextual information).

Rights (source = MIDAS) – the rights associated with the digital object itself (copyright, access rights, reproduction rights).

Is Representation Of – this is the relation between a digital resource and the heritage asset that it represents, e.g. a digital image is a representation of a monument. Give the id number of the target record or a URI.

Has Part – this is the relation between a digital resource and other digital resources which it contains, e.g. a 3D model may be related to some text documents and sound files. There is no ordered sequence between the parts. Give the id number of the target record or a URI.

Is Part of – this is the relation between a digital resource and the digital resources of which it forms part, e.g. an image is part of a book. There is no ordered sequence between the parts. Give the id number of the target record or a URI.

Next In Sequence – this is the relation between a digital resource and the next digital resource in the sequence, e.g. the pages in a book or an ordered sequence of 3D models showing the change of a monument over time. Give the id number of the target record or a URI.

Is Previous in Sequence – this is the relation between a digital resource and the previous digital resource in the sequence, e.g. the pages in a book or an ordered sequence of 3D models showing the change of a monument over time. Give the id number of the target record or a URI.

Is Derivative Of – this is the relation between a Digital Resource and the digital resource which it is a derivative (or version) of. Give the id number of the target record or a URI.

Has Derivative – this is the relation between a Digital Resource and a derivative (or version) of the resource. Give the id number of the target record or a URI.

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 specialisation of Is Derivative Of. Give the id number of the target record or a URI.

DC:Relation – this is for general relations from the Digital Resource. Give the id number of the target record or a URI.

3.3   Activity

This is information about the events or activities that the monument has taken part in, for example: Field investigation; Research and analysis; Creation; Change in use; Historical events, etc. Source = MIDAS + POLIS DTD + CIDOC CRM-DIG.

The information includes:

Record information (source = MIDAS) – The ID, language, creation information and other metadata describing the record. The ID element of this information block holds the ID assigned by the content provider, cf. section 8.

Appellation (source = MIDAS) (global) – This is the name and ID number of the event.

Description (source = MIDAS) (global) – of the event or activity that took place. Free-text

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.

Event type (source = MIDAS – 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. (Not repeating)
- Term
- Namespace – this is the name/location of the controlled vocabulary from which the term is taken.

Actors (source = MIDAS) (global) – the people or organisations involved in this event, may be repeated.

Temporal (source = MIDAS) (Global) – the date or time span of the event.

Spatial (source = MIDAS) (Global) – the location or area covered by the event.

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.

>   **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.

>   **Start date** – the start date of the specific activity

>   **End date** – the end date of the specific activity

>   **Methods** (source = LIDO) – 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.

Techniques – the techniques used in this specific activity. Use of a controlled vocabulary is recommended.

Materials (source = LIDO) – the materials used during the event method. Use of a controlled vocabulary is recommended.

Equipment– the equipment used during the event method.

Assessments (source = MIDAS) – assessments made of the monument during the event, e.g. of the condition of the monument. Use of a controlled vocabulary is recommended.
- Term
- Namespace – this is the name/location of the controlled vocabulary from which the term is taken.

Occurred At – can be used to associate an Activity to the time span that overlaps with the occurrence of that activity. Give the id number of the target record or a URI.

Happened At – can be used to relate an event to the place where it happened. Give the id number of the target record or a URI.

Was Present At – can be used to relate an Activity to the Actors who were present. Give the id number of the target record or a URI.

Has Part – this is the relation to an Activity and shorter Activity which forms part of the overall event, e.g. a war consists of a series of battles, a landscape survey consists of a series of geophysical surveys and sample excavations. Give the id number of the target record or a URI. Give the id number of the target record or a URI.

Is Part of – this is the relation between an Activity and the general activity of which it forms a part of, e.g. a battle is part of a war. Give the id number of the target record or a URI.

Has Created (source = CRD 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.

DC:Relation – this is for general relations from the Activity, e.g. to publications. Give the id number of the target record or a URI.

4. Global types

The following types are used globally across the CARARE schema to define its elements.

4.1   Record information

Basic administrative information about the record:

- ID – i.e. the local ID number in the content providers' information system; it is mandatory to provide an ID which must be unique within the collection; but the id may follow any schema.
- Source – of the record (the name of the organisation that maintains the record)
- Country – in which the head office of the organisation that maintains the record is based.
- Creation – when created and who by;
    - Contacts
    - Date
- Update – the date of the last update to the record and who by;
    - Contacts
    - Date
- Language (of the metadata record). Specifies the default language of the record; deviations in particular elements are specified using the xml:lang attribute where allowed. Specified (like the xml:lang attribute) using ISO 639-1:2002, i.e. standard two letter language codes (en, fr, etc.).
- Metadata Rights – statement about any rights to the metadata, include a link to a licence online if appropriate.

4.2   Appellation

- Name – this is the name of the entity. The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred.

- ID – an identifier of an object. An attribute type should be accompany this sub-element denoting the type of the identifier (URI, ISBN, etc,) The element may be repeated.

4.3   Rights

Information about the rights associated with the object, metadata and the digital surrogate being harvested into the service environment based on MIDAS Heritage. The information includes:

- Copyright credit line – a statement about the rights holder and rights dates

- Access rights – a statement about the access rights to the content.

- Reproduction rights – a statement about the reproduction rights including contact information

- License – a URI indicating a license or conditions for the use of the object or data, e.g. this could be a page on the content providers website which includes information about copyright, access rights and reproduction rights or a link to a Creative

Commons license[1] or the public domain mark[2]. Use as a supplement to the information above. It is always recommended that the Copyright elements are given when known.

- Europeana Rights – one of the 12 rights statements used by Europeana in its portal. See: http://pro.europeana.eu/web/guest/available-rights-statements. Required for content being provided to Europeana.

4.4   Temporal

Information about the date and/or period of an entity.
- Time span
  - o   start date – the earliest date in the range (literal)
  - o   end date – the latest date in the range (literal)
  - o   Date range qualifier – the nature of the time span given (e.g. throughout, at some time during, etc.) Use of a controlled vocabulary is recommended.

- Period name – the name given to the period in history when something occurred. The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred.

- Display date – a free text field used to display the date or period for users (e.g. early 19th century, 1950s). The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred.

- Scientific Date – date according to scientific dating methods, e.g. '1250 bp +/-30 PBN-1675', recorded precisely as received from the specialist.

- Scientific Date Method – e.g. 'radiocarbon dating'. Use of a controlled vocabulary is recommended, and the vocabulary used may be indicated using an attribute. No particular common vocabulary is recommended for use.

4.5   Spatial

(source = MIDAS). Information about locations or positions in space.
- **Location set**
  - o   **Named location** – the name of a place or location which is relevant to the entity being described, for example 'Lake Windermere'.
    - ▪   The element may be repeated using the XML:lang attribute if the element value is available in alternate languages.
    - ▪   A preferred/alternate attribute may be used to indicate which value is preferred.
    - ▪   Use of a controlled vocabulary such as http://www.geonames.org/ is recommended, and the vocabulary used may be indicated using an attribute.
  - o   **Address** – the postal address.
  - o   **Geopolitical area** – the name of an administrative region which does not form part of the address, for example Scotland, England, Tuscany etc. May also be used for a historical geopolitical area, or an administrative unit (e.g. as defined in the INSPIRE directive).

---

[1] http://creativecommons.org/about/licenses/
[2] http://creativecommons.org/publicdomain/mark/1.0/

- - - Use of a controlled vocabulary such as http://www.geonames.org/ is recommended, and the vocabulary used may be indicated using an attribute.

  - - The type of the geopolitical area may be indicated using an attribute, for example Country, District, Region, City.

  - **Cadastral reference**

  - **Historical name**. The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred.

- **Geometry**
  - **Spatial reference system** – Use of a controlled vocabulary is recommended, and the vocabulary used may be indicated using an attribute (e.g. WGS84). The ESPG spatial coordinate reference is recommended for use. (Mandatory where coordinates are provided)

  - **Quickpoint** (this is the centre point)
    - X (according to the spatial coordinate system in use)
    - Y(according to the spatial coordinate system in use)
    - Z (height)
    - Accuracy – free-text, the accuracy of the point in metres.
  - **Bounding box**
    - Min X (according to the spatial coordinate system in use)
    - Min Y (according to the spatial coordinate system in use)
    - Max X (according to the spatial coordinate system in use)
    - Max Y (according to the spatial coordinate system in use)
    - Min Z (height)
    - Max Z (height)
    - Accuracy

  - **Other geometries** (optional)
    - Type (Point, Line, Polygon) (repeating group to allow a series of points to be recorded)
      - X (according to the spatial coordinate system in use)
      - Y (according to the spatial coordinate system in use)
      - Z (height)
    - Accuracy – free-text, the accuracy of the point in metres.

  - **Entity**: GML, Well-known text (WKT).

  - **Stored precision**, delivery precision (the precision of a coordinate as stored in the system, and as delivered to users).

  - **Area**: units

- **Representations** – how a feature is represented on a map

- **3D Cartesian coordinates** – use for the local coordinates within a 3D model,
  - **Cartesian coordinate system** (this is the local coordinate system in use within the particular software)
  - **Viewpoints** – a viewpoint within the 3D model, i.e. where the camera is positioned in the digitisation process
    - X
    - Y
    - Z

4.6   Actors

(source = MIDAS + elements from LIDO).

- ID
- Name (the name of the person or organisation)
- ActorType (source = LIDO – indicates whether the actor is an individual, a group of individuals or an organisation.
- Roles – the roles of the actor (creator, custody, repository, curator, architect, sculptor, photographer, compiler, etc.) Use of a controlled vocabulary such as Getty Arts and Architecture thesaurus is recommended, and the vocabulary used may be indicated using an attribute.
- Contacts – contact information if known
- Vital dates (source = LIDO):
  - o date of birth, if known.
  - o date of death if known.
  - o Display date
- Place of birth – Use of a controlled vocabulary such as http://www.geonames.org/ is recommended, and the vocabulary used may be indicated using an attribute.
- Place of death – Use of a controlled vocabulary such as http://www.geonames.org/ is recommended, and the vocabulary used may be indicated using an attribute.
- Place of activity – Use of a controlled vocabulary such as http://www.geonames.org/ is recommended, and the vocabulary used may be indicated using an attribute.
- Biographical note – The element may be repeated using the XML:lang attribute if the element value is available in alternate languages. A preferred/alternate attribute may be used to indicate which value is preferred.

4.7   Contacts

(source = MIDAS).
Information about how a person or organisation can be contacted

- Name – title, first name, last name, other name
- Role – the particular role played by the person or organisation
- Organisation
- Address – the postal address of the person or organisation
- Phone
- Fax
- Email

4.8   Address

The postal address for a building, contact, etc.

- Building name
- Number in road – the number in a road or street used to identify a property
- Road name
- Town or city
- Postcode or zipcode
- Locality –
- Admin area – the name by which an administrative area is known, e.g. Shropshire
- Country

4.9   Publication statement

- Publisher
- Place – of publication
- Date – of publication

5. References

ATHENA Europe, 2009, LIDO - Lightweight Information Describing Objects, http://www.athenaeurope.eu/

CIDOC, 2010, CIDOC CRM Home Page: http://cidoc.ics.forth.gr/index.html (accessed 12 May 2010).

CIDOC, 2008, CIDOC Archaeological Sites working group webpage: http://cidoc.mediahost.org/wg_archaeological_sites%28en%29%28E1%29.xml (accessed 11 May 2010).

DCMI Metadata Terms, http://dublincore.org/documents/dcmi-terms/

English Heritage, 2007, MIDAS Heritage: the UK historic environment data standard, Second Edition, English Heritage, London. http://www.english-heritage.org.uk/ (accessed 12 May 2010).

European Commission, INSPIRE directive, http://inspire.jrc.ec.europa.eu/

FISH, 2010, FISH: the Forum on Information Standards in Heritage website: http://www.fish-forum.info/(accessed 12 May 2010).

PARADOTEA project, POLIS DTD, http://www.ics.forth.gr/CULTUREstandards/paradotea.htm

UK GEMINI standard, http://www.gigateway.org.uk/metadata/standards.html