



Project Acronym: Europeana v2
Grant Agreement number: 270902
Project Title: Europeana Version 2

D7.6: Report on Model refinements for the Europeana Data Model

Revision	FINAL
Date of submission	30 May 2014
Author(s)	Marlies Olenky, Humboldt-Universität zu Berlin; Valentine Charles, Europeana Foundation
Dissemination Level	[Public]

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
Draft 0.1	21/02/2014	Marlies Olensky	HUB	First draft
Draft 1.0	14/03/2014	Marlies Olensky	HUB	Restructuring according to suggestions from Valentine Charles and Antoine Isaac
Draft 1.1	24/03/2014	Valentine Charles	Europeana Foundation	Addition of subsections
Draft 1.2	26/03/2014	Marlies Olensky	HUB	Few editorial changes
Draft 1.3	27/03/2014	Valentine Charles	Europeana Foundation	Proof-reading
Draft 1.4	27/03/2014	Robina Clayphan	Europeana Foundation	Proof-reading
Final 1.0	28/03/2014	Marlies Olensky	HUB	Final adaptations and addition of DDB-section
Final 1.1	02/04/2014	Marlies Olensky	HUB	Few editorial changes
Final 1.2	16/04/2014	Marlies Olensky	HUB	Final editorial changes

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1. Introduction	4
2. Important contributors to the EDM development.....	4
3. Refinements of the EDM model for domain-specific profiles	5
3.1. Identifying a ProvidedCHO	5
3.2. Creating URIs for the EDM entities	5
3.3. Contextualising information in EDM.....	7
3.4. Modelling FRBR in EDM.....	8
3.5. Other modelling issues.....	8
4. Refinements and changes in EDM.....	9
4.1. Representation of hierarchical objects in EDM	9
4.2. Profiles for representing collections, datasets and organizations.....	10
5. Re-use and extensions of EDM.....	14
5.1. Within the Europeana network.....	14
5.1.1. Europeana Fashion profile	14
5.1.2. DM2E Model	14
5.1.1. German Digital Library (DDB).....	15
5.1.2. The Polymath Virtual Library	16
5.1.3. BiographyNet project	16
5.2. Outside the Europeana network	17
5.2.1. Smithsonian American Art Museum	17
5.2.2. DPLA model	18
6. Ongoing and future work.....	18
6.1. Recommendations from the Task force on EDM mappings, refinements and extensions	18
6.2. EDM Roadmap	19
6.3. Profile for representing sounds	20
7. References	21
ANNEX.....	22

1. Introduction

The Europeana v2.0 project started in October 2011. In the course of this project the Europeana Semantic Elements were transitioned to EDM.

This synthesizing report documents refinements for the Europeana Data Model (EDM) that have been carried out in the course of the Europeana v2.0 project. The EDM has been further developed, extended and re-used for prototyping and delivery of EDM data by Europeana data providers and aggregators. In addition further work was carried out within domain-specific projects as well as in the EuropeanaTech Task Forces¹. This report summarises these activities that have been going on within the wider Europeana Network and provides the 'big picture' of model refinements over the course of the project time.

This report contains the results of the Task Force on EDM mappings, refinements and extensions which was mostly launched to gather information from the Europeana community on their EDM refinements and to be reported in this deliverable. It is also closely related to the D5.3 Further development of the Europeana Data Model (Clayphan, 2014) from Work package 5. It reuses parts of the final reports of two of the EuropeanaTech (WP7) Task Forces: Task Force on hierarchical objects and Task Force on EDM/FRBRoo.

2. Important contributors to the EDM development

The Europeana Data Model (EDM) plays now an important role for Europeana's network of projects and data providers. The further development of Europeana's technical infrastructure in 2013 has allowed the ingestion of more than 6 mio objects in EDM also referencing to Linked Open Data vocabularies. The MIMO and the CARARE projects were the two projects delivering first EDM. The mapping of their data to EDM has allowed Europeana to identify what were the key benefits of EDM in terms of data modelling and the issues that needed to be addressed in the EDM documentation.

MIMO was a two-year-project that acted as a consortium of some of Europe's most important musical instruments museums. MIMO was the first project delivering its metadata using the EDM rather than the Europeana Semantic Elements (ESE). MIMO took the advantage of the richness provided by EDM for delivering to Europeana richer data in terms of descriptive information and links to external resources (Linked Open Data vocabularies). In addition to the descriptive metadata for their musical instruments (defined as Cultural Heritage Object (CHO) by the project), MIMO provided extra information about the digital representations (or Web resource) of the musical instruments and integrated information from its own vocabularies and Linked Open Data vocabularies using the specific EDM contextual entities².

CARARE was a three-year project making the digital content from the archaeology and architecture heritage domain available to Europeana. CARARE took advantage of EDM to model the network of connections existing between the heritage assets themselves (such as monuments, buildings) 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. The CARARE case study³ illustrates a few challenges but also characteristics data providers can face when mapping their data to EDM.

¹ List of all EuropeanaTech task forces at: <http://pro.europeana.eu/network/task-forces/overview>

² The full case study can be accessed here: <http://pro.europeana.eu/web/guest/mimo-edm>

³ The full case study can be accessed here: <http://pro.europeana.eu/carare-edm>.

3. Refinements of the EDM model for domain-specific profiles

The prototyping work undertaken by data providers when mapping their metadata to EDM highlighted some challenges which needed to be addressed in the main EDM documentation. This chapter summarises discussions about the "abstract" model which lead to solutions of how to organise data in specific situations.

3.1. Identifying a ProvidedCHO

One of the main issues for data providers is the definition of the Cultural Heritage Object (CHO) mapped to the class *edm:ProvidedCHO*. EDM allows the distinction between "works", which are expected to be the focus of users' interest, and their digital representations. Therefore, data providers are asked to define the focus of the description according to their represented domain. For instance, for the CARARE⁴ project (CARARE, 2013) representing the archaeological and architectural domain, the CHO can be:

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

For the DM2E project⁵ (DM2E, 2014a) a CHO could be a book but also a page or a paragraph within the book. The variation in the level of granularity from one data provider to another is related to their domain-specific (digital humanities) or technical (annotations of the metadata) requirements.

The Europeana Libraries⁶ project discussed how library metadata could be best represented in EDM taking in consideration monographs, multi-volumes works, serials but also the distinction between born-digital and digitised objects. One of the recommendations of the project was that: "a clear definition of what an *edm:ProvidedCHO* is should be provided". In the library domain the guiding model is FRBR and, in this case, the group 1 entities (Work, Expression, Manifestation and Item.) There is a desire to describe both the edition level and Item level for rare and unique objects (e.g. a copy of a book but it was owned by a significant person, or it has significant notes in the margins.) It would be useful if Europeana could define a standardised way to represent the FRBR entities in EDM. (Europeana Libraries, 2012). This question was followed-up by the Task Force on EDM-FRBRoo (see Chapter 3.4).

EDM also enables the provision of multiple digital resources per CHO and allows the different resources to be connected to each other. The global context of a ProvidedCHO is therefore enriched. The MIMO project for instance provided their CHO (musical instruments) in association with audio recording, images and video of the related CHO.

3.2. Creating URIs for the EDM entities

Another challenge for data providers converting their data to EDM is the creation of identifiers for the various entities that need to be described in EDM. Since it relies on the

⁴ <http://www.carare.eu/>

⁵ <http://dm2e.eu/>

⁶ <http://www.europeana-libraries.eu/>

principles of the Semantic Web, EDM requires a URI for each described entity. The creation of unique identifiers per EDM class enables the representation of individual information sources while representing the relationships between them.

Most of the data providers already have URIs available in their metadata for the ProvidedCHO and for the WebResources. If this is not the case they usually generate those URIs from their local identifiers. However, data providers do not always have URIs for other entities such the *ore:Aggregation* as this resource is a very specific one to the EDM model. The project CARARE created new identifiers for the *ore:Aggregation* class. The CARARE aggregation identifiers are web-enabled, in the sense that they redirect to a landing page that CARARE creates for each object (cf. Figure 1). The landing page provides a unique identifier that can be used in the data. The MIMO project solved the issue by providing a local identifier for each CHO.

<http://store.carare.eu/landing-pageha.php?id=iid:2920150&eid=HA:6161>

Figure 1. Landing page used by CARARE as Identifier of the Aggregation class

In EDM, there are also properties that do not require the use of URIs (e.g. *edm:dataProvider*, *edm:provider*) and others that do (e.g. *edm:isShownAt*, *edm:rights*). MIMO decided on each case individually and decided in the case of *edm:dataProvider* and *edm:provider* to use literals as Europeana does not use a controlled list of data providers (yet). In some cases, MIMO had to create new URIs and paid particular attention to make them robust.

For CARARE, each print, map, or book about an archaeological place counts as a separate object provided to Europeana. Therefore, these information sources also needed unique identifiers which now allow explicit links between the heritage assets and their related objects

using for example the property *edm:isRepresentationOf*. The distinction in EDM between the real world object and the digital representation is also important when dealing with rights metadata, as they might have different, or even contradictory, rights statements which determine the conditions for re-use of the content. In EDM, the class *edm:Place* allows the archaeological domain a semantically richer way to represent information about geographic locations. CARARE uses the property *dcterms:spatial* to link the Place class to the Provided CHO.

3.3. Contextualising information in EDM

EDM allows the description of individual entities related to a Cultural Heritage Object such as place, agent, time and concept. Data providers use the EDM constructs to provide richer data and align their metadata to domain specific vocabularies.

This chapter includes parts from a case study (MIMO, 2012) and parts from the Task Force report on EDM mappings, refinements and extensions (Charles & Olensky, 2014):

In the context of the project, MIMO created three domain-specific thesauri available in six languages and represented in SKOS. The first two thesauri define concepts for musical instruments using the MIMO instrument keywords vocabulary and the Hornbostel-Sachs musical classification system. MIMO has re-used the elements offered by EDM to provide their rich information in their mappings. MIMO used the EDM contextual entity *skos:Concept* to describe all entities from knowledge organisation systems like thesauri, classification schemes, including some place gazetteers or person authority files (cf. Figure 2).

```
<skos:Concept rdf:about="http://www.mimodb.eu/InstrumentsKeywords/2308">
  <skos:prefLabel xml:lang="en">Square pianoforte</skos:prefLabel>
</skos:Concept>
<skos:Concept rdf:about="http://www.mimodb.eu/HornbostelAndSachs/6458">
  <skos:prefLabel xml:lang="en">314.122-4-8 True board zithers with
resonator box (box zither) sounded by hammers or beaters, with
keyboard</skos:prefLabel>
</skos:Concept>
```

Figure 2. Representation of concepts by MIMO using the *skos:Concept* class

MIMO also created an authority list for instrument makers. To represent them, MIMO used the class *edm:Agent* which defines people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible (cf. Figure 3).

```
<edm:Agent rdf:about="http://www.mimo-db.eu/InstrumentMaker/Person/3487">
  <skos:prefLabel>Christian Salomon Wagner</skos:prefLabel>
</edm:Agent>
```

Figure 3. Representation of an agent by MIMO using the *edm:Agent* class

A loss of the rich MIMO metadata only occurred for one specific class. As a key concept in the LIDO format is the event and the *edm:Event* class is not part of the first implementation of EDM, information related to events was lost during the mapping. Only information related to the creation events has been kept in the description.

3.4. Modelling FRBR in EDM

The EDM – FRBRoo Application Profile Task Force (EFAP-TF)⁷ was launched in response to the recommendations from the deliverable D3.4 from Europeana V1.0. This deliverable asked for an application profile that would allow a better representation of the FRBR group 1 entities: work, expression, manifestation and item. Additionally, it was to be conceived as an application profile of FRBRoo where each intellectual contribution (e.g., in the publication process) and the related activity are treated as entities in their own right, and does not depend too much on the notion of a bibliographic record. As a starting point they suggested the mapping of FRBRoo and EDM offered by the CIDOC CRM working group. The aim of the EFAP-TF was to extend, correct or restrict this suggested mapping and provide examples for the use of the combined EDM and FRBRoo namespaces. In order to find an adequate FRBR application profile for Europeana, the examples were first modeled in full FRBRoo. Then, the following principles were applied in simplifying the FRBRoo model, such that the meaning necessary to answer the relevant research questions was preserved:

- Work is only employed where it is really needed. As long as there is only one Expression, no Work is needed.
- R21 created (was created through)(F29 Recording Event:F26 Recording) is replaced by R17 created (was created by)(F28 Expression Creation:F2 Expression)
- Performance Plan is identified by performed expression; R25 performed (was performed in)(F31 Performance:F25 Performance Plan) can point to more objects of type F25 (Performance Plan)
- In F33 Reproduction Event the link produced is replaced by P108 has produced (was produced by)(E12 Production:E24 Physical Man-Made Thing)
- R6 to be replaced by P128
- E56 language is the unique range of P72 has language. The latter maps dc:language, which has literal as range. Therefore E56 Language maps implicitly to the range values of dc:language..
- F6 concept is replaced by E55 Type since the examples do not exhibit other cases of use of F6 Concept.

The final Task Force report delivered combined models in terms of properties and classes of EDM and FRBRoo illustrated by sample data. Smaller groups have worked on three different examples that are described in section 3 of the final report. The report also provides principles for modeling and mapping rules based on the experiments of the working groups.

This summary is copied from the final Task Force report (Doerr et al., 2013).

3.5. Other modelling issues

Other modelling issues were reported during the preparation of EDM data by data providers or with projects working on domain-specific needs. For instance EDM allows non-compliance with the DCMI specifications of ranges for the dcterms properties which caused some lengthy discussions in projects such as Europeana Libraries. It was recommended the situation should be rectified as soon as it is feasible.

A lot of projects have also pointed out the mapping difficulties introduced by the absence of the Event class in the current implementation of EDM by Europeana. The Europeana Libraries identified it as a key class to represent metadata such as the place of publication (Europeana Libraries, 2012). CARARE also highlighted that a lot of metadata relevant to researchers are more difficult to represent without the Event class.

⁷ <http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+Force+EDM+FRBRoo>

The last modelling issue concerns the way the rights information is currently modelled in EDM. In the portal Europeana currently uses the rights statement attached to *edm:Aggregation*, but it would prefer this to be at the level of the web resource. It is highly desirable that rights statements could be attached to each web resource as there may be several of them and they are likely to carry different rights. Projects support Europeana in its intention to use rights statements at this more granular level. Therefore Europeana has made this a mandatory property in the profile (but not at the Web Resource level).

4. Refinements and changes in EDM

This chapter summarises development efforts from the Europeana community as result of validation efforts to fit the Europeana Data Model to community-specific needs.

In order to meet the specific needs of their domain, data providers have created refinements or application profiles of EDM. They take the form of a set of guidelines or rules that are applied to the classes or properties because existing elements in the source format are used in a more specific sense than the ones that correspond best to them in the original EDM. Usually this means that the property or class being refined will be used in a narrower, but still compatible, sense compared to that defined by EDM. Some of the application profiles are highlighted in this report and Table 1 gives an overview.

Table 1. List of refinements / application profiles collected by the Task Force on mappings, refinements and extensions

EDM refinements / application profiles	Characteristics
CARARE 2.0 schema	EDM refinement for archaeology and architectural heritage domain
DPLA data model	Refinement for DPLA services
Europeana Libraries	Refinement for the library-domain
Europeana v2.0 Task Force on hierarchical objects	EDM Refinement for hierarchical objects

4.1. Representation of hierarchical objects in EDM

The final report of the Task Force on hierarchical objects⁸ illustrates the diversity of definitions concerning hierarchical entities. It also proposes solutions on how to represent these different cases by using the Europeana Data Model (EDM). The report concludes with the following 11 recommendations on how to model hierarchical objects in Europeana:

- 1. Granularity of hierarchy description:** The Task Force assumes that the Data Provider makes the initial choice on the granularity of the description of hierarchical objects, within the limits of relevance for Europeana. The fact that a provided CHO is in a hierarchy does not mean that all parts should be provided CHOs for Europeana. It is important to note that the granularity chosen when modelling the data might induce some issues for an end-user when searching for objects. Additionally, the Task Force feels that more guidance from Europeana in terms of content strategy would be helpful for providers willing to contribute complex objects.

⁸ <http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Taskforce+on+hierarchical+objects>

2. **Granularity of hierarchy description:** The Task Force favours the description of “semantic” hierarchy over the description of hierarchy representing the “physical” structure of a provided CHO (e.g. the book’s page structure vs. say its chapters).
3. **Sequence vs. succession:** the Task Force noticed that the example given for the property *edm:isSuccessorOf* doesn’t reflect the semantics of the property. Europeana should change it to avoid confusion with the property *edm:isNextInSequence*.
4. **EDM properties for ProvidedCHO:** The Task Force recommends making *edm:isNextInSequence* and *dcterms:isPartOf* repeatable to allow representing that a provided CHO belongs to different hierarchies. Data Providers should however note that this situation may raise display issues when retrieving such objects from a search interface.
5. **EDM properties for WebResource:** The Task Force recommends the implementation of the property *dcterms:isPartOf* for the *edm:WebResource* class to allow for a flexible description of a hierarchy of WebResources.
6. **EDM properties for WebResource:** the property *dcterms:hasPart* should only allow for references as values, when applied to this class of resources.
7. **EDM properties for contextual resources:** We recommend the implementation of *dcterms:hasPart* and *dcterms:isPartOf* for the *edm:Agent* class.
8. **EDM properties for contextual resources:** We recommend the implementation of the property *edm:isNextInSequence* for the *edm:Place*, *edm:TimeSpan* and *edm:Event* classes in order to support ordered hierarchies of contextual resources.
9. **Representation of “missing parts”:** In order to alleviate the issue of “missing parts” in a hierarchy, the Task Force proposes to supply the CHO to Europeana as a contextual resource.
10. **Representation of sequences:** *edm:isNextInSequence* should be used to describe a sequence between two resources in a hierarchy. When needed, the Task Force allows for a borderline usage (semantically speaking) of *edm:isNextInSequence* to indicate a sequence between two non-consecutive resources when the resource in-between is missing.
11. **Display of hierarchical objects:** The ability of navigation in hierarchical object is crucial for the user when displaying hierarchical objects.

This summary is copied from the final Task Force report (Charles et al., 2013).

The project Europeana 1914-1918 encountered some challenges while mapping the data to EDM concerning the hierarchies of objects. The contributions of the public consisted of a story that may be accompanied by one or several digitised objects (which themselves could consist of more than one item such as a letter with multiple pages). The current EDM implementation does not support the modelling of these hierarchical relationships. The project made a suggestion on how to model storytelling data in EDM, which has been used in the Task Force on hierarchical objects (Charles et al, 2013).

4.2. Profiles for representing collections, datasets and organizations

Europeana Foundation has created a series of profiles to represent specific metadata related to collections, datasets and organizations.

Collection profile. In the aggregation model used to submit data to Europeana the totality of a curated collection submitted by a data provider may be difficult to find in a large, aggregated dataset. “Collection structures provide the organizational and intellectual context important to researchers, and collection descriptions provide information needed by users for interpreting the relevance and significance of individual items for their purposes” (Wickett et al, 2013). The collection profile is therefore defined to enable the description of collections in

the context of the Europeana Data Model by reusing and extending the classes and properties available. It will allow data providers to supply descriptions of the identifiable collections they have submitted to Europeana. The profile draws on the work cited above and defines the classes and properties that can be used to represent collections in the Europeana system. In the context of this profile the Collection class represents the original collection of cultural heritage objects (probably physical objects but also born digital where appropriate). This definition is visualised in the diagram below in the relationship between the ProvidedCHO and the Collection (cf. Figure 4).

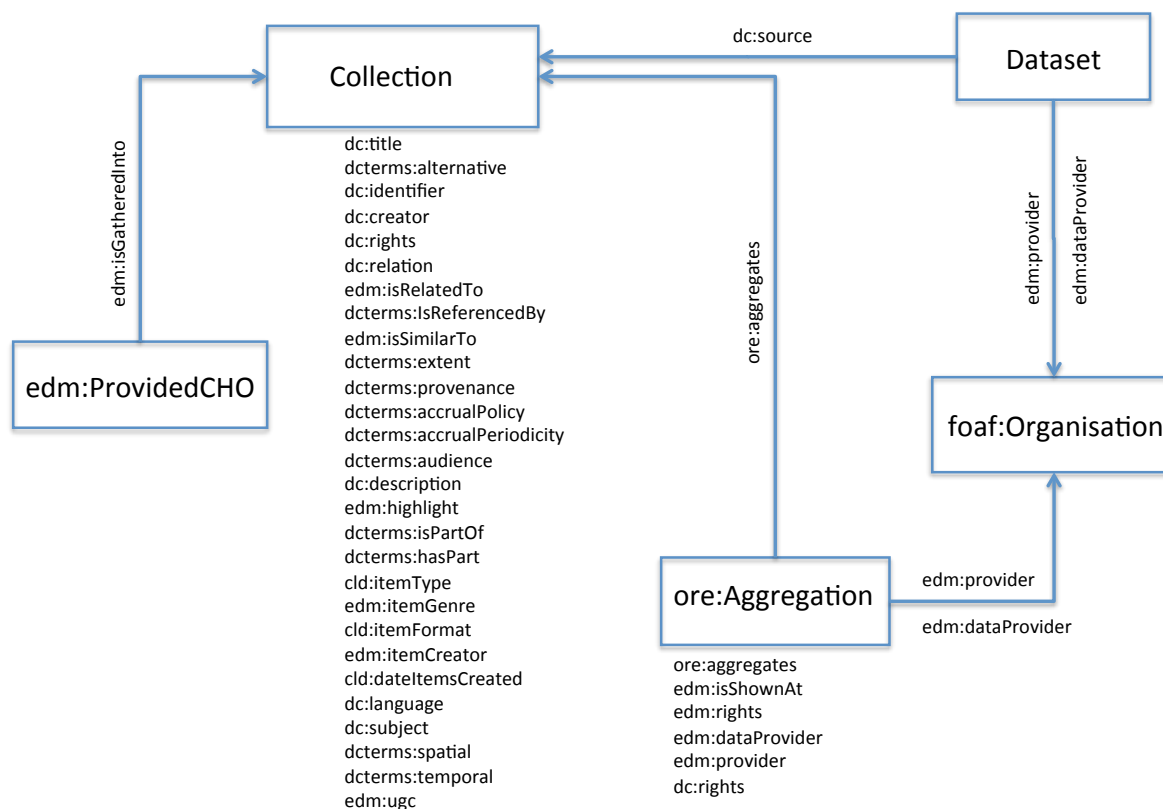


Figure 4. Overview of collection classes and properties

Dataset profile. A dataset for Europeana is an Information Package (as defined in the OAIS reference model⁹) or a collection of data. A Europeana dataset can be about a certain topic, originate from a certain source or process and is aggregated by a certain custodian. Europeana receives datasets from their providers and data providers. In some cases, a Europeana dataset can be the representation of one existing collection held by an institution. In other cases the correspondence between datasets and collections will not be one-to-one, because aggregators may put several collections together into one dataset for submission to Europeana (cf. Figure 5).

⁹ http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57284

The dataset profile defines the elements that will be used to represent a dataset ingested by Europeana. This profile will be mainly used to disseminate dataset level information via the API. The profile is based on the dataset description elements defined in:

- Data Catalog Vocabulary (DCAT) <http://www.w3.org/TR/vocab-dcat/>
- The VoID vocabulary <http://www.w3.org/TR/void/>
- Asset Description Metadata Schema (ADMS) <http://www.w3.org/TR/2013/NOTE-vocab-adms-20130801/>

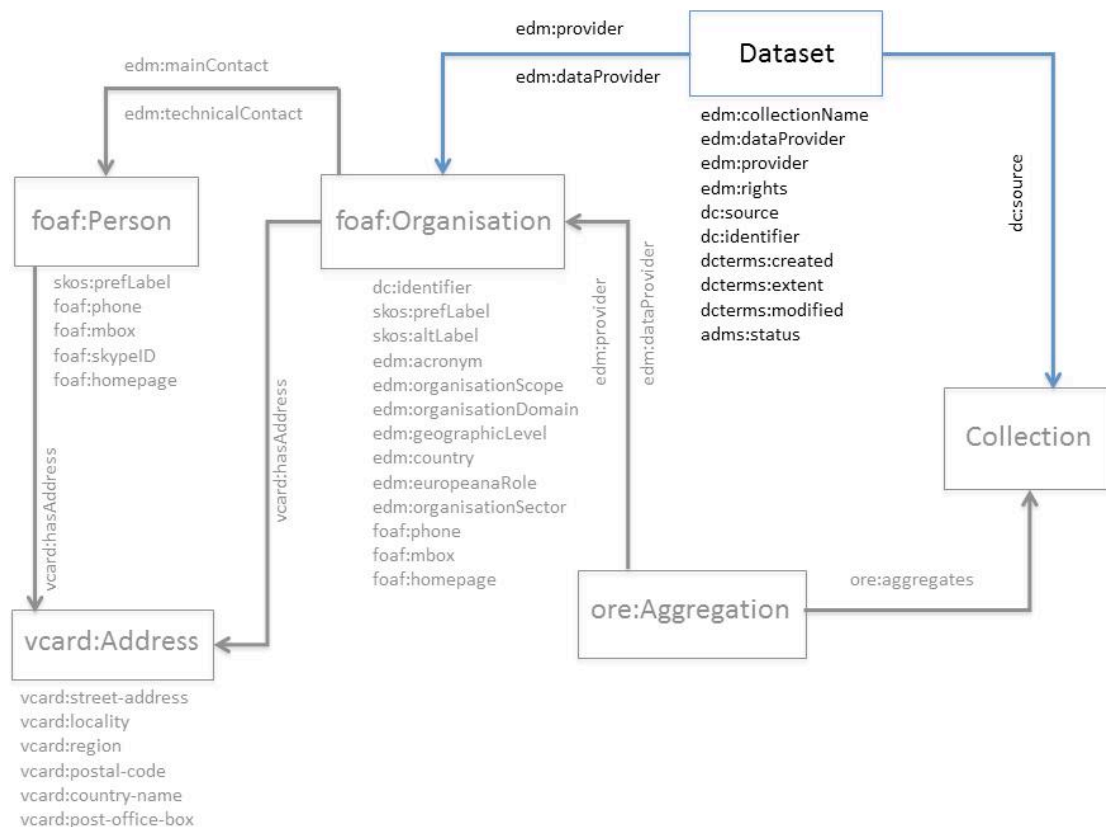


Figure 5. Overview of the Dataset class, its properties and relationships

Organization profile. In the context of Europeana, an organization is defined by any data-contributing organization. Europeana distinguishes the Provider, which is usually an aggregator organization that collects, formats and manages data, from the Data Provider, which is an organization directly submitting data to Europeana or submitting data through an aggregator (cf. Figure 6). In order to provide functionality based on the names of organizations providing data to Europeana it is necessary for their names to be recorded in a controlled fashion. The organization profile will be used to support authority control of (Data) Providers' names and lists the data elements that can be used to describe data providers and aggregators. It is based on the description elements defined for the W3C Org and vCard ontologies, the Institution Entity in the MICHAEL Data Model¹⁰ and the elements that Europeana is already gathering in its Customer Relationship Management system.

¹⁰ http://www.michael-culture.eu/documents/MICHAELDataModelv1_0.pdf, section 6.2

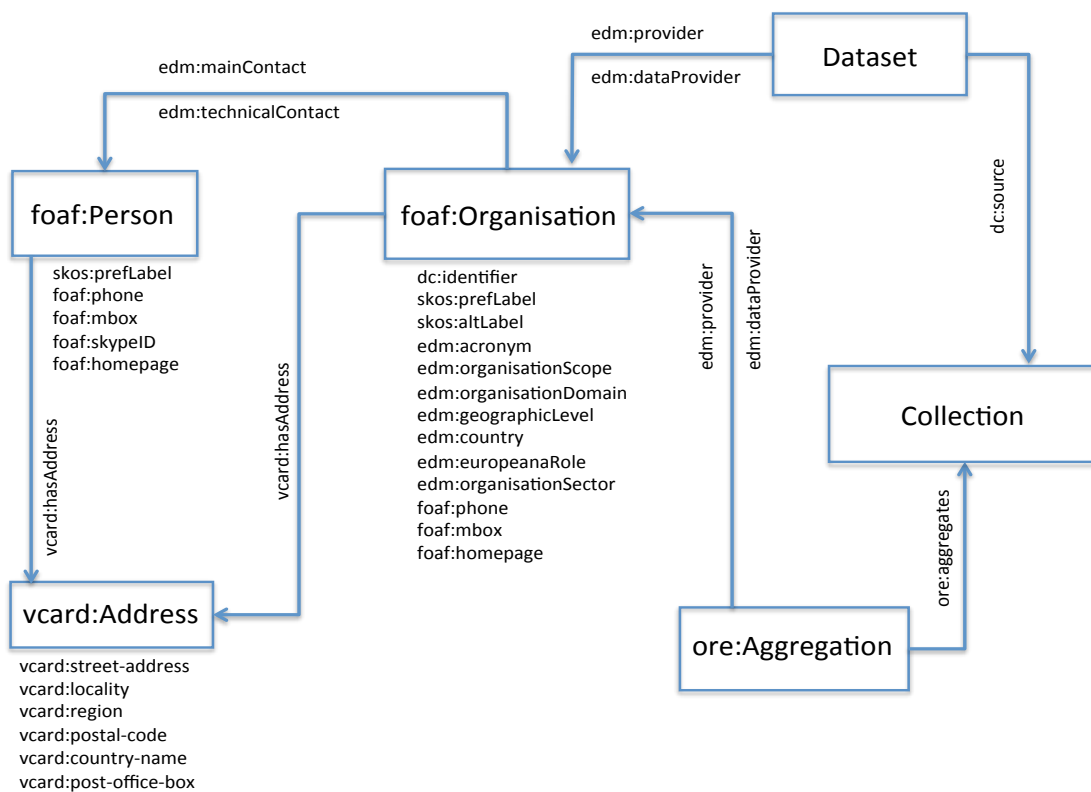


Figure 6. Overview of Organization and Person properties

5. Re-use and extensions of EDM

5.1. Within the Europeana network

5.1.1. Europeana Fashion profile

The Europeana Fashion project developed an EDM application profile, called EDM Fashion Profile (EDM-FP), which mainly builds on EDM but also integrates elements from other metadata standards for their specific needs. EDM-FP has specialized EDM for the following properties (Table 2, Vanstappen, 2012):

Table 2: EDM-FP specific properties defined as subproperties of EDM

Additional properties in EDM-FP	Subproperty of
gr:color	dc:description
edmf:isSubjectOf	edm:isRelatedTo (is the reverse of dc:subject and is semantically equal to the CIDOC-CRM concept P129 About (isSubjectOf))
edmf:localType	dc:type
edmf:technique	dc:description
rdaGr2:biographicalInformation	dc:description
foaf:based_near	dc:description
edmf:nationality	dc:description
skos:definition	skos:note
Not mappable to EDM	
foaf:homepage	
rdaGr2:gender	

5.1.2. DM2E Model

The DM2E model is a specialization and extension of/to the EDM for handwritten and then digitized manuscripts. In the course of the DM2E project a lot of modelling decisions had to be taken that included the method of reusing external resources, decisions on the granularity of the specialization and instruments of documentation (DM2E, 2014a). Like in other domains the metadata formats used in the manuscript domain can vary from data provider to data provider (e.g. TEI, METS/MODS, MARC21 or provider-specific schema). The biggest difference to the EDM model is that the DM2E model uses named graphs (extension of the RDF data model) and that they do not use proxies but use the VOID vocabulary for provenance tracking. Other than that, they have added properties, classes, resource definitions, restrictions and recommendations that were needed to represent the manuscripts as adequately as possible (DM2E, 2014a). The following table lists the DM2E properties and classes that have been added as direct subproperties/classes of the EDM (DM2E, 2014b):

Table 3: DM2E specific classes and properties defined as subproperties/classes of EDM

DM2E Property	Subproperty of (in EDM)
dm2e:publishedAt	dc:coverage, dcterms:coverage
dm2e:artist	dc:creator
dm2e:composer	dc:creator
dm2e:writer	dc:creator
dm2e:condition	dc:description
dm2e:watermark	dc:description
dm2e:illustration	dc:description
dm2e:restoration	dc:description
dm2e:cover	dc:description
dm2e:levelOfGenesis	dc:description
dm2e:callNumber	dc:identifier
dm2e:influencedBy	dc:relation
dm2e:subtitle	dc:title
dm2e:genre	dc:type, dc:description
dm2e:writtenAreaDimension	dcterms:extent
dm2e:pageDimension	dcterms:extent
dm2e:origin	dcterms:provenance
dm2e:modeOfAcquisition	dcterms:provenance
dm2e:refersTo	dcterms:references
dm2e:printedAt	dcterms:spatial
dm2e:shelfmarkLocation	edm:currentLocation
dm2e:copyist	edm:hasMet
dm2e:owner	edm:hasMet
dm2e:previousOwner	edm:hasMet
dm2e:sponsor	edm:hasMet
dm2e:studentOf	edm:hasMet
dm2e:wasStudiedBy	edm:hasMet
dm2e:wasTaughtBy	edm:hasMet
dm2e:principal	edm:hasMet
dm2e:hasAnnotatableVersionAt	edm:hasView
dm2e:mentioned	edm:isRelatedTo
dm2e:misattributed	edm:isRelatedTo
dm2e:honoree	edm:isRelatedTo
DM2E Class	Subclass of (in EDM)
dm2e:Collection	edm:NonInformationResource
dm2e:Work	skos:Concept
dm2e:Paragraph	skos:Concept
dm2e:Publication	skos:Concept

The first operational version of the model was created in April 2013 and since then iteratively refined. The latest version dates to February 2014.

5.1.1. German Digital Library (DDB)

The German Digital Library (Deutsch Digitale Bibliothek, DDB) is the central access point to digital cultural and scientific objects in Germany. The DDB is a national initiative and contributes to Europeana. The objects are also cross-domain (archives, libraries, museums, research-, monument protection- and media institutions), interdisciplinary and the target groups are manifold (culturally interested public, research, education and economy).

Currently the DDB provides access to around 5.7 mio objects. In order to make metadata from the DDB compatible with EDM, an application profile was created. In particular, EDM is used as the facets in their expert search. Table 4 shows the facets and the respective elements in EDM. EDM is also used to model hierarchical structures of compound library objects (format: METS-MODS and MARCXML) and archival inventories (format: EAD-DDB) as well as events (using edm:Event and LIDO vocabulary). In addition, DDB uses EDM for data exchange over OAI-PMH and Open API. In contrast to other application profiles, it is not used to represent the cultural objects or the data providers.

Table 4: EDM use in the DDB facet search

Facet in the DDB	Elements from EDM
Time	edm:TimeSpan
Location	edm:Place
Person/Organisation	edm:Agent
Keyword	skos:Concept
Language	dcterms:LinguisticSystem
Media type	skos:Concept
Sector	skos:Concept
Data provider	edm:Agent

5.1.2. The Polymath Virtual Library

The Polymath Virtual Library¹¹ brings together the works of the most important Hispanic polymaths and establishes semantic relationships between them, expressing the different schools of thought, from Seneca to Octavio Paz in Hispania. The aim is to bring together information, data, digital texts and web resources about Spanish, Hispano-American, Brazilian and Portuguese polymaths from all times.

The Polymath Virtual Library decided to use the resources made available by Europeana through its API and by the Linked Open Data cloud. The semantic relationships between authors and their works have been described using the Europeana Data Model, in addition to other schemas.

The Polymath Virtual Library uses all properties defined in the EDM specifications, especially *edm:Agent* not just the properties defined for the first EDM implementation-phase. For example, this is the case of *edm:isRelatedTo* that establishes generic relationships between people or between people and institutions. *edm:isRelatedTo* is included in their EDM resources, as well as the *owl:sameAs* property that links to similar resources from other contexts. The most important relationships for the Polymath Virtual Library of Polygraph are exemplified by these two properties.

5.1.3. BiographyNet project

The Biography Portal of the Netherlands provides a knowledge base of 75.000 Dutch people. The project wants to develop a demonstrator that will allow using this knowledge base for historical research by interlinking those people to each other. In their data model, they reuse elements from EDM and other data models.

Through the Open Archives Initiative's 'Object Re-use & Exchange' ontology, EDM allows conflicting metadata for one and the same ProvidedCHO. The BiographyNet project reuses

¹¹ http://www.w3.org/2005/Incubator/ld/XGR-ld-usecase-20111025/#Use_Case_Polymath_Virtual_Library

this in their schema¹². They define their *bgn:Person* as *edm:providedCHO* (as a subclass of the *ore:Proxy* class) making their model compatible with EDM. This decision also incorporates all associated predicate relations (Ockeloen et al., 2013).

5.2. Outside the Europeana network

5.2.1. Smithsonian American Art Museum

Szekely et al. (2013) report on their experiences while publishing the data from the Smithsonian American Art Museum (SAAM) to the web. They report that the most challenging part was to select and extend the ontologies they would use in their project. While they actually developed their own ontology (SAAM ontology), they considered EDM and CIDOC CRM in their data modelling. Yet, they found that both did not fully comply with their needs and that is why they extended EDM and added essential subclasses and properties. An additional reason for them to choose EDM was that it “maximizes compatibility with a large number of existing museum LOD datasets” (Szekely et al., 2013). Also, they incorporated classes and properties from schema.org in order to represent geographical data (city, state, country) and they used the RDA Group 2 Elements for biographical information. Figure 7 illustrates the SAAM ontology and shows where the ontology reuses elements from EDM.

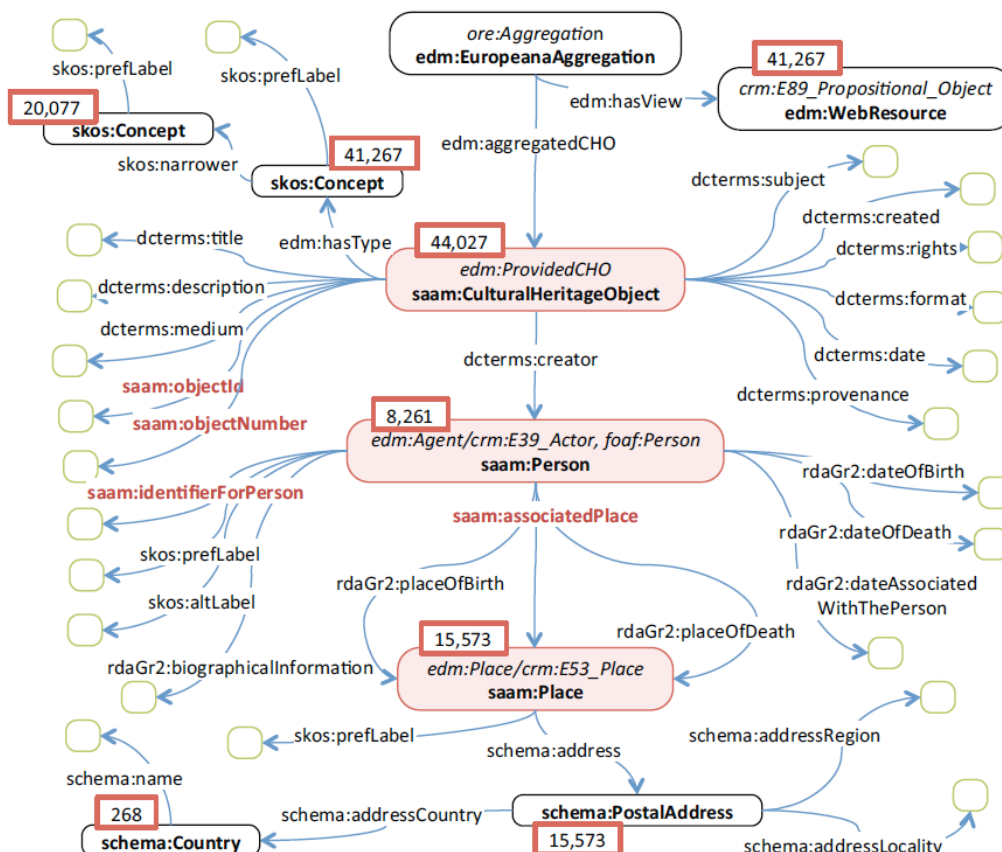


Figure 7. The SAAM ontology. Named ovals represent classes, un-named green ovals represent literals, arcs represent properties, boxes contain the number of instances generated in the SAAM

¹² http://www.biographynet.nl/wp-content/uploads/2013/07/BiographyNet_schema_with_provenance_V1_2.png

dataset, italicized text shows superclasses, all properties in the saam namespace are subproperties of properties in standard vocabularies. (Szekely et al., 2013)

5.2.2. DPLA model

The Digital Public Library of America (DPLA) has developed their DPLA Metadata Application Profile which is based on the Europeana Data Model but also incorporates specific requirements from the DPLA community. Their current version (V3, Feb 2013)¹³ reuses EDM specific classes and properties as well as definitions from the following sources (DPLA, 2013):

- Resource Description Framework (RDF) and the RDF Schema (RDFS)
- OAI Object Reuse and Exchange (ORE)
- Dublin Core namespaces (dcelements, dcterms, and dcmitype)
- The Basic Geo (WGS84 lat/long) Vocabulary¹⁴

Their core classes are *edm:WebResource*, *ore:Aggregation* and *dpla:SourceResource*. The *dpla:SourceResource* is a subclass of *edm:ProvidedCHO* which includes all what DPLA calls source resources which are equivalent to EDM's cultural heritage objects. Attributes that belong to the actual object and not the digital representation are attached to *dpla:SourceResource*. They also added a class *dpla:Place* which holds properties like *dpla:city*, *dpla:state*, etc. that are expressed as sub-properties of *dcterms:spatial*. Also they added a property for coordinates which does not have an adequate mapping to EDM. In their next version they plan to re-use Geonames for geographical properties. In the class *ore:Aggregation* they have added a property *dpla:originalRecord* which is defined as subproperty of *ore:aggregates* and should be used for describing the complete original record.

6. Ongoing and future work

6.1. Recommendations from the Task force on EDM mappings, refinements and extensions

Since the creation of the EDM many mappings, refinements and extensions have been produced without a central reference that would bring them together in a single overview. The Task Force has answered this need and collected mappings to and extensions of EDM, documenting and summarising work from different partners and projects into an overall 'big picture'.

In general, the analysis of the mappings, refinements and extensions as well as the results from the survey highlights a few recommendations for data providers, or related to the EDM model itself.

First, data providers can learn a lot by sharing their mappings, refinements and extensions and by looking at the efforts done within the Europeana network. The collected documentation provides some patterns that could help data providers to tackle the various challenges highlighted in the sections above: creation of identifiers, creation of rich contextual resources, etc. The documentation on extensions and refinements should also encourage data providers to use specialisations of EDM or to create their own application profiles. In the longer run, Europeana could also support extensions defined by data providers. Ideally extensions should be handled following RDF patterns where a general

¹³ <http://dp.la/info/wp-content/uploads/2013/04/DPLAMetadataApplicationProfileV3.pdf>

¹⁴ <http://www.w3.org/2003/01/geo/>

level and a more specific level can co-exist within the same database. It would leave data consumers with different interests to choose the level that fits their needs best.

Some changes done in the EDM model could facilitate the mapping process, such as the support of the *edm:Event* class. The implementation of this class would allow data providers working with LIDO for instance to improve their mappings and to provide richer metadata.

The work on the Task Force has highlighted the need to coordinate and collect the mappings, refinements and extensions produced by Europeana data providers and related projects. Mappings, refinements or extensions to/of EDM are very relevant for the different actors contributing metadata to Europeana as they are a means to guarantee good interoperability of the metadata and a high level of data quality. Sharing the documentation related to these techniques and specifications is crucial for metadata interoperability within the Europeana ecosystem.

Based on the analysis done in the report, the Task Force provides a few recommendations for data providers and projects doing similar interoperability efforts. The mappings, refinements and extensions can be represented via different means: spreadsheets, transformation files, etc. However, the Task Force has identified some key elements that are required for the re-use and the understanding of those mappings, refinements and extensions by a third party.

Documentation about original metadata and mappings to EDM should always include:

- A semantic definition of each metadata element
- Information about how to handle the cardinality constraints and occurrences of each metadata element when mapping metadata
- Constraints related to the structure of some metadata elements (elements that are refinements of others such as *dc:coverage* and *dcterms:spatial* in Dublin Core, or hierarchical parent-child relationships)
- Constraints on the value of an element (e.g.: literals vs URIs, use of a controlled vocabulary).

In general the documentation of an Application profile should re-state the recommendations provided at schema-level.

- Information related to the context of production of this mapping (version date, domain represented, etc.)

An extension document should contain the same elements as the general documentation with additional information about:

- The namespaces extending the target metadata schema and how they have been integrated
- Additional constraints if required.
- The motivations of the extension (particular technical requirements, domain-specific needs, users requirements).

This summary is copied from the final task force report (Charles & Olensky, 2014).

6.2. EDM Roadmap

The efforts made by data providers and projects with EDM are summarised in an EDM roadmap maintained by the Europeana Foundation. This roadmap identifies the ongoing activities related to EDM and plans its future implementation in the core Europeana services. The current state of the roadmap is provided in the Annex. Europeana will primarily focus its development activities in 2014 around the extension of EDM for the support of annotation,

the development of an EDM profile for sound materials and the representation of rights information in the context of the content-reuse framework¹⁵.

6.3. Profile for representing sounds

The Europeana Sound project¹⁶ has just started to work on a EDM model refinement for sound. The project will provide recommendations for the refinement of EDM for audio and audio-related objects to ensure richer metadata description and enhanced discovery. This work will be based on the work done within the EUScreen and the Europeana Film Gateway (EFG) projects.

¹⁵ <http://pro.europeana.eu/web/europeana-creative/extended-europeana-licensing-framework>

¹⁶ <http://pro.europeana.eu/web/europeana-sounds>

7. References

- BiographyNET. <http://www.biographynet.nl/>
- CARARE / EDM case study (2012). Retrieved January 27, 2014 from <http://pro.europeana.eu/carare-edm>
- Charles, V. et al. (2013). Recommendations for the representation of hierarchical objects in Europeana. Retrieved January 27, 2014 from <http://pro.europeana.eu/documents/468623/4a6eb2ec-4cc6-48b1-8824-92a1e564a279>
- Charles, V. & Olensky, M. (2014). Report on Task force on EDM mappings, refinements and extensions. Retrieved January 27, 2014 from <http://pro.europeana.eu/web/network/europeana-tech-/wiki/Main/Task+force+on+EDM+mappings+refinements+and+extensions>
- Clayphan, R. (2014). D5.3. Further development of the Europeana Data Model. To be published.
- DM2E (2014a). Dublin Core Wiki. <http://wiki.dublincore.org/index.php/DM2E>
- DM2E (2014b). DM2E Model V 1.1 Specification http://wiki.dm2e.eu/wiki/images/8/8f/DM2E_Model_V1.1_Rev1.3_Final_Specification.pdf
- Doerr, M., Gradmann, S., LeBoeuf, P., Aalberg, T., Bailly, R. & Olensky, M. (2013). Final Report on EDM – FRBRoo Application Profile Task Force. Retrieved January 27, 2014 from <http://pro.europeana.eu/documents/468623/1760978/TaskfoApplication+Profile+EDM-FRBRoo>
- DPLA. <http://dp.la/info/developers/map>
- Europeana Libraries (2012). D5.1 Report on the alignment of library metadata with the European Data Model (EDM). Version 2.0 (2012). <http://www.europeana-libraries.eu/documents/868553/245c353b-ec25-480f-96d4-fc2e0a58d445>
- EDM collection profile (2014). Retrieved March 20, 2014 from <http://pro.europeana.eu/documents/900548/90890f64-777b-40d6-af78-593864a41072>
- EDM dataset profile (2014). Retrieved March 20, 2014 from <http://pro.europeana.eu/documents/900548/3345fdd3-9d81-4828-a1f6-da099a394e4e>
- EDM organisation profile (2014). Retrieved March 20, 2014 from <http://pro.europeana.eu/documents/900548/2f1b6cfd-b803-4f01-8128-6749b59cd2da>
- MIMO / EDM case study (2012). Retrieved January 27, 2014 from <http://pro.europeana.eu/mimo-edm>
- Ockeloen, N., Fokkens, A., ter Braake, S., Vossen, P., de Boer, V., Schreiber, G. & Legêne, S. (2013) BiographyNet: Managing Provenance at multiple levels and from different perspectives. In: Proceedings of the Workshop on Linked Science (LiSC) at ISWC 2013, Sydney, Australia, October 2013.
- Polymath Virtual Library and EDM case study (2012). Retrieved March 19, 2014 from <http://pro.europeana.eu/polymath-edm>
- Szekely, P., Knoblock, C., Yang, F., Zhu, X., Fink, E., Allen, R., & Goodlander, G. (2013). Connecting the Smithsonian American Art Museum to the Linked Data Cloud. In P. Cimiano, O. Corcho, V. Presutti, L. Hollink, & S. Rudolph (Eds.), *Lecture Notes in Computer Science. The Semantic Web: Semantics and Big Data* (pp. 593–607). Berlin, Heidelberg: Springer.
- Vanstappen, H. (2012). Europeana Data Model – Fashion Profile (EDM-fp)
- Wickett, K.M., Isaac, A., Fenlon, K., Doerr, M., Meghini, C., Palmer, C.L & Jett, J. (2013). Modeling cultural collections for digital aggregation and exchange environments. Retrieved March 24, 2014 from <http://hdl.handle.net/2142/45860>

