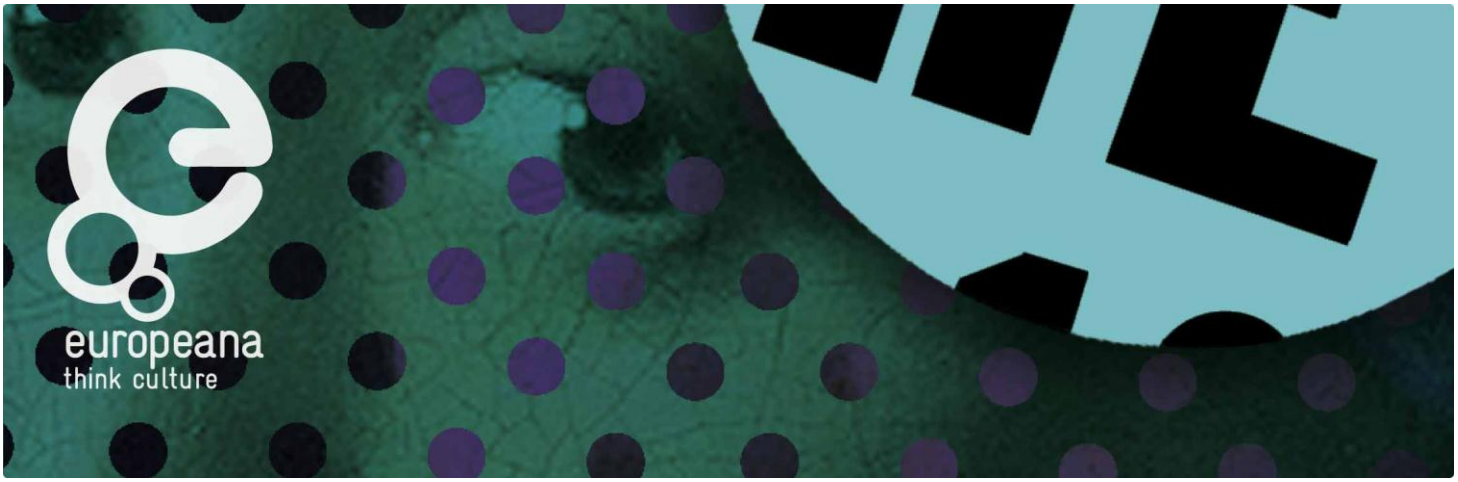


Report on Task force on EDM mappings, refinements extensions

Contributors

Alessia Bardi, HOPE
Antoine Isaac, Europeana Foundation
Amy Rudersdorf, DPLA
Erik Mannens, DCA
Esther Guggenheim, Judaica and DM2E
Eva Coudyzer, Linked Heritage and AthenaPlus
Gerda Koch, OpenUp!
Henk Vanstappen, Europeana Fashion
Jean-Pierre Evain, EUScreen
Kate Fernie, CARARE and 3DIcon
Kerstin Arnold, APEx
Marko Knepper, Judaica and DM2E
Pierfrancesco Bellini, ECLAP
Rodolphe Bailly, MIMO
Regine Stein, Linked Heritage, Partage Plus, Europeana Inside and AthenaPlus
Sam Coppens, DCA
Stefanie Ruehle, German Digital Library and Europeana Libraries
Stefanie Gehrke, Europeana Regia
Steffen Hennicke, Digital Manuscripts to Europeana (DM2E) project.

Authors

Valentine Charles, Europeana Foundation
Marlies Olensky, Humboldt-Universität zu Berlin

1. Introduction

The Europeana Data Model (EDM) plays an important role in Europeana's network of projects and providers. The number of contributing data providers is steadily increasing within this network and with this also the number and variety of metadata formats, schemas, etc. that Europeana has to integrate. Many projects and partners have been working on mappings as a way to enable metadata interoperability with other Europeana metadata, making decisions based on domain-specific requirements and possibilities offered by the EDM model. In this process, some providers have minimised the loss of metadata from their original metadata format to EDM by further extending and refining it. Experiments with EDM have shown that the model can be used to describe metadata at various levels of granularity. Data providers either directly map their metadata to EDM according to the guidelines (Europeana, 2013) or create specialisations of EDM for representing specific use-cases.

Since the creation of the EDM many of these mappings, refinements and extensions have been created 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'.

2. The Europeana Data Model

The *Europeana Data Model* (EDM) is the current data model used by Europeana to collect, connect, represent and enrich the metadata descriptions provided by Europeana data providers. It allows the articulation of descriptions and links between provided cultural heritage objects (CHO) (painting, book, archaeology site, recordings, films…) and their digital representations; as well as between objects and metadata records describing them. EDM is based on the principles of the Semantic Web and provides more opportunities to provide rich and interlinked metadata. Defining mappings between original data models and EDM is crucial for data providers wishing to deliver metadata to Europeana.

The main requirements identified for the development of EDM included:

- Distinction between a “provided item” (a painting, a book…) and its digital representations;
- Distinction between an item and the metadata record describing it;
- Allowing aggregation of multiple records for the same item, containing potentially contradictory statements about it;
- Provide support for contextual resources, including concepts from controlled vocabularies.

These principles underlie key decisions when mapping the metadata to EDM. EDM provides a selection of Dublin Core properties (such as *dc:title*, *dc:creator*) and additional EDM-specific properties for describing objects. It enables the use of Dublin Core properties with fully-fledged resources, rather than literals. New EDM properties have been created to provide more semantics

to the data: to allow the representation of more specific relationships between objects such as the "aboutness" of a link or the similarities between objects, etc. EDM also supports contextual resources which help building a "semantic layer" (Gradmann, 2010; EuropeanaConnect, 2011) including concepts from "value vocabularies" or knowledge organisation systems (KOS) like thesauri, authority lists and classifications on top of the aggregated objects (Isaac et al., 2011).
By creating different layers of description, EDM provides different levels of granularity and enables the specification of domain-specific application profiles. When mapping their metadata to EDM, Europeana data providers consider the following types of resources:

- Provided Cultural Heritage Objects or CHOs (*edm:ProvidedCHO*) denote the original objects—either physical (painting, book, etc.) or born-digital (3D model), which are the focus of description and search in Europeana. The choice in granularity of description chosen for the ProvidedCHO belongs to the data provider, within the limits of relevance set by Europeana.
- Web Resources (*edm:WebResource*) are a digital representation of the provided cultural heritage object, published on the web.
- Aggregations (*ore:Aggregation*) group the Provided CHO and the Web Resource(s) into one bundle, and information on the aggregation process is also recorded (e.g., the provider of the metadata) as shown in Figure 1.
- EDM also defines contextual resources that can be used to provide more information related to the object (e.g., *edm:Agent, edm:Place, edm:Concept, edm:TimeSpan*).

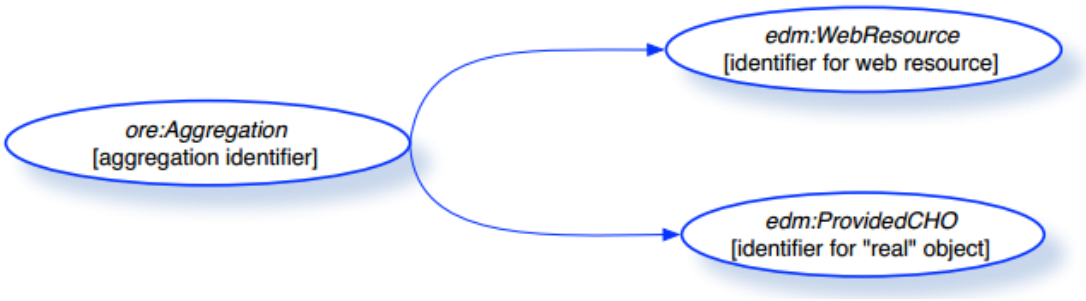


Figure 1. EDM core classes for data providers

3. Collection and classification of mappings, refinements and extensions

3.1. Gathering of the mappings, refinements and extensions

The objective of the Task Force was to collect the various mappings that have been created by data providers and projects working with Europeana. A survey was distributed to a group of selected people and to the wider network. The aim of the survey was to collect key metadata that could be used to describe the mappings.
The following information was collected:

- Name of the project and domain represented
- Type of mapping (name of the standard representing the original metadata)
- Date of creation of this mapping as well as the last version
- Mapping document: the mapping can be a text document, an excel table or an XSLT, etc.
- Have you extended or specialised EDM? If so, please specify.
- Would your project benefit from an implementation by Europeana (beyond mere storage, i.e., for display or search) of your EDM extension?
- Background information linked to the creation of the mapping (short note on their specific requirements supported by the mapping. Strong choices made when doing the mapping should be written here.)
- Any other comments

In the process of gathering and describing the contributions collected by the Task Force we articulated the following (non-exclusive) categories:

- Mappings to EDM or to an extension of EDM
- Refinements of EDM or EDM Application Profiles
- Extensions of EDM

The categories are further defined in the following sections and do not exclude each other.

3.2. What is a mapping to EDM?

A mapping is any kind of specification (e.g., a concordance table) of relationships and equivalences between two metadata formats or models, e.g., between EDM and another metadata schema. A mapping defines the relationships between elements of two metadata schemas on a structural and a semantic level (Haslhofer & Klas, 2010). In this report, we consider the aspect of schema-level correspondences as well as the instance transformation as being a defining part of the mapping. The instance transformation specifies functions that allow interoperability of the content values (e.g. combining the first and last name into one field that is called creator).This definition of a metadata format mapping can also be referred to as crosswalk (Chan & Zeng, 2006a; NISO 2004). The goal of crosswalks is to enable elements defined in one metadata format to be available to communities using related metadata formats. Figure 2 illustrates an example mapping table.

Applicable class	EDM mandatory properties	SoundCloud	Example of a value for processing	Remarks
Aggregation	<edm:dataProvider>	Nederlands Instituut voor Beeld en Geluid		This is a default value.
Aggregation	<edm:isShownAt>	[permalink_url]	[permalink_url]: http://soundcloud.com/beeldengeluid/het-overslaan-van-golven-over	
Aggregation	<edm:isShownBy>	[id]	[id]: 49084894	Based on this ID the API should be able to return a streaming url/player embed. Or maybe there is a default method to parse one?
Aggregation	<edm:provider>	Digitale Collectie		This is a default value.
Aggregation	<edm:rights>	[license]	[license]: cc-by-sa	Based on the value SoundCloud returns, a CC license url can be parsed (SoundCloud uses v 3.0
Aggregation	<edm:aggregatedCHO>			
ProvidedCHO	<dc:title>	[title]	[title]: Het overslaan van golven over een pier en misthoornsignalen	
ProvidedCHO	<dc:description>	[description]	[description] Westerschelde: overslaan van golven over een pier, met misthoornsignalen op de achtergrond	
ProvidedCHO	<dc:subject/>	[tag_list]	[tag_list]: "Eigen Opnames" zeeën Nederland golfbrekers golven misthoorns Westerschelde rivieren Zeeland geo:lat=51.44289 geo:lon=3.46507	Everything, excluding the values that include "geo:lat" and "geo:lon" or terms from "Geografische namen" axis of our GTAA thesaurus. (NOTE: GTAA terms can probably be

Figure 2. Example of mapping table

In order to provide metadata to Europeana, data providers need to conform to the EDM model as implemented by Europeana. Mappings either refer to the full EDM specification (Europeana, 2012) or the current EDM implementation (Europeana, 2013), which is a subset of the full specification. The current EDM implementation defines mandatory properties but also leaves out some of the classes and properties defined in the full specification. In some cases a mapping also includes correspondences between the source metadata fields and new elements extending EDM that have been defined as part of the mapping effort. Note that these extensions are not yet supported for aggregation into Europeana. Table 1 lists the mappings collected by the Task Force in alphabetical order by source format; table 2 lists the mappings collected by the Task Force in alphabetical order by target format.

Table 1. List of mappings collected by the Task Force in alphabetical order by source format

Source format or source model	Target format or model
ABCD(EFG) (OpenUp! format)	EDM
Denkxweb	DDB data model (DDB extension to EDM)
DIFDC	DDB data model (DDB extension to EDM)
Dublin Core	DPLA data model (extension to EDM)
	EDM
EAC-PF (APEX EAC-CPF)	EDM
EAD	DDB data model (DDB extension to EDM)
	DM2E data model (DM2E extension to EDM)
	EDM
EAD (APEX EAD)	EDM
ECLAP	EDM
FRBRoo	EDM
GMA	EDM

HOPE data model	EDM
LIDO	EDM
	DDB data model (DDB extension to EDM)
MAB2	DM2E data model (DM2E extension to EDM)
MARC21	EDM
MARCXML	DDB data model (DDB extension to EDM)
	DPLA data model (extension to EDM)
	DM2E data model (DM2E extension to EDM)
METS (APEX METS)	EDM
METS/MODS	EDM
	DPLA data model (extension to EDM)
	DDB data model (DDB extension to EDM)
	DM2E data model (DM2E extension to EDM)
	EDM (obsolete version)
METS/MODS (HidaXML+MAB)	EDM
MODS	DPLA data model (extension to EDM)
	EDM
PICA3	EDM
Proprietary Formats (DM2E)	DM2E data model (DM2E extension to EDM)
Proprietary Formats (DPLA)	DPLA data model (extension to EDM)
Proprietary formats (MARC & DC)	EDM
SCRAN (proprietary XML)	EDM
SoundCloud	EDM
TEI	EDM
TEI P5	DM2E data model (DM2E extension to EDM)
UNIMARC	EDM

Table 2. List of mappings collected by the Task Force in alphabetical order by target format

Target format or model	Source format or source model
DDB data model (DDB extension to EDM)	Denkxweb
	DIFDC
	EAD
	LIDO
	MARC XML
	METS/MODS
DM2E data model (DM2E extension to EDM)	EAD
	MAB2
	MARCXML
	METS/MODS
	Proprietary Formats (DM2E)
	TEI P5
DPLA data model (extension to EDM)	DC

	MARC XML
	MODS
	METS/MODS
	Proprietary Formats (DPLA)
EDM	ABCD(EFG) (OpenUp! format)
	Dublin Core
	EAC-PF (APEX EAC-CPF)
	EAD
	EAD (APEX EAD)
	ECLAP
	FRBRoo
	GMA
	HOPE data model
	LIDO
	MARC21
	METS (APEX METS)
	METS/MODS
	METS/MODS (HidaXML+MAB)
	MODS
	PICA3
	Proprietary formats (MARC & DC)
	SCRAN (proprietary XML)
	SoundCloud
	TEI
	UNIMARC
EDM (obsolete version)	METS/MODS

3.3. What is an EDM refinement / application profile?

A refinement is any kind of specialisation of EDM to meet specific needs of the data provider. Often a set of guidelines or rules 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. Refer to section 4.1.1 of the report for examples of refinements. Strictly speaking, an “extension” can also be seen as a “refinement”, but one that requires the addition of sub-classes or sub-properties to capture the precise usage in the source format. For the sake of this Task Force we distinguish between these two categories.

The definition of an EDM refinement corresponds to the definition of an application profile in the literature. An application profile is a set of metadata properties, policies and guidelines defined for a specific application with specific requirements. An application profile (DCMI, 2005):

- describes what a community wants to accomplish with its application (Functional Requirements);

- characterizes the types of things described by the metadata and their relationships (New Domain Model or existing one);
- enumerates the metadata elements to be used and the rules for their use (Guidelines)
- defines the machine syntax that will be used to encode the metadata.

Heery & Patel (2000) state the following defining aspects for an application profile:

- May draw on one or more existing namespaces
- Introduce no new metadata elements
- May specify permitted schemes and values
- Can refine standard definitions

In the context of this Task Force an application profile for one element set can qualify as a mapping. According to the literature (e.g. Heery & Patel, 2000), an application profile should not declare any new metadata elements. Yet, in some cases the definition of a new application profile will coincide with the creation of an extension to capture all the details required. In this case an extension can be part of an application profile[1]. Table 3 lists the refinements / application profiles collected by the Task Force.

Table 3. List of refinements / application profiles collected by the Task Force

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

3.4. What is an extension to EDM?

Every EDM extension is also a type of refinement of EDM, but is characterised by any addition of classes or properties (implemented by XML attributes and elements) to the original EDM specifications. An EDM extension is required when existing EDM classes and properties cannot represent the semantics of providers' metadata with sufficient details. EDM extensions are created to represent metadata at a different level of granularity required for the representation of specific metadata, from specific domains. They define a new set of classes and properties that are declared as specialisations of the existing ones. These new (sub-)classes or (sub-)properties can be taken from other existing namespaces but can also be newly created as part of an ontology or schema. These new elements specify constraints applicable to the whole model. Refer to section

[1] The Scholarly Works application profile illustrates this situation. A specific extension for eprints was created within the application profile:
http://www.ukoln.ac.uk/repositories/digirep/index/EPrints_Application_Profile

4.2 of the report for examples of extensions. Table 4 lists the extensions collected by the Task Force in alphabetical order.

Table 4. List of extensions collected by the Task Force in alphabetical order

Extension of EDM	**Characteristics**
DM2E data model	subproperties specific to manuscripts and addition of new subclasses
DPLA data model	classes and properties specific to the Digital Public Library of America
EBUCore	subproperties specific to the audiovisual / television /radio community
EDM-Fashion Profile	subproperties and subclasses specific to the fashion domain
EUScreen	subproperties specific to the audiovisual / television /radio community
German Digital Library data model	subproperties specific to the German Digital Library
Modeling Cultural Collections for Digital Aggregation and Exchange Environments https://www.ideals.illinois.edu/handle/2142/45860	EDM application profile for representing collections
EDM Paths	subclasses and subproperties specific for enrichment annotations

4. Results from the mapping analysis

The mappings, refinements and extensions defined earlier in this document have been collected and organised in a spreadsheet[2]. It provides further details on the various types of documentation collected by the Task Force and a direct access to the technical specifications of those mappings: http://pro.europeana.eu/documents/468623/89d5472f-4a7c-41fb-8fa6-7400c3fe3058

The further analysis of the corpus of mappings, refinements and extensions has allowed the identification of a set of characteristics (e.g. represented domain, strong choices made during the mapping process) which will be described in section 4.1 and 4.2.

4.1. Details on the mapping choices identified in the documentation

The survey highlighted that providers had to cope with some particular issues when defining the mapping from their metadata format to EDM. In order to solve those issues, providers took specific

[2] The working version of the spreadsheet is available at https://docs.google.com/spreadsheet/ccc?key=0AqVfQb4_fRp1dFlXdVFGX1NmOEVoZFB4THY1amdtYWc&usp=sharing#gid=0

decisions on how to implement the mapping. Those decisions can be categorized as (Chan & Zeng, 2006b):

- Schema (or ontology) level decisions: decisions that affect any records formatted using the schema.
- Instance level decisions: decisions that affect specific (sets of) metadata records to be transformed with the mapping.

Note that in both cases the analysis done by the Task Force can include element-based and value-based approaches.

4.1.1. Schema-level decisions

When mapping their metadata to EDM, data providers first need to compare the semantics of their source and target metadata elements sets. Therefore, the mapping decisions involve choices related to the semantics, i.e. the meaning of the metadata elements and their refinements, and to the declarations and instructions on the use of the values declared in the EDM metadata schema (such as the *edm:type* values).

Schema- level recommendations in EDM guidelines

The EDM mapping guidelines for data providers as well as the EDM specifications define rules on how the metadata should be formed and a mapping should take this into account. For instance, the EDM documentation provides a description of the semantics and the content of each metadata element used in the schema, and recommendations on how to use it. Figure 3 shows an example of how to use *dc:language* in Europeana, which recommends the use of the ISO 639 standards for the representation of language codes.

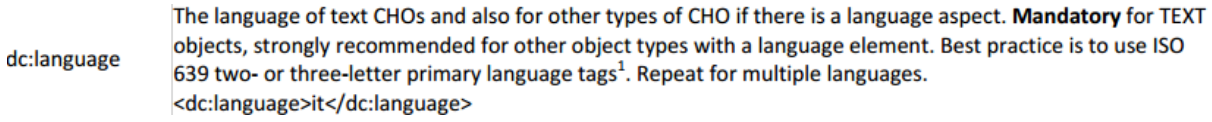


Figure 3. Example of schema-level recommendation in the EDM mapping guidelines

Definition of the Cultural Heritage Object in EDM

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 CARARE [3] project (Isaac, Charles & et al, 2013) representing the archaeological and architectural domain, the CHO is an immoveable Heritage Asset such as a monument, building or another real world object. For the DM2E project [4] (Iwanowa, Dröge, & Hennicke, 2013) a CHO could be a book but also a page or a paragraph within the book as shown in Figure 4. 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.

[3] http://www.carare.eu/

[4] http://dm2e.eu/

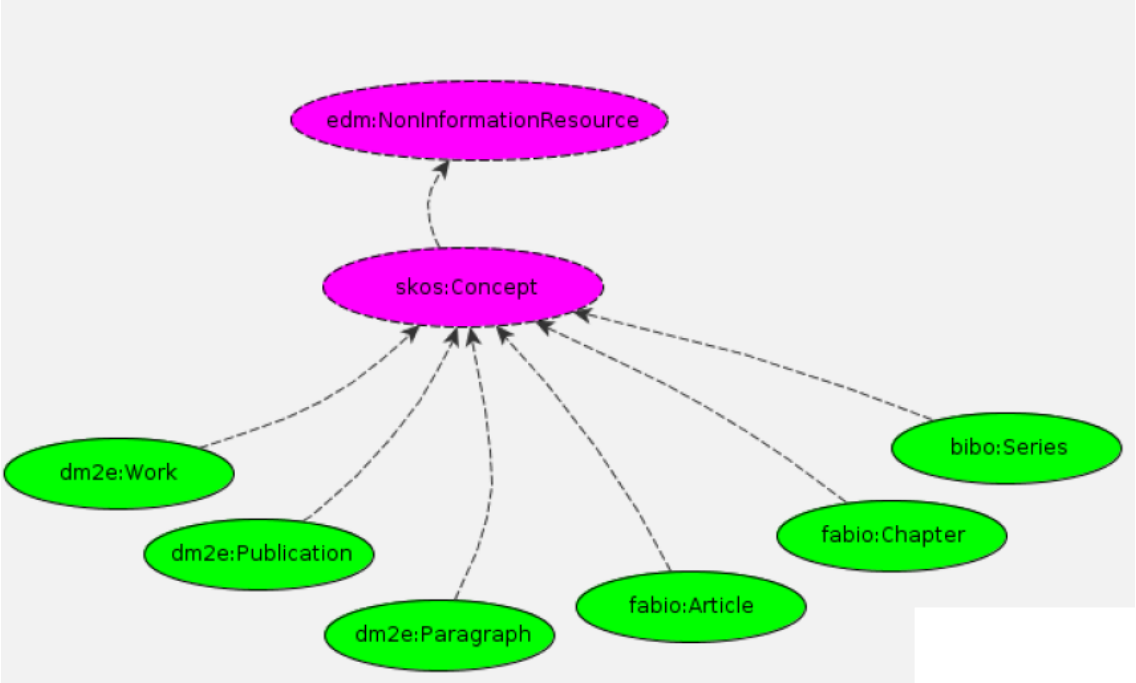


Figure 4. Levels of granularity defined by DM2E to describe the edm:ProvidedCHO

EDM schema specifications: the example of (ordered) hierarchies

EDM defines some elements that need to be used strictly according to the specifications when doing the mappings. This is the case for elements that represent ordered hierarchies. In order to represent hierarchies EDM provides a series of metadata elements that need to be used in a particular pattern to be valid. For example, EDM specifies that the vertical relationships between a whole ProvidedCHO and its parts can be expressed with two specific properties:

- has-part relation (*dcterms:hasPart* property) for top-down relationships;
- is-part-of relation (*dcterms:isPartOf* property) for bottom-up relationships.

The horizontal relationships between the parts of a resource (as given for instance by the consecutive numbering of the parts or by pagination) can be expressed with the is-next-in-sequence-to property (*edm:isNextInSequence*). It needs to be used in a specific direction: it relates a part in a sequence with the part immediately preceding it. For appropriately rendering hierarchical relationships, these elements require the use of URIs (the ones of the ProvidedCHOs involved in the relationships) instead of literals (i.e., mere labels).

Note that in some cases, like HOPE, the sequential relationships are represented by linking an object to the next one (which is the inverse relationship of *edm:isNextInSequence*). The issue concerning the direction of the relationships had to be solved in order to provide Europeana with correct sequences without implementing an extension to EDM. The solution devised was to process HOPE records and to calculate the inverse relationships and map the calculated values into *edm:isNextinSequence*.

Figure 5 shows another example of MODS-METS data represented in EDM. The example comes from the German Digital Library which re-uses relational metadata elements in its extension. For instance two ProvidedCHOs in a parent-child relationship are related using the *dcterms:isPartOf* metadata element[5].

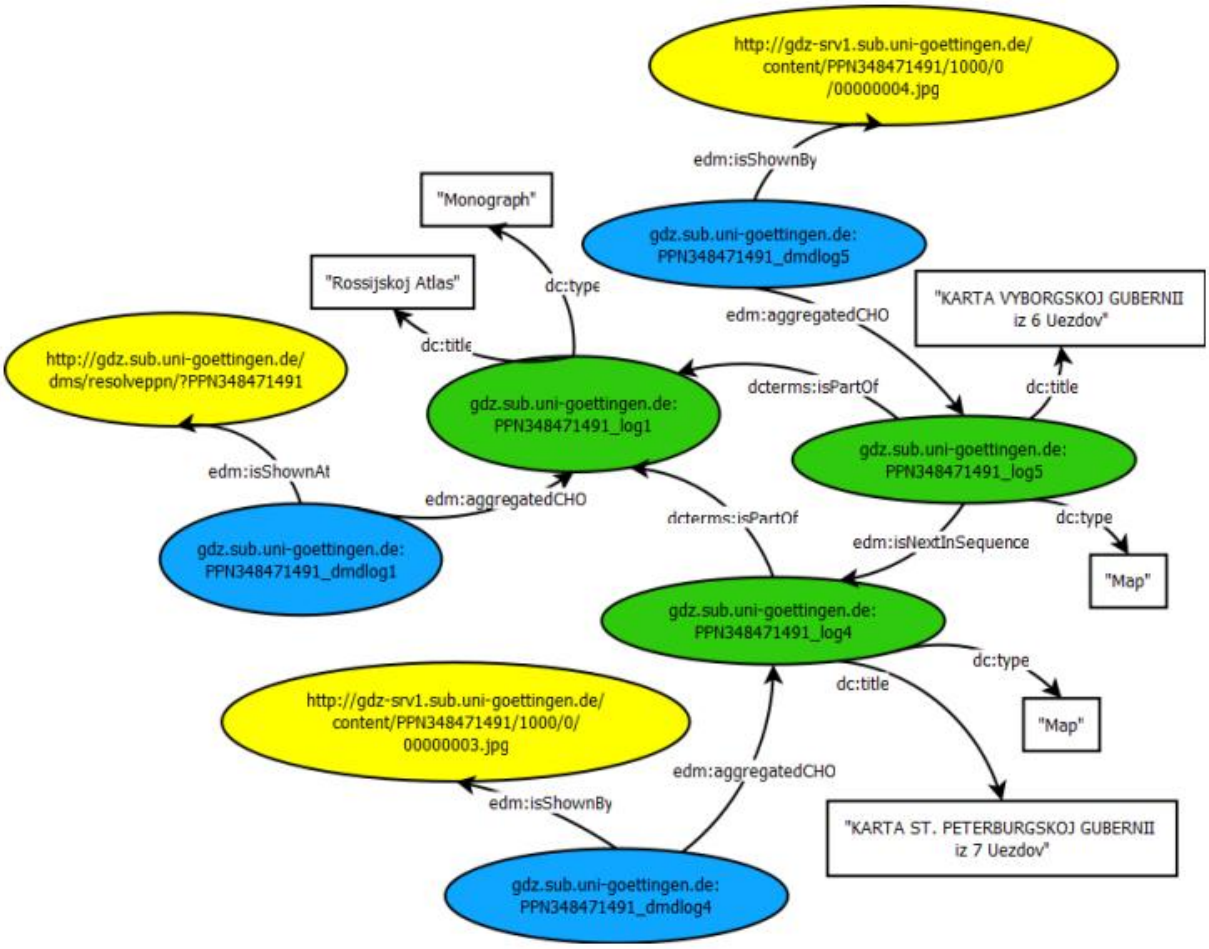


Figure 5. Example of MODS-METS data represented in EDM

Richer metadata via the use of contextual resources

A final noticeable aspect in all the mappings collected is the effort made by data providers to provide rich data using EDM classes. EDM allows the description of individual entities related to a Cultural Heritage Object such as place, agent, time and concept. The mapping of the data to the contextual resources involves decisions related to the use of relevant vocabularies available in the institutions or on the Web. Data providers use the EDM constructs to provide richer data and align their metadata to domain specific vocabularies.

The MIMO project has provided a great deal of information on resources related to their musical instruments (defined as CHOs by the project) in addition to the descriptive metadata. In the context

[5] See more examples at http://pro.europeana.eu/documents/468623/5e3d852c-0e24-4a17-bc55-1bda1388c8bb

of the project, MIMO created a domain-specific thesaurus 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 6).

```
<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 6. 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 7).

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

Figure 7. Representation of an agent by MIMO using the edm:Agent class

Similarly to MIMO, the Partage Plus[6] and Europeana Photography[7] projects have used the EDM contextual resources to provide new reference resources from their vocabularies.

Other projects decided to link their existing data to existing resources available in the Semantic Web. Figure 8 shows an example from Judaica Europeana. The resource URIs used in *dc:creator* in the EDM metadata point to the RDF representation of the GND[8] authority database. GND is a widely used external authority file available as linked open data and maintained by the German National Library.

[6] See Deliverables of WP3 Metadata Enrichment at http://www.partage-plus.eu/en/contents/12,Deliverables+and+documents

[7] See D4.1 - EuropeanaPhotography Vocabulary Definition at http://www.europeana-photography.eu/index.php?en/115/deliverables

[8] http://www.dnb.de/EN/gnd

```
<edm:ProvidedCHO rdf:about="#ubffm:item:urn:nbn:de:hebis:30:2-523">
    <dc:title>Ms. hebr. oct. 6 - Nof 'ets ḥayim</dc:title>
    <dc:creator rdf:resource="http://d-nb.info/gnd/119264528"/>
    <dc:creator rdf:resource="http://d-nb.info/gnd/124255957"/>
    <dc:contributor rdf:resource="http://d-nb.info/gnd/13766365X"/>
    <dc:description>[S.l.] : 1728</dc:description>
```

Figure 8. References to the GND person authorities in the Judaica Europeana data.

The metadata are further enriched with additional references from Wikipedia[9], DBpedia[10] and VIAF[11] (cf. Figure 9).

```
<foaf:page rdf:resource="http://de.wikipedia.org/wiki/Isaak_Luria"/>
<owl:sameAs rdf:resource="http://dbpedia.org/resource/Isaac_Luria"/>
<owl:sameAs rdf:resource="http://viaf.org/viaf/120700932"/>
```

Figure 9. Additional references to the Wikipedia, DBpedia and VIAF entities

4.1.2. Record-level decisions

After having considered the EDM schema-level recommendations, data providers need to also consider the record-level recommendations. Those are crucial when starting the conversion of the metadata. The decisions made at the record-level have a strong impact on the quality of the metadata after their conversion.

Creating identifiers (URIs)

One of the first challenges for data providers when mapping their metadata for future conversion is the creation of identifiers for the various entities that need to be described in EDM. Since it relies on the principles of the Semantic Web, EDM requires a URI for each described entity.

Most of the data providers already have URIs available in their metadata for the *edm:ProvidedCHO* and for the WebResources. If this is not the case they usually generate those URIs from their local identifiers. The project OpenUp! has for instance solved the issue by concatenating the values of a few metadata elements to provide a URI for the *edm:ProvidedCHO* (cf. Figure 10).

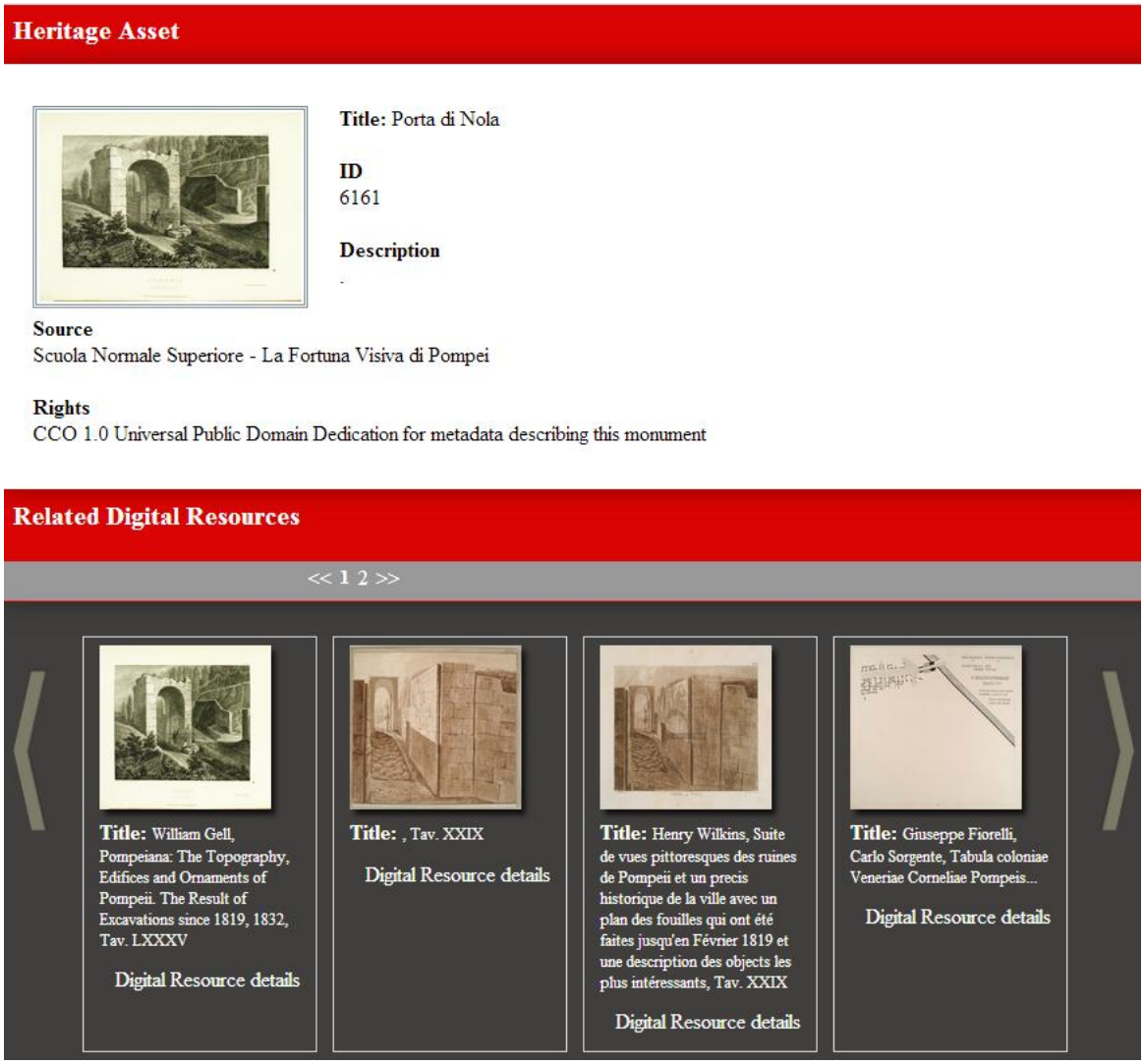
/DataSets/DataSet/Units/Unit/SourceInstitutionID /DataSets/DataSet/Units/Unit/SourceID /DataSets/DataSet/Units/Unit/UnitID	edm:ProvidedCHO rdf:about=”URI”

Figure 10. Creation of the ProvidedCHO URI by the project Open-Up!

However, data providers do not always have URIs for the *ore:Aggregation* as this resource is a very specific one to the EDM model. The project CARARE created new identifiers for the

[9] http://en.wikipedia.org/
[10] http://dbpedia.org/
[11] http://viaf.org/

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 11). The landing page provides a unique identifier that can be used in the metadata.



http://store.carare.eu/landing-pageha.php?id=iid:2920150&eid=HA:6161
Figure 11. Landing page used by CARARE as Identifier of the Aggregation class

The HOPE project creates persistent identifiers (PIDs) through the Handle System for each HOPE entity. PIDs are persistent and resolvable, and they are used as identifiers for *edm:ProvidedCHO* and *edm:WebResource*. As there is no equivalent entity in HOPE that maps to the *ore:Aggregation*, a new URI is created based on the PID used as identifier of the aggregated CHO. Other projects such as Linked Heritage, AthenaPlus or Partage Plus, that use LIDO as their metadata schema, have a step-by-step approach resulting in either:

- A traditional mapping approach: if a resource in LIDO already has a URI, a respective EDM resource is created and the source URI is described using the relevant metadata element. Or

- A mapping involving metadata creation: if the URI is not in the original metadata but is required by EDM, URIs are created while transforming the metadata.

Converting values

Another challenge lies in the different syntax of data models used by data providers and EDM. This issue is crucial since it can quickly result in a loss of metadata. When data providers use metadata elements with strings (e.g. subject fields in MARC where the values refer to a specific vocabulary), the mapping to EDM will require the conversion of this string into a URI. A similar conversion process is required when data providers use metadata elements with an identifier standing for a URI. In this case a conversion of the ID into a URI is necessary. Figure 12 illustrates this situation: an identifier referring to a GND resource is available in the data, but needs to be converted into a URI to be mapped as *edm:Agent* resource. Without this conversion process the reference to GND would be lost since the mapping would handle it as a simple string and not a rich resource.

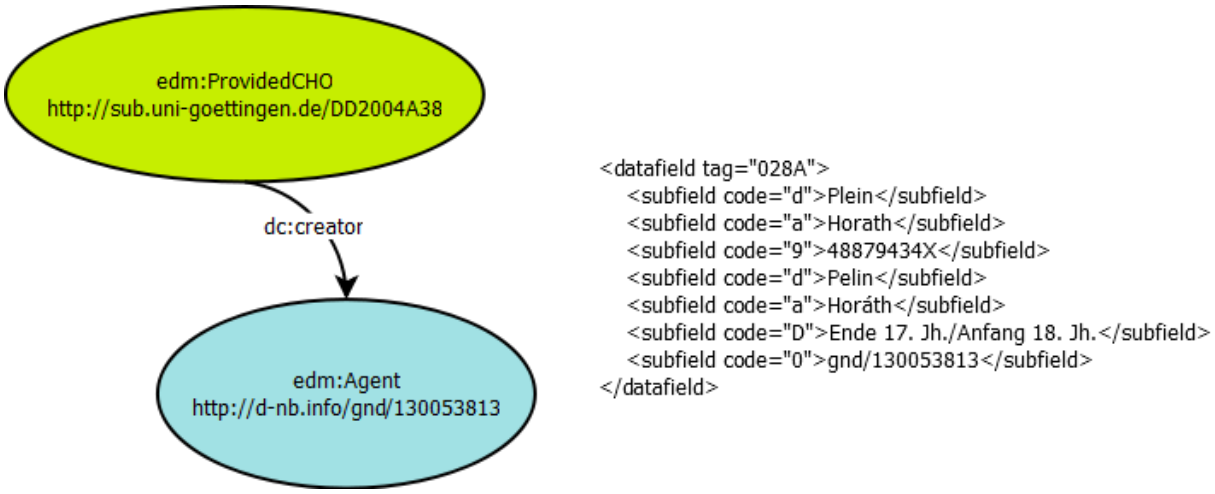


Figure 12. Conversion of an identifier referring to GND into a proper URI

More complex cases can be found when metadata needs to be converted from metadata schemas represented in XML to metadata schemas in RDF. Metadata schemas represented in XML often use attributes to qualify some of the metadata elements. EDM relies on the RDF model, which is based on binary relations that are used to link two individuals or an individual and a value. However there are cases which require the representation of more than one relation between two individuals. Representing attributes that add “shades” of meaning to relations require complex “reification” of these relations, as individual resources. Hence, the mapping of attributes is sometimes difficult.

For instance, the MARC record below (Figure 13) represents relations among multiple resources, such as a CHO, a creator and the creation event of the CHO, In addition a set of subfield codes (or attributes) is used to represent the relations between those resources.

100 1#$a Morgan, John Pierpont,**$d**1837-1913,**$e**collector.

Figure 13 Example[12] of a creator name associated with a role in a MARC record

The representation of the attributes in RDF can be done only if those attributes are converted into individual resources (called also reification[13]) as shown in Figure 14.

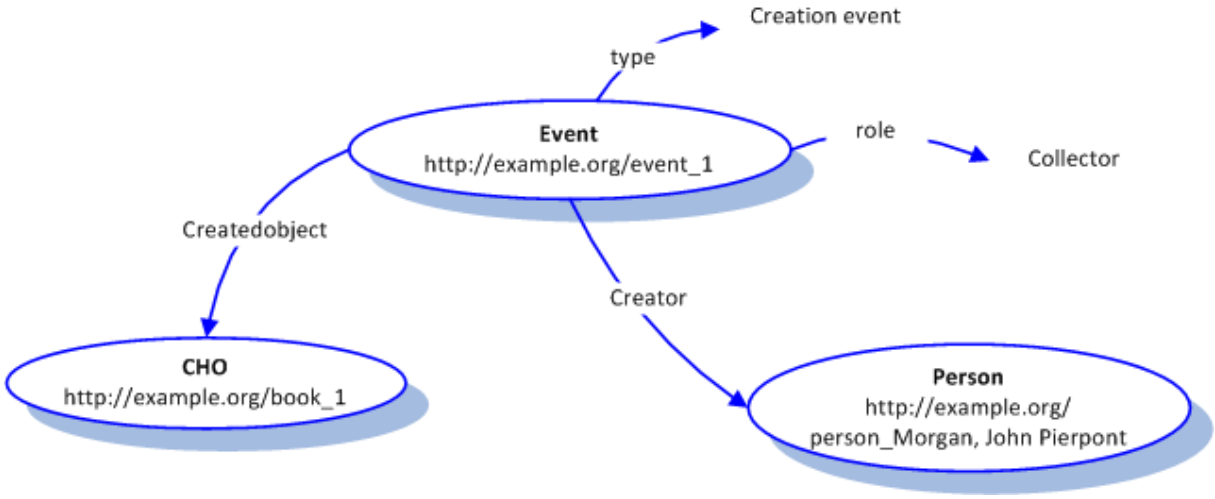


Figure 14 Example of reification of XML like attributes

In order to be able to represent the different resources and their relations in RDF, a new event class needs to be created which is then linked to the other resources described in out example. New properties are also created to describe the relationships between the different nodes. The semantics of the original attributes are therefore kept in the mapping.

Creating Europeana-specific metadata

Finally, the record-level mapping can also involve the creation of new metadata that cannot be found in the source metadata but are required in EDM. The *edm:rights* metadata element is a good example. Data providers need to indicate via a URI the type of licence that applies to the Web resources they provide to Europeana. Europeana provides a list of values data providers have to choose from. In this case, the metadata element and the related value need to be added during the metadata conversion process. A similar process happens with the *edm:type* value.

The analysis of the mappings reveals that lot of decisions need to be taken at the record level and sometimes even during the metadata conversion. These processes are quite often responsible for introducing metadata quality issues such as embedded html tags, no separation of multiple values in metadata elements as well as the loss of references to controlled vocabularies. They will usually result in a loss of semantics.

4.2. Details about the EDM extensions recorded so far

Extensions address the needs of specific communities and the selection of additional elements is done according to these needs. The creation of extensions can therefore be a way to compensate

[12] Example from http://www.loc.gov/marc/bibliographic/bd100.html

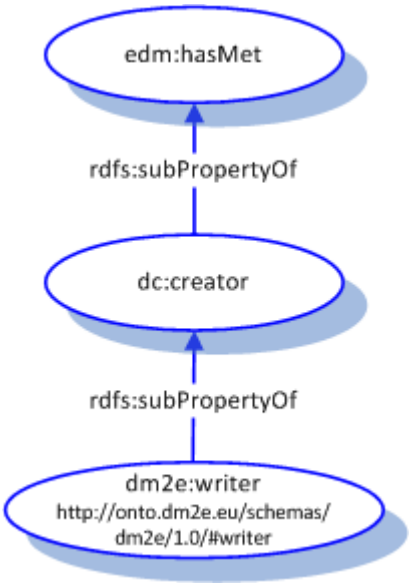
[13] For more details see http://www.w3.org/TR/swbp-n-aryRelations/

the loss of finer-grained metadata which would be present in the source metadata but not mappable to EDM. The extensions collected by the Task Force always start from the EDM model to which other element sets are added in order to cater to local and specific needs. It is important to note that the elements in these extensions can almost always be mapped to a more generic property in EDM. The extensions collected by the Task Force are for the most part classes or properties re-used from existing namespaces. Table 5 lists ontologies re-used in EDM extensions. Some projects have decided to create their own properties and have therefore declared a new namespace for it.

Table 5. List of ontologies re-used in the EDM extensions and their characteristics

Extension of EDM	**Re-used ontologies**	**Scope**
DM2E data model	Friend-of-a-Friend (FOAF) http://xmlns.com/foaf/0.1/	Types of agents
	Publishing Roles Ontology (PRO) http://purl.org/spar/pro/	Roles of agents in the publication process
	VIVO http://vivoweb.org/ontology/core#	Types of agents
	FRBR-aligned Bibliographic Ontology (FaBiO) http://purl.org/spar/	Detailed semantics on bibliographic entities
	Citation Typing Ontology (CiTO) http://purl.org/spar/cito	Detailed semantics on bibliographic entities
	Bibliographic Ontology (BIBO) http://purl.org/ontology/	Detailed semantics on bibliographic entities
	DM2E own namespace http://onto.dm2e.eu/schemas/dm2e/1.1/	Detailed semantics on bibliographic entities
DPLA data model	DPLA own namespace http://dp.la/about/map/	Additional subclasses for the DPLA service such as dpla:SourceResource, dpla:Place
	DCMI type http://purl.org/dc/dcmitype/	Class dcmitype:Collection
EDM-Fashion Profile	BIO vocabulary (BIO) http://vocab.org/bio/0.1/	Terms about people and their backgrounds with some crossover into genealogical information
	GoodRelations (GR) http://www.heppnetz.de/ontologies/goodrelations/v1	Vocabulary for e-commerce. Used for properties like colour.
	MARC Code List for Relators (Mrel) http://id.loc.gov/vocabulary/relators/	List of relator terms and associated codes to designate the relationship between a name and a resource bibliographic records.
	Friend-of-a-Friend (FOAF) http://xmlns.com/foaf/0.1/	
	RDVocab http://rdvocab.info/	Vocabulary defining the Resource Description and Access (RDA). RDA is a standard for cataloguing and is intended for use by libraries and other cultural organizations such as museums and archives

	Europeana Fashion own namespace http://www.europeanafashion.eu/edmfp	
EUScreen	EBUCore properties	
German Digital Library data model	DDB own namespace http://www.ddb.de/	Additional properties to define hierarchies. Most of the added properties are required for the DDB portal.
Modeling Cultural Collections for Digital Aggregation and Exchange Environments	Collection description terms (CDL) http://purl.org/cld/terms	Collection specific properties



To be integrated to EDM, the properties from the new element sets should be declared as specialisations of the properties used in the "standard" EDM. The specialisation of the EDM classes and properties is done by using constructs from RDF Schema[14]:

- *rdfs:subClassOf* to state that all the instances of one class are instances of another
- *rdfs:subPropertyOf* to state that all resources related by one property are also related by another

This principle of specialisation, as shown in Figure 15, allows the co-existence between a generic level (the EDM classes and properties) and a specific level (DM2E classes in the example below). In Figure 15, *dm2e:writer* is a specialisation of the more generic property *dc:creator*.

Figure 15. Principle of ontology specialization based on the RDFs properties. This figure is based on some properties available in the Europeana Data Model.

5. Further recommendations for the EDM model and conclusion

In general, the analysis of the mappings, refinements and extensions as well as the results from the survey highlighted 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

[14] http://www.w3.org/TR/rdf-schema/

could also support extensions defined by data providers. Ideally extensions should be handled following RDF patterns where a general level and a more specific level can co-exist within the same database. It would leave data consumers with different interests to choose the levels 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 specifications 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).

The work of the Task Force was limited to the identification, collection and documentation of the mappings, refinements and extensions but there is definitely a need to continue these efforts.

Further work could include the analysis of the mappings in order to further define patterns and providing more recommendations on how to map metadata to EDM or defining profiles of EDM. On a practical front, this work could give rise to the creation of a database and a simple search interface, which would allow finding the mappings to EDM for individual elements from other schemas. It would enable the re-use of the mappings, refinements and extensions created by the Europeana Network.

6. References

Chan, L. M., & Zeng, M. L. (2006a). Metadata interoperability and standardization-A study of methodology, Part I. D-Lib Magazine, 12(6), 3. Retrieved from http://www.dlib.org/dlib/june06/chan/06chan.html (December 21, 2013).

Chan, L. M., & Zeng, M. L. (2006b). Metadata interoperability and standardization-A study of methodology, Part II. D-Lib Magazine, 12(6), 4. Retrieved from http://www.dlib.org/dlib/june06/zeng/06zeng.html (December 21, 2013).

DCMI. (2005). DCMI Glossary. Retrieved from http://dublincore.org/documents/usageguide/glossary.shtml (January 6, 2014)

EuropeanaConnect. (2011). Europeana Semantic Data Layer. Retrieved from http://www.europeanaconnect.eu/results-and-resources.php?page=1 (January 21, 2014)

Europeana (2012). Definition of the Europeana Data Model elements, version 5.2.4. Retrieved from http://pro.europeana.eu/edm-documentation (January 8, 2014)

Europeana (2013). Europeana Data Model–Mapping Guidelines v2.0. Retrieved from http://pro.europeana.eu/edm-documentation (January 8, 2014)

Gradmann, S. (2010). Knowledge = Information in Context: on the Importance of Semantic Contextualisation in Europeana. Retrieved from http://pro.europeana.eu/c/document_library/get_file?uuid=cb417911-1ee0-473b-8840-bd7c6e9c93ae&groupId=10602 (January 21, 2014)

Haslhofer, B., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. ACM Computing Surveys (CSUR), 42(2), 7. Retrieved from http://eprints.cs.univie.ac.at/79/1/haslhofer08_acmSur_final.pdf (December 21, 2013).

Heery, R., & Patel, M. (2000). Application profiles: mixing and matching metadata schemas. *Ariadne*, Issue 25. Retrieved from http://www.ariadne.ac.uk/issue25/app-profiles (December 21, 2013).

Iwanowa, J., Dröge, E. & Hennicke, S. (2013). DM2E Model v1.0. Retrieved from http://onto.dm2e.eu/schemas/dm2e/1.0/ (July 12, 2013)

Isaac, A., Charles, V., Fernie, K., Dallas, C., Gavrilis, D., Angelis, S.(2013). Achieving Interoperability between the CARARE Schema for Monuments and Sites and the Europeana Data Model. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/3676 (January 8, 2014)

Isaac, A., Waites, W., Young, W. & Zeng, M. (2011). Library Linked Data Incubator Group. Datasets, Value Vocabularies, and Metadata Element Sets. Retrieved from http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/ (January 21, 2014)

NISO (National Information Standards Organization). (2004). *Understanding metadata*. Bethesda, MD: NISO Press. Retrieved from http://www.niso.org/standards/resources/UnderstandingMetadata.pdf (December 21, 2013).