

|  |
| --- |
| HOPEHeritage of the People’s EuropeGrant agreement No. 250549 |

|  |
| --- |
| Details on the mapping from the HOPE Data Model to the Europeana Data Model |

|  |  |
| --- | --- |
| Version | 0.3 |
| Date | 07/10/2013 |
| Status | Final |
| Authors | Alessia BARDI (CNR) |

Revision history

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Authors** | **Modifications** |
| 0.1 | 16/04/2012 | Alessia BARDI (CNR) | Creation |
| 0.2 | 04/05/2012 | Alessia BARDI (CNR) | Completed description of the first version of the mapping from HOPE to EDM |
| 0.3 | 07/10/2013 | Alessia BARDI (CNR) | Updated mapping tables according to the feedback received from the Europeana Ingestion Team |

**Table of contents**

1. Introduction 4

2. Terms and Definitions 5

3. HOPE and Europeana Data Models 6

3.1. The HOPE Data Model 6

3.2. The Europeana Data Model 6

4. Mapping the HOPE Data Model to EDM 8

4.1. Mapping HOPE entities into EDM entities 8

4.2. Mapping details 9

**Tables**

Table 1 Mapping cross-domain properties of DescriptiveUnits 10

Table 2 Mapping DescritiveUnit: generic profile 14

Table 3 Mapping DescriptiveUtit: archive profile 14

Table 4 Mapping DescriptiveUnit: library profile 15

Table 5 Mapping DescriptiveUnit: visual profile 18

Table 6 Mapping DescriptiveUnit: audio/visual profile 19

Table 7 Mapping DigitalResource to WebResource 20

# Introduction

This document describes how metadata records compliant to the HOPE Data Model are mapped into records of the Europeana Data Model.

*For questions feedback, please contact:*

Alessia Bardi (CNR-ISTI)

Email alessia.bardi@isti.cnr.it

Skype alessiabardi

# Terms and Definitions

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Aggregator | An aggregator in the context of Europeana is an organisation that collects metadata from a group of content providers and transmits them to Europeana. Aggregators gather material from individual organisations, standardise the file formats and metadata, and channel the latter into Europeana according to the Europeana guidelines and procedures. Aggregators also support the content providers with administration, operations and training. |
| Content | Content is digital information with a reference to an individual object of the real world or is born digital. Examples: photographs, videos, letters, censorship documents, etc.  |
| Content provider | Institution that delivers content and metadata to the HOPE system. |
| HOPE Metadata Schema | The schema for (meta)data representation defined by HOPE |
| HOPE-ST | The HOPE Support Team. Members of the team are in charge of the assistance of the Providers. Contacts: hope-support@research-infrastructures.eu |
| Local schema | The schemata for (meta)data representation used by the content providers |
| Metadata | Metadata are “data about data” which are extracted from the content providers´ local databases. They describe content and can be produced by authority files or controlled vocabularies. Examples: filmographic data, temporary and spatial data, etc. |
| OAI-PMH | Open Archive Initiative - Protocol for Metadata Harvesting |

# HOPE and Europeana Data Models

The HOPE Aggregator transforms the heterogeneous XML metadata records collected from into homogenous records compliant to the HOPE metadata schema. XML field content is also harmonized according to standard formats.

The metadata records in the HOPE format are then provided to third-party applications via OAI-PMH and SRW.

Europeana will harvest HOPE metadata records via the Aggregator OAI-PMH interface.

In order to ensure the correctness and completeness of the provided records, a mapping from the HOPE Data Model and the Europeana Data Model (EDM) must be established.

In the following, the main entities of the HOPE Data Model and EDM are briefly presented.

Finally, a mapping from the HOPE Data Model to EDM is proposed.

## The HOPE Data Model

The main entities of the HOPE Data Model are:

* *DescriptiveUnit*: represents a metadata description of a real world object. DescriptiveUnits can be linked with containment relationships (hence they can establish hierarchies) and sequential relationships. DescriptiveUnits can belong to one of the following sub-entities, based on the domain of the object they describe:
	+ libraryUnit: library domain
	+ archiveUnit: archival domain
	+ visualUnit: visual domain, i.e., pictures, photos, drawings, etc.
	+ audioVisualUnit: audio/video domain, i.e., music, films, etc.
	+ dublinCoreUnit: generic domain that can be adopt when the more specific ones cannot be applied
* *DigitalResource*: represents a web accessible digital object representing (a part of) the real world object described by a DescriptiveUnit. One DescriptiveUnit can be linked to several DigitalResource (for example: pictures of the front and the back of a building). DigitalResources can be linked with each other via sequential relationships, hence establishing a “reading path” of digital resources (for example: scanned pages of a book linked to the DescriptiveUnit of the book and ordered based on page numbers). Each DigitalResource has two web links: the derivative2 links to a low-resolution artefact of the digital object; derivative3 links to a thumbnail of the digital object.

## The Europeana Data Model

The main entities of the Europeana Data Model are:

* *ProvidedCHO*: entity that represents the real-world cultural heritage object. Properties of this entity describe the real object, not its digital representations. ProvidedCHOs can be linked by containment and sequential relationships.
* *WebResource*: entity for a digital object, which is a digital representation of a ProvidedCHO. The only properties that can be set are dc:rights and edm:rights.
* *Aggregation*: this entity groups one ProvidedCHO together with its WebResources. The aggregation is linked to its WebResources via ‘hasView’ relationships and to the provided CHO via the ‘aggregatedCHO’ relationship. Properties of an aggregation include:
	+ ens:isShownBy: the URL of a web view of the object
	+ ens:isShownAt: the URL of a web view of the object in full information context (i.e., a landing page)
	+ edm:object: the URL of a representation of the object that Europeana will use to generate previews. It can be the same as isShownBy. It can be the URL to the thumbnail.

# Mapping the HOPE Data Model to EDM

The mapping proposed in this chapter is based on the information detailed in the D2.2 deliverables, the mapping sheets, and the content of the following guides and documents provided by the Europeana Ingestion Team and available online on the Europeana web site (http://pro.europeana.eu):

* Europeana Data Model Definition v.5.2.3, <http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22>
* Europeana Data Model Primer, <http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>
* Europeana Data Model Mapping Guidelines, <http://pro.europeana.eu/documents/900548/ea68f42d-32f6-4900-91e9-ef18006d652e>
* EDM Case Study: MIMO and EDM,
<http://pro.europeana.eu/mimo-edm>
* EDM Case Study: Mapping EAD to EDM, <http://pro.europeana.eu/web/guest/ead-edm>
* Europeana Rights Guidelines, http://pro.europeana.eu/documents/900548/0d423921-23e0-45fa-82a4-2ac72b3e6f38

## Mapping HOPE entities into EDM entities

Figure 1 sketches the mapping logic among HOPE and EDM entities. In summary:

* One HOPE DescriptiveUnit contains descriptive information of a real world object. Consequently it can be mapped into one EDM ProvidedCHO.
The identifier of the EDM Provided CHO is the persistent identifier URL of the HOPE DescriptiveUnit.
The HOPE isContainedBy relationship is mapped into the EDM isPartOf relationship.
The HOPE isNextInSequence-1 will be mapped into the EDM isNextInSequence (in HOPE the current record links to the next record, in EDM the current record links to the previous one).
* One HOPE DigitalResource contains information about a digital representation of a real world object or of a part of a real world object. Consequently, it can be mapped into one EDM WebResource.
EDM does not currently allow to specify sequential relationships between WebResources[[1]](#footnote-1), hence the relationship isNextInSequence between DigitalResource is not mapped.
The identifier of a WebResource is the persistent identifier URL of the HOPE DigitalResource which resolves to a derivative (file representation) of the resource.
* EDM requires at least one Aggregation to aggregate a ProvidedCHO.
One Aggregation can aggregate one and only one ProvidedCHO, together with its digital representations (via the relationships: isShownBy, isShownAt, object hasView). As a consequence, the mapping of a DescriptiveUnit generates not only a ProvidedCHO, but also the corresponding Aggregation.
The Aggregation links to the WebResources generated from the DigitalResources representing the DescriptiveUnit.
The identifier of an Aggregation is the persistent identifier (not the whole URL, which is instead used for the ProvidedCHO) of the HOPE DescriptiveUnit.
The Aggregation’s isShownBy and object link to the first WebResource of the object.
Further WebResources (i.e., the DescriptiveUnit is represented by several DigitalResources) are linked to the Aggregation via the relationship hasView.

## Mapping details

isShownBy

object

hasView

isPartOf

isNextInSequence

ProvidedCHO
(DU)

aggregatedCHO

Aggregation
(DU)

WebResource
(DR)

Figure HOPE Data Model: mapping logic to EDM

isNextInSequence

🡨 represents

isRepresentedBy 🡪

isNextInSequence

isContainedBy

Descriptive
Unit

Digital
Resource

In the following, the list of HOPE properties (version 1.1.3 of the HOPE XML schema[[2]](#footnote-2)), together with its mapping to Europeana (EDM version 5.2.3[[3]](#footnote-3)) is provided.

Based on those correspondences, an XML style sheet will be implemented to transform HOPE metadata records into EDM records.

Table 1 shows the properties that are shared among the five different profiles (sub-classes) of the HOPE DescriptiveUnit entities. From Table 2 to Table 6, properties of each sub-class are presented according to the following order: generic profile, archive profile, library profile, visual profile, audio/visual profile. Finally, Table 7 presents the property mapping for the DigitalResource entity.

Table Mapping cross-domain properties of DescriptiveUnits

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| DescriptiveUnit Property | Mandatory in HOPE/empty allowed/mult | EDM property | EDM entity (ProvidedCHO or Aggregation) | Mandatory in EDM/mult | Comments |
| persistentID | y 1 | rdf:about | Aggregation | y 1 | The about attribute is created taking the last part of the PID (NA/id) so that the generated Aggregation and ProvidedCHO have correctly two different identifers. |
| persistentID | y 1 | rdf:about | ProvidedCHO | y 1 | The persistentID is used as it is. |
| persistentID | y 1 | edm:aggregatedCHO/@rdf:resource | Aggregation | y 1 | link between the aggregation and its aggregated CHO |
| localID | y 1 | dc:identifier | ProvidedCHO | n \* |  |
| localIDCollectionItem | n \* | dc:identifier | ProvidedCHO | n \* |  |
| localIDParent | y empty 1 | X | X | X | X |
| isContainedBy | y empty 1 | dcterms:isPartOf | ProvidedCHO | n \* |  |
| localIDNext | y empty 1 | X | X | X | X |
| isNextInSequence-1(to be stored in du/misc/miscfield[@key = ‘comesAfter’]/@value) | y empty 1 | edm:isNextInSequence | ProvidedCHO | n 1 | We shall compute the inverse relationship of the HOPE isNextInSequence, since EDM wants the id of the preceding resource.  |
| aggregator | n \* | X | X | X | X |
| isSuppliedBy | y empty 1 | edm:dataProvider | Aggregation | y 1 |  |
| “HOPE – Heritage of the People’s Europe” | NA | edm:provider | Aggregation | y 1 |  |
| isRepresentedBy/persistentID | n \* | edm:hasView | Aggregation | n \* | references to WebResources id (rdf:about) |
| //digitalResource[1]/derivative2/@PID  | y 1 | edm:isShownByedm:object (for TEXT and IMAGE only) | Aggregation | n 1n 1 | HOPE descriptive units are compound objects, that is they are a composition of digital objects. We take the first derivative2 as the ‘main’ representative to sue in the edm:isShownBy.  |
| //digitalResource[1]/derivative3/@PID | y 1 | edm:object (for AUDIO and VIDEO only) | Aggregation | n 1 |  |
| landingPage | y empty 1 | edm:isShownAt | Aggregation | n 1 | One of isShownAt or isShownBy has to be provided to Europeana |
| thumbnail | y empty 1 | X | X | X | We can’t pass our thumbnail as edm:object: Europeana requires image files with a better resolution (see //digitalResource[1]/derivative2/@PID and //digitalResource[1]/derivative3/@PID) |
| metadataLanguage/@normalised | y empty 1 | X | X | X | No corresponding field in EDM. |
| title[@label=’title’] | n \* | dc:title | ProvidedCHO | n \* | Either dc:title or dc:description must be provided to Europeana |
| title[@label=’title’]/@language | n | dc:title/@xml:lang | ProvidedCHO | n |  |
| title[@label=’title’]/@script | n | X | X | X | No script attribute in EDM. |
| creator/value | n \* | dc:creator | ProvidedCHO | n \* |  |
| isCreatedBy | UNUSED | X | X | X | X |
| isCreatedIn | UNUSED | X | X | X | X |
| contributor/value | n \* | dc:contributor | ProvidedCHO | n \* |  |
| hasContributionsBy | UNUSED | X | X | X | X |
| publisher/value | n \* | dc:publisher | ProvidedCHO | n \* |  |
| isPublishedBy | UNUSED | X | X | X | X |
| isPublishedIn | UNUSED | X | X | X | X |
| description | n \* | dc:description | ProvidedCHO | n \*  |  |
| description/@language | n | dc:description/@lang | ProvidedCHO | n |  |
| description/@script | n | X | X | X | No script attribute in EDM. |
| spatialCoverage | n \* | dcterms:spatial | ProvidedCHO | n \* | One of subject, coverage or spatial must provided to Europeana |
| associatedPlace | UNUSED | X | X | X | X |
| depictedPlace | UNUSED | X | X | X | X |
| temporalCoverage | n \* | dcterms:temporal | ProvidedCHO | n \* |  |
| associatedEvent | UNUSED | X | X | X | X |
| depictedEvent | UNUSED | X | X | X | X |
| subject | n \* | dc:subject | ProvidedCHO | n \* | One of subject, coverage or spatial must provided to Europeana |
| associatedAgent | UNUSED | X | X | X | X |
| depictedAgent | UNUSED | X | X | X | X |
| associatedConcept | UNUSED | X | X | X | X |
| depictedConcept | UNUSED | X | X | X | X |
| language/@normalised | n \* | dc:language | ProvidedCHO | n \* |  |
| rights | n \* | dc:rights | ProvidedCHO  | n \* |  |
| rights/@language | n | dc:rights/@xml:lang | ProvidedCHO | n |  |
| rights/@script | n | X | X | X | No script attribute in EDM. |
| rights[@label=’europeana rights] | y 1 | edm:rights | Aggregation | y 1 | Value from Europeana Rights Guidelines. The value refers to the rights of the digital representation of the CHO. |
| provenance | n \* | dcterms:provenance | ProvidedCHO | n \* |  |
| provenance/@language | n | dcterms:provenance/@xml:lang | ProvidedCHO | n  |  |
| provenance/@script | n | X | X | X | No script attribute in EDM. |
| associatedHopeTheme | n \* | dc:subject | ProvidedCHO | n \* |  |
| relation | n \* | dc:relation | ProvidedCHO | n \* |  |
| relation/@language | n | dc:relation/@language | ProvidedCHO | n |  |
| relation/@script | n | X | X | X | No script attribute in EDM. |
| resolveURL | UNUSED | X | X | X | X |
| descriptionLevel/@normalised | y 1 | dc:type | ProvidedCHO | n \* |  |
| europeanaType | n 1 | edm:type | ProvidedCHO | y 1 | Europeana Type. One of: 3D, TEXT, IMAGE, SOUND, VIDEO. |

Table Mapping DescritiveUnit: generic profile

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dublinCoreUnit Property | Mandatory in HOPE / empty allowed/ mult | EDM property of ProvidedCHO | Mandatory/mult in EDM | Comments |
| extent | n \* | dcterms:extent | n \* |  |
| extent/@language | n | dcterms:extent/@language | n |  |
| extent/@script | n | X | X | No script attribute in EDM. |
| medium | n \* | dcterms:medium | n \* |  |
| medium/@language | n | dcterms:medium/@language | n |  |
| medium/@script | n | X | X | No script attribute in EDM. |
| date[@label=’date’]/@normalised | n \* | dc:date | n \* |  |

Table Mapping DescriptiveUtit: archive profile

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| archiveUnit | Mandatory in HOPE / empty allowed/ mult | EDM property of ProvidedCHO | Mandatory/mult in EDM | Comments |
| date[@label=’date of creation’]/@normalised | n \* | dcterms:created | n \* |  |
| appearanceOfMaterial | n \* | dc:description | n \* |  |
| appearanceOfMaterial/@language | n | dc:description/@xml:lang | n |  |
| appearanceOfMaterial/@script | n | X | X | No script attribute in EDM. |
| genreOfFonds | n \* | dc:type | n \* | Mapping suggested by Valentine Charles from Europeana |
| genreOfFonds/@language | n | dc:type/@xml:lang | n 1 |  |
| genreOfFonds/@script | n | X | X | No script attribute in EDM. |
| extent | n \* | dcterms:extent | n \* |  |
| extent/@language | n | dcterms:extent/@language | n |  |
| extent/@script | n | X | X | No script attribute in EDM. |
| physicalCharacteristics | n \* | dc:description | n \* |  |
| physicalCharacteristics/@language | n | dc:description/@xml:lang | n 1 |  |
| physicalCharacteristics/@script | n | X | X | No script attribute in EDM. |

Table Mapping DescriptiveUnit: library profile

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| libraryUnit  | Mandatory in HOPE/empty allowed/mult | EDM property of ProvidedCHO entity | Mandatory in EDM | Comments |
| date[@label=’date of publication’]/@normalised | n \* | dcterms:issued | n \* |  |
| date[@label=’date of printing or manufacture or engraving]/@normalised | n \* | dc:date | n \* |  |
| generalMaterialDesignation | n \* | dcterms:medium | n \* |  |
| generalMaterialDesignation/@language | n | dcterms:medium/@xml:lang | n 1 |  |
| generalMaterialDesignation/@script | n | X | X | No script attribute in EDM. |
| specificMaterialDesignation | n \* | dc:type | n \* |  |
| specificMaterialDesignation/@language | n | dc:type/@xml:lang | n 1 |  |
| specificMaterialDesignation/@script | n | X | X | No script attribute in EDM. |
| extent | n \* | dcterms:extent | n \* |  |
| extent/@language | n | dcterms:extent/@language | n |  |
| extent/@script | n | X | X | No script attribute in EDM. |
| publicationFrequency | n \* | ? | ? | Should be dcterms:accrualsPeriodicity, but it is not included in EDM. |
| publicationFrequency/@language | n | ? | ? |  |
| publicationFrequency/@script | n | X | X | No script attribute in EDM. |
| dimensions | n \* | dcterms:extent | n \* |  |
| otherPhysicalDetails | n \* | dc:description | ? |  |
| otherPhysicalDetails/@language | n | dc:description/@xml:lang | ? |  |
| otherPhysicalDetails/@script | n | X | X | No script attribute in EDM. |
| title[@label=’title proper] | n \* | dc:title | n \* |  |
| title[@label=’title proper’]/@language | n  | dc:title/@xml:lang | n |  |
| title[@label=’title proper’]/@script | n | X | X | No script attribute in EDM. |
| title[@label != ‘title’ and @label != ‘title proper’] | n \* | dcterms:alternative | n \* | Include: subtitles, translation, parallel titles, edition statement, etc |
| title[@label != ‘title’ and @label != ‘title proper’]/@language | n | dcterms:alternative/@xml:lang | n |  |
| title[@label != ‘title’ and @label != ‘title proper’]/@script | n | X | X | No script attribute in EDM. |

Table Mapping DescriptiveUnit: visual profile

| visualUnit | Mandatory in HOPE/empty allowed/mult | EDM property of ProvidedCHO entity | Mandatory/mult in EDM | Comments |
| --- | --- | --- | --- | --- |
| date[@label=’object production date’]/@normalised | n \* | dcterms:created | n \* |  |
| objectName | n \* | dc:type | n \* |  |
| objectName/@language | n | dc:type/@xml:lang | n |  |
| objectName/@script | n | X | X | No script attribute in EDM. |
| technique | n \* | ? | ? | Seems to be unmappable (spectrum:technique) |
| technique/@language | n | ? | ? |  |
| technique/@script | n | X | X | No script attribute in EDM. |
| material | n \* | dcterms:medium | n \* |  |
| material/@language | n | dcterms:medium/@xml:lang | n |  |
| material/@script | n | X | X | No script attribute in EDM. |
| dimensions | n \* | dcterms:extent | n \* |  |
| technicalAttribute | n \* | ? | ? | I couldn’t find a suitable mapping (spectrum:technical attribute) |
| technicalAttribute/@language | n  | ? | ? |  |
| technicalAttribute/@script | n | X | X | No script attribute in EDM. |

Table Mapping DescriptiveUnit: audio/visual profile

| audioVisualUnit | Mandatory in HOPE/empty allowed/mult | EDM property of ProvidedCHO entity | Mandatory/mult in EDM | Comments |
| --- | --- | --- | --- | --- |
| instantiationType | n \* | dc:type | n \* |  |
| instantiationType/@language | n | dc:type/@xml:lang | n |  |
| instantiationType/@script  | n | X | X | No script attribute in EDM. |
| carrierType | n \* | dcterms:medium | n \* |  |
| carrierType/@language | n | dcterms:medium/@xml:lang | n |  |
| carrierType/@script | n | X | X | No script attribute in EDM. |
| originalDuration | n \* | dcterms:extent | n \* |  |
| originalLenght | n \* | dcterms:extent | n \* |  |
| aspectRatio | n \* | dc:format | n \* |  |
| colourSystem | n \* | dc:format | n \* |  |
| gauge | n \* | dcterms:extent | n \* |  |
| soundSystem | n \* | dc:format | n \* |  |

Table 7 presents the mapping from the HOPE DigitalResource entity to the EDM WebResource entity. EDM property with a star (\*), are in the EDM model but they are not yet included in the EDM schema. For this reason, a mapping is provided but it is not going to be implemented, otherwise records can’t be validated against the current EDM XML schema. As a result, the only properties of a WebResource are rdf:about for the link to the resource, edm:rights to link to an item of the Europeana rights vocabulary, and dc:rights for rights statements. However the latter will not be used, as the HOPE DigitalResource records are defined to include only one rights property to be mapped into the Europeana rights.

Table Mapping DigitalResource to WebResource

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DigitalResource | Mandatory in HOPE/empty allowed/mult | EDM property of WebResource entity | Mandatory/mult in EDM | Comments |
| persistentID | y 1 | rdf:about | n 1 | The PID must resolve to something even when the Aggregator generates one. |
| localID | y 1 | X | X | X |
| localIDNext  | y empty 1 | X | X | X |
| derivative2/@PID | y 1 | dcterms:isFormatOf (\*) | n \* |  |
| derivative3/@PID | y 1 | dcterms:isFormatOf (\*) | n \* |  |
| transcription | y empty 1 | dcterms:isFormatOf (\*) | n \* |  |
| transcription/@PID | n | dcterms:isFormatOf (\*) | n \* |  |
| isNextInSequence-1(to be pre-calculated and saved into dr/misc/miscfield[@key=’comesAfter’]/@value) | y empty 1 | edm:isNextInSequence (\*) | n 1 | In HOPE the property contains the id of the next DR, in EDM the id of the preceding. Hence we should compute the inverse relation. This task will be added to ISTI TODO list if and only if an EDM schema supporting the feature will be adopted by the Europeana portal during the life of the HOPE project. |
| language/@normalised | n \* | X | X | Not available in EDM |
| type | y 1 | dc:format (\*) | n \* |  |
| rights | y empty 1 | edm::rights | n 1 |  |
| contentProvider | y empty 1 | X | X | X |
| represents | y 1 | X | X | X |

1. Europeana Data Model Mapping Guidelines, <http://pro.europeana.eu/documents/900548/ea68f42d-32f6-4900-91e9-ef18006d652e>, pages 14-16. [↑](#footnote-ref-1)
2. http://node1.d.hope.research-infrastructures.eu/hopeSchema1.1.3.xsd [↑](#footnote-ref-2)
3. http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22 [↑](#footnote-ref-3)