

HOPE
Heritage of the People's Europe
Grant agreement No. 250549

Deliverable D2.2

The Common HOPE Metadata Structure, including the Harmonisation Specifications

Deliverable number	D2.2
Version	1.1
Authors	Bert Lemmens – Amsab-ISG Joris Janssens – Amsab-ISG Ruth Van Dyck – Amsab-ISG Alessia Bardi – CNR-ISTI Paolo Manghi – CNR-ISTI Eric Beving – UPIP Kathryn Máthé – KEE-OSA Katalin Dobó – KEE-OSA Armin Straube – FES-A
Delivery Date	31/05/2011
Dissemination Level	Public



Table of Contents

Table of Contents	3
1. Introduction	8
1.1 Task Description	8
1.2 Approach	10
1.2.1 Task breakdown	10
1.2.2. Dublin Core Application Profile Guidelines	14
1.2.3. Used Terminology	15
1.2.4. Structure of the Document	17
1.3 Glossary	18
1.3.1 Hope glossary	18
1.3.2 Technical Glossary	18
2. Common HOPE Metadata Structure	24
2.1 Functional requirements	24
2.1.1 General	24
2.1.1.1 What does HOPE want to accomplish with the HOPE Aggregator?	24
2.1.1.2 What the HOPE Aggregator will not attempt to do?	26
2.1.2. Aggregation	26
2.1.2.1 What are the key characteristics of the HOPE metadata?	27
2.1.2.2 How do the characteristics of the HOPE metadata affect the design of the Common HOPE Metadata Structure?	28
2.1.3. Dissemination	30
2.1.3.1 What are the key characteristics of the HOPE users?	30
2.1.3.2 How does the HOPE System want to serve its users?	31
2.1.3.3 Listing the Functional Requirements of the Common HOPE Metadata Structure	33
2.2 Transformation Procedure	36
2.2.1. Definitions	36
2.2.2 Transformation Procedure	37
2.2.3. Transformation Issues regarding Semantic Interoperability	40
2.3 HOPE Data Model	42
2.3.1 Introduction	42

2.3.2 HOPE Data Model	55
2.3.2.1 Diagram	55
2.3.2.2 Entities	56
2.4 HOPE Domain Profiles	119
2.4.1 Introduction	119
2.4.1.1 Purpose of domain profiles	119
2.4.1.2 Definition of domain profiles	120
2.4.2 HOPE Domain Profiles	122
2.4.2.1 HOPE archive profile	124
2.4.2.2 HOPE library profile	127
2.4.2.3 HOPE visual profile	131
2.4.2.4 HOPE audio-visual profile	135
2.4.2.5 HOPE Dublin Core profile	138
2.5 HOPE XML Schema	141
2.5.1 Introduction	141
2.5.2 XML Schema	142
2.5.2.1 Enumerations	142
2.5.2.2 Domain	143
2.5.2.3 Bi-/one-directional relationships	143
2.5.3 Structure of the XML Schema	144
2.5.3.1 Entity level	144
2.5.3.2 Complex types	145
3. Harmonisation Requirements	153
3.1 Local and Persistent Identifiers	153
3.1.1 Introduction	153
3.1.1.1 Objective	153
3.1.1.2 Survey Results	154
3.1.2. General requirements	154
3.1.2.1 General requirement 01: identifiers and URLs	154
3.1.2.2. General requirement 02 : persistence of identifiers and URLs	155
3.1.2.3. General requirement 03: uniqueness of identifiers	155
3.1.2.4. General requirement 04: identifiers (except PIDs) should not be URLs	156
3.1.3. PID services and resolver services	156

3.1.3.1. What is a PID service?	156
3.1.3.2 Requirements for the usage of PIDs in the HOPE System - Glossary	159
3.1.3.3 Requirements for the HOPE PID Service	161
3.1.4. Scenarios and requirements for creation, supply and update of identifiers by content providers	161
3.1.4.1 Structure of scenarios	162
3.1.4.2 Requirements for metadata records (i.e. descriptive units) and landing pages	162
3.1.4.3 Requirements for digital files	164
3.1.4.4 Requirements for CP authority records	172
3.2. Granularity	175
3.2.1 Survey Results	175
3.2.2 Europeana and Granularity	176
3.2.3 Recording Granularity in the Common HOPE Metadata Structure	177
3.2.4 Hierarchical Descriptions in the Descriptive Unit Entity	178
3.2.4.1 Elements	178
3.2.4.2 Hierarchical structures	182
3.2.5 Compound Digital Objects in the Digital Resource Entity	194
3.2.5.1 Elements	198
3.2.5.2 Acceptance Levels	200
3.3. Descriptive and Administrative Metadata	203
3.3.1 Mandatory, Recommended & Optional Elements	203
3.3.1.1 HOPE Aggregator	204
3.3.1.2 Europeana	205
3.3.1.3 HOPE@Europeana:	207
3.3.1.4 IALHI Portal	207
3.3.2. Metadata Normalisation	213
3.3.2.1 Dates	214
3.3.2.2 Languages	216
3.3.2.3 Scripts	218
3.3.2.4 Description Level	219
3.3.2.5 Rights	220
3.3.2.6 Type	221
4. Case Studies Semantic Enrichment	223

4.1. Common HOPE Authorities	223
4.1.1. Survey results	223
4.1.1.1 Content Providers Survey	223
4.1.1.2 Additional Survey	225
4.1.2. Mapping and merging procedure	227
4.1.2.1 HOPE Authority File Manager Tool	227
4.1.2.2 <i>Agents</i>	230
4.1.2.3 <i>Places</i>	231
4.1.2.4 <i>Concepts</i>	232
4.1.2.5 <i>Events</i>	232
4.1.3 HOPE Themes	233
4.2. Multilingual Access	235
4.2.1 Introduction	235
4.2.1.1 Sources of information	235
4.2.1.2 Purpose of this document	235
4.2.1.3 Building a multilingual controlled vocabulary	237
4.2.2 Hope Metadata: Language and Scripts, Translations and Transliterations	238
4.2.2.1 Languages / translations	238
4.2.2.2 Scripts / transliterations	239
4.2.3 Resources, Tools and Services	241
4.2.3.1 Introduction	241
4.2.3.2 Availability of controlled vocabularies and tools	242
4.2.3.3 MultiMatch project	243
4.2.3.4 EuropeanaConnect	244
4.2.3.5 Hitime	246
4.2.3.6 Multilingual vocabularies for subjects / topical terms	247
4.2.3.7 Conclusion	249
4.2.4 Recording Multilingual HOPE Metadata	252
4.2.4.1 Translation issues	252
4.2.4.2 Transliteration issues	255
4.2.4.3 Issues related to the use of special characteristics and diacritics	259
4.2.5 Enrichment: Building Multilingual HOPE Metadata	262
Labels	262

Linked Data	262
Online Translation Tools	262
Transliterations	262
Multilingual vocabularies	263
Persons & Places	263
HOPE Historical Themes	264
BIBLIOGRAPHY	266
APPENDIX A, B, C	266

1. Introduction

This chapter includes the description of the tasks 2.3 and 2.4 as presented in the HOPE Description of Work and the way these tasks have been approached.

1.1 Task Description

This deliverable includes the specifications for the Common HOPE Metadata Structure and the subsequent harmonization requirements. These specifications and requirements have been drafted within the scope Tasks 2.3 and 2.4 of the HOPE Description of Work. The task description of these two tasks reads as follows:

T2.3 Specification of the common metadata structure

This task deals with the HOPE data model at the more syntactical level.

T2.3.1

An inventory will be made of the metadata standards and schemes applied locally by the participants, the level of granularity of the descriptive units and the local practices for producing previews/thumbnails (low res scans, video stills, etc.) and the use of unique identifiers for referential integrity. These will be matched with the Europeana specifications (metadata scheme, previews/thumbnails sizes and formats and the object link) in order to assess the range of interoperability issues and the extent of the harmonization and enrichment work that needs to be done.

T2.3.2

This task will then specify a common metadata structure for interoperability within HOPE and full-interoperability with Europeana. This will include:

- Specification of the overall metadata framework
- Agreed standards for metadata (descriptive, admin and structural)
- Agreed levels of descriptive granularity per content object type (collection, archive, book, periodical, image, etc.)
- Agreed standards for digital content representations (thumbnails, preview, low-res, etc.) per content object type
- Agreed common format for cross-searching
- Requirements for resolvable identifiers (incl. requirements for ensuring that the identifiers are compatible with different resolution solutions in case of migration from one solution to another)

The specifications of the common HOPE metadata structure (incl. acceptance criteria) will be part of a WP2 deliverable and integrated in the HOPE wiki.

T2.4 Specification of harmonisation requirements.

This task involves all issues that need to be addressed to ensure optimal harmonisation and enrichment of the metadata at the semantic level.

T2.4.1

This task will establish agreement on metadata encoding schemes and metadata cleansing practices to be carried out (either by WP3 local implementations or by the semi-automatic data curation tool in WP4). The schemes of particular importance to HOPE concern relevant entities to social history such as events, persons, organisations, places and time expressions/periodisation. There are as yet no authoritative semantic schemes for such entities, but based on the local source data (subject headings, thesauri, classifications) this task will establish agreement on how to map such data to the common list of major historical events (see task 1.6). This task will also look into the practical possibilities to merge the local authority lists into a common list of persons/organisations and the requirements for the management of Authority Files, which will be used in WP4 as well.

T2.4.2

This task will establish agreement on recording metadata in foreign languages and how to tackle local practices of transliterations and translations. This task will work together with CNR-ISTI (WP4 leader) and look into the effects of applying tools and translation resources developed by MultiMATCH to reach agreement on the practicalities of implementing such solutions.

The outcomes of this task are agreed harmonisation specifications (incl. acceptance criteria) and they will be part of a WP2 deliverable and will feed into the HOPE BPN wiki.

1.2 Approach

This chapter describes the approach of T2.3 and T2.4 to the specification of the HOPE Common Metadata Structure and the subsequent harmonisation requirements.

1.2.1 Task breakdown

Between M4 (August 2010) and M10 (February 2011), the task partners involved in Task 2.3 and Task 2.4 have specified the common HOPE metadata structure and its formal representation in the HOPE Data Model and the HOPE XML Schema. Hence, these tasks have also specified the requirements for the normalisation of the Content Providers metadata, the supply and maintenance of local and persistent identifiers, and the implementation of hierarchically structured metadata and compound objects. These tasks have also looked into the possibilities and feasibility of creating common HOPE authority lists and multilingual access to the HOPE metadata. The results of these tasks are published in D2.2 (M11, March 2011), which is now lying in your hands.

Since both tasks 2.3 and 2.4 heavily depend on each other and seek agreement on a wide range of issues, WP2 decided to make a combined roadmap and divide the task descriptions into 11 sub-tasks.

T2.3 Specification of the common metadata structure This task deals with the HOPE data model at the more syntactical level.

T2.3.2 This task will then specify a common metadata structure for interoperability within HOPE and full-interoperability with Europeana (**sub-task 5**). This will include:

- Specification of the overall metadata framework (**sub-task 1**)
- Agreed standards for metadata (descriptive, admin and structural) (**sub-task 2**)
- Agreed levels of descriptive granularity per content object type (collection, archive, book, periodical, image, etc.) (**sub-task 3**)
- Agreed standards for digital content representations (thumbnails, previews, low-res, etc.) per content object type. (**sub-task 4**)
- Agreed common format for cross-searching (**sub-task 6**)
- Requirements for resolvable identifiers (incl. Requirements for ensuring that the identifiers are compatible with different resolution solutions in case of migration from one solution to another) (**sub-task 7**)

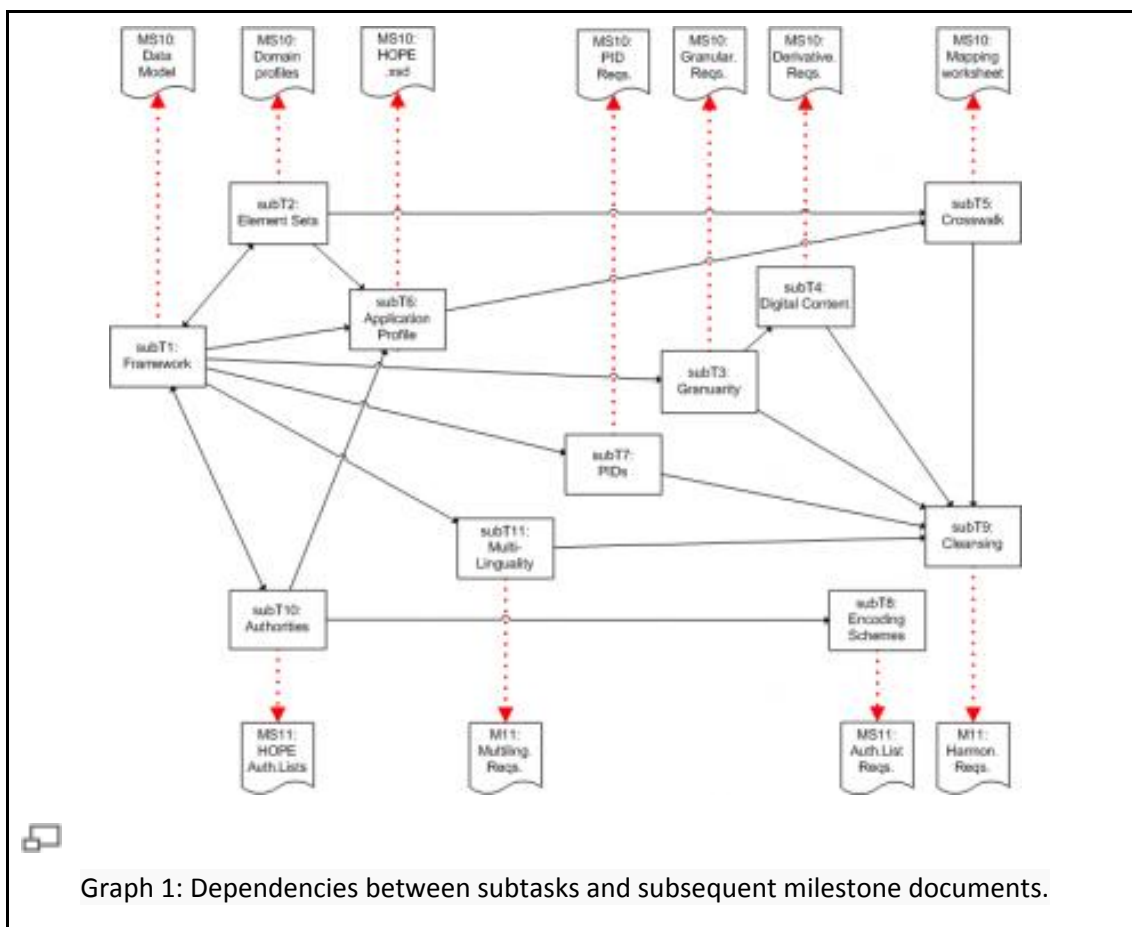
T2.4 Specification of harmonisation requirements. This task involves all issues that need to be addressed to ensure optimal harmonisation and enrichment of the metadata at the semantic level.

T2.4.1 This task will establish agreement on metadata encoding schemes (**sub-task 8**) and metadata cleansing practices (**sub-task 9**) to be carried out (either by WP3 local implementations or by the semi-automatic data curation tool in WP4). The schemes of

particular importance to HOPE concern relevant entities to social history such as events, persons, organisations, places and time expressions/periodisation. There are as yet no authoritative semantic schemes for such entities, but based on the local source data (subject headings, thesauri, classifications) this task will establish agreement on how to map such data to the common list of major historical events (see task 1.6). This task will also look into the practical possibilities to merge the local authority lists into a common list of persons/organisations and the requirements for the management of Authority Files (**sub-task 10**).

T2.4.2 this task will establish agreement on recording metadata in foreign languages and how to tackle local practices of transliterations and translations (**sub-task 11**). This task will work together with CNR-ISTI (WP4 leader) and look into the effects of applying tools and translation resources developed by multiMATCH to reach agreement on the practicalities of implementing such solutions. The outcomes of this task are agreed harmonisation specifications (incl. acceptance criteria) and they will be part of a WP2 deliverable and will feed into the HOPE BPN wiki.

The dependencies between these 11 subtasks are show in the graph below. The corresponding MS10 and MS11 documents report on all sub-tasks and have been incorporated in this deliverable.



Each sub-task dealt with a concrete issue the HOPE partners had to agree on. For each sub-

task we specified objectives, actions and distribution of work. The objectives for each task have been listed below.

- **Sub-task 1**

Agreement on the general principles and standards for the specification of the common HOPE metadata structure, including:

- the relations between the local metadata structures, the common HOPE metadata structure, and Europeana, the IALHI Portal and Social Sites (such as Flickr, Youtube, Scribd)
- the data model and format representing the common HOPE metadata structure, and
- the metadata standards/namespaces to be integrated in the HOPE data model.

The results of this subtask are incorporated in chapters 2.1 till 2.5.

- **Sub-task 2**

This sub-task is a further specification of the HOPE data model and seeks agreement on the element sets (domain profiles) for descriptive, administrative and structural metadata.

Based on the results of the Content Providers Survey and best practice within each 'community, this sub-task identifies encoding standards for accommodating rich metadata about archival, library, museum and audio-visual collections.

The results of this subtask are incorporated in chapter 2.4.

- **Sub-task 3**

This sub-task seeks agreement on the required/advised granularity levels for archival, library and visual and audio-visual descriptions, including the identification of all subsequent harmonisation issues. These requirements have to match with the granularity requirements for Europeana.

The results of this subtask are incorporated in chapter 3.3.

- **Sub-task 4**

This sub-task seeks agreement on the requirements for supplying derivatives. This agreement includes formats and resolutions for thumbnails and previews. These requirements have to match with the granularity requirements for Europeana.

The results of this subtask are incorporated in chapter 3.2.

- **Sub-task 5**

This task seeks agreement on the procedure of mapping the HOPE profile to the EDM profile, in order to make HOPE interoperable with Europeana.

This sub-task involves the integration of EDM mappings already developed by other Europeana related projects.

The results of this subtask are incorporated in chapter 2.5.

- **Sub-task 6**

This sub-task deals with the specification of a HOPE format, i.e. a serialized format of the common HOPE metadata structure, specified in sub-tasks 1, 2 and 10.

This HOPE format will be used for storing HOPE metadata in the HOPE aggregator, and for supplying metadata to the Labour History Portal.

The results of this subtask are incorporated in chapter 2.5.

- **Sub-task 7**

This sub-task specifies the requirements unique identifiers, including the requirements to make some of these identifiers resolvable. These requirements depend on used PID technology and involvement in HOPE compliant repositories.

The results of this subtask are incorporated in chapter 3.1.

- **Sub-task 8**

This task will specify the procedure for enrichment of the HOPE metadata.

This procedure includes 'encoding' of common authority lists, using the D-NET infrastructure.

The results of this subtask are incorporated in chapter 4.1.

- **Sub-task 9**

This sub-task seeks agreement on the manual and semi-automated harmonisation work to be executed in order to make the original metadata compliant with the HOPE application format.

The results of this subtask are incorporated in chapter 3.3.

- **Sub-task 10**

This sub-task is a further specification of the HOPE data model and seeks agreement on the specification of separate entities for persons, organisations, subjects, geographical locations, and time/period expressions.

The results of this subtask are incorporated in chapter 2.3.

- **Sub-task 11**

This sub-task seeks agreement on the way the HOPE aggregator deals with multilingual metadata, e.g. translation of element labels, free-text metadata and controlled vocabularies.

This sub-task also seeks agreement on the scope of the translation and transliteration work to be done by WP3, as well as the implementation of

[MultiMatch](#)¹ tools by WP4.

The results of this subtask are incorporated in chapter 4.2.

1.2.2. Dublin Core Application Profile Guidelines

In order to organise the findings of the 11 subtask into a deliverable specifying a common metadata structure and its subsequent requirements, T2.3 and T2.4 have used [Dublin Core Application Profile Guidelines](#)² as guideline for this document.

A Dublin Core Application Profile (DCAP) is a framework for developing an application profile for a specific purpose. It seeks to address the needs of the metadata in this specific application, while ensuring semantic interoperability with other applications on the basis of globally defined vocabularies and models. The guidelines are an initiative of the Dublin Core Metadata Initiative, however a profile does not require the use of metadata terms defined by DCMI. The [guidelines document](#)³ was used by HOPE as a reference in the process of writing the specifications for the Common HOPE Metadata Structure.

The DCAP guidelines document defines four parts of a profile.

“... a Profile:

- **PART 1** *describes what a community wants to accomplish with its application (Functional Requirements);”*

In this deliverable document we covered this under:

- **2.1 Functional Requirements** (based on the work performed under sub-task 1)
- **2.2 Transformation Procedure** (based on the work performed under sub-task 1)
- **PART 2** *“characterizes the types of things described by the metadata and their relationships (Domain Model);”*

In this deliverable document we covered this under:

¹See <http://www.multimatch.org/>

²See: <http://dublincore.org/documents/profile-guidelines/>

³Dublin Core Application Profile Guidelines, <http://dublincore.org/documents/profile-guidelines/>

- **2.3 HOPE Data Model** (based on the work performed under sub-task 1)
- **PART 3** “enumerates the metadata terms to be used and the rules for their use (Description Set Profile and Usage Guidelines); and”

In this deliverable document we covered this under:

- **2.4 HOPE Domain Profiles** (based on the work performed under sub-task 2)
- **4.1 HOPE Authorities** (based on the work performed under sub-task 10)
- **3.1-3.3 Harmonisation Requirements** (based on the work performed under sub-task 3, 4, 7 and 11)
- **PART 4** “defines the machine syntax that will be used to encode the data (Syntax Guidelines and Data Formats).”

In this deliverable document we covered this under:

- **2.5 HOPE schema** (based on the work performed under sub-task 5 and sub-task 6)

1.2.3. Used Terminology

Some concepts which are used throughout the different chapters are specific for the deliverable D2.2, so they are not present in the general HOPE glossary. To avoid any ambiguity we will elicit them here.

Basically, it involves following concepts:

- The **Common HOPE Metadata Structure** as such is a structured set of metadata elements, accommodating descriptive, administrative as well as structural metadata about HOPE collections.
- A **Domain Profile** is a subset of metadata elements, borrowed from a domain-specific metadata standard, accommodating descriptive, administrative as well as structural metadata about HOPE collections that belong to one specific domain. HOPE provides 5 subsets, one for each domain type and each related with a specific metadata standard.
 - the archive profile is based on the APEnet/EAD standard;
 - the library profile is based on the MARC21 bibliographic standard;
 - the audio-visual profile is based on the EN15907 standard;
 - the visual domain profile is based on the LIDO standard;
 - the ‘generic’ Dublin Core profile is based on the Dublin Core standard.

HOPE Domain Profiles are used as an intermediate metadata structure when mapping local metadata elements to the common HOPE Metadata Structure.

- The high level **HOPE Data Model** is an abstraction of the Common HOPE Metadata Structure and Domain Profiles in terms of information about their physical or conceptual entities and their relationships. The HOPE data model is a high-level design of the Aggregators metadata architecture and is the reference document for the implementation of the Common HOPE Metadata Structure in the D-NET infrastructure.
- The **HOPE Format** is a serialized representation of the Common HOPE Metadata Structure using XML, in order to be processed by the HOPE aggregator. The metadata, encoded in the HOPE Format, is validated by the **HOPE Schema**, which is specified using the XML Schema language, also known as an XML Schema Definition (XSD).
- The HOPE Glossary 3.0 defines a **PID** as: “An identifier that is globally unique and permanently identifies a given item. In HOPE, a PID is a Uniform Resource Identifier (URI) that can be expressed using two syntaxes: PID-URN: Identifier only without any web location, and PID-URL: Identifier with the Web location of the root resolver. PIDs must be persistently resolvable on the Internet and are always associated with a Resolve URL.” In this deliverable a PID is consistently used in the meaning of a PID-URL. On a PID-URL the HOPE Glossary 3.0 states: “When PIDs are used as URLs, they take the form of a URL-address. The PID-URL is based on the address of the root site of the PID resolver service. Example (Handle system): <http://hdl.handle.net/1066/1112>”.

These concepts related to each other as follows:

- The Common HOPE Metadata Structure is a metadata structure, used for the storage of metadata in the HOPE System. The common HOPE metadata structure will act as a pivot or switchboard between local metadata structures and metadata structures used by discovery services harvesting metadata from the HOPE aggregator.
- In addition to the Common HOPE Metadata Structure, HOPE specified five domain-specific metadata structures, named domain profiles. These profiles will create an intermediary step in the transformation process from local metadata structures to the common HOPE Metadata Structure.
- The Common HOPE metadata structure and the domain profiles themselves can be considered as a series of discrete lists of logical units, capturing information about a HOPE resource in a human readable form, and should be distinguished from the corresponding HOPE Data Model, HOPE Format and the five domain-specific

Formats, which are technical representations of the Common HOPE Metadata Structure and the domain profiles.

1.2.4. Structure of the Document

This deliverable consists of three chapters.

- The first chapter contains the functional requirements and detailed specifications of the Common HOPE Metadata Structure and the Domain Profiles, as well as the corresponding Data Model and HOPE Format. This chapter includes contributions by AMSAB-ISG, FES and CNR-ISTI and refers to D1.2.
- The second chapter contains the subsequent harmonisation requirements, as to local and persistent identifiers, granularity, and descriptive and administrative metadata. This chapter includes contributions by AMSAB-ISG, UPIP and KEE-OSA.
- The third and final chapter contains two case studies as to the enrichment of metadata, i.e. the development of common HOPE authorities and enabling multi-lingual access to the HOPE Metadata. This chapter includes contributions by AMSAB-ISG, CNR-ISTI and KEE-OSA.

1.3 Glossary

1.3.1 Hope glossary

The general HOPE Glossary V2.0 is available as a separate publication from T2.1.

1.3.2 Technical Glossary

This technical WP2 Glossary is an addition to the general HOPE Glossary. It seeks to clarify the specific technical terms used in deliverable D2.2. This glossary will define these terms both in their general meaning, and in the sense in which they should be conceived in the deliverable.

If you are looking for terms that are not present in this technical Glossary, please consult the general HOPE Glossary v.2.

<p>Collection Item</p>	<p>A Collection Item is the original analogue or born-digital object.</p> <p>It can consist of a single object, such as an archival document, a publication, a photograph, a movie.</p> <p>A Collection Item can consist of a set of single objects that are created as a whole.</p> <p>Collection Items can be grouped in series, such as archival series, periodicals, <i>Formal Collections</i>, TV programs.</p>	<p>Source: HOPE</p> <p><i>In Europeana terms a Collection Item is denoted as a Cultural Heritage Object</i></p>
<p>Content Providers' Metadata</p>	<p>Content Providers' Metadata is metadata provided by Content Providers to the Aggregator.</p> <p>Content Providers' Metadata is compliant with a set of technical requirements in order to cope with low-level interoperability issues (agreements on protocols and export format) and to be harvested and transformed by the Aggregator.</p> <p>Content Providers' Metadata has not</p>	<p>Source: HOPE</p> <p>See also:</p>

		See also: <i>Digital Resource Entity</i>
Digital File	<p>A Digital File is an image, text, film or sound recording, encoded as a binary computer file. A Digital File contains a version of a <i>Digital Representation</i>.</p> <p>A digital master file is a version of a <i>Digital Representation</i>, created for the purpose of reuse and re-expression. From a single master file, multiple versions can be derived for specific uses, i.e. derivative files.</p> <p>A digital derivative file is a version of a Digital Representation, created for a specific use, such as printing, viewing of the Digital Representation.</p>	<p>Source: HOPE</p> <p>See also: <i>Digital Representation,</i> <i>Digital Object,</i> <i>Digital Resource Entity</i></p>
Digital Object	<p>The expression 'Digital Object' usually refers to a <i>Digital File</i> or a number of <i>Digital Files</i> that contains or contain <i>Digital Representations</i> of a Collection Item.</p> <p>In the HOPE Data Model, a Digital Object is the name for one or more <i>Digital Resource Entities</i> that are associated with a <i>Descriptive Unit Entity</i>.</p> <p>A Digital Object represents one single <i>Collection Item (Descriptive Unit Entity)</i>, even if that item is represented by multiple images (<i>Digital Resource Entities</i>). If different representations exist for a single Collection Item, e.g. different pictures for each side of a coin, then the whole of these representations is what is called the Digital Object, i.e. the Digital Object of the coin is the collection of all the information on the pictures of both sides of the coin.</p> <p>A simple Digital Object consists of exactly one <i>Digital Resource Entity</i>.</p>	<p>Source: HOPE</p> <p>See also: <i>Digital Representation,</i></p>

	A Compound Digital Object consists of multiple <i>Digital Resource Entities</i> .	<i>Digital File</i> , <i>Digital Resource Entity</i>
Digital Representation	<p>A Digital Representation is a single digital image or audio-visual / sound recording, that provides a single, unique rendition of a Collection Item</p> <p>In the HOPE system, a Digital Representation can have multiple versions:</p> <ul style="list-style-type: none"> • a Master • a High-resolution Derivative • a Low-resolution Derivative • a Preview (image, AV clip, sound fragment) • a Thumbnail (image) • an OCRed text <p>Each version is expressed in a <i>Digital File</i>.</p>	<p>Source: HOPE</p> <p>See also: <i>Digital Object</i>, <i>Digital File</i>, <i>Digital Resource Entity</i></p>
Digital Resource Entity	<p>A Digital Resource Entity contains information about a <i>Digital Representation of a Collection Item</i>.</p> <p>A Digital Resource Entity is associated with a Descriptive Unit Entity that describes the represented <i>Collection Item</i>. A <i>Descriptive Unit</i> can hold associations to more than one Digital Resource Entity, e.g. as with a coin description being associated to two pictures, one for each side of the coin.</p> <p>The Digital Resource Entity contains Structural and Administrative Metadata about digital master and derivative files, stored in a Local or Shared Object Repository.</p> <p>If a representation exists in more than one file, then the Digital Resource Entity should contain information on all such files, e.g. if a picture of the flip side of a coin exists in one TIFF, three JPEGs and an OCR-layer in text format, information on all five files should be contained in one Digital Resource Entity.</p>	<p>Source: HOPE</p> <p>See also: <i>Digital Representation</i>,</p>

		<i>Digital File,</i> <i>Digital Object</i>
Formal Collection	<p>A series of <i>Collection Items</i> that have been created and/or published as one Collection.</p> <p>A Formal Collection can be a series of photographs created and/or published by one photographer, a series of prints created by one artist.</p>	Source: HOPE See also: <i>Informal Collection</i>
HOPE Metadata	<p>HOPE Metadata is metadata represented in the HOPE Data Model.</p> <p>The Aggregator creates HOPE Metadata by:</p> <ul style="list-style-type: none"> • Transforming <i>Content Providers' metadata</i> from the CPs' data model to the HOPE domain profiles' data models. • Transforming metadata from the HOPE Domain Profiles' data models to the HOPE Data Model. 	Source: HOPE See also: <i>Content Providers' Metadata</i>
Informal Collection	<p>A series of <i>Collection Items</i> that have been accumulated into one Collection.</p> <p>An Informal Collection can be a series of photographs collected by a person or an institution</p>	Source: HOPE See also: <i>Formal Collection</i>
Mapping	<p>A Mapping is a set of "rules", which specify how the values contained by the elements of a source <i>Metadata Record</i> should move to the elements of a target metadata record.</p> <p>A Mapping can also express how such values should be converted, refined, corrected, removed or merged with other values to obtain new values.</p> <p>In the HOPE system, a Mapping is defined by:</p>	Source: HOPE

	<ul style="list-style-type: none"> ○ The Content Provider for the Mappings from the data model of the different Collections to the Domain Profiles. ○ WP2 for the Mappings of the Domain Profiles to the HOPE Data Model. 	<p>See also: <i>Crosswalk</i></p> <p>Use for: <i>Mapping Rules</i></p>
Mapping Rules	Use: <i>Mapping</i>	
Mapping Worksheet	A Mapping Worksheet is a document in which the <i>Mapping Rules</i> are being defined.	Source: HOPE See also: <i>Mapping</i>
Metadata Record	<p>A Metadata Record is a set of Metadata Elements that contains information about one or more <i>Collection Items</i>.</p> <p>A Metadata Record can encode one or more entities of the HOPE Data Model.</p> <p>In the HOPE System, Metadata Records are supplied as XML files, grouped in Data Sets.</p>	Source: HOPE See also: <i>Collection Item</i>
XML Transformation	<p>An XML Transformation is the process of applying a set of <i>Mapping Rules</i> to one or more source <i>Metadata Records</i> to obtain one or more target <i>Metadata Records</i>.</p> <p>In the HOPE System, this process consists of a sequence of steps, including granularity mapping steps, semantics/structural mapping steps and validation steps.</p>	Source: HOPE See also: <i>Crosswalk</i>

2. Common HOPE Metadata Structure

This chapter contains the specifications of the Common HOPE Metadata Structure and the corresponding HOPE Data Model, XML Schema and Domain Profiles.

The Common HOPE Metadata Structure is a metadata structure, used for the transformation, harmonisation and storage of metadata in the HOPE System. This metadata structure acts as an intermediary/pivot between the local metadata structures, used by HOPE content providers (CPs), and metadata structures used by the targeted discovery services. As such, the common HOPE metadata structure ensures the semantic interoperability of metadata between content providers and discovery services.

2.1 Functional requirements

Authors:	Bert Lemmens – Amsab-ISG
Internal reviewers:	Armin Straube – FES Afelonne Doek – KNAW-IISG Alessia Bardi – CNR-ISTI Eric Beving – UPIP

This chapter describes the functional requirements for the design of the Common HOPE Metadata Structure and the corresponding Data Model, and XML Schema.

2.1.1 General

The Common HOPE Metadata Structure will be implemented by the **HOPE Aggregator**, i.e. the entity within the HOPE System that is responsible for harvesting, harmonisation, enrichment, storage and dissemination of the HOPE metadata.

This paragraph outlines what the HOPE Aggregator, according to the Description of Work aims for, and how the Common HOPE Metadata Structure may contribute in realising these aims.

2.1.1.1 What does HOPE want to accomplish with the HOPE Aggregator?

The Description of Work identifies three aims in which the HOPE Aggregator has an important role:

- HOPE is an **aggregation platform** for the Social History knowledge domain.
 “The [HOPE] project aims to improve access to the vast amount of highly significant

but scattered digital collections on social history in Europe, by creating a digital library on social history.” (DoW:3)

Expected outcome:

“A social history metadata aggregator that collects the available metadata and preview/thumbnails and links to the digital objects and ensures interoperability with the Europeana platform. The aggregator supplies this data to Europeana by use of the Europeana metadata ingest tools.” (DoW B2.1:8)

The Common HOPE Metadata Structure must provide the Aggregator with a semantic framework that is able to accommodate all metadata about the collections supplied to the HOPE Aggregator.

The detailed functional requirements for the ‘aggregation’ function of the Common HOPE Metadata Structure are listed in paragraph 2.1.1.2.

- HOPE is a **dissemination platform** for the Social History knowledge domain. The HOPE project aims for a “new web model for d2d logistics. The institutional website is no longer the exclusive place to reach Internet users. It has been observed by others before us: discovery on the web happens elsewhere. Users flock around Google-type facilities to search and discover information resources. The web environment imposes a new model for d2d logistics (discovery to delivery).” (DoW:3)

Expected outcome:

- “The aggregation of existing digital cultural content to make it searchable and accessible through Europeana.
- The upgraded Labour History Portal (aka IALHI Portal) by use of web services based on Europeana and DRIVER technologies.
- The aggregation of HOPE Metadata by other search facilities. “ (DoW B2.1:8)

The Common HOPE Metadata structure must provide the HOPE Aggregator with a semantic framework that allow the IALHI Portal, the Europeana Portal and a series of other discovery services to harvest the HOPE Metadata in the required metadata format.

The detailed functional requirements for the ‘aggregation’ function of the Common HOPE Metadata Structure are listed in paragraph 2.1.2.2.1.

- HOPE defines a **standardised interoperability framework** for the Social History knowledge domain. The HOPE project aims for resolving the “lack of interoperability of solutions across collections and services: What are the commonly agreed, pan-European metadata level interoperability standards for distributed digital data collections? How to cope

with semantic and linguistic differences across Europe? “ (DoW:3)

Expected outcome:

Improved quality of [...] the metadata [...] through sharing best practices in [...] metadata harmonisation [...] The BPN will seek to adhere and to contribute to Europeana guidelines and solutions through the Europeana open source and community space. (DoW B2.1:8)

The Common HOPE Metadata Structure must provide the HOPE Aggregator with a standardised interoperability framework, ensuring the semantic interoperability of the HOPE metadata between the HOPE Aggregator and the targeted discovery services, encouraging the Content Providers to use metadata standards for the creation, encoding and exchange of metadata, and reducing dependencies on system architectures on different levels of the HOPE System.

The aim for using metadata standards in the HOPE System applies to the aggregation, as well as the dissemination function of the the Common HOPE Metadata Structure and will be discussed in the paragraphs below.

2.1.1.2 What the HOPE Aggregator will not attempt to do?

The HOPE Aggregator has an important role in the Aggregation and Dissemination process of the HOPE Metadata. Evidently, in between these two processes, the HOPE Aggregator will also act as metadata repository for the HOPE Metadata.

However, it is important to realise that the HOPE Aggregator must not be conceived as a fully-fledged ‘collection management system’. Although many of the metadata standards discussed below have their origins in enabling a standardised registration and management of cultural heritage collections, the Common HOPE Metadata Structure will deliberately not accommodate particular metadata elements that are considered best practice in collection registration and management.

- The Common HOPE Metadata Structure does not accommodate administrative metadata on the creation, maintenance and integrity of the ‘item descriptions’.
- The Common HOPE Metadata Structure does not record administrative metadata about the ‘original materials’, such as information about the acquisition, condition and conservation of the items.

The common HOPE Metadata Structure should not be seen as a data architecture facilitating the long-term preservation of Social History metadata. A more appropriate image is conceiving the Common HOPE Metadata Structure as pivot or switchboard between the Local Systems, which are responsible for ensuring preservation and maintenance of the original materials and the corresponding metadata, and a series of discovery services or web portals, that facilitate the dissemination of the metadata.

2.1.2. Aggregation

This paragraph will further detail the functional requirements regarding the aggregation function of the Common HOPE Metadata Structure.

2.1.2.1 What are the key characteristics of the HOPE metadata?

The Common HOPE Metadata Structure must provide the Aggregator with a semantic framework that is able to accommodate all metadata about the “highly significant but scattered” collections supplied to the HOPE Aggregator.

Seeking an understanding of the variety and complexity of the collections supplied by the Content Providers, the HOPE Content Providers Survey⁴ addressed several questions regarding the encoding, granularity and language of the collections, as well as the institutional context in which the metadata has been created.

The Survey provided T2.3 with following conclusions regarding the key characteristics of the HOPE metadata.

- HOPE Metadata is **domain-specific**:

The results of this survey underlined that the Content Providers’ metadata has a distinct domain-specific character. The CPs’ metadata is ‘domain-specific’ in the sense that item descriptions are recording information that is specific for the type of material (e.g. works of art, book, a series of letters) or the institutional context in which the description has been made (museum description, library description, archival description).

Archival items tend to contain detailed information on the provenance of the item, but few information about physical characteristics. Visual items, such as photographs or posters, tend to contain extensive information about the materials and techniques used for the production of the item.

An analysis of the material type and institutional context of the HOPE collections allowed to specify four domains for categorising HOPE collections:

- archival collections
- library collections
- visual collections
- audio-visual collections

Each domain imposes particular requirements as to the semantics of the metadata, granularity of the descriptions, and the relationship with digital objects. In addition to these four domains a fifth ‘cross-domain’ profile was specified, particularly for Content Providers supplying metadata using Dublin Core:

- Dublin Core collections

⁴Survey was held in M3-M4 (July-August 2010). The Survey has been published as milestone MS8 in M4.

- **Idiosyncratic vs. standard encoding**
Half of the HOPE metadata is encoded using published standards, but another half is encoded using XML Schemas that are specific to the CPs local system (idiosyncratic metadata).
- **Hierarchical metadata:**
Three quarters of the HOPE metadata consist of archival and bibliographic metadata, which is likely to have a hierarchical structure. However, in the survey, only 7,3% of the metadata has been denoted by CPs as having a multilevel structure. When comparing the number of metadata records and digital files from Table0 with the indicated levels of granularity in the survey, it is likely that 40% of the metadata might have a multilevel structure which has not been identified by the survey. This particularly applies to HOPE bibliographic metadata on periodicals.
- **Compound objects:**
The HOPE Content Providers Survey also studied the relations between metadata records and digital objects. The ingestion of metadata and content in Europeana's Rhine release required a one-to-one relation between metadata record and a corresponding digital file. However, for more than half of the HOPE sub-collections (68 out of 129 sub-collections covering 41,2% of the metadata) the number of digital files exceeds the number of metadata records. This might indicate that a significant part of the HOPE metadata relates to compound digital objects, i.e. digital objects that each consist of multiple digital files.
- **Multilingual metadata:**
HOPE Metadata has been recorded in 8 different languages. Almost half of the descriptions are in German and only 14,2% are in English.
More than 80% of the HOPE metadata records are unilingual. The remaining metadata is bilingual, mostly Dutch/English.

2.1.2.2 How do the characteristics of the HOPE metadata affect the design of the Common HOPE Metadata Structure?

An analysis of the key characteristics of the HOPE metadata (domain-specific, idiosyncratic, hierarchical, compound objects, multilingual), and the aim providing a standardised interoperability framework (ensure interoperability, encourage standards, reduce dependencies), lead to the following list of functional requirements for the Common HOPE Metadata Structure.

req.1. Use open, well-established metadata standards.

The Common HOPE Metadata Structure and the subsequent Data Model, Domain Profiles

and XML Format should use open, well-established metadata standards. They should:

- use web-standards for the exchange of data, preferably the protocols endorsed by the W3C.
- use a 'flexible' but standardised data model, that can easily accommodate metadata from the identified domains, as well as from new domains of future HOPE sub-collections.
- enforce the use of encoding standards already used by the Content Providers and verify with current best-practices in the domain and in the Europeana community.
- enforce the use of cataloguing standards already used by the Content Providers. Endorsement of cataloguing rules may improve the semantic and syntactic consistency of the HOPE Metadata.

req. 2. Accommodate domain-specific metadata.

The data model and the subsequent Common HOPE Metadata Structure, and the XML Format should accommodate specific metadata elements for each of the four domains, identified by the Content Providers Survey, in one information space. The Common HOPE Metadata Structure should enable cross-searching the entire HOPE Metadata, while retaining the domain-specific context of the metadata.

req. 3. Accommodate hierarchically structured descriptions.

The data model and the subsequent Common HOPE Metadata Structure, and the XML Format should accommodate hierarchically structured descriptions for each of the four domains. The targeted description levels may include:

- fonds, series, files, items for multi-level archival descriptions
- titles and issues of periodicals; monographs and series of monographs
- formal or informal collections of single objects
- episodes of (radio or TV) broadcasts and documentaries

req. 4. Accommodate compound objects.

The data model and the subsequent Common HOPE Metadata Structure, and the XML Format should accommodate one-to-many relations between metadata records and digital files. Collections with compound object(s) may include:

- archival series with multiple digitized files
- publications digitized page by page
- cinematographic works that, for technical reasons, consist of multiple audio-visual files.

req. 5. Accommodate multilingual metadata.

The data model and the subsequent Common HOPE Metadata Structure, and the XML Format should record the language of the metadata record, as well as the language of translated metadata elements.

2.1.3. Dissemination

This paragraph will further detail the functional requirements regarding the dissemination function of the Common HOPE Metadata Structure.

2.1.3.1 What are the key characteristics of the HOPE users?

Before we can determine the functional requirements for the dissemination function of the Common HOPE Metadata Structure we need to get to know HOPE targeted user groups and their characteristics.

Task 1.1 has defined three user groups that convey the targeted users for the HOPE System:

Social history researchers

Research community, interested in sources for research, documentation, education in the field of social and labour history. This target group may include:

- **IALHI members**, interested in sources about and institutions involved with Social History. They are looking for specialised search features on other institutions collections to gather information about their work and for quotation or use (exhibitions, publications, etc.) by copy/download or request.
- **University students**, interested in sources for university research. They are looking for specialised search features on digitised contents for their academic works, using the information (metadata and images) by copy/download for citation/illustration.
- **University teachers/researchers**, interested in sources for academic research. They are looking for specialised search features on digitised contents for education, presentations and scholarship publications, using the information (metadata and images) by copy/download for citation/illustration.
- **Amateur historians**, interested in sources for private research. They are looking for specialised search features on aspects of local history, family history, specialist themes, using the information (metadata and images) by copy/download for citation/illustration.
- **Journalists**, interested in sources for journalistic research. They are looking for thematic search features for quote or illustration on journalistic works.
- **Cultural sector** (Archives, Libraries, Museums), interested in knowledge/sources on other European cultural institutions. They are looking for specialised search features on other institutions collections to improve their work and for use (exhibitions, publications, etc.) by copy/download or request.
- **Political/Unions Militants**, interested in sources on Social History, political parties and unions or social movements. They are looking for thematic search on areas related to their activities, using the information (metadata and images) by copy/download for citation/illustration.

The preferential channel for HOPE, targeting the Social-Research Community is the IALHI Portal.

General Public

General Public interested in European heritage. This user group involves:

- **General public**, interested in European cultural heritage. Informed citizens searching digitised European cultural and historical information to improve their knowledge about themes, personalities or events.
- **School students**, interested in sources for school works. They are searching information and digitised contents that could help them preparing school works, using the information (metadata and images) by copy/download.
- **Cultural sector** (Archives, Libraries, Museums) interested in knowledge/sources on other European cultural institutions. Searching on other institutions collections to gather information about their work and for use (exhibitions, publications, etc.) by request.

The preferential channel for HOPE, targeting the general public is the Europeana Portal.

Social Sites Users

General public interested in audiovisual sources for re-use and for sharing with communities. This user group may include:

- **General public**, interested in access on audiovisual and information sharing. They are looking for thematic search motivated by external information and open to related suggestions.
- **Cultural sector** (Archives, Libraries, Museums), interested in sharing and promoting information. They are looking for thematic search about other institutions presentations and collections and sharing their own work.
- **Amateur historians**, interested in access on audiovisual sources. They are looking for thematic search on specific social sites, downloading documents to their personal collection and sharing information with other users, namely blogs and social networks.

The preferential channel for HOPE, targeting social sites users are social sites, such as Flickr, YouTube, Scribd.

2.1.3.2 How does the HOPE System want to serve its users?

Based on these three user groups, T1.2 has developed 9 personas, identifying the HOPE user for searching metadata and digital content in the HOPE System.

These requirements have been listed in D1.1. This paragraph summarizes the functional requirements set as to the dissemination of metadata stored by the Aggregator.

Req. 1. HOPE system MUST provide users access to digital objects and metadata within the limits specified by content providers.

Req. 2. The users SHOULD be clearly informed of intellectual property rights and access restrictions, including use conditions applicable for each item of the collections/sub-collections made available.

Req. 3. HOPE system MUST provide users with both simple and advanced search functions.

Req. 4. Boolean logic operators and diacritic insensitive search MUST be implemented.

Req. 5. HOPE system MUST provide users with a clear and intuitive interface to browse metadata and content by institution, collection/sub-collection and themes.

Req. 6. Metadata and hierarchical structures for archival, bibliographic, audiovisual and museum units MUST be supported and presented in a meaningful way, unambiguously identifying the context of each item and which parts of each collection/sub-collection have digitized objects and are accessible online.

Req. 7. Relationships between collections and contextual information provided by the HOPE System (e.g. timeline) MUST be explored and presented to the user as a browsing possibility.

Req. 8. Users MUST be able to sort and filter search and browsing result sets by relevance, date, themes, language, media type, collections/sub-collections and institutions.

Req. 9. The HOPE system must support common access formats, which allow the user to interact with digital objects, reuse them in another environment. These access formats include different media type digital objects for viewing/listening (via online reader/player), printing and downloading.

Req. 10. Users MUST be able to request high-resolution and print copies of digital objects from the Content Provider.

Req. 11. HOPE system MUST be able to provide different access levels, based on roles controlled by an IAA (Identification, Authentication, Authorization) system, to specific users (e.g. anonymous, content providers and archive owners) and in conditions determined by content providers via the access condition workflow.

Req. 12. If provided with a transcript of a digital object, the HOPE system COULD have an interface that allows users to visualize both item and transcript at the same time.

Req. 13. If metadata in Cyrillic characters is made available by content providers, the system MUST also support transliteration to latin alphabet.

2.1.3.3 Listing the Functional Requirements of the Common HOPE Metadata Structure

An analysis of the key characteristics of the HOPE users (general public, social sites users, social-history research community) and the Use Cases, leads to the following list of functional requirements for the Common HOPE Metadata Structure.

Administrative & Contextual Metadata Elements

Following the Use Cases specified by T1.4, following metadata elements must/should be included in the HOPE Data Model:

- access restrictions for metadata records and digital objects
- use rights for the original resources
- hierarchical relationships between metadata records
- description level of metadata records
- relationships between metadata records and the corresponding digital content
- media types for digital content
- derivative types for digital content
- transliterated values and language elements

Metadata elements required for search and browse features

These requirements basically apply to search and retrieval of the HOPE Metadata via the HOPE Search API. The use cases specify three main access levels for the HOPE metadata:

- date – time span (i.e. a timeline)
- free text
- specific fields

The use cases do not exhaustively list the metadata elements to be included in the 'advanced search'. However, several requirements refer to specific search/browse features:

- searching by place
- searching by date
- searching by institution
- searching by collection
- searching by HOPE theme
- searching by language
- browsing through hierarchy
- browsing by media type

Export / harvesting of HOPE Metadata

T1.1 indicates for each user group the preferential channel for disseminating the HOPE Metadata. T1.7 will develop dissemination profiles for each of these dissemination channels. These profiles will define the scope and content of the metadata to be supplied. As for now, this chapter lists the requirements as defined by the Description of Work.

IALHI Portal

The IALHI Portal addresses particularly the Social History Research Community. The XML Format, used to store metadata in the HOPE Aggregator, will also act as an application profile for the IALHI Portal. The requirements for this application profile will be specified by WP3. T1.7 will specify a specific content profile for the IALHI Portal.

"The content profile of the Labour History Portal is international of scope and is not restricted to digital collections only. The data set supplied to the Labour History Portal will be much larger than the one supplied to Europeana, because it will also include the metadata records of non-digitised collections belonging to the HOPE participants. Digitised collections form a relatively small part of the whole labour history collection. Therefore it is important for researchers in this domain to have a comprehensive search facility; that does not exclude analogue materials." (DoW:7)

WP1 will provide a content profile for the Labour History Portal (T1.7). The portal will be updated by WP3 (T3.7) embedding the HOPE search API (T3.6).

Europeana Portal

The main service harvesting metadata from the HOPE Aggregator is the Europeana Portal. In order to make metadata interoperable between HOPE and Europeana, the HOPE Aggregator transforms HOPE metadata into Europeana compliant metadata using the EDM schema. This schema will be based on the Europeana Data Model (EDM) which will be published in the first half of 2011.

The EDM format is the successor of ESE and will provide:

- a cross-domain encoding format, based on the Europeana Data Model
- the possibility to accommodate rich, community-specific metadata
- the possibility to query metadata using semantic search and browse tools developed for the Europeana Portal.
- better support for hierarchical descriptions

Requirements for supplying metadata to Europeana are published in the Europeana Aggregator's Handbook. Currently only the requirements for the Rhine release are available. The requirements for the Danube release will be published in January 2011 at the earliest.

Social Sites

The preferential channel for HOPE, targeting the Social Sites Users are sites, such as Flickr, YouTube, Scribd. This channel basically focuses on the dissemination of digital content, provided with a

concise set of metadata, giving the semantic context for the digital content.

The HOPE Aggregator will enable the export of metadata and digital objects to a series of social sites application profiles. The Hope Aggregator will encode metadata elements in the format specified by the targeted Social Site. WP1 will provide a dissemination profile for the targeted Social Sites (T1.7).

2.2 Transformation Procedure

Authors:	Bert Lemmens – Amsab-ISG
Internal reviewers:	Armin Straube – FES Afelonne Doek – KNAW-IISG Alessia Bardi – CNR-ISTI Eric Beving – UPIP

This chapter provides you with definitions for the Common HOPE Metadata Structure and the corresponding Data Model, Domain Profiles and XML Schema, and a brief introduction on how these concepts fit into the transformation and harmonisation procedure for the HOPE Metadata.

2.2.1. Definitions

The following terms are used in the sense as defined here:

- The **Common HOPE Metadata Structure** as such is a structured set of metadata elements, accommodating descriptive, administrative as well as structural metadata about HOPE collections.
- A **Domain Profile** is a subset of metadata elements, borrowed from a domain-specific metadata standard, accommodating descriptive, administrative as well as structural metadata about HOPE collections that belong to one specific domain. HOPE provides 5 subsets, one for each domain type and each related with a specific metadata standard.
 - the archive profile is based on the APEnet/EAD standard;
 - the library profile is based on the MARC21 bibliographic standard;
 - the audio-visual profile is based on the EN15907 standard;
 - the visual domain profile is based on the LIDO standard;
 - the 'generic' Dublin Core profile is based on the Dublin Core standard.

HOPE Domain Profiles are used as an intermediate metadata structure when mapping local metadata elements to the common HOPE Metadata Structure.

- The high level **HOPE Data Model** is an abstraction of the Common HOPE Metadata Structure and Domain Profiles in terms of information about their physical or conceptual entities and their relationships. The HOPE data model is a high-level design of the Aggregators metadata architecture and is the reference document for

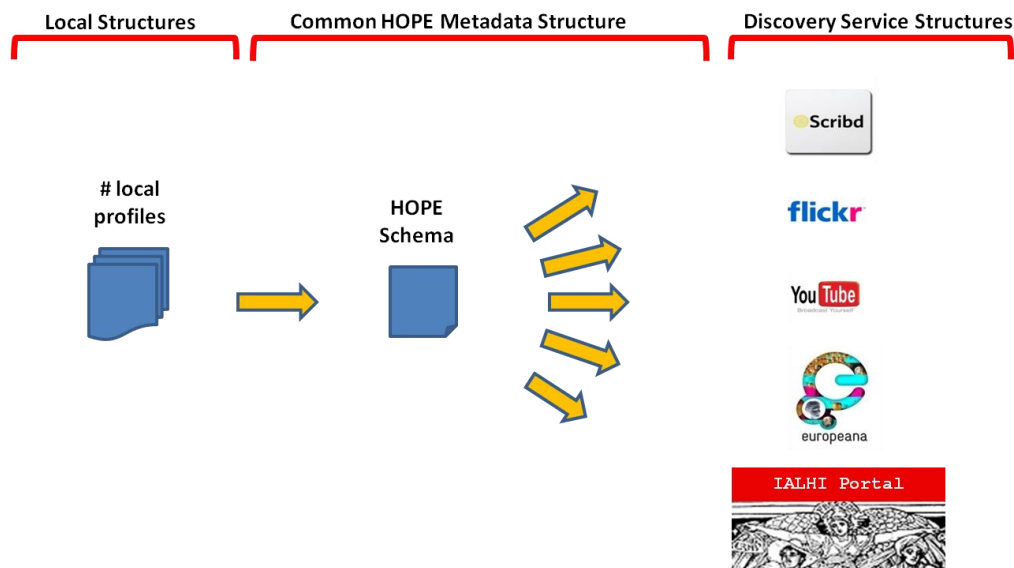
the implementation of the Common HOPE Metadata Structure in the D-NET infrastructure.

- The **HOPE Format** is a serialized representation of the Common HOPE Metadata Structure using XML, in order to be processed by the HOPE aggregator. The metadata, encoded in the HOPE Format, is validated by the **HOPE Schema**, which is specified using the XML Schema language, also known as a XML Schema Definition (XSD).

2.2.2 Transformation Procedure

The Common HOPE Metadata Structure is conceived as a pivot or switchboard between local and discovery service structures. Hence it acts as turning point in a series of transformations. These transformations are roughly divided in two phases:

- the transformations from multiple local metadata structures to one Common Metadata Structure, and
- the transformation from one Common Metadata Structure to a defined number of discovery service metadata structures.



The first phase in the transformation process seeks to gather the wide variety of metadata coming from approximately 150 HOPE collections into one single metadata structure. This transformation is facing two major thresholds.

- The Content Provider's Metadata has a distinct **cross-domain character**. The CPs' Metadata contains elements that are specific for the domain the metadata belongs to. Archival records have specific elements for recording the provenance of the material. Library records have an extensive set of title statements. Visual records

have specific fields for recording information about material and technique used for creating the item. The challenge is to map this metadata in a consistent way into one single, generic metadata structure.

- Moreover, the Content Provider's metadata has a distinct **idiosyncratic character**. Even within metadata records belonging to one domain, there is a wide variety of metadata structures. Half of the HOPE metadata is encoded using published encoding standards, but another half is encoded using XML Schemas that are specific to the CPs' local system.

The second phase in the transformation process will take care of the transformation of the metadata records from the HOPE format to a target format which allows a series of discovery services to disseminate the HOPE Metadata. Currently, three services have been identified as targets for this transformation phase:

- The Europeana Data Model format will service the Europeana Portal.
- The different Social Sites formats will be used by a series of Social Sites, such as Flickr, Youtube, Scribd.
- A specific case is the IALHI Portal, which will be serviced by the HOPE format.

Domain Profiles

In order to overcome the two thresholds identified in the first part of the transformation procedure, an intermediate stage has been defined, which consists of a transformation of the local metadata structures into 5 domain profiles. This implies that CPs will map their metadata first to a domain profile. Then the Aggregator will transform the Content Providers metadata to the Common HOPE Metadata Structure.

A domain profile is a structured set of domain-specific metadata elements, which are derived from an agreed-on, domain-specific metadata standard. The domain profiles are based on three parameters: the most commonly used metadata standards among Content Providers, best practice metadata standards for each domain, and metadata standards currently used within the Europeana community. T2.3 specified the following five domain profiles and their corresponding metadata standards:

- The **Archive Profile** accommodates archival records.
The elements are borrowed from EAD 2002 and ISAD(G).
- The **Library Profile** accommodates library records.
The elements are borrowed from MARC21 Bibliographic and ISBD 2007 Consolidated
- The **Visual Profile** accommodates visual records.
The elements are borrowed from LIDO and Spectrum 2.0.
- The **Audio-Visual Profile** accommodates audio-visual records.
The elements are borrowed from the EFG Schema and the EN15907 data model.
- The **Dublin Core Profile** accommodates cross-domain metadata records, created using a set of Dublin Core elements.

Since the domain profiles are based on the most commonly used standards among the HOPE Content Providers, the first mapping from local metadata structure to domain profile should easily overcome the threshold of the idiosyncrasies in the HOPE metadata.

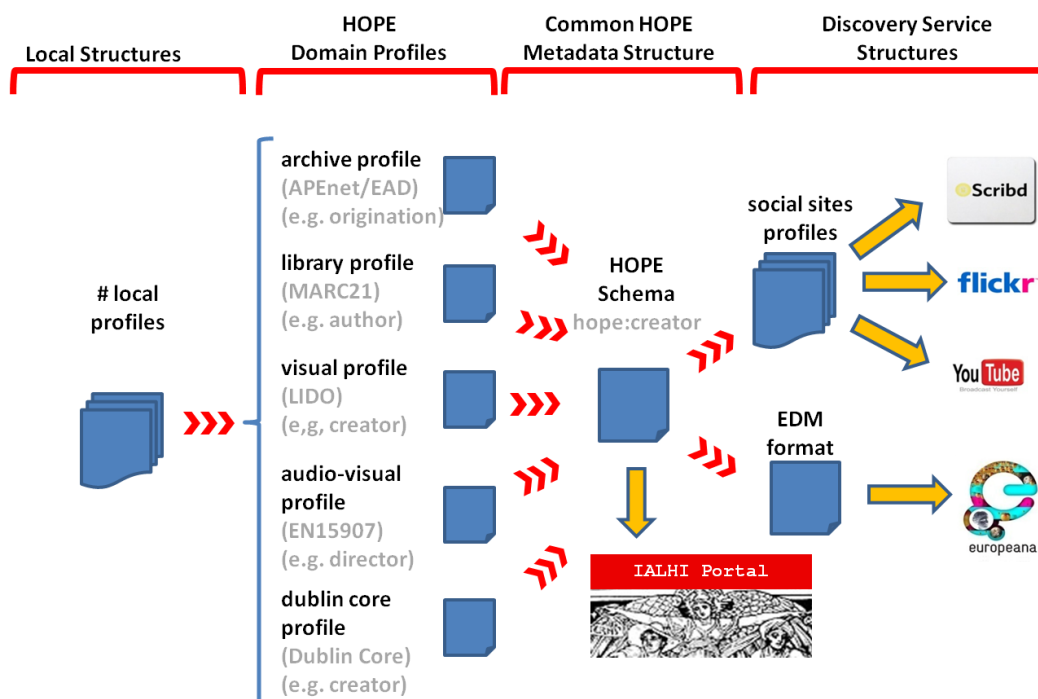
For detailed discussion of the domain profiles, see chapter 2.4 HOPE domain profiles.

Next phase is to transform the metadata, accommodated by the five domain profiles, to a single “generic” element set, i.e. the actual common HOPE metadata structure. The use of this intermediate stage, enables T2.3 to specify five default mappings, one for each of the domain profiles, which ensures a consistent mapping of the domain-specific metadata into the generic element set. This transformation is performed by the Aggregator.

A detailed discussion about the elements of this “generic” element set can be found in the Data Model chapter, including how the Data Model is able to retain the original semantic and syntactical expressivity of the Content Provider’s metadata.

The last phase of the transformation procedure deals with the mappings from the common HOPE metadata structure to the metadata structures of the Discovery Services.

- The IALHI Portal will use the common HOPE metadata structure as an application profile, which means no further transformation of the metadata is required.
- The Europeana portal will use the EDM format to be published in January 2011.
- The targeted Social Sites require a mapping from the Common HOPE Metadata Structure to a set of Social Sites specific formats.



Hence, the complete Transformation procedure is organised into 5 different transformation pipelines, one for each content domain. For each typology, the workflow has the following three phases:

- Transformation of Content Providers' local metadata structure to a HOPE Domain Profile.
- Transformation of the HOPE domain profile to the HOPE metadata structure.
- Transformation of the HOPE metadata structure to a Discovery Service metadata structure.

CPs are involved only in the first transformation step, where they must provide the input required to map their metadata structures onto the structures of the domain profiles. The other transformations are internal and based on crosswalks specified by the HOPE Project. The transformations are established by the Aggregator.

2.2.3. Transformation Issues regarding Semantic Interoperability

In general, "transformation" is the solution to the interoperability problems occurring between a *content provider* and a *consumer* (both information systems) willing to exchange content (metadata and objects) while supporting different data models. In HOPE, providers are CPs and the consumer is the aggregator. The interoperability problem must be tackled at two levels: **low-level** and **high-level**, described below.

- **Low-level interoperability: protocols and export formats.**
Provider and consumer system must agree on an access protocol and export format. In HOPE the low-level interoperability issue has been solved by assuming CPs must support the OAI-PMH protocol and metadata is encoded and exported by CPs as XML files of a given XML schema, called metadata records.
Metadata records are sets of metadata elements encoded as XML files. One record can encode one entity of a data model (e.g. HOPE metadata structure files in the aggregator: one file describes only one entity, descriptive unit, digital resources, etc with references to others) or more entities of a data model together with their relationships (e.g., EAD files in some CPs and in the HOPE archive domain profile: one file describes different and interlinked data model entities such as collections, fonds, subfonds, items, etc. all).
- **High-level interoperability: transformation process**
Based on the low-level solution (protocol and export format), the consumer fetches metadata from the provider and converts it to make it obey to the consumer data

model. When the export format is XML, this transformation process generally consists in providing a set of **XML transformation rules** to be applied to one or more *source* XML records (representing entities of the source data model) to obtain one or more *target* XML records (representing entities of the target data model). This set of rules may include:

- *Unpackaging rules*: these are often very specific to the context and must be implemented by means of XSLT scripts or actual programming languages code; in HOPE we are using XSLT at the moment.
- *Packaging rules*: these are often very specific to the context and must be implemented by means of programming language code.
- *Semantic/structural rules*: these specify how a machine should move values of the elements of one source record onto the elements of one target record. Rules can also express how, before being moved, such values should be converted (from-to vocabularies), refined, corrected, removed or merged with other values to obtain new values. Mapping rules can be implemented in several ways, by XSLT scripts or more complex programming languages.
- *Validation rules*: rules that specify which transformation steps should be applied to a source record based on some predicates applied to the values of its elements. These can be implemented in several ways, by XSLT scripts or more complex programming languages.

XML transformation rules are defined by aggregator administrators, instructed by the specifications (crosswalks) provided by domain experts, who are the only ones capable of defining proper semantics/structural correspondence.

The **XML transformation process** consists of a sequence of “source set-to-target set” steps of three kinds: *granularity mapping steps*, *semantics/structural mapping steps* and *validation steps*.

- **Granularity mapping steps** are of two typologies:
 - “1toN steps”: one source record is converted into N target records, based on a given *unpackaging transformation rule*. In HOPE, this is the case for EAD records of the archive domain profile, each transformed into a number of HOPE entity records.
 - “Nto1 steps”: N source records are converted into one target record, based on a given *packaging transformation rule*. In HOPE there is no such case at the moment.
- **Semantics/structural steps**: one source record is converted into one target record, based on given *semantic/structural transformation rules*.
- **Validation steps**: one source record is matched against given **validation rules**, and delivered to different steps based on the result. It can also be deleted or left pending, waiting for further analysis by administrator’s for example.

2.3 HOPE Data Model

Authors:	Bert Lemmens – Amsab-ISG Joris Janssens – Amsab-ISG Alessia Bardi – CNR-ISTI Paolo Manghi – CNR-ISTI
Internal reviewers:	Armin Straube – FES Afelonne Doek – KNAW-IISG Alessia Bardi – CNR-ISTI Eric Beving – UPIP

This chapter contains the detailed specification of the HOPE Data Model. The introduction includes an explanation of the rationale used to design the HOPE Data Model.

2.3.1 Introduction

This paragraph describes the approach T2.3 has followed when specifying the HOPE Data Model. The main source of inspiration has been the Europeana Data Model Specifications v5.2 (EDM), since several of the functional requirements for this data model are the same for the HOPE Data Model, i.e. accommodating cross-domain metadata, hierarchical descriptions and compound objects. Moreover, the Europeana Portal is one of the main discovery services targeted by the HOPE aggregator, so the resemblance with the EDM is expected to ensure a simple transformation to the forthcoming EDM format.

Apart from EDM, a series of other metadata standards have been incorporated in the HOPE Data Model. First of all Dublin Core Qualified for recording descriptive metadata, which is also an important component of the EDM, and PREMIS for accommodating metadata about digital resources.

Further the functional requirements for the Common HOPE Metadata Structure have been the source for adding a series of HOPE specific elements and attributes.

The following paragraphs will discuss the Europeana Data Model and the specification process of the entities, elements and attributes of the HOPE Data Model.

Europeana Data Model

The EDM has been developed as a data model facilitating the Danube release of the Europeana portal. The EDM is the successor of the Europeana Semantic Elements Specifications (ESE) and solves much of the shortcomings of ESE, such as handling complex digital objects, accommodating rich community-specific metadata and hierarchical descriptive units. The Danube release has been scheduled for May 2011, at the moment that HOPE will start preparing the ingestion of metadata.

EDM is a major step in the development of the data model of the Europeana portal and establishes the semantic framework for the future development of the Europeana portal. Europeana v1.0 announced that future releases of the portal and of the data model will consist of smaller and more regular updates, which will further specify the current framework. (source WP3 meeting, Europeana plenary October 2010)

Using EDM as a reference for the HOPE data model ensures that the HOPE aggregator will easily comply with the future development of the Europeana Portal. Compliancy will be ensured regarding the following domains:

- The current EDM introduces a number of classes that distinguish between six non-information resources: agents, events, places, physical things, concepts and time spans. These classes establish a semantic framework for the creation of common authority lists. Future releases of the EDM will specify element sets for accommodating descriptive metadata about agents, events, places etc. The implementation of published controlled vocabularies validating these classes is currently also being studied, in particular the use of the Virtual International Authority File (VIAF) for personal and organisation names. Implementing these classes in the HOPE data model not only ensures the interoperability with Europeana's authority lists, but it also allows to use the authoritative schemes and vocabularies developed within the Europeana community for enriching the HOPE metadata.
- As for descriptive metadata, the current EDM incorporates the ESE set as a default metadata profile. This enables Europeana to migrate Rhine metadata to Danube. In the near future, it will be possible to ingest ESE structured data in the Danube release, but Europeana considers this option to be temporary. In a first stage, it will thus still be possible for HOPE to ingest ESE encoded metadata in the Danube release. (source WP3 meeting, Europeana plenary October 2010)
- Next to the ensured interoperability, the EDM also provides specific solutions for issues related to the cross-domain nature of the HOPE metadata. Since Europeana is a 'cross-domain network' as well, EDM also has to enable cross-searching diverse community-specific metadata. The HOPE Data Model can benefit from the development of EDM. EDM has specified a set of generic properties which currently combines two different namespaces. On the one hand, EDM integrates the ESE/DC element set as a series of properties. Within EDM, the ESE properties are basically used in an 'object-centric' approach which directly links the features of the described object to the ore:proxy (the entity representing the description of the object). On the other

hand, EDM introduces a set of ENS properties which allow for an event-centric approach that first groups the features of the described object per event "in the history of the object", and then links the events to the ore:proxy. These properties are particularly relevant for event-centric metadata formats such as LIDO (the Europeana best-practice encoding format for museum metadata). Europeana v1.0 WP3 has produced several test mappings for archival, bibliographic, museum and audio-visual descriptions, using ESE as well as ENS properties. However, the final guidelines for mapping to EDM are still to be published.

- Unlike ESE, EDM allows for hierarchical links between descriptive units using the dc:hasPart and dc:isPartOf properties. This feature is particularly relevant for accommodating multi-level archival descriptions, but also for describing formal or informal collections of museum objects and titles and issues of publications. As for bibliographic resources, EDM will in the future also incorporate FRBR entities such as Work, Expression, Manifestation and Item. As for multi-level archival descriptions, APENet has made test mappings between EAD and EDM, which were used for the specification and serialisation of the HOPE archive domain profile.

Entities

This paragraph describes the rationale used for the specification of the 13 entities of the HOPE Data Model.

In the HOPE System, entities are considered the distinct separate units of information that are exchanged between local systems, the HOPE Aggregator and the discovery services. In this respect, the transformation process consists of identifying these entities in the Content Providers metadata and mapping them in the HOPE Data Model. The paragraph below describes how these entities are modeled in the HOPE Data Model.

- *Descriptive Unit & Digital Resource*

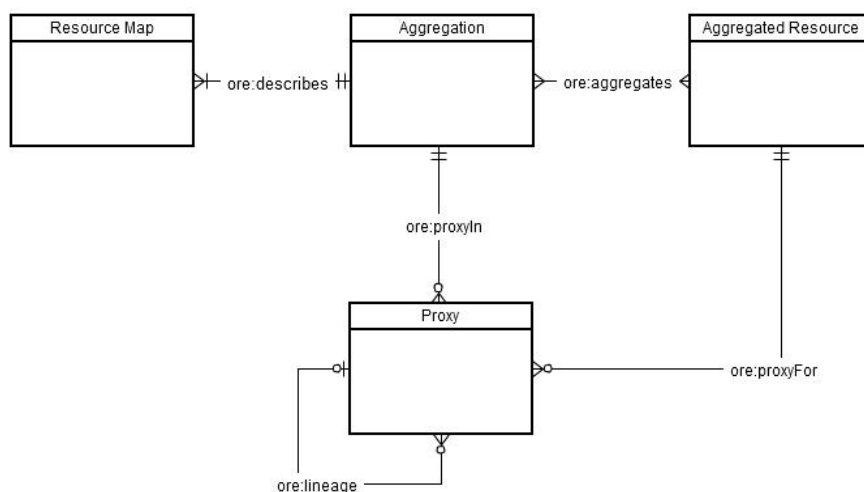
These entities are the core entities of the HOPE data model, accommodating information about respectively the description of the collection item and the digital content representing the collection item. They originate from an analysis of the implementation of OAI-ORE entities and properties by the EDM.

OAI-ORE is a standardised protocol⁵ for the description and exchange of aggregations of web resources. The ORE Model defines 4 entities:

- **ore:aggregation**: a place holder for a set of resources.
- **ore:aggregatedResource**: any resource part of an aggregation
- **ore:resourceMap**: a resource describing an aggregation based on a set of assertions.

⁵For background information and examples on OAI-ORE, refer to the official ORE primer, <http://www.openarchives.org/ore/1.0/primer>

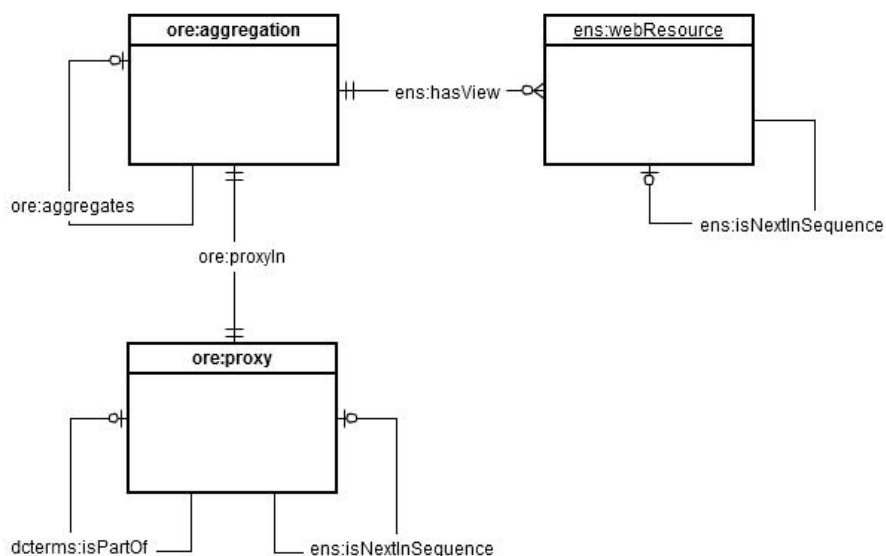
- **ore:proxy**: a virtual resource acting as a proxy for a certain aggregated resource in the context of a certain aggregation.



The core of EDM⁶ is organised around three entities, borrowed from the OAI-ORE standard.

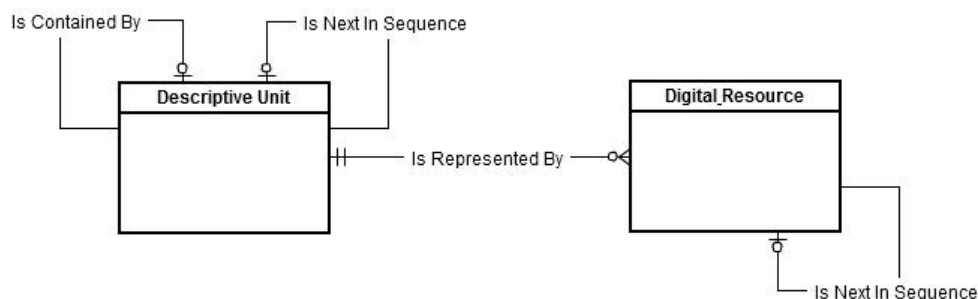
- **ore:aggregation**: in EDM used to represent complex constructs of web resources and proxies.
- **ens:webResource**: in EDM used for web representations of the corresponding collection item.
- **ore:proxy**: in EDM used to convey metadata about the corresponding collection item.

⁶For background information and examples on OAI-ORE, refer to the official ORE primer, http://www.version1.europeana.eu/c/document_library/get_file?uuid=718a3828-6468-4e94-a9e7-7945c55eec65&groupId=10605



In EDM, aggregations are basically used to distinguish between proxies and web resources about the same collection item (cf. EDM uses the word Cultural Heritage Object), but coming from different sources (i.e. different Europeana Content Providers). Since this kind of disambiguation is not a requirement for the HOPE System, the HOPE Data Model will specify only entities for the metadata description (cf. ore:proxy) and the corresponding digital resource (cf. ens:webResource). So the ore:aggregation entity is cut from the HOPE model.

- **Descriptive Unit:** represents an entity that metadata record about one or more materials that are contained by a *Collection*. The HOPE Data Model allows to associate Descriptive Units entities in order to create hierarchies of Descriptive Units, or to create a sequence of Descriptive Units.
- **Digital Resource:** contains information about a digital representation, i.e. each single digital image or audio-visual / sound recording, that provides a single, unique rendition of a collection item. The HOPE Data Model allows to associate Digital Resource entities in order to create a sequence of Digital Resources.

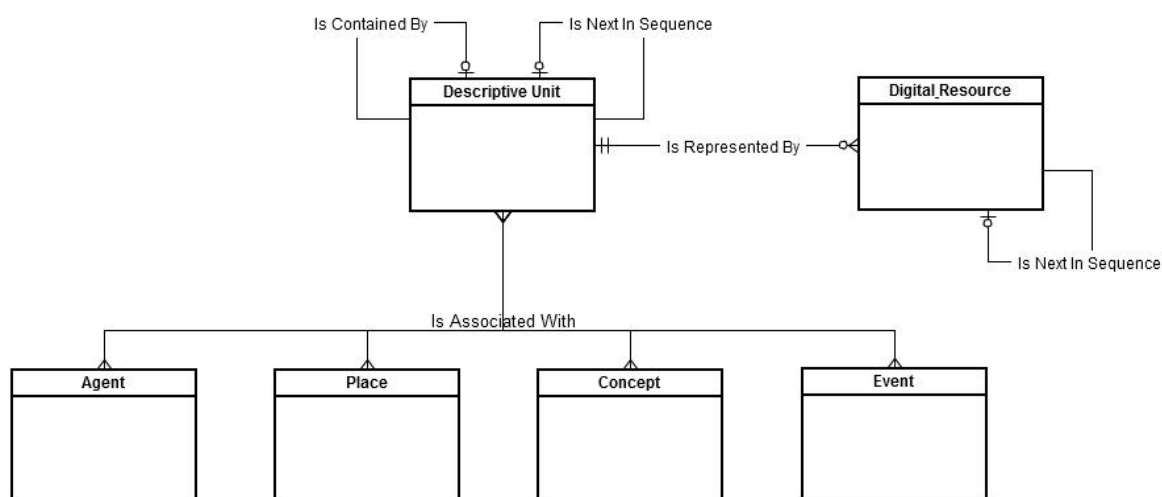


A Descriptive Unit entity CAN be associated with one or more Digital Resources. A Digital Resource entity MUST be associated with one single Descriptive Unit.

- *EDM Authorities:*

In order to improve cross-searching the entire HOPE Metadata, the HOPE Data Model borrows four more entities from EDM, representing:

- **agent:** people, individually or in group
- **place:** an 'extent' in space
- **concept:** ideas, notions
- **event:** a set of coherent phenomena or cultural manifestations, bounded in time and space.



A Descriptive Unit entity CAN be associated with one or more Agent, Place, Concept or Event entities.

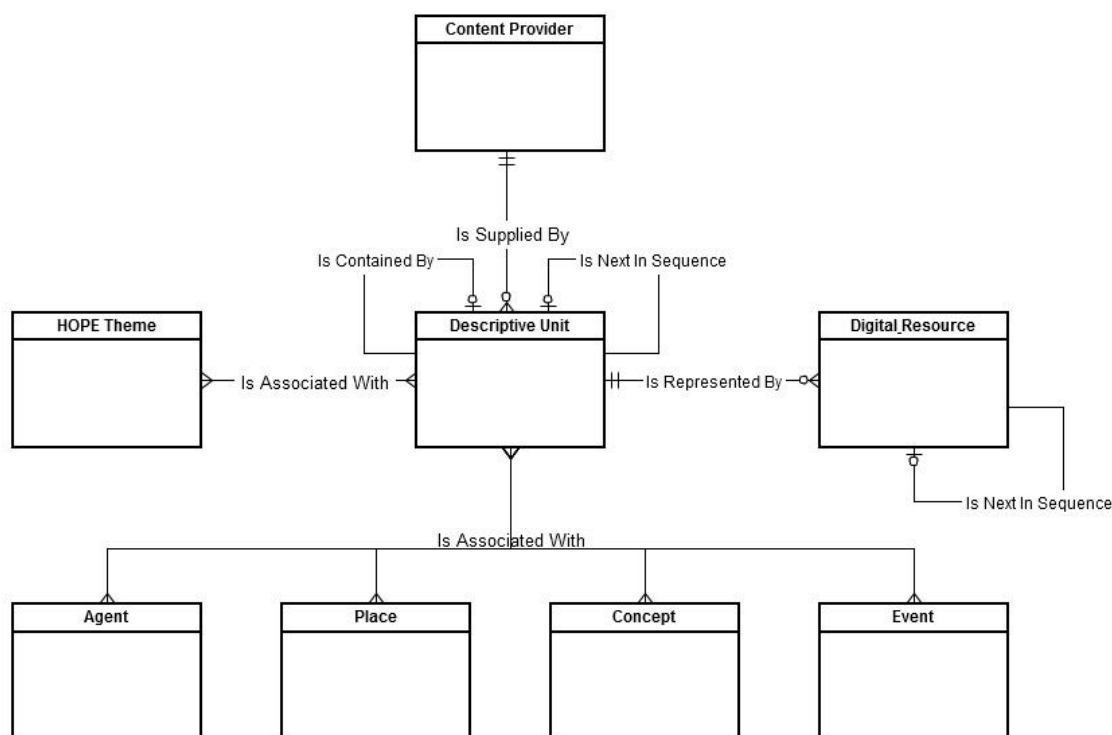
An Agent, Place, Concept or Event entity CAN be associated with one or more Descriptive

Unit entities.

- *HOPE Authorities:*

Following the functional requirements, the HOPE Data Model required two HOPE specific entities, representing:

- a **Content Provider**: the institution supplying metadata and content to the HOPE Aggregator
- a **HOPE Theme**: a thematic heading specific to the fields of social and labour history.



A HOPE Theme entity CAN be associated with one or more Descriptive Unit entities. A Descriptive Unit entity CAN be associated with one or more HOPE Theme entities. A Content Provider entity CAN be associated with one or more Descriptive Unit entities. A Descriptive Unit entities MUST be associated with one Content Provider entity.

Metadata Elements

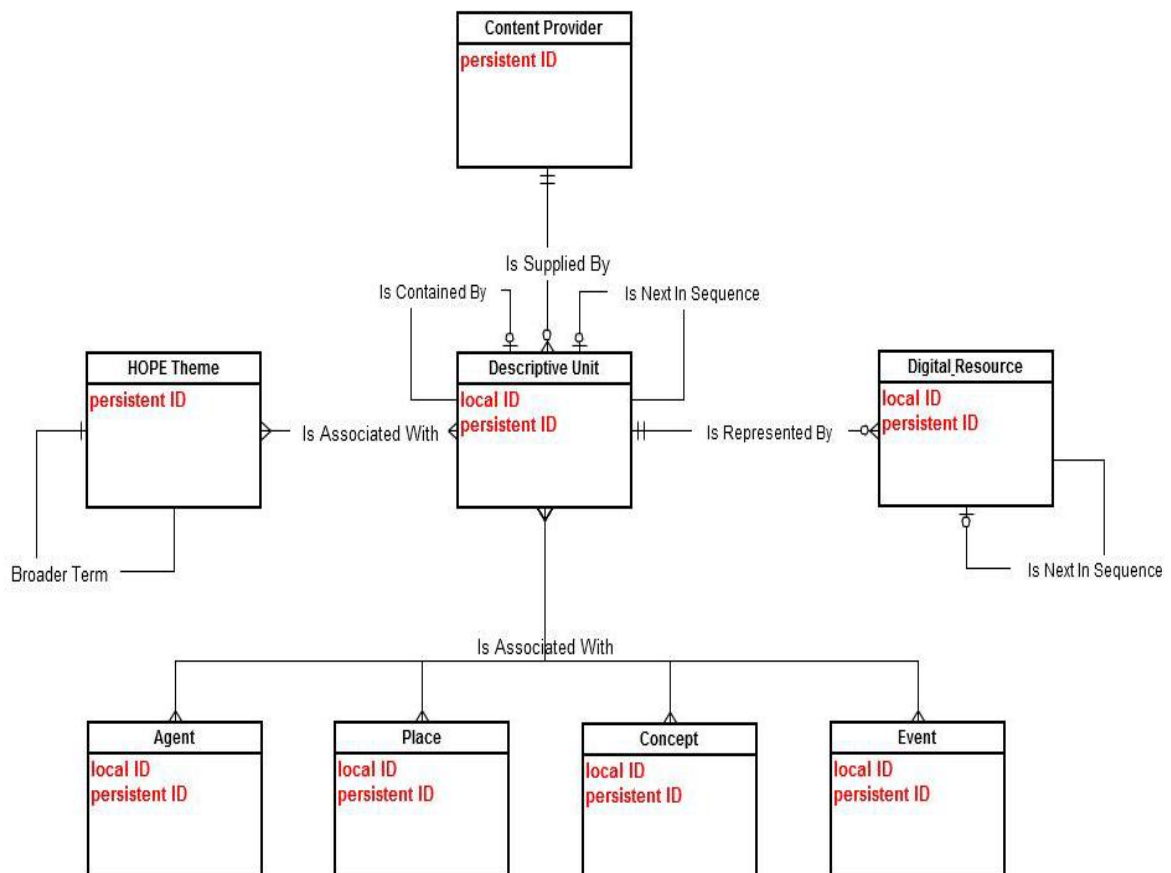
This paragraph describes the rationale used for the specification of the metadata elements

in the HOPE Data Model.

- *Local and persistent identifiers:*

In the HOPE Data Model, each entity is uniquely identified by a persistent identifier. Entities for which the information is supplied by the Content Provider also record a local identifier, which allows to identify the analogous entity in the local system.

- **local identifiers:** an unambiguous reference to an entity in the local system of the Content Provider.
- **persistent identifiers:** a globally unique, resolvable URL that ensures an unambiguous reference to a resource in and outside the HOPE system.

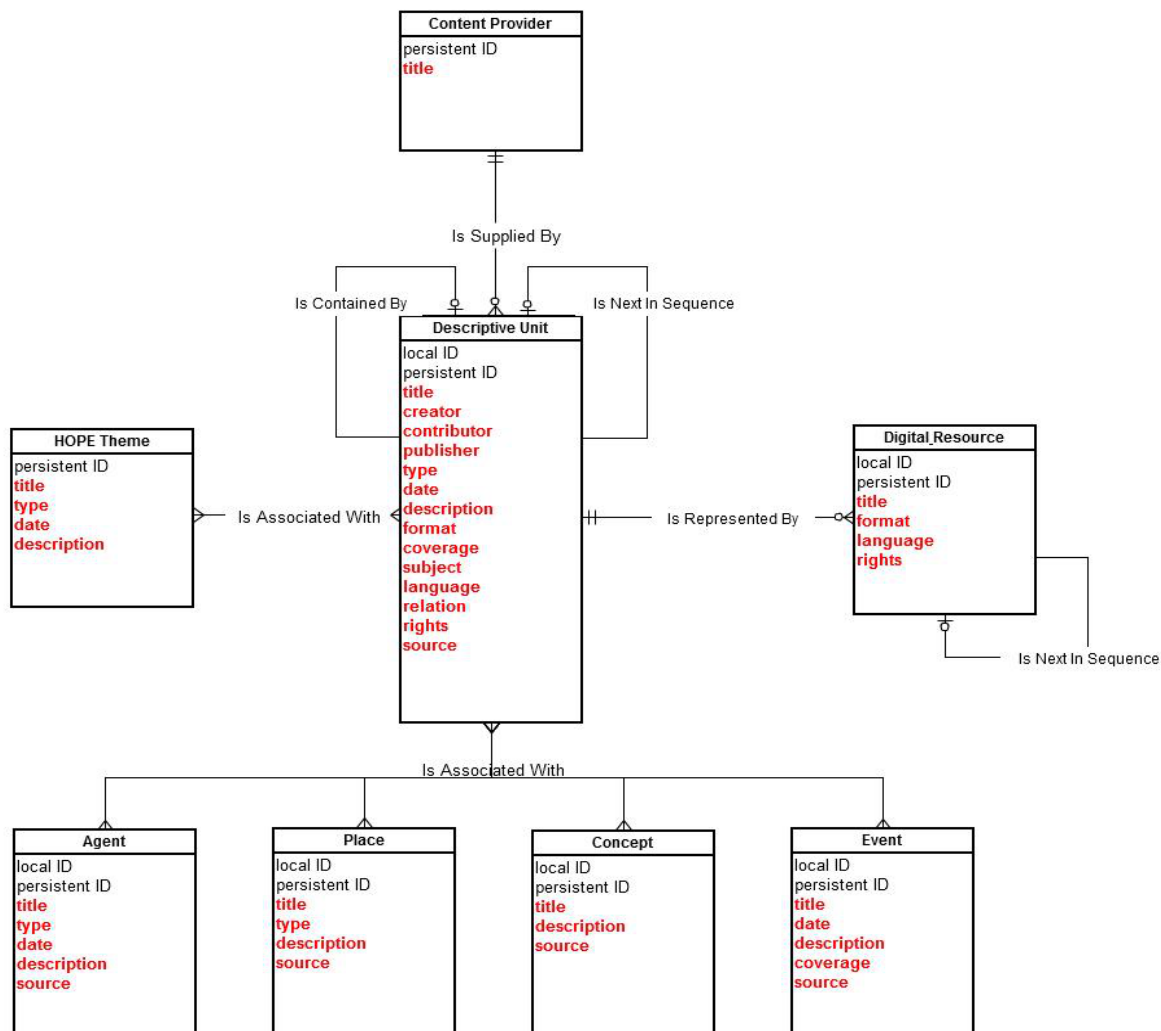


Descriptive Unit, Digital Resource, Agent, Place, Concept and Event entities are all entities that are identified in the Content Providers metadata and mapped into separate entities in the HOPE metadata. Hence, for each of these entities, the HOPE Data Model records a local and persistent identifier.

The HOPE and Content Provider entity are entities created by the HOPE System. Hence, for these entities, the HOPE Data Model only records a persistent identifier.

- *Descriptive metadata:*

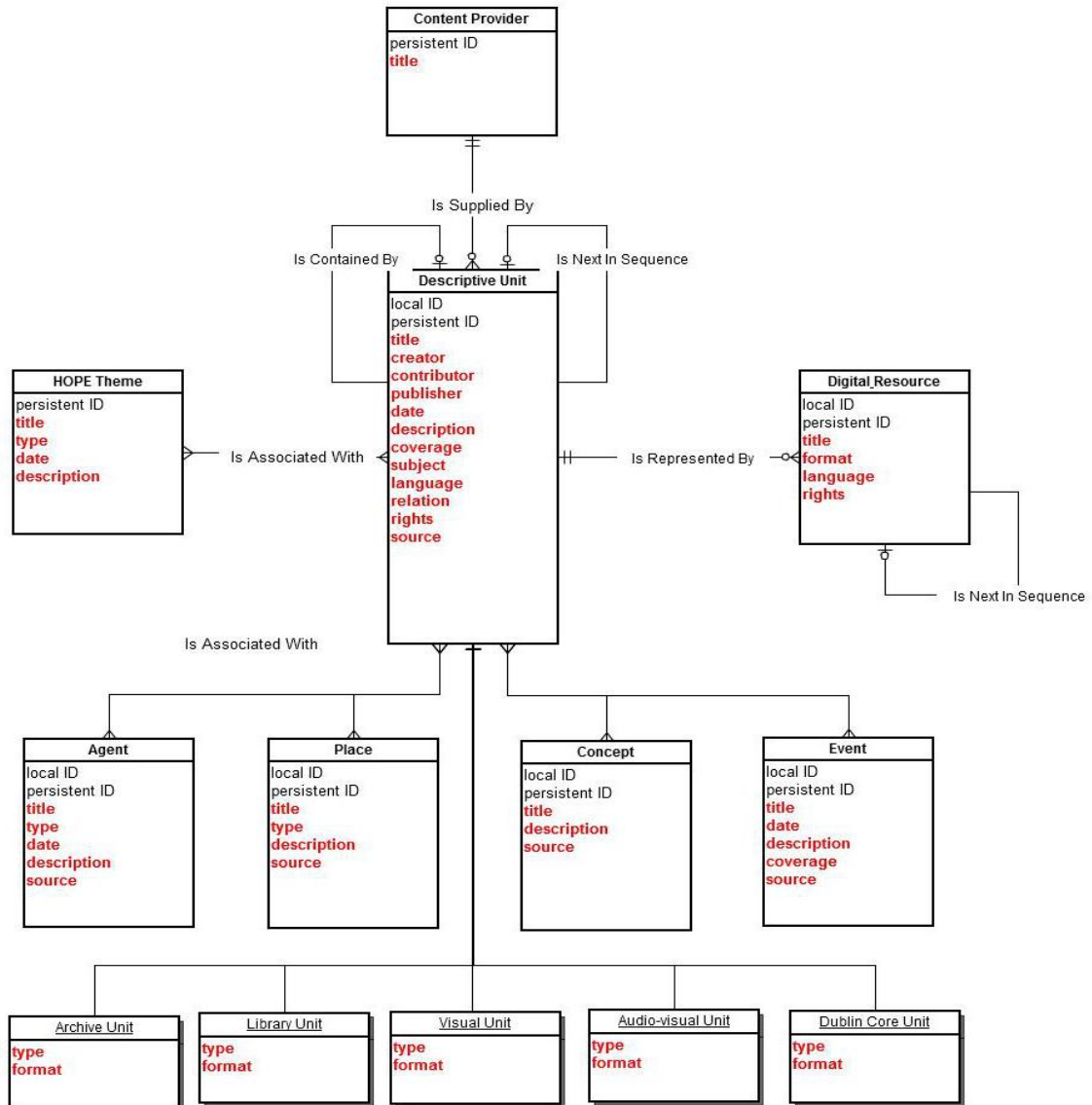
For recording descriptive metadata, the HOPE Data Model uses by default DC Qualified metadata elements. The diagram below shows the 'native' DC elements used for each entity. Please note that the dc:identifier element is represented by the local and persistent identifiers.



Sub-entities:

In the Descriptive Unit entity, each instance of a Dublin Core element can hold a value that is related with one of the domain profiles. This way, creating one index for every Dublin Core element enables cross-domain searching over all the instances of this element. However, for some Dublin Core elements this might be not desirable and for some Dublin Core elements searching for each domain separately is preferred. This particularly applies to domain-specific elements which correspond with dc:format and dc:type, such as material designation for the library domain, or object name for the visual domain. These domain-specific elements have been specified in their own domain-specific sub-entity.



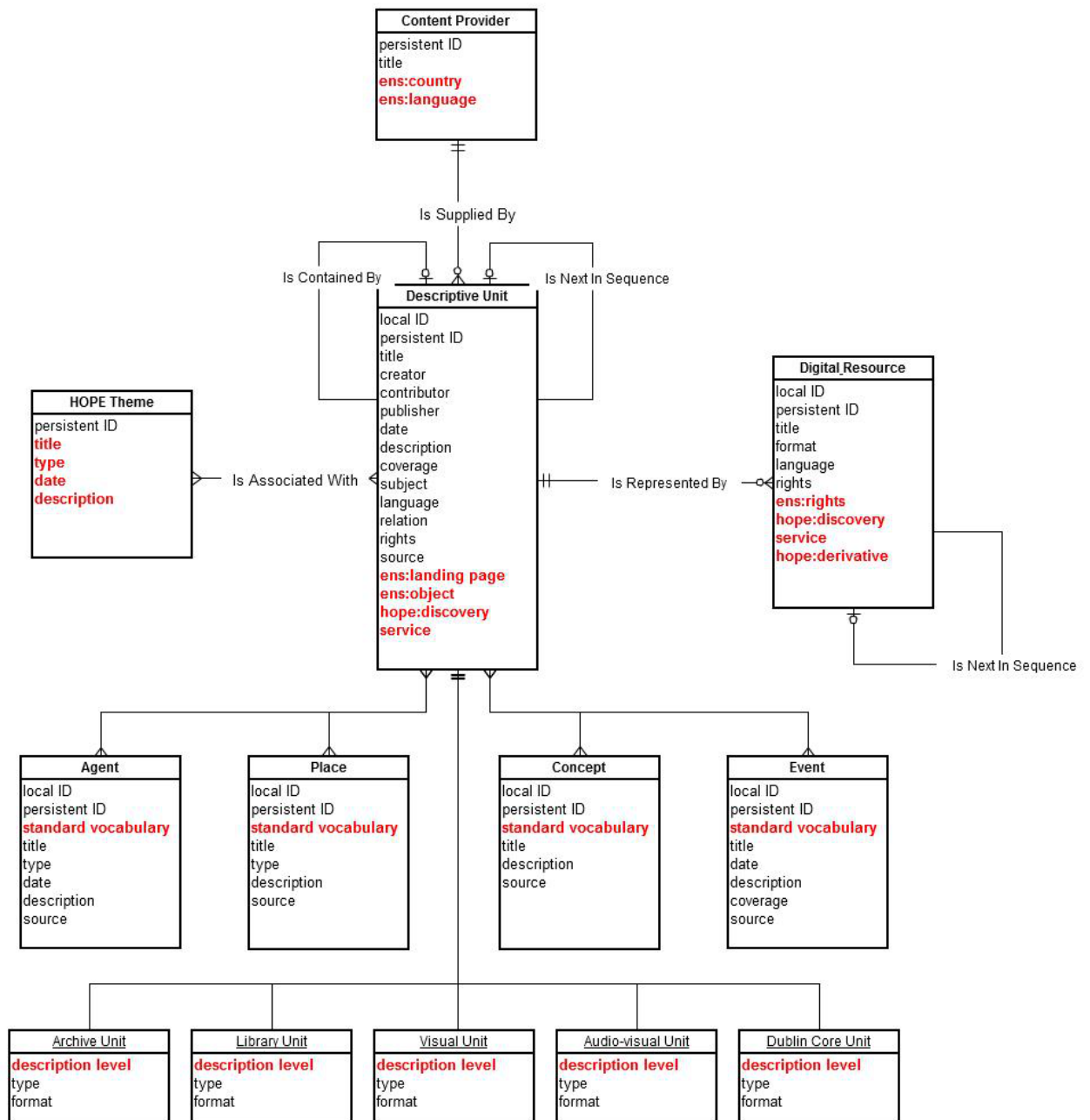


Each sub-entity holds the domain-specific elements which are unique in the context of their own domain. But each sub-entity also inherits all the elements from the associated Descriptive Unit. Thus a sub-entity and its associated Descriptive Unit can be considered as one entity.

- *Administrative metadata:*
Finally, the HOPE Data Model includes a series of mandatory administrative metadata elements that are required by Europeana and/or HOPE.



Definitions, cardinality and detailed specifications of these elements are listed in the next section.



Attributes

In order to retain the domain-specific characteristics of each metadata element, the HOPE Data Model specifies a set of attributes that record the domain-specific context for each

element.

- **label:** contains a domain-specific label to allow discovery services to display the corresponding value with a domain-specific label.
- **encoding:** contains a reference to a metadata data standard that is the source from which the HOPE element has been derived. This attribute contains a reference to an element name, including the namespace. The metadata standards include EAD, MARC21 Bibliographic, LIDO, EN15907, DC Qualified, EDM v5.2.1, PREMIS. Elements that are specific for HOPE refer to the HOPE namespace.
- **cataloguing:** contains a reference to a metadata data standard that represents the cataloguing rules, used for creating the recorded metadata. This attribute contains the name of the cataloguing rules and the identification number of the element. The used cataloguing rules include ISAD(g) and ISBD 2007 Consolidated.

example:

```
<dc:creator encoding="ead:origination - name of person" cataloguing="ISAD(G) 3.2.1" label="origination">Sir Joris Janssens</dc:creator>
```

This way, the Dublin Core element enables integrating the HOPE Metadata in one information space, while the attributes ensure retaining the domain-specific character of the metadata.

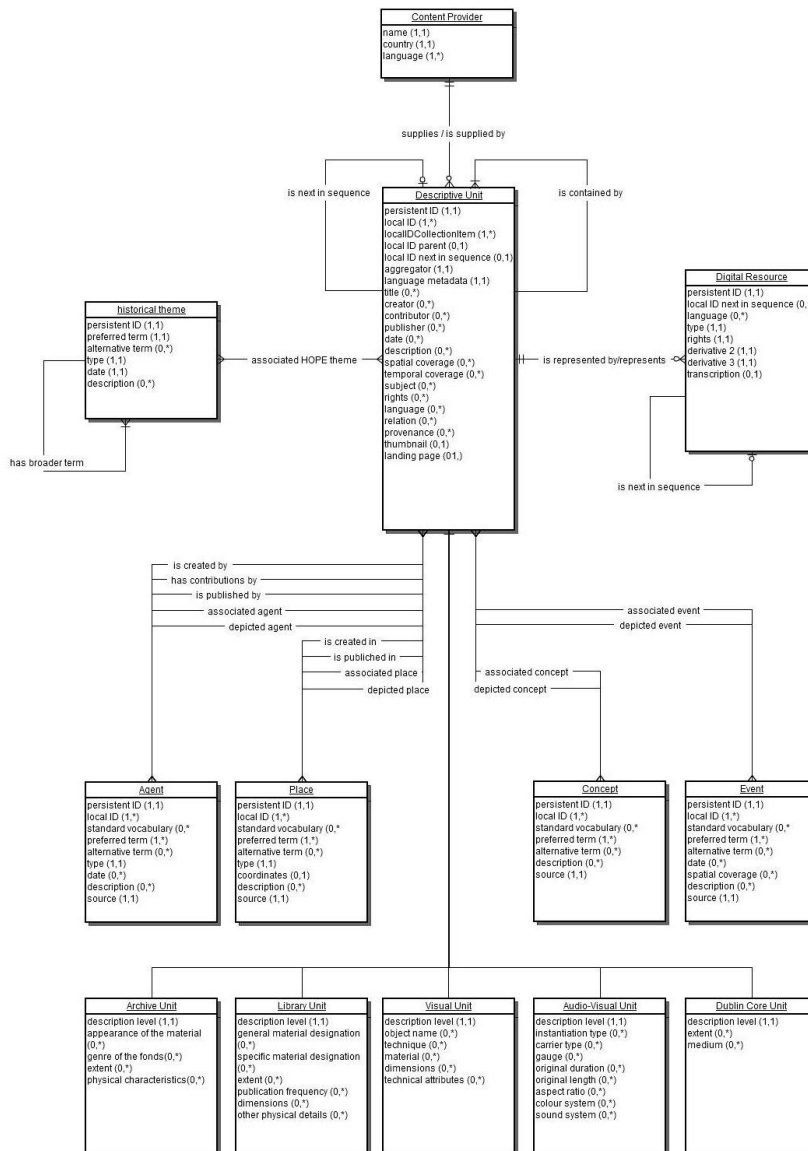
The HOPE Data Model specifies more attributes, but these will be discussed below in the actual specifications.

2.3.2 HOPE Data Model

This Section contains the detailed specifications of the HOPE Data Model.

2.3.2.1 Diagram

The HOPE Data Model is visualised below as a diagram representing the abstract structure of the information accommodated by the HOPE Aggregator. This diagram seeks to capture the units of information identified in the HOPE System and how they relate to one another.



2.3.2.2 Entities

- The HOPE Data Model includes 13 entities.

These entities represent the units of information that are exchanged between the different actors in the d2d chain.

Eight entities accommodate metadata supplied by the Content Providers:

- *Content Provider*:
Information about the institution supplying metadata and digital content to the aggregator.
- *Descriptive Unit*:
Information about one or more collection items, made available via the aggregator. This entity accommodates information that is available for all kinds of collection items, irrespective of the domain they belong to (archive, library, visual, audio-visual).
- Sub-entities:
However, the data model also records information that is specific for a particular domain the collection item belongs to, using four domain-specific 'sub-entities'.

These entities accommodate 'domain-specific' properties **and** the properties in the related Descriptive Unit. Hence:

- *Archival Unit*: accommodates all information about archival materials.
- *Library Unit*: accommodates all information about library materials.
- *Visual Unit*: accommodates all information about visual materials
- *Audio-Visual Unit*: accommodates all information about audio-visual materials

A fifth 'sub-entity' has been added for recording information about 'generic' materials.

- *Dublin Core Unit*: accommodates all information that is created using the 'generic' Dublin Core standard.
- Digital resource:
Information about the available digital content for the corresponding Descriptive Unit.

Following five entities are created by HOPE and enable cross searching the HOPE metadata.

- *HOPE Theme*: information about a thematic heading specific to the fields of social and labour history.
- *Agent*: information about a person or organisation associated with a Descriptive Unit.

- *Place*: information about a geographical location associated with a Descriptive Unit.
 - *Concept*: information about a concept associated with a Descriptive Unit.
 - *Event*: information about an event associated with a Descriptive Unit
- Each entity is characterized by a set of **elements**, which contain the information about the entity. Each element accommodates one or more values, i.e. a full text description, a controlled vocabulary term, a URL directing to an external resource, or an identifier of the entity.

The specifications for the elements below contain for each element the following information:

- A **definition** of the semantics of the element.
 - The **cardinality** of the element:
 - 0,1: the element is optional and can have only one instance
 - 1,1: the element is mandatory and can have only one instance
 - 0,*: the element is optional, and can have multiple instances.
 - 1,*: the element is mandatory and can have multiple instances.
 - The **value space** of the element, i.e. the type of value recorded for the element, as well as the associated attributes recording administrative metadata about the element. (see below).
 - The **constraints** for the element, including the controlled values that are allowed for the value space.
 - **Remarks** as to the use and syntax of the element.
 - A reference to the location of the element in the **HOPE Schema**.
- Each entity is also characterized by a set of **relationships**, defining the entities that are linked to each other. Each relationship accommodates the persistent ID of a related entity and often additional administrative metadata about the relationship in the attributes. (see below).
 - Each element or relationship is characterized by a set of attributes, that accommodate administrative metadata about the value or identifier recorded in the value space. The value space specification lists for each element and relationship the attributes that are available.

Following '**general attributes**' are available for most of the descriptive metadata elements.

- **normalised**: contain a normalised version of the term, recorded in the value space.

- **local ID:** contains an identifier, used for identifying an entity or online resource in the local system of the content provider.
- **persistent ID:** contains a globally unique, resolvable identifier, used for identification of the term in and outside (i.e. the web) the HOPE system.
- **label:** contains a label that allows discovery services to display the given value with a domain-specific label.
- **encoding:** contains a reference to a metadata data standard that is the source from which the HOPE metadata element has been derived. This attribute contains a reference to an element name, including the namespace. The metadata standards include EAD, MARC21 Bibliographic, LIDO, EN15907, DC Qualified, EDM v5.2.1, PREMIS. Elements that are specific for HOPE refer to the HOPE namespace.
- **cataloguing:** contains a reference to a metadata data standard that represents the cataloguing rules for creating the recorded metadata. This attribute contains the name of the cataloguing rules and the identification number of the element. The used cataloguing rules include ISAD(g) and ISBD 2007 Consolidated.
- **language:** contains a reference to the language, used for a translated value
- **script:** contains a reference to the script, used for a transliterated value.

Following attributes are available only for specific elements:

- **date range type:** Available for dates of agents and events only. Indicates whether the date is the start (from) or the end (to) of a date range.
- **derivative type:** Available for the derivative element only. indicates the type of derivative, i.e. derivative 1 (high resolution), derivative 2 (low resolution), derivative 3 (preview) or OCRed text.
- **primary:** Available for the derivative 3 element only. Boolean value, indicating which thumbnail is used as a Europeana Portal Image.
- **baseURL:** Available for the vocabulary standard element only. Contains the base URL of the PID recorded in the value space.
- **source:** Available for the vocabulary standard element only. Contains the name of the data value standard for the ID in the value space.
- **access:** Available for the derivative element only. Indicates whether the derivative is available for dissemination

Content Provider

The content provider entity accommodates all information about the institution that provides the HOPE aggregator with metadata and digital content.

Elements:

- **name:**
 - Name of the institution, representing the content provider.
 - occurrences: 1,1
 - value space: record (value:literal; label:constant value; encoding:constant value)
 - constraints:
 - label="data provider";
encoding="ens:dataProvider"
 - HOPE Schema: /hopeEntity/contentProvider/name

- **country:**
 - Country where the content provider is located.
 - occurrences: 1,1
 - value space: record (value:literal; normalised:controlled value; label:constant value; encoding:constant value)
 - constraints:
 - normalised=ISO 3166-1 (country codes);
label="europeana country";
encoding="ens:country"
 - HOPE Schema: /hopeEntity/contentProvider/country

- **language:**
 - Official language(s) of the country where the content provider is located.
 - occurrences: 1,*
 - value: record (value:literal; normalised:controlled value; label:constant value; encoding:constant value)
 - constraints:
 - normalised= ISO-639-3 (language codes);
label="europeana language";
encoding="ens:language"
 - HOPE Schema: /hopeEntity/contentProvider/language

Relationships:

- **supplies:**
 - this relationship links the content provider with the Descriptive Unit entities it supplies to the HOPE Aggregator.
 - occurrences: 0,*
 - range: collection

- value: record(value:URI, label:constant value)
- remarks:
 - value=PID Descriptive Unit;

label="HOPE collections"
- HOPE Schema: see Is Supplied By (uni-directional)

Descriptive Unit

The Descriptive Unit entity contains information about one or more *Collection Items*. A Descriptive Unit Entity can describe a single *Collection Item*, such as an archival document, a publication, a photograph or a movie. But it can also describe a series of *Collection Items*, such as an archival fonds, series or file, series of monographs, a periodical, collection of photographs, a series of TV broadcasts. The Descriptive Unit Entity accommodates information about all kinds of *Collection Items*, irrespective of their domain (archive, library, visual, audio-visual).

Domain -specific information is recorded in one of the five specific sub-entities of the Descriptive Unit Entity: archive unit, library unit, visual unit, audio-visual unit and Dublin Core unit.

A Descriptive Unit Entity basically accommodates Descriptive, Administrative and Structural Metadata, supplied by the CP via *Metadata Records* in a data set.

Descriptive Unit entities provide the semantic context for the associated *Digital Resources* that represent the *Collection Item*, and for its child Descriptive Units.

Elements:

- **persistent identifier descriptive unit:**
 - a globally unique, resolvable identifier, identifying the Descriptive Unit entity in and outside the HOPE System.
 - occurrences: 1,1
 - value: record (value:URI, label: constant value, encoding:literal)
 - constraints:
 - if domain="Archive", then
 - label="persistent identifier";

encoding="ead:c@id"
 - if domain="Library", then
 - label="persistent identifier";

encoding="marc:001"
 - if domain ="Visual", then
 - label="persistent identifier";

encoding="lido:lidoRecID"
 - if domain ="Audio-Visual", then
 - label="persistent identifier";

encoding="tba"

- this element is used for recording local identifiers describing the hierarchy of the descriptions of a collection.
- HOPE Schema: /hopeEntity/entity/Domain/descriptiveUnit/localIDParent
- **local identifier next descriptive unit in sequence:**
 - local identifier that relates two descriptive units that are 'ordered' parts of the same parent descriptive unit, such that one descriptive unit comes immediately after the other descriptive unit.
 - occurrences: 0,1
 - value: record (value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="next in sequence";
 - encoding="ens:isNextInSequence"
 - remarks:
 - useful for having a series of descriptive units displayed in a particular order.
 - HOPE Schema: /hopeEntity/descriptiveUnit/localIDNext
- **aggregator:**
 - Name of the organisation aggregating the collection.
 - occurrences: 0,*
 - value: record (value:constant value; label:constant value; encoding:constant value)
 - constraints:
 - label="aggregator";
 - encoding="ens:provider"
 - remarks:
 - This element contains first of all the name of the aggregator for the HOPE project (i.e. HOPE). This value will be mapped to the ens:provider element in the EDM. But the element can also contain the name of an aggregator supplying metadata to the HOPE aggregator.
 - HOPE Schema: /hopeEntity/descriptiveUnit/aggregator
- **metadata language:**
 - Primary language of the metadata describing the collection item.
 - occurrences: 1,1
 - value:(value:literal; normalised: controlled value; encoding:literal; cataloguing:literal; label:literal)
 - constraints:

- label="title",
encoding="lido:titleSet/appellationValue",
cataloguing="spectrum:title"
 - if domain="Audio-Visual", then
 - label="title";
encoding="EN15799:title/title"
 - label="translated title";
encoding="EN15799:title/title relationship="translation"
 - if domain:"Dublin Core", then
 - label="title";
encoding="dc:title"
 - language=ISO-639-3 (language codes);
script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/descriptiveUnit/title
- **creator:**
 - A person or organisation primarily responsible for making the collection item.
 - occurrences: 0,*
 - value: record(value:literal; label:literal, encoding:literal, cataloguing:literal, place:record)
 - constraints:
 - if domain="Archive", then
 - label="origination";
encoding="ead:origination,
cataloguing="ISAD(G):2.1"
 - label="origination - name of person"¹;
encoding="ead:origination/corpname@authfilenumber"
 - label="origination - name of corporation"¹;
encoding="ead:origination/persname@authfilenumber"
 - if domain="Library", then
 - label="statement of responsibility";
encoding="marc:245\$c";
cataloguing="ISBD:1.5"
 - label="author - person;
encoding="marc:100\$a"
 - label="author - corporate;
encoding="marc:110\$a"
 - label="author - meeting;
encoding="marc:111\$a"

- HOPE Schema: /hopeEntity/descriptiveUnit/contributor
- **publisher:**
 - A person or organisation primarily responsible for making the cultural heritage object available.
 - occurrences: 0,*
 - value: record(value:literal; label:literal; encoding:literal; cataloguing:literal; place:record; notes:record)
 - constraints:
 - if domain="Library", then
 - label="publisher"¹;
 - encoding="marc:260\$b;
 - cataloguing="ISBD 4.2"
 - label="printer or manufacturer or engraver"¹
 - encoding="marc:260\$f";
 - cataloguing="ISBD 4.6"
 - if domain="Audio-Visual", then
 - label="distributor";
 - encoding="EN15907:publicationEvent/HasAgent[activity="Distributor"]/Agent/Name
 - label="broadcaster";
 - encoding="EN15907:publicationEvent/HasAgent[activity="Broadcaster"]/Agent/Name"
 - if domain="Dublin Core", then
 - label="publisher";
 - encoding="dc:publisher
 - remarks:
 - if CP use a authority file for the data in this field and they can provide a local identifier for each term in this field, the metadata may be mapped to the 'is published by' element, which can be used for metadata enrichment.
 - HOPE Schema: /hopeEntity/descriptiveUnit/publisher
 - sub-elements:
 - **place:**
 - A geographical location, indicating where the Collection Item was made available.
 - occurrences: 0,*
 - value space: record(value:literal; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Library", then

- label="place of publication";
encoding="marc:260\$a";
cataloguing="ISBD 4.1"
 - label="place of printing or manufacture or engraving";
encoding="marc:260\$e";
cataloguing="ISBD 4.5"
 - if domain="Audio-Visual", then
 - label="publication location";
encoding="EN15907:Publicationevent/location"
 - if domain="Dublin Core", then
 - label="publisher";
encoding="dc:publisher"
 - remarks:
 - if CP use a authority file for the data in this field and they can provide a local identifier for each term in this field, the metadata may be mapped to the 'is published in' element, which can be used for metadata enrichment.
 - HOPE Schema: /hopeEntity/descriptiveUnit/publisher/place
 - **notes:**
 - Additional information on the publisher.
 - occurrences: 0,*
 - record(value:literal; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Library", then
 - label="notes on publication";
encoding="marc:260\$b";
cataloguing="ISBD 7.4"
 - label="notes on printing or manufacture or engraving";
encoding="marc:260\$f";
cataloguing="ISBD 7.4"
 - HOPE Schema: /hopeEntity/descriptiveUnit/publisher/notes
 - **date:**
 - A date associated with an event in the life cycle of the described collection item.

- occurrences: 0,*
- value: record(value:literal; label:literal; encoding:literal, cataloguing:literal; normalised:controlled value)
- constraints:
 - if domain="Archive", then
 - label="date of creation";
encoding="ead:unitdate";
cataloguing="ISAD(G):1.3"
 - if domain="Library", then
 - label="date of publication";
encoding="marc:260\$c";
cataloguing="ISBD 4.4"
 - label="date of printing or manufacture or engraving";
encoding="marc:260\$g";
cataloguing="ISBD 4.7"
 - if domain="Visual", then
 - label="object production date";
encoding="lido:eventDate";
cataloguing="spectrum:object production date"
 - if domain="Audio-Visual", then
 - label="publication date";
encoding="EN15907:Production event/date"
 - label="production date";
encoding="EN15907:Publication event/publicationDate + PublicationType="release""
 - label="broadcast date";
encoding="EN15907:Publication event/publicationDate + PublicationType="broadcast""
 - if domain="Dublin Core", then
 - label="date";
encoding="dc:da"
 - normalised=ISO 8601:2004, Data elements and interchange formats - Information interchange - Representation of dates and times
- HOPE Schema: hopeEntity/descriptiveUnit/date
- **description:**
 - An account of the collection item.
 - occurrence: 0,*
 - value: record(value:literal; label:literal, encoding:literal, cataloguing:literal, language:controlled vocabulary, script:controlled vocabulary)
 - constraints:

- if domain="Archive", then
 - label="appraisal information";
encoding="ead:appraisal";
cataloguing="ISAD(G):3.2"
 - label="accruals";
encoding="ead:accruals";
cataloguing="ISAD(G):3.3"
 - label="arrangement";
encoding="ead:arrangement";
cataloguing="ISAD(G):3.4"
 - label="general notes";
encoding="ead:note";
cataloguing="ISAD(G):6.1"
 - label="other descriptive data";
encoding="ead:odd";
cataloguing="ISAD(G):6.1"
 - label="other finding aids";
encoding="ead:otherfindaid";
cataloguing="ISAD(G):4.5"
 - label="content summary";
encoding="ead:scopecontent";
cataloguing="ISAD(G):3.1"
- if domain="Library", then
 - label="abstract";
encoding="marc:520\$a"
 - label="table of contents";
encoding="marc:505\$a"
- if domain="Visual", then
 - label="brief description";
encoding="lido:objectDescriptionSet";
cataloguing="spectrum:brief description"
- if domain="Audio-Visual", then
 - label="synopsis";
encoding="EN15907:contentDescription/descriptionText +
contentDescription/descriptionType="synopsis""
 - label="translated synopsis";
encoding="EN15907:contentDescription/descriptionText +
contentDescription/descriptionType="synopsis"@xml:lang=
"[language code]""

- label="shotlist";
encoding="EN15907:contentDescription/descriptionText + contentDescription/descriptionType="shotlist""
 - if domain="Dublin Core", then
 - label="description";
encoding="dc:description"
 - language=ISO-639-3 (language codes);
script=ISO 15924, Codes for the representation of names of scripts
 - HOPE Schema: /hopeEntity/descriptiveUnit/description
- **spatial coverage:**
 - the spatial extent or scope of the collection item.
 - occurrence: 0,*
 - value: record(value:literal; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Archive, then
 - label="index of places"¹;
encoding="ead:controlaccess/geogname"
 - if domain="Library", then
 - label="subject - geographic"¹;
encoding="marc:651\$a"
 - if domain="Visual, then
 - label="associated place"¹;
encoding="lido:eventPlace";
cataloguing="spectrum:associated place"
 - label="content - place"²;
encoding="lido:subjectPlace";
cataloguing="spectrum:content-place"
 - if domain="Audio-Visual", then
 - label="subject - place"¹;
encoding="EN15907:subject relationship [relationship type="Place"]"
 - label="country of reference"¹;
encoding:"EN15907:countryOfReference"
 - if domain="Dublin Core", then
 - label="spatial coverage";
encoding="dcterms:spatial"
 - remarks:
 - values for this element can be used to enrich the vocabulary on places.
 - values for following qualifiers relate with subsequent relationships:
 - ¹: associated place

- ²: depicted place
 - HOPE Schema: /hopeEntity/descriptiveUnit/spatialCoverage
- **temporal coverage:**
 - the temporal extent or scope of the collection item.
 - occurrence: 0,*
 - value: record(value:literal; label:literal, encoding:literal, cataloguing:literal)
 - constraints:
 - if domain="Library", then
 - label="subject - temporal";
 - encoding="marc:648\$a"
 - if domain="Visual", then
 - label="associated event name";
 - encoding="lido:eventName";
 - cataloguing="spectrum:content-event name"
 - label="content - event name";
 - encoding="lido:subjectEvent";
 - cataloguing="spectrum:associated event name"
 - label="associated date";
 - encoding="lido:eventDate";
 - cataloguing="spectrum:associated date"
 - label="content - date";
 - encoding="lido:subjectDate";
 - cataloguing="spectrum:content-date"
 - if domain="Audio-Visual, then
 - label="year of reference";
 - encoding="EN15907:yearOfReference"
 - if domain="Dublin Core", then
 - label="spatial coverage";
 - encoding="dcterms:temporal"
 - HOPE Schema: hopeEntity/descriptiveUnit/spatialCoverage
- **subject:**
 - the topic of the collection item.
 - occurrence: 0,*
 - value: record(value:literal; label:literal, encoding:literal, cataloguing:literal)
 - constraints:
 - if domain="Archive", then
 - label="index of names";
 - encoding="ead:controlaccess/name"

- label="index of persons"¹;
encoding="ead:controlaccess/persname"
- label="index of corporations"¹;
encoding="ead:controlaccess/corpname"
- label="index of families";
encoding="ead:controlaccess/famname"
- label="index of subjects"³;
encoding="ead:controlaccess/subject"
- label="index of materials";
encoding="ead:controlaccess/materials"
- if domain="Library", then
 - label="subject - person"¹;
encoding="marc:600\$a"
 - label="subject - corporate"¹;
encoding="marc:610\$a"
 - label="subject - topic"³;
encoding="marc:655\$a"
 - label="subject - genre";
encoding="marc:650\$a"
- if domain="Visual", then
 - label="associated person"¹;
encoding="lido:subjectActor";
cataloguing="spectrum:associated person"
 - label="content – person"²;
encoding="lido:subjectActor";
cataloguing="spectrum:content-person"
 - label="associated organisation"¹;
encoding="lido:subjectActor";
cataloguing="spectrum:associated organisation"
 - label="content – organisation"²;
encoding="lido:subjectActor";
cataloguing="spectrum:content-organisation"
 - label="associated concept"³;
encoding="lido:subjectConcept";
cataloguing="spectrum:associated concept"
 - label="content – concept"⁴;
encoding="lido:subjectConcept";
cataloguing="spectrum:content-concept"
 - label="associated object";
encoding="lido:subjectObject";
cataloguing="spectrum:associated object"

- label="content – object";
encoding="lido:subjectObject";
cataloguing="spectrum:content-object"
 - if domain="Audio-Visual, then
 - label="subject - person"¹;
encoding="EN15907:subject relationship [relationship type="Person"]"
 - label="subject - organisation"¹;
encoding="EN15907:subject relationship [relationship type="Organisation"]"
 - label="subject - genre"³;
encoding="EN15907:subject relationship [relationship type="Genre"]"
 - label="subject - topic"³;
encoding="EN15907:subject relationship [relationship type="Topic"]"
 - if domain="Dublin Core", then
 - label="subject";
encoding="dc:subject"
 - remarks:
 - values for this element can be used to enrich the vocabularies on agents, places topics and events.
 - values for following qualifiers relate with subsequent relationships:
 - ¹: associated agent
 - ²: depicted agent
 - ³: associated topic
 - ⁴: depicted topic
 - HOPE Schema: /hopeEntity/descriptiveUnit/subject
- **relation:**
 - textual description of a related resource in or outside HOPE information system.
 - occurrences: 0,*
 - value: record(value:literal; label:literal, encoding:literal, cataloguing:literal)
 - constraints:
 - if domain="Archive", then
 - label="location of originals";
encoding="ead:originalsloc";
cataloguing="ISAD(G):5.1"

- label="alternative form available";
encoding="ead:altformavail";
cataloguing="ISAD(G):5.2"
 - label="related material";
encoding="ead:relatedmaterial";
cataloguing="ISAD(G):5.3"
 - label="separated material";
encoding="ead:separatedmaterial";
cataloguing="ISAD(G):5.4"
 - label="bibliography";
encoding="ead:bibliography";
cataloguing="ISAD(G):6.1"
 - label="other finding aids";
encoding="ead:othfindaid";
cataloguing="ISAD(G):4.5"
 - if domain="Dublin Core", then
 - label="relation";
encoding="dc:relation"
 - HOPE Schema: /hopeEntity/descriptiveUnit/relation
- **rights:**
 - textual information about rights held in and over the collection item.
 - occurrences: 0,*
 - value: record (value:literal, label:literal, encoding:literal, cataloguing:literal)
 - constraints:
 - if domain:"Archive", then
 - label="conditions governing access";
encoding="ead:accessrestrict";
cataloguing="ISAD(G):4.1"
 - label="conditions governing use";
encoding="ead:userrestrict";
cataloguing="ISAD(G):4.2"
 - if domain:"Audio-Visual", then
 - label="access conditions";
encoding="EN15907:IPR Registration/nameOfApplicant"
 - label="IPR registration";
encoding="EN15907:accessConditions"
 - if domain="Dublin Core", then
 - label="rights";
encoding="dc:rights"

- HOPE Schema: /hopeEntity/descriptiveUnit/rights

- **language:**
 - language of the collection item.
 - occurrences: 0,*
 - value:(value:literal; normalised: controlled value; encoding:literal; cataloguing:literal; label:literal)
 - constraints:
 - if domain="Archive", then
 - label="language of the described material";
encoding="ead:langmaterial";
cataloguing="ISAD(G):4.3"
 - if domain="Audio-Visual", then
 - label="original language";
encoding="language"
 - label="language used";
encoding="language@usage="language used"
 - label="subtitle language";
encoding="language@usage="subtitle"
 - normalised=ISO 639-3
 - HOPE Schema: /hopeEntity/descriptiveUnit/language

- **provenance:**
 - A statement of any changes in ownership and custody of the cultural heritage object since its creation that are significant for its authenticity, integrity, and interpretation.
 - occurrences: 0,*
 - value: record(value:literal; label:literal, encoding:literal, cataloguing:literal)
 - constraints:
 - if domain="Archive", then
 - label="archival history";
encoding="ead:custodhist";
cataloguing="ISAD(G):2.3"
 - label="immediate source of acquisition";
encoding="ead:acqinfo";
cataloguing="ISAD(G):2.4"
 - label="repository";
encoding="ead:repository"
 - if domain="Visual", then

- label="owner";
encoding="lido:eventActor";
cataloguing="spectrum:owner"
 - if domain="Audio-Visual", then
 - label="holding institution";
encoding="EN15907:holdingInstitution"
 - if domain="Dublin Core", then
 - label="provenance";
encoding="dcterms:provenance"
 - HOPE Schema: /hopeEntity/descriptiveUnit/provenance
 - **thumbnail:**
 - The persistent identifier of a thumbnail representing the described Collection Item in the Europeana Portal. By default the europeana portal image is the thumbnail of the first digital resource related with the descriptive unit.
 - occurrences: 0,1
 - value: record(value:URI; label:constant value; encoding:constant value)
 - constraints:
 - value=PID;
label="europeana portal image";
encoding="ens:object"
 - HOPE Schema: hopeEntity/descriptiveUnit/europeanaPortallImage
 - **landing page:**
 - A globally unique, resolvable URL, identifying a webpage in the Content Provider's local website, displaying the digital object in its full information context.
 - occurrences: 1,1
 - value: record(value:URL; label:constant value; encoding:constant value)
 - constraints:
 - value=PID;
label:"landing page";
encoding="ens:isShownAt"
 - HOPE Schema: hopeEntity/descriptiveUnit/landingPage

Relationships:

- **is supplied by:**
 - this relationship links the descriptive unit entity with the content provider, supplying the descriptive unit to the HOPE aggregator.
 - occurrences: 1,1

- range: content provider
- value:URI
- remarks:
 - Used for defining subsets of descriptive units per Content Provider
- HOPE Schema: hopeEntity/descriptiveUnit/isSuppliedBy

- **is contained by:**
 - This relationship links a descriptive unit with another constituent descriptive unit that 'contains' the descriptive unit.
 - occurrences: 0,1
 - range: descriptive unit
 - value:URI
 - remarks:
 - used for recording hierarchically structure metadata
 - related with literal value 'original parent identifier'.
 - HOPE Schema: /hopeEntity/en/descriptiveUnit/isContainedBy

- **is next in sequence:**
 - This relationship links two descriptive units that are 'ordered' parts of the same constituent descriptive unit, such that one descriptive unit comes immediately after the other descriptive unit.
 - occurrences: 0,1
 - range: descriptive unit
 - value:URI
 - remarks:
 - used for having a series of descriptive units displayed in a particular order.
 - related with original next identifier in sequence
 - HOPE Schema: /hopeEntity/descriptiveUnit/isNextInSequence

- **is represented by:**
 - This relationship links the Descriptive Unit entity with one or more Digital Resource entities representing the described Collection Item.
 - occurrences: 0,*
 - range: digital resource
 - value: URI
 - remarks:
 - used to retrieve all digital files that represent the described resource.
 - HOPE Schema: hopeEntity/descriptiveUnit/isRepresentedBy

- **is created by:**

- This relationship links an Descriptive Unit with an Agent that created the Collection Item.
- occurrences: 0,*
- range: agent
- value: record(value:literal; persistentID:URI; localID:literal; label:literal, encoding:literal, cataloguing:literal)
- constraints:
 - if domain="Archive", then
 - label="origination - name of person"¹;
encoding="ead:origination/persname@authfilenumber";
cataloguing="ISAD(G):2.1"
 - label="origination - name of corporation"¹;
encoding="ead:origination/corpname@authfilenumber";
cataloguing="ISAD(G):2.1"
 - if domain="Library", then
 - label="author - person"¹; encoding="marc:100\$a";
cataloguing="ISBD 1.5"
 - label="author - corporate"¹; encoding="marc:110\$a";
cataloguing="ISBD 1.5"
 - if domain="Visual", then
 - label="object production person"¹;
encoding="lido:eventActor"; cataloguing="spectrum:object
production person"
 - label="object production organisation"¹;
encoding="lido:eventActor"; cataloguing="spectrum:object
production organisation"
 - if domain="Audio-Visual", then
 - label="director - person"¹; encoding="EN15907:Production
event/HasAgent[activity="Director"]/Agent/Name
 - label="director organisation"¹;
encoding="EN15907:Production
event/HasAgent[activity="Director"]/Agent/Name
 - label="producer - person"¹; encoding="EN15907:Production
event/HasAgent[activity="Producer"]/Agent/Name
 - label="production company"¹;
encoding="EN15907:Production
event/HasAgent[activity="Production
company"]/Agent/Name
- remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the agent

- label="publisher"¹; encoding="marc:260\$b"; cataloguing="ISBD 4.2"
 - label="printer or manufacturer or engraver"¹; encoding="marc:260\$f"; cataloguing="ISBD 4.6"
 - if domain="Audio-Visual", then
 - label="distributor"¹; encoding="EN15907:Publication event/HasAgent[activity="Distibutor"]/Agent/Name"
 - label="broadcaster"¹; encoding="EN15907:Publication event/HasAgent[activity="Broadcaster"]/Agent/Name"
 - remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the agent
 - value=name of the agent that will be harmonised when merging authority records
 - values for this relationship are recorded as a literal in the contributor element, as a backup of the initial name for the agent.
 - HOPE Schema: /hopeEntity/descriptiveUnit/isPublishedBy
- **associated agent:**
 - This relationship links a descriptive unit with an agent that is the topic the collection item is about.
 - occurrences: 0,*
 - range: agent
 - value: record(value:literal; localID:literal; persistentID:URI; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Archive", then
 - label="index of persons"¹; encoding="ead:controlaccess/persname"
 - label="index of corporations"¹; encoding="ead:controlaccess/corpname"
 - if domain="Library"; then
 - label="subject - person"¹; ; encoding="marc:600\$a"
 - label="subject - corporate"¹; encoding="marc:610\$a"
 - if domain="Visual", then
 - label="associated person"¹; encoding="lido:eventActor"; cataloguing="spectrum:associated person"
 - label="associated organisation"¹; encoding="lido:eventActor"; cataloguing="spectrum:associated organisation"
 - if domain="Audio-Visual", then

- label="subject - person"¹; encoding="EN15907:subject relationship [relationship type="Person"]"
 - label="subject - organisation"¹; encoding="EN15907:subject relationship [relationship type="Organisation"]"
 - remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the agent
 - value=name of the agent that will be harmonised when merging authority records
 - values for this relationship are recorded as a literal in the subject element, as a backup of the initial name for the agent.
 - HOPE Schema: hopeEntity/descriptiveUnit/associatedAgent
- **depicted agent:**
 - This relationship links a descriptive unit with an agent that is depicted by the collection item.
 - occurrences: 0,*
 - range: agent
 - value: record(value:literal; localID:literal; persistentID:URI; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Visual", then
 - label="content – person"; encoding="lido:subjectActor"; cataloguing="spectrum:content - person"
 - label="content – organisation"; encoding="lido:subjectActor"; cataloguing="spectrum:content - organisation"
 - remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the agent
 - value=name of the agent that will be harmonised when merging authority records
 - values for this relationship are recorded as a literal in the subject element, as a backup of the initial name for the agent.
 - HOPE Schema: hopeEntity/descriptiveUnit/depictedAgent
- **is created in:**
 - This relationship links a descriptive unit with a geographical location where the collection item was created.
 - occurrences: 0,*

- range: place
- value: record(value:literal; localID:literal; persistentID:URI; label:literal; encoding:literal; cataloguing:literal)
- constraints:
 - if domain="Visual", then
 - label="object production place"; encoding="lido:eventPlace"; cataloguing="spectrum:object production place"
 - if domain="Audio-Visual", then
 - label="production location"; encoding="EN15907:Production event/location"
- remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the place
 - value=name of the place that will be harmonised when merging authority records
 - values for this relationship are recorded as a literal in the creator element, as a backup of the initial name for the place.
- HOPE Schema: hopeEntity/descriptiveUnit/isCreatedIn
- **is published in:**
 - This element relates an descriptive unit with a geographical location where the collection item was published.
 - occurrences: 0,*
 - range: place
 - value: record(value:literal; localID:literal; persistentID:URI; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Library", then
 - label="place of publication"; encoding="marc:260\$a"; cataloguing="ISBD 4.1"
 - label="place of printing or manufacture or engraving"; encoding="marc:260\$e"; cataloguing="ISBD 4.1, ISBD 4.5"
 - if domain="Audio-Visual", then
 - label="publication location"; encoding="EN15907:Publicationevent/location"
 - remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the place
 - value=name of the place that will be harmonised when merging authority records

- constraints:
 - if domain="Visual", then
 - label="content - place"; encoding="lido:subjectPlace"; cataloguing="spectrum:content-place"
- remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the place
 - value=name of the place that will be harmonised when merging authority records
 - values for this relationship are recorded as a literal in the subject element, as a backup of the initial name for the place.
- HOPE Schema: hopeEntity/descriptiveUnit/depictedPlace
- **associated concept:**
 - This relationship links an descriptive unit with a concept that is the topic of the collection item.
 - occurrences: 0,*
 - range: concept
 - value: record(value:literal; localID:literal; persistentID; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Archive", then
 - label="index of subjects"; encoding="ead:controlaccess/subject"
 - if domain="Library", then
 - label="subject - genre"; encoding="marc:655\$a"
 - label="subject-topic"; encoding="marc:650\$a"
 - if domain="Visual", then
 - label="associated concept"; encoding="lido:subjectConcept"; cataloguing="spectrum:associated concept"
 - if domain="Audio - Visual", then
 - label="subject - genre"; encoding="EN15907:subject relationship [relationship type="Genre"]"
 - label="subject-topic"; encoding="EN15907:subject relationship [relationship type="Topic"]"
 - remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the concept

- **associated event:**
 - This relationship links a descriptive unit with an event that is the topic of the collection item.
 - occurrences: 0,*
 - range: HOPE event
 - value: record(value:literal; localID:literal; persistentID:URI; event; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Visual, then
 - label="associated event";
 - encoding="lido:eventName";
 - cataloguing="spectrum:associated event name"
 - remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the event
 - value=name of the event that will be harmonised when merging authority records
 - values for this relationship are recorded as a literal in the subject element, as a backup of the initial name for the event.
 - HOPE Schema: /hopeEntity/descriptiveUnit/associatedEvent

- **depicted event:**
 - This relationship links a descriptive unit with an event that is depicted by the collection item.
 - occurrences: 0,*
 - range: HOPE event
 - value: record(value:literal; localID:literal; persistentID:URI; label:literal; encoding:literal; cataloguing:literal)
 - constraints:
 - if domain="Visual" then
 - label="content - event name"; encoding="tba";
 - cataloguing="spectrum:content-event name"
 - remarks:
 - persistentID=PID
 - localID=local identifier of the authority record for the event
 - value=name of the event that will be harmonised when merging authority records
 - values for this relationship are recorded as a literal in the subject element, as a backup of the initial name for the event.
 - HOPE Schema: /hopeEntity/descriptiveUnit/depictedEvent

Archive Unit

The Archive Unit entity contains all information about archival materials. This class is a subclass of Descriptive Unit and inherits all its properties. The Archive Unit contains additional properties that are specific for describing archival materials.

Elements:

- **description level:**
 - the level of arrangement of the descriptive unit.
 - occurrences: 1,1
 - value: record(value:literal; normalised:controlled value; label:constant value, encoding:constant value, cataloguing:constant value)
 - constraint:
 - label="description level";
normalised="collection", "mid-level", "item"
encoding="ead:c@level";
cataloguing="ISAD(G):1.4"
 - remarks:
 - A domain-specific controlled vocabulary for audio-visual resources is to be specified by HOPE. Standard ISAD(G) levels include: (sub)fonds, (sub)series, files, items.
 - HOPE Schema: /hopeEntity/descriptiveUnit/archiveProfile/descriptionLevel

- **appearance of the material:**
 - information about the appearance or construction of the described materials.
 - occurrences: 0,*
 - value: record (value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraint:
 - label="appearance of the material";
encoding="ead:physdesc";
cataloguing="ISAD(G):1.5"
 - remark:
 - this element can be considered a super property of genre of the fonds, extent and physical characteristics (cf. EAD)
 - HOPE Schema:
/hopeEntity/descriptiveUnit/archiveProfile/appearanceOfMaterial

- **genre of the fonds:**

- information about specific types of archival materials, e.g. collections, or about specific physical characteristics.
 - occurrences: 0,*
 - value: record (value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraint:
 - label="genre of the fonds";
encoding="ead:physdesc/genreform";
cataloguing="ISAD(G):1.5"
 - remark:
 - this element can be considered a sub property of appearance of the material (cf. EAD)
 - HOPE Schema: /hopeEntity/descriptiveUnit/archiveProfile/genreOfFonds
- **extent:**
 - information about the quantity of the materials being described or an expression of the physical space they occupy.
 - occurrences: 0,*
 - value: record (value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="extent";
encoding="ead:physdesc/extent";
cataloguing="ISAD(G):1.5"
 - remark:
 - this element can be considered a sub property of appearance of the material (cf. EAD)
 - HOPE Schema: /hopeEntity/descriptiveUnit/archiveProfile/extent
- **physical characteristics:**
 - A description of important physical conditions or characteristics that affect the storage, preservation, or use of the materials described.
 - occurrences: 0,*
 - value: record (value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="physical characteristics";
encoding="ead:phystech";
cataloguing="ISAD(G):4.4"
 - remark:

- this element can be considered a sub property of appearance of the material (cf. EAD)
- HOPE Schema:
/hopeEntity/descriptiveUnit/archiveProfile/physicalCharacteristics

Library Unit

The Library Unit entity contains all information about library materials. This class is a subclass of Descriptive Unit and inherits all its properties. The Library Unit contains additional properties that are specific for describing library materials.

Elements:

- **description level:**
 - the level of arrangement of the descriptive unit.
 - occurrences: 1,1
 - value: record(value:literal; normalised:controlled value; label:constant value, encoding:constant value)
 - constraint:
 - label="description level";
 - normalised="collection", "mid-level", "item"
 - encoding="hope:descriptionLevel"
 - remarks:
 - A domain-specific controlled vocabulary for library resources is to be specified by HOPE.
 - suggestion: series, periodicals, monographs, issues
 - HOPE Schema: /hopeEntity/descriptiveUnit/libraryProfile/descriptionLevel

- **general material designation:**
 - The characteristics and general material type of content of the resource
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="general material designation";
 - encoding="marc:245\$k";
 - cataloguing="ISBD 1.2"
 - remarks:
 - Development of a controlled vocabulary for this element is desirable.
 - Terms should preferably matched with a controlled vocabulary

- HOPE Schema:
/hopeEntity/descriptiveUnit/libraryProfile/generalMaterialDesignation
- **specific material designation:**
 - Physical description of the described item, including its extent, dimensions, and such other physical details as a description of any accompanying materials and unit type and size.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="specific material designation";
encoding="marc:300";
cataloguing="ISBD 5.1.1"
 - HOPE Schema:
/hopeEntity/descriptiveUnit/libraryProfile/specificMaterialDesignation
- **extent:**
 - Number of pages, volumes, etc. of the physical or digital unit or units constituting the resource, adding other measures of extent as appropriate.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="extent";
encoding="marc:300\$a";
cataloguing="ISBD 5.1.3"
 - HOPE Schema: /hopeEntity/descriptiveUnit/libraryProfile/extent
- **publication frequency:**
 - Complete statement, exclusive of dates, of the current publication frequency.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="extent";
encoding="marc:310\$a";
cataloguing="ISBD 7.0.2"

- HOPE Schema:
/hopeEntity/descriptiveUnit/libraryProfile/publicationFrequency
- **dimensions:**
 - Dimensions of the resource, such as size and duration.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="dimensions";
encoding="marc:300\$c";
cataloguing="ISBD 5.3"
 - HOPE Schema: /hopeEntity/descriptiveUnit/libraryProfile/dimensions
- **other physical details:**
 - Other physical characteristics of the resource, such as method of production, colour, material from which the resource is made.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="other physical details";
encoding="marc:300\$b";
cataloguing="ISBD 5.2"
 - HOPE Schema:
/hopeEntity/descriptiveUnit/libraryProfile/otherPhysicalDetails

Visual Unit

The Visual Unit entity contains all information about visual materials. This class is a subclass of Descriptive Unit and inherits all its properties. The Visual Unit contains additional properties that are specific for describing visual materials.

Elements:

- **description level:**
 - the level of arrangement of the descriptive unit.
 - occurrences: 1,1
 - value: record(value:literal; normalised:controlled value; label:constant value, encoding:constant value, cataloguing:constant value)
 - constraint:

- occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="material";
 - encoding="lido:eventsMaterialTech";
 - cataloguing="spectrum:material"
 - remarks:
 - Development of a controlled vocabulary for this element is desirable.
 - Terms should preferably be matched with a standard vocabulary.
 - HOPE Schema: /hopeEntity/descriptiveUnit/visualProfile/material
- **dimensions:**
 - The measurement of a dimension of the object.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="dimensions";
 - encoding="lido:eventsMeasurements";
 - cataloguing="spectrum:dimension"
 - HOPE Schema: /hopeEntity/descriptiveUnit/visualProfile/dimensions
- **technical attribute:**
 - The name of a technical attribute possessed by an object which can be described and quantified.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value, cataloguing:constant value)
 - constraints:
 - label="technical attribute";
 - encoding="TBA";
 - cataloguing="spectrum:technical attribute"
 - HOPE Schema: /hopeEntity/descriptiveUnit/visualProfile/technicalAttribute

Audio-Visual Unit

The Audio-Visual Unit entity contains all information about audio-visual materials. This class

is a subclass of Descriptive Unit and inherits all its properties. The Audio-Visual Unit contains additional properties that are specific for describing audio-visual materials.

Elements:

- **description level:**
 - the level of arrangement of the descriptive unit.
 - occurrences: 1,1
 - value: record(value:literal; normalised:controlled value; label:constant value, encoding:constant value)
 - constraint:
 - label="description level";
normalised="collection", "mid-level", "item"
encoding="hope:descriptionLevel"
 - remarks:
 - A domain-specific controlled vocabulary for audio-visual resources is to be specified by HOPE.
 - HOPE Schema:
/hopeEntity/descriptiveUnit/audioVisualProfile/descriptionLevel

- **instantiation type:**
 - A term or phrase describing the item type relative to the duplication process
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="instantiation type";
encoding="EN15907:instantiation type"
 - remarks:
 - suggested controlled vocabulary form EN15907:
"original negative", "dupnegative", "positive", "original positive (reversal film)", "dupe positive", "Lavender", "image negative", "sound negative", "non-film analogue carrier", "non-film digital carrier".
 - HOPE Schema:
/hopeEntity/descriptiveUnit/audioVisualProfile/instantiationType

- **carrier type:**
 - The type of physical carrier used for the manifestation. Digital manifestations may be bound to a physical carrier (such as DVD) or exist without a defined carrier (such as online streaming media, podcasts, etc.).
 - occurrences: 0,*

- value: record(value:literal, label:constant value, encoding:constant value)
- constraints:
 - label="carrier type";
 - encoding="EN15907:format/carrierType"
- HOPE Schema: /hopeEntity/descriptiveUnit/audioVisualProfile/carrierType

- **gauge:**
 - The width of the film stock or other carrier (such as magnetic tape) used for the manifestation. Should include value and unit (e.g. "35 mm", "1/2 in.")
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="gauge";
 - encoding="EN15907:format/Gauge"
 - HOPE Schema: /hopeEntity/descriptiveUnit/audioVisualProfile/gauge

- **original duration:**
 - The running time of the first known manifestation of a cinematographic work, measured in minutes and second.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="original duration";
 - encoding="EN15907:extentUnit="min."|"sec.""
 - HOPE Schema: /hopeEntity/descriptiveUnit/audioVisualProfile/originalDuration

- **original length:**
 - The total physical, logical or temporal extent of the manifestation of a cinematographic work, or of the item from a manifestation. Physical extent is the total length of the medium carrying the manifestation or item. Logical extent is the number of discrete units, e.g. image frames, bytes (octets), etc., of which the manifestation or item is composed. Temporal extent is the running time under normal or specific playback conditions (e.g. a specific frame rate).
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="original length";
 - encoding="EN15907:extent unit="ft."|"m""

- HOPE Schema:
/hopeEntity/descriptiveUnit/audioVisualProfile/originalLength
- **aspect ratio:**
 - The ratio between width and height of the image. For anamorphic formats, this element should contain the aspect ratio of the projected image rather than that on the carrier. Some aspect ratios are known not only by numeric values, but also by names (e.g. "full frame", "Cinemascope").
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="aspect ratio";
encoding="EN15907:format/aspectRatio"
 - HOPE Schema: /hopeEntity/descriptiveUnit/audioVisualProfile/aspectRatio
- **colour system:**
 - A name for the colour system or process.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="colour system";
encoding="EN15907:format/colour/colourSystem"
 - HOPE Schema:
/hopeEntity/descriptiveUnit/audioVisualProfile/colourSystem
- **sound system:**
 - The name of the system by which sound is recorded either on the carrier, on a separate medium, or as part of the digital encoding (e.g. Western Electric (Westrex) Movietone, RCA Photophone, Tobis, Dolby Digital, etc.).
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="sound system";
encoding="EN15907:format/soundSystem/systemName"
 - HOPE Schema: /hopeEntity/descriptiveUnit/audioVisualProfile/soundSystem

Dublin Core Unit

The Dublin Core Unit entity contains all items that are described using the generic Dublin

Core metadata standard. Because of the simple and generic metadata structure, this information does not fit well in one of the other domain-specific units. This entity is a sub-entity of Descriptive Unit and inherits all its properties.

Elements:

- **Level of description:**
 - the level of arrangement of the descriptive unit.
 - occurrences: 1,1
 - value: record(value:literal; normalised:controlled value; label:constant value, encoding:constant value)
 - constraint:
 - label="description level";
 - normalised="collection", "mid-level", "item"
 - encoding="hope:descriptionLevel"
 - remarks:
 - A domain-specific controlled list for Dublin Core described materials is to be specified by HOPE
 - HOPE Schema: /hopeEntity/descriptiveUnit/dublinCoreProfile/descriptionLevel

- **extent:**
 - The size or duration of the resource.
 - occurrence: 0,*
 - value: record(value:literal, label:literal, encoding:literal, cataloguing:literal)
 - Constraints
 - label="extent";
 - encoding="dcterms:extent"
 - Remarks:
 - domain-specific data about the extent is primarily stored in the sub-classes, which allows for setting requirements for the syntax of the value or introduce domain-specific controlled vocabularies.
 - HOPE Schema: /hopeEntity/descriptiveUnit/dublinCoreProfile/extent

- **medium:**
 - The material or physical carrier of the resource.
 - occurrence: 0,*
 - value: record(value:literal, label:literal, encoding:literal, cataloguing:literal)
 - Constraints
 - label="medium";
 - encoding="dcterms:medium"
 - remarks:

- domain-specific data about the medium is primarily stored in the sub-classes, which allows for setting requirements for the syntax of the value or introduce domain-specific controlled vocabularies.
- HOPE Schema: /hopeEntity/descriptiveUnit/dublinCoreProfile/medium

Digital resource

The Digital Resource entity contains information about a *Digital Representation* of one or more *Collection Items*, described by an associated Descriptive Unit entity. A Digital Representation is a single digital image or audio-visual / sound recording, that provides a single, unique rendition of a Collection Item.

In the HOPE system, a Digital Representation can have multiple versions, each expressed in a *Digital File*.

The Digital Resource Entity contains Structural and Administrative Metadata about following versions:

- **Derivative 2:**
i.e. a low-resolution derivative, which is primarily used for display of the Digital Representation by the targeted Discovery Services. The Digital Resource entity may record a resolve URL, local ID or persistent ID for a low-resolution derivative stored by a Local or Shared Object Repository.
- **Derivative 3:**
i.e. a thumbnail, which is primarily used for preview of the Digital Representation in search results generated by the targeted Discovery Services. The Digital Resource entity may record a resolve URL, local ID or persistent ID for a thumbnail derivative stored by a Shared Object Repository or the Aggregator.
- **Transcriptions:**
i.e. manually created or automatically generated (e.g. using OCR) transcriptions of the intellectual content of the Digital Representation. The Digital Resource entity may record the transcription as plain text in the value space, or it may record a resolve URL, local ID or persistent ID for a text file containing the transcription, and stored by a Local or Shared Object Repository.

Apart from information about multiple version files, the Digital Resource entity also contains following administrative information: language of the digital representation, rights held in and over the digital representation, and the Europeana type of the digital representation.

A *Descriptive Unit* entity can have associations with more than one Digital Resource Entity, e.g. as with a description of a coin, being associated to two images, one for each side of the coin. In this case the Descriptive Unit entity, containing the description of the coin, is associated with two Digital Resource entities, one for each Digital Representation.

Elements:

- **persistent identifier digital resource:**
 - the globally unique, resolvable URL, identifying the Digital Representation in- and outside the HOPE information system.
 - occurrences: 1,1
 - value: record(value:URI; label:literal; encoding:literal)
 - constraints:
 - label="persistent identifier";
encoding="premis:objectIdentifier"
 - HOPE Schema: /hopeEntity/persistentID

- **local identifier next digital resource in sequence:**
 - local identifier of a digital representation that comes immediately after the current digital representation, as part of an ordered series.
 - occurrences: 0,1
 - value: record(values:literal, label:constant value, encoding:constant value)
 - constraints:
 - label="next in sequence";
encoding="premis:relatedObjectSequence"
 - remarks:
 - useful for having a series of digital representations displayed in a particular order.
 - HOPE Schema: /hopeEntity/digitalResource/localIDnext

- **language:**
 - Language(s) of the intellectual content of the Digital Representation.
 - occurrences: 0,*
 - value:record(value:literal; label:constant value; normalised:controlled value; encoding:literal)
 - constraints:
 - label="language digital content";
encoding="premis:significantProperties";
normalised=ISO-639-3:language codes
 - HOPE Schema: /hopeEntity/digitalResource/language

- **type:**
 - The Europeana material type of the Digital Representation, which Europeana uses to categorize search results by text, image, sound or video.
 - occurrences: 1,1
 - value: record(value:controlled value; label:constant value, encoding:constant value)
 - constraints:

label="thumbnail derivative"
encoding="premis:relatedObjectIdentification"
primary="true; false"

- remarks:
 - The primary attribute is used to indicate whether the thumbnail is the preferred thumbnail that is to be mapped to the element Thumbnail in the corresponding Descriptive Unit. Only one thumbnail should be primary. If no thumbnail is primary, then the system will pick one to use with the descriptive unit. If there are several primary thumbnails, then the system will pick one of those.
 - EDM mapping: ens:object
- HOPE Schema: /hopeEntity/digitalResource/derivative3
- **Transcription:**
 - Transcription of the textual content of a digital representation.
 - occurrences: 0,1
 - value: record(value:literal; persistentID:URI; localID:literal; resolveURL:URI; label:constant value; encoding:constant value)
 - constraints:
 - persistentID=persistent identifier of a text file containing the transcription
 - localID=local identifier of a text file containing the transcription
 - resolve URL=location of a text file containing the transcription
 - label="transcription"
 - encoding="premis:relatedObjectIdentification"
 - HOPE Schema: /hopeEntity/digitalResource/transcription

Relationships:

- **represents:**
 - This relationship links the Digital Resource entity with the Descriptive Unit entity, describing the represented Collection Item.
 - occurrences: 1,1
 - range: descriptive unit
 - value:URI
 - HOPE Schema: /hopeEntity/digitalResource/represents
- **is next in sequence:**
 - This relationship links two Digital Resource entities that are part of an ordered series, such that one Digital Resource comes immediately after the other Digital Resource.
 - occurrences: 0,1
 - range: digital resource
 - value:URI

- remarks:
 - only used for digital object that consist of an ordered series of digital resources.

HOPE Schema: hopeEntity/digitalResource/isNextInSequence

Agent

Agent entity accommodates information about persons or organisations that are associated with HOPE collections. This information is retrieved from the metadata records, supplied by the Content Providers, but will be enriched by the HOPE Aggregator.

Elements:

- **persistent identifier agent**
 - globally unique, resolvable URL, identifying a vocabulary term denoting a person or organisation.
 - occurrences: 1,1
 - value: record (value:URI, label: constant value)
 - constraints:
 - label="persistent identifier"
 - HOPE Schema: /hopeEntity/persistentID

- **local identifier agent**
 - an unambiguous reference to a vocabulary term denoting a person or organization in the local system of the Content Provider
 - occurrences: 1,*
 - value: record (value:literal, label: constant value)
 - constraints:
 - label="local identifier"
 - remarks:
 - This field contains the local ID of the vocabulary list where the preferred term originates from, and the local IDs all terms from other vocabulary list that have been merged with this list.
 - HOPE Schema: /hopeEntity/localID (1,1) → contains the local ID of the native term.
 - HOPE Schema: /hopeEntity/agent/alternateID (1,*) → contains the local IDs of the native terms and all merged terms.

- **standard vocabulary:**
 - persistent ID of a corresponding term for the Agent in a published data value standard, such as VIAF.
 - occurrences: 0,*
 - value: record(value:literal, source:literal, baseURL:literal)
 - constraints:

- **value=persistent ID;**
 - source=name or abbreviation of the standard vocabulary;
 - baseURL: baseURL prefix for the persistent ID
- HOPE Schema: /hopeEntity/agent/standardVocabulary

- **preferred term:**
 - The accepted or most desirable term for the Agent, used in HOPE system.
 - occurrences: 1,*
 - value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
 - constraints:
 - label="name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/agent/preferredTerm

- **alternative term**
 - An equivalent term of the preferred term for the agent, used in HOPE system.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
 - constraints:
 - label="alternative name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/agent/alternativeTerm

- **type:**
 - The nature of the agent
 - occurrences: 1,1
 - value: controlled value
 - constraints:
 - value="person"; "corporation"
 - HOPE Schema: /hopeEntity/agent/type

- **date:**
 - dates of birth and/or dead of the person, or start and/or end date of activity of the organisation.
 - occurrences: 0,*

- value: record(value:literal, normalised:controlled value, range type:controlled value)
- constraints:
 - normalised=ISO 8601:2004
 - range type="from"; "to"
- HOPE Schema: /hopeEntity/agent/date

- **description:**
 - An account of the agent
 - occurrences: 0,*
 - value: record(value:literal; label:literal; encoding:literal; cataloguing:literal; language:controlled value, script:controlled value)
 - constraints:
 - if domain="Archive", then
 - label="history of the originator"
 - encoding="ead:biohist"
 - cataloguing="ISAD(G):2.2"
 - otherwise
 - label="description"
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/agent/description

- **source:**
 - name of the content provider, who provided the metadata about the Agent
 - occurrences: 1,1
 - value:record(value:literal, label:constant value)
 - constraints:
 - label="source"
 - HOPE Schema: /hopeEntity/agent/source

Relationships:

- **has created:**
 - This relationship links a Descriptive Unit with an Agent that created the Collection Item.
 - occurrences: 0,*
(inverse relationship is created by)

- **has contributed to:**
 - This relationship links a Descriptive Unit with an Agent that contributed to the Collection Item
 - occurrences: 0,*
(inverse relationship has contributions by)

- **has published:**
 - This relationship links a Descriptive Unit with an Agent that published the Collection Item.
 - occurrences: 0,*
(inverse relationship is published by)

- **inverse associated agent:**
 - This relationship links a Descriptive Unit with an Agent that is the topic of the Collection Item.
 - occurrences: 0,*
(inverse associated agent)

- **inverse depicted agent:**
 - This relationship links a Descriptive Unit with an Agent that is depicted by the Collection Item.
 - occurrences: 0,*
(inverse depicted agent)

Place

Place entity contains information about geographical locations that are associated with HOPE collections. This information is retrieved from the metadata records, supplied by the Content Providers, but will be enriched by the HOPE Aggregator.

Elements:

- **persistent identifier place**
 - a globally unique, resolvable URL, identifying a vocabulary term denoting a geographical location.
 - occurrences: 1,1
 - value: record (value:URI, label: constant value)
 - constraints:
 - label="persistent identifier"
 - HOPE Schema: /hopeEntity/persistentID

- **local identifier place**

- an unambiguous reference to a vocabulary term denoting a geographical location in the local system of the Content Provider.
 - occurrences: 1,*
 - value: record (value:literal, label: constant value)
 - constraints:
 - label="local identifier"
 - remarks:
 - This field contains the local ID of the vocabulary list where the preferred term originates from, and the local IDs all terms from other vocabulary list that have been merged with this list.
 - HOPE Schema: /hopeEntity/localID (1,1) → contains the local ID of the native term.
 - HOPE Schema: /hopeEntity/place/alternateID (1,*) → contains the local IDs of the native terms and all merged terms.
- **standard vocabulary:**
 - persistent ID for a corresponding term for the Place in a published data value standard, such as geonames.org.
 - occurrences: 0,*
 - value: record(value:literal, source:literal, baseURL:literal)
 - constraints:
 - value=persistent ID;
 - source=name or abbreviation of the standard vocabulary;
 - baseURL: baseURL prefix for the persistent ID
 - HOPE Schema: /hopeEntity/place/standardVocabulary
 - **preferred term:**
 - The accepted or most desirable term for the Place used in HOPE system.
 - occurrences: 1,*
 - value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
 - constraints:
 - label="name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/place/preferredTerm
 - **alternative term**
 - An equivalent term of the preferred term for the Place, used in HOPE system.

- occurrences: 0,*
- value: record(value:literal, label:constant value, language:controlled value, script: controlled value)
- constraints:
 - label="alternative name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
- HOPE Schema: /hopeEntity/place/alternativeTerm

- **type:**
 - The nature of the Place.
 - occurrences: 0,1
 - value: controlled value
 - constraints:
 - value="local"; "regional"; "national"; "supra national"
 - HOPE Schema: /hopeEntity/place/type

-

- **coordinates**
 - geographical longitude and latitude, defining the geographical location of the Place on the Earth.
 - occurrences: 0,1
 - value:record(value:literal, normalised:controlled value)
 - constraints:
 - controlled value=EPSG 4326
 - HOPE Schema: /hopeEntity/place/coordinates

- **description:**
 - An account of the Place.
 - occurrences: 0,*
 - value: record(value:literal; label:literal; language:controlled value; script:controlled value)
 - constraints:
 - label="description"
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/place/description

- **source:**
 - name of the content provider, who provided the metadata about the Place.

- occurrences: 1,1
- value:record(value:literal, label:constant value)
- constraints:
 - label="source"
- HOPE Schema: /hopeEntity/place/source

Relationships:

- **is creation place of**
 - This relationship links a Descriptive Unit with a geographical location where the Collection Item was created.
 - occurrences: 0,*
(inverse creation place)
- **is publication place of:**
 - This relationship links a Descriptive Unit with a geographical location where the Collection Item was published.
 - occurrences: 0,*
(inverse publication place)
- **inverse associated place:**
 - This relationship links a Descriptive Unit with a geographical location that covers the spatial extent or scope of the Collection Item.
 - occurrences: 0,*
(inverse associated place)
- **inverse depicted place:**
 - This relationship links a Descriptive Unit with a geographical location that is depicted by Collection Item.
 - occurrences: 0,*
(inverse depicted place)

Concept

Concept entity contains information about topical terms that are associated with HOPE collections. This information is retrieved from the metadata records, supplied by the Content Providers, but will be enriched by the HOPE Aggregator.

Elements:

- **persistent identifier concept**
 - a globally unique, resolvable URL, identifying a vocabulary term denoting a concept.
 - occurrences: 1,1
 - value: record (value:URI, label: constant value)
 - constraints:
 - label="persistent identifier"
 - HOPE Schema: /hopeEntity/persistentID

- **local identifier place**
 - an unambiguous reference to a vocabulary term denoting a concept in the local system of the Content Provider.
 - occurrences: 1,*
 - value: record (value:literal, label: constant value)
 - constraints:
 - label="local identifier"
 - remarks:
 - This field contains the local ID of the vocabulary list where the preferred term originates from, and the local IDs all terms from other vocabulary list that have been merged with this list.
 - HOPE Schema: /hopeEntity/localID (1,1) → contains the local ID of the native term.
 - HOPE Schema: /hopeEntity/concept/alternateID (1,*) → contains the local IDs of the native terms and all merged terms.

- **standard vocabulary:**
 - persistent ID of a corresponding term for the Concept in a published data value standard, such as LCSH.
 - occurrences: 0,*
 - value: record(value:literal, source:literal, baseURL:literal)
 - constraints:
 - value=persistent ID;
source=name or abbreviation of the standard vocabulary;
baseURL: baseURL prefix for the persistent ID
 - HOPE Schema: /hopeEntity/concept/standardVocabulary

- **preferred term:**

- The accepted or most desirable term for the Concept, used in HOPE system.
- occurrences: 1,*
- value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
- constraints:
 - label="name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
- HOPE Schema: /hopeEntity/concept/preferredTerm

- **alternative term**
 - An equivalent term of the preferred term for the Concept, used in HOPE system.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, language:controlled value, script: controlled value)
 - constraints:
 - label="alternative name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/concept/alternativeTerm

- **description:**
 - An account of the Concept.
 - occurrences: 0,*
 - value: record(value:literal; label:literal; language:controlled value; script:controlled value)
 - constraints:
 - label="description"
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/concept/description

- **source:**
 - name of the content provider, who provided the metadata about the Concept.
 - occurrences: 1,1
 - value:record(value:literal, label:constant value)
 - constraints:
 - label="source"

- HOPE Schema: /hopeEntity/concept/source

Relationships:

- **inverse associated concept:**
 - This relationship links an object with a concept that is the topic of the object.
 - occurrences: 0,*
(see associated concept relationship under object)
- **inverse depicted concept:**
 - This relationship links an object with a concept that is depicted by the object.
 - occurrences: 0,*
(see depicted concept relationship under object)

Event

Event entity contains information about a phenomena or cultural manifestations, bounded in time and space that are associated with HOPE collections. This information is retrieved from the metadata records, supplied by the Content Providers, but will be enriched by the HOPE Aggregator.

Elements:

- **persistent identifier event**
 - unique, resolvable identifier, identifying the Event in the HOPE information system.
 - occurrences: 1,1
 - value: record (value:URI, label: constant value)
 - constraints:
 - label="persistent identifier"
 - HOPE Schema: /hopeEntity/persistentID
- **local identifier event**
 - a globally unique, resolvable URL, identifying a vocabulary term denoting an event.
 - occurrences: 1,*
 - value: record (value:literal, label: constant value)
 - constraints:
 - label="local identifier"
 - remarks:

- This field contains the local ID of the vocabulary list where the preferred term originates from, and the local IDs all terms from other vocabulary list that have been merged with this list.
- HOPE Schema: /hopeEntity/localID (1,1) → contains the local ID of the native term.
- HOPE Schema: /hopeEntity/event/alternateID (1,*) → contains the local IDs of the native terms and all merged terms.

- **standard vocabulary:**
 - persistent ID of a corresponding term for the Event in a published data value standard.
 - occurrences: 0,*
 - value: record(value:literal, source:literal, baseURL:literal)
 - constraints:
 - value=persistent ID;
 - source=name or abbreviation of the standard vocabulary;
 - baseURL: baseURL prefix for the persistent ID
 - HOPE Schema: /hopeEntity/event/standardVocabulary

- **preferred term:**
 - The accepted or most desirable term for the Event, used in HOPE system.
 - occurrences: 1,*
 - value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
 - constraints:
 - label="name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/event/preferredTerm

- **alternative term**
 - An equivalent term of the preferred term for the Event, used in HOPE system.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
 - constraints:
 - label="alternative name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)

- HOPE Schema: /hopeEntity/event/alternativeTerm

- **date:**
 - start and/or end date of the Event.
 - occurrences: 0,*
 - value: record(value:literal, label:literal, normalised:controlled value, range type:controlled value)
 - constraints:
 - label="date"
normalised=ISO 8601:2004
range type="from"; "to"
 - HOPE Schema: /hopeEntity/event/date

- **spatial coverage:**
 - the location where the event took place.
 - occurrences: 0,*
 - value: record(value:literal, label:location, persistentID:URI)
 - constraints:
 - label="location"
persistentID=PID Place
 - HOPE Schema: /hopeEntity/event/spatialCoverage

- **description:**
 - An account of the Event
 - occurrences: 0,*
 - value: record(value:literal; label:literal; encoding:literal; cataloguing:literal; language:controlled value, script:controlled value)
 - constraints:
 - label="description"
language=ISO-639-3 (language codes);
script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/event/description

- **source:**
 - name of the content provider, who provided the metadata about the Event
 - occurrences: 1,1
 - value:record(value:literal, label:constant value)
 - constraints:
 - label="source"

- HOPE Schema: /hopeEntity/event/source

Relationships:

- **inverse associated event:**
 - This relationship links a Descriptive Unit with a event that is the topic of the Collection Item.
 - occurrences: 0,*
(inverse associated event)
- **inverse depicted event:**
 - This relationship links a Descriptive Unit with an event that is depicted by the Collection Item.
 - occurrences: 0,*
(inverse depicted event)

HOPE Theme

HOPE Theme entity contains information about thematic headings specific to the fields of social and labour history that are associated with HOPE collections. This information is retrieved from the metadata records, supplied by the Content Providers, but will be enriched by the HOPE Aggregator.a

Elements:

- **persistent identifier HOPE theme**
 - the globally unique, resolvable URL, identifying a vocabulary term denoting a HOPE theme.
 - occurrences: 1,1
 - value: record (value:URI, label: constant value)
 - constraints:
 - label="persistent identifier"
 - HOPE Schema: /hopeEntity/persistentID
- **preferred term:**
 - The accepted or most desirable term for the theme used in HOPE system.
 - occurrences: 1,1
 - value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
 - constraints:
 - label="name";
language=ISO-639-3 (language codes);
script=ISO15924 (Codes for the representation of names of scripts)

- HOPE Schema: /hopeEntity/hopeTheme/preferredTerm

- **alternative term**
 - A translation of the preferred term for the HOPE Theme, used in HOPE system.
 - occurrences: 0,*
 - value: record(value:literal, label:constant value, language: controlled value, script: controlled value)
 - constraints:
 - label="alternative name";
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/hopeTheme/alternativeTerm

- **date:**
 - date associated with the HOPE Theme.
 - occurrences: 1,1
 - value:controlled value
 - constraints:
 - value=ISO 8601:2004
 - HOPE Schema: /hopeEntity/hopeTheme/date

- **description:**
 - An account of the HOPE Theme
 - occurrences: 0,*
 - value: record(value:literal; label:literal; encoding:literal; cataloguing:literal; language:controlled value, script:controlled value)
 - constraints:
 - label="description
 - language=ISO-639-3 (language codes);
 - script=ISO15924 (Codes for the representation of names of scripts)
 - HOPE Schema: /hopeEntity/hopeTheme/description

Relationships:

- **inverse associated HOPE theme:**
 - This relationship links an object with a HOPE theme that is the topic of the object.
 - occurrences: 0,*

- range: Descriptive Unit
(inverse associated HOPE theme)

- **broader term:**
 - This relationship links a historical theme with another higher-level historical theme that includes the historical theme.
 - occurrences: 0,1
 - range: HOPE theme
 - value: URI
 - constraints:
 - value=PID HOPE Theme

2.4 HOPE Domain Profiles

Authors:	Armin Straube – FES-A Bert Lemmens – Amsab-ISG Alessia Bardi – CNR-ISTI Paolo Manghi – CNR-ISTI
Internal reviewers:	Armin Straube – FES Afelonne Doek – KNAW-IISG Alessia Bardi – CNR-IST

This chapter describes the five domain profiles, used as an intermediate format when mapping local metadata to the HOPE Schema.

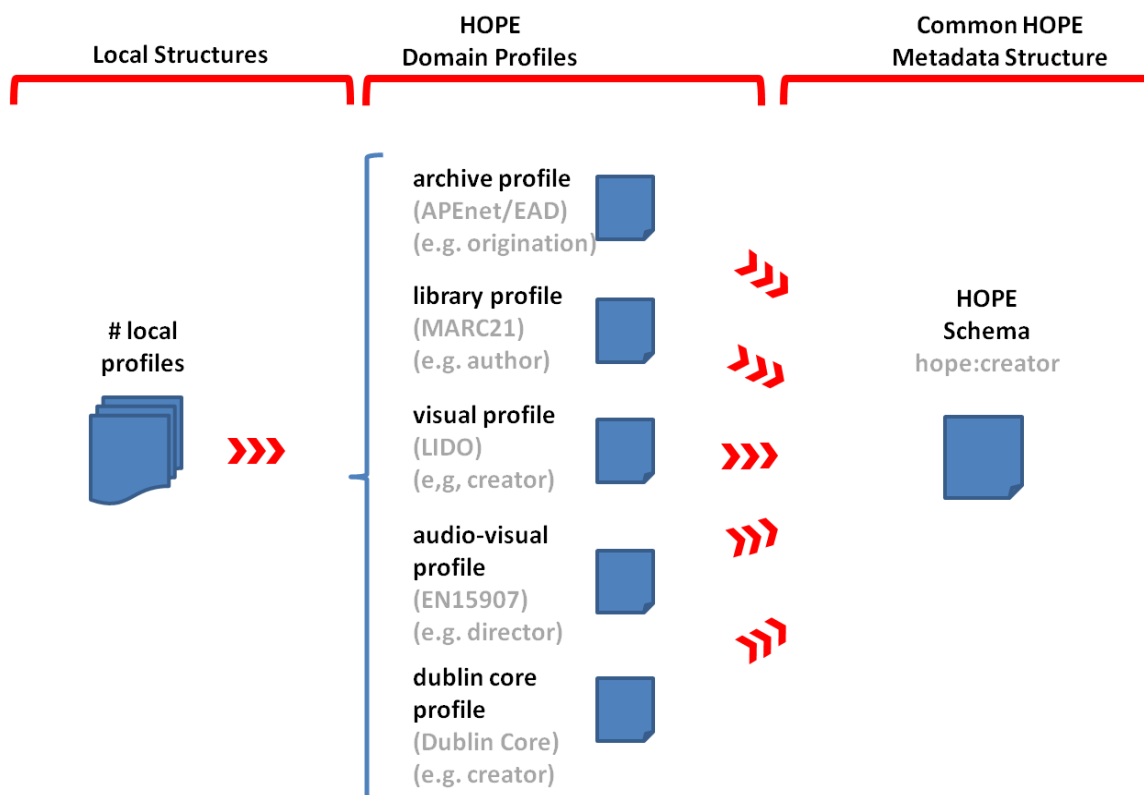
2.4.1 Introduction

2.4.1.1 Purpose of domain profiles

The findings of the content provider survey revealed two main thresholds for the mapping of the metadata from the Content Providers into a common HOPE metadata structure:

- The **cross-domain nature** of Content Providers metadata, i.e the fact that the metadata describes archival, as well as library, visual and audio-visual resources.
- The **idiosyncratic** nature of the Content Providers metadata, i.e the fact that metadata is not encoded according to a published encoding standard (e.g. EAD, MARC, DC, etc.). This means that even within one domain, Content Providers use a wide variety of encoding schemas to prepare their data sets.

In order to overcome these two thresholds, HOPE has specified five domain profiles, which establish an intermediary stage in the mapping process from the local metadata structure of the Content Providers metadata to the Common HOPE Metadata Structure.



By mapping local metadata to one of these profiles, CPs will first overcome the idiosyncratic threshold of their metadata and make the metadata interoperable within its domain. In a second step the HOPE metadata is mapped from these five standard profiles to the common HOPE metadata structure, overcoming the cross-domain threshold.

2.4.1.2 Definition of domain profiles

A domain profile is a subset of *Metadata Elements*, derived from a domain-specific metadata standard. HOPE provides five subsets: four domain-specific profiles and a fifth 'cross-domain' profile, particularly for Content Providers supplying metadata using Dublin Core.

The domain profiles are based on following metadata standards:

- The archive profile is based on the APEnet/EAD encoding standard and the related ISAD(G) cataloguing standard;
- the library profile is based on the MARCXML Schema encoding standard and the related MARC21 bibliographic cataloguing standard;
- the visual domain profile is based on the LIDO encoding standard and the related Spectrum information units.
- the audio-visual profile is based on the EFG Schema and the related EN15907 data model.
- the 'generic' Dublin Core profile is based on the Dublin Core Qualified standard.

These five domains have been identified by the HOPE Content Providers Survey. Initially the survey started from the four domains or communities identified by Europeana:

- museums,
- archives,
- libraries and
- audio-visual collections.

The names of these domains primarily refer to institutions and a particular type of resources.

In order to relieve the strong institutional connotation of the term, HOPE discarded the name 'museum collection' and replaced it by 'visual collection', since many HOPE content providers have photographic collections described with metadata, which fits very well in the museum domain, but are held by institutions which consider themselves as libraries or archives.

Please note that the Content Provider does not need to choose one single domain profile for all its collections. The Content Provider can choose for each collection separately the domain profile that is most appropriate. This means that the type of domain profile does not have to correspond with the type of material or institution. HOPE agreed that, when choosing a particular domain profile for a specific collection, the encoding and cataloguing standard already used by the Content Provider may take priority.

For example: A Content Provider currently using MARC21 for encoding audio-visual collections may choose for a library profile. A Content Provider currently using EAD for encoding library resources may choose for an archive profile.

In both examples the mapping procedure to the domain profiles is relatively easy. However in these cases, Content Providers should be aware that their metadata will be presented in the IALHI and Europeana portal in a domain-specific context that corresponds to the chosen domain profile, rather than on the actual type of the material.

Europeana encourages Content Providers to establish agreement on the use of a common domain-specific encoding standard for collections from each of these four different domains. A common metadata standard for each of these domains would simplify the integration of rich, domain-specific metadata in the Europeana Data Model (EDM). Currently there is agreement in two domains on using a common encoding standard: APENet/EAD in the archive domain and LIDO in the Museum domain. HOPE has incorporated the agreements reached so far and studied the current best practices in the visual and audio-visual domain.

The four domain-specific collection types have been defined as follows:

Archive collection	<ul style="list-style-type: none"> ● Archive collections typically accommodate (large collections of) materials created or received by a person, family or
--------------------	---



	<p>organisation/institution, public or private, in the conduct of their affairs</p> <ul style="list-style-type: none"> ● Because of the extent of an archival collection, archival descriptions typically refer to multiple documents regarding the same event or case. ● Archival records are typically organised in a multi-level descriptions. Lower levels inherit the characteristics of the the higher level descriptions.
Library collection	<ul style="list-style-type: none"> ● Library collections typically accommodate 'published' materials, e.g. monographs, periodicals, leaflets, posters, etc. ● The metadata typically consists of title descriptions, which may include additional information about the "copy-at-hand". ● Bibliographic records can have a hierarchical structure including series, titles and item descriptions.
Visual collection	<ul style="list-style-type: none"> ● visual collections typically accommodate sets of single objects, e.g. photographs, posters, prints, objects etc. ● The metadata typically holds item descriptions for each single object ● Visual records typically hold rich metadata on single objects, sometimes gathered in formal/informal collections, e.g. a series of prints, a cycle of paintings.
Audio-visual collection	<ul style="list-style-type: none"> ● Audio-visual collections typically accommodate multimedia resources, such as film, video, broadcasts, sound recordings. ● The metadata typically holds title descriptions and metadata on the related events, such as releases and broadcasts ● Audio-visual records typically require a distinction between the creation and the manifestation of the resource

After the first test mappings in December 2010, a fifth domain profile has been added, i.e. the Dublin Core Profile. Some Content providers supply collections encoded using Dublin Core Simple. These collections involved different types of materials (archive, as well as library and audio-visual). Though, the generic nature and the limited number of metadata elements specified by Dublin Core made it difficult to map this metadata in one of the four domain-specific domain profiles. Therefore, a fifth profile has been added, based on the Dublin Core Simple element set. This profile may accommodate archival, as well as library, visual and audio-visual material.

2.4.2 HOPE Domain Profiles

This chapter contains the rationale, used when specifying the five domain profiles. For each domain profile, the chapter includes:

- **Survey results:** a brief discussion of the results of the content provider survey relevant to the specific domain profile.
- **Encoding standard:** a brief discussion of the encoding standard on which the domain profile is based.
- **Cataloguing standard:** The logical framework given in the next paragraph sets the domain profile in a wider context. The reason for this is the use of idiosyncratic element sets by many content providers. The metadata structure of many collections supplied to Hope by the content providers has few resemblance with the established metadata standards. The specification of an additional community-specific cataloguing standard will provide guidance for choosing the appropriate domain profile, as well as mapping to the appropriate metadata elements.
- **Europeana Best Practice:** The last paragraph in the description of the HOPE domain profiles will deal with Europeana best practices. Several Europeana related projects have worked on reaching agreement on the use of common namespaces and 'metadata dialects' to establish semantic interoperability within their network. Some of these projects have specified an element set for one particular type of metadata, e.g. for museum or for audio-visual resources. Within the museum and archive community in particular there is a growing consensus to use a common profile (LIDO and EAD respectively) for supplying metadata to Europeana. The interoperability of the HOPE metadata will benefit from incorporating these best practices in its own common HOPE metadata structure.

Following table contains an overview of the metadata standards T2.3 agreed on for specifying the domain profiles.

	encoding standard	cataloguing standard	Europeana Best practice
Archive Profile	EAD v. 2002	ISAD(G) 2nd ed. (2000)	APEnet
Library Profile	MARCXML Schema v.1.2 (2009)	ISBD consolidated 2007	/
Visual Profile	LIDO v1.0 (2010)	Spectrum v3.1 (2007)	ATHENA
Audio-Visual Profile	EFG XML Schema (2009)	/	EN15907

Dublin Core Profile	DC Qualified	/	ESE v3.3.1
---------------------	--------------	---	------------

2.4.2.1 HOPE archive profile

This chapter describes the rationale used for the specification of the HOPE Archive Profile.

Survey results

The main findings regarding archival collections within the HOPE project according to the content provider survey are:

- 9 content providers (46% of the metadata records) indicated they have one or more archival collections.
- 5 content providers (86,4% of the archive metadata records) do not use a metadata standard for encoding this archive metadata, but one or more “idiosyncratic” sets of metadata elements.
- 3 CPs (9,7% of the archive metadata) use EAD 2002 and 1 CP (0,15% of the archive metadata) uses Dublin Core.
- One CP (3,4% of the archive metadata) uses the ISAD(G) cataloguing standard as a basis for its idiosyncratic element set.

These numbers indicate that the large majority of the HOPE metadata on archive resources are encoded with idiosyncratic element sets. Only a small minority uses established archive metadata standards such as ISAD(G) or EAD.

A number of the collections marked as archival collections in the survey can also be mapped to the visual profile. These collections have idiosyncratic metadata and the content are photographs.

For a listing of all HOPE Collections for which a mapping using an archive profile is considered most appropriate, see Appendix A: HOPE Archive Collections.

T2.3 decided to use EAD as a source for the defining the subset of archival metadata elements. EAD is a well-established encoding standard in the archive domain, and is also used by the APENet project, building a european archival metadata aggregator that will be linked with the Europeana Portal. In order to guide Content Providers with idiosyncratic data sets, the archive profile will use ISAD(G) as a framework for grouping the EAD based elements in logical units.

Encoding standard

The encoding standard used to define the HOPE archive profile is Encoded Archival Description version 2002 (EAD). EAD is a data structure standard used to encode finding aids for use in a networked (online) environment. It reflects the hierarchical nature of archival collections and it provides a structure for describing the whole of a collection, as well as its

components. EAD comprises a Document Type Definition (DTD) for encoding archival finding aids that is written following the syntactic rules of Standard Generalized Markup Language (SGML) and Extensible Markup Language (XML). EAD is considered as a 'de facto' standard for encoding archival resources. EAD originated in 1993, at the University of California, Berkeley. The project's goal was to create a standard for describing collections held by archives and special collections, similar to the MARC standards for describing regular books.

EAD is currently maintained by the Society of American Archivists (SAA) and the Library of Congress (LoC). The SAA Encoded Archival Description Working Group is responsible for updating and editing the EAD DTD and tag library. The Network Development and MARC Standards Office of the LoC serves as the maintenance agency for online EAD documentation, including storage and delivery of electronic files and maintenance of the EAD web site.

The current EAD version 2002 established compatibility with the ISAD(G).

Reference documents:

- [official EAD website Library of Congress](http://www.loc.gov/ead/index.html)⁷
- [EAD website Society of American Archivists](http://www.archivists.org/saagroups/ead/index.html)⁸
- [EAD tag library \(pdf, 800 kB\)](http://archivists.org/publications/epubs/EAD2002-TL-05-03.pdf)⁹

EAD does not specify the "intellectual content" but defines the encoding designations. It eases the ability to exchange finding aids among institutions and allows users to find out about collections in distant places. Moreover, The EAD XML Schema is flexible, hence it allows to encode the same information in several ways.

Using EAD as a common metadata structure therefore requires the specification of the "intellectual content" accommodated by the selected metadata elements. This is particularly useful when mapping idiosyncratic element sets to EAD.

Cataloguing standard

The logical framework of the HOPE archive profile is provided by the General International Standard for Archival Description (ISAD(G)). ISAD(G) is a set of cataloguing rules, providing "general guidance for the preparation of archival descriptions. It is to be used in conjunction with existing national standards or as the basis for the development of national standards." (ISAD(G): 1)

⁷Official EAD website Library of Congress, <http://www.loc.gov/ead/index.html>

⁸EAD website Society of American Archivists, <http://www.archivists.org/saagroups/ead/index.html>

⁹EAD tag library, <http://archivists.org/publications/epubs/EAD2002-TL-05-03.pdf>

ISAD(G) has been developed by the International Council on Archives Ad Hoc Commission on Descriptive Standards (ICA/DDS). The first version has been published in 1994. The permanent ICA Committee on Descriptive Standards (ICA/CDS) is responsible for maintaining ISAD(G) and published a second version in 2000.

ISAD(G) defines 26 elements that may be combined to constitute the description of an archival entity. The structure and content of the information in each of these elements should be formulated in accordance with applicable national rules. As general rules, these are intended to be broadly applicable to descriptions of archives regardless of the nature or extent of the unit of description.

The elements are organized into seven areas of descriptive information:

- Identity Statement Area: where essential information is conveyed to identify the unit of description
- Context Area: where information is conveyed about the origin and custody of the unit of description
- Content and Structure Area: where information is conveyed about the subject matter and arrangement of the unit of description
- Condition of Access and Use Area: where information is conveyed about the availability of the unit of description
- Allied Materials Area: where information is conveyed about materials having an important relationship to the unit of description
- Note Area: where specialized information and information that cannot be accommodated in any of the other areas may be conveyed
- Description Control Area: where information is conveyed on how, when and by whom the archival description was prepared

The ISAD(G) areas and elements will be used as a framework for organising the selected EAD elements in logical information units.

The Archive Profile will include following areas:

- Identity Statement Area
- Context Area
- Content and Structure Area
- Condition of Access and Use Area
- Allied Materials Area
- Notes Area

For a listing of all domain-specific elements contained by the HOPE Archive Profile, see Appendix B: Archive Profile

Reference documents:

- [ISAD\(G\): General International Standard Archival Description Second Edition \(2000\)](#)¹⁰

Europeana best practice

The HOPE archive profile will be compliant with the specifications set up by the Archives Portal Europe network (APEnet). APEnet is a European sister project of HOPE, which runs from January 2009 till January 2012. APEnet is network of fourteen European National Archives in close cooperation with the Europeana initiative. APEnet is building a common gateway for archival resources, but also aims at bringing national archival collections to Europeana.

For this purpose APEnet has specified an EAD profile as a common metadata format. This profile is compliant with the EAD 2002 specifications, but has been extended with labels for the display of metadata in its own portal, as well as metadata standards for the validation of particular metadata elements.

APEnet is currently working on a mapping of APEnet EAD elements to EDM.

Reference documents:

- [APEnet project website](#)¹¹
- [APEnet project page on EAD](#)¹²
- [APEnet EADprofile: specification EAD elements](#)¹³
- [Overview APEnet EAD profile](#)¹⁴

2.4.2.2 HOPE library profile

This chapter describes the rationale used for the specification of the HOPE Library Profile.

Survey results

The main findings regarding library collections according to the content provider survey are:

¹⁰ISAD(G): General International Standard Archival Description Second Edition (2000), http://www.ica.org/sites/default/files/isad_g_2e.pdf

¹¹APEnet project website, <http://apenet.nac.kei.pl/index.php?lang=en>

¹²APEnet project page on EAD, http://apenet.nac.kei.pl/index.php?option=com_content&view=article&id=94&Itemid=150&lang=de

¹³APEnet EADprofile: specification EAD elements, http://apenet.nac.kei.pl/images/stories/apenet/all/technics/100809_ead_finding_aid_guide_apenet.pdf

¹⁴Overview APEnet EAD profile, http://apenet.nac.kei.pl/images/stories/apenet/all/technics/100809_ead_finding_aid_profile_apenet.pdf

- 8 content providers (26,7% of the metadata records) indicated they have one or more library collections.
- 4 content providers (AMSAB, FMS, IISG, OSA representing 13,5% of the library metadata records) do not use a metadata standard for encoding library metadata, but one or more “idiosyncratic” sets of metadata elements. AMSAB uses the ISBD cataloguing rules for the specification of its element set.
- 3 content providers (FES library, CGIL, IISG representing 84,8% of the library metadata records) use a metadata standard from the "MARC family" for encoding library metadata.
 - FES library uses MAB, the German national library standard based on MARC21, as an export format (84,3% of the library metadata records).
 - CGIL uses MODS, a subset of MARC fields that uses language-based tags rather than numeric ones, as an export format.(CGIL, representing 0,39% of the library metadata records)
 - IISG uses native MARCXML (0,09% of the library metadata records)
- UPIP (MSH) uses EAD as an export format (1,58% of the library metadata records), but the metadata is native ISBD.
- UPIP (BDIC) uses Dublin Core Simple (0,09% of the library metadata records)

Since 84,3% of the HOPE bibliographic metadata is supplied by FES Library, the major part of the HOPE bibliographic metadata is encoded using MAB. However, MAB is only used by one CP. The most 'commonly used' metadata standard can be denoted as the "MARC family", including native MARCXML and MARC21-related MAB and MODS.

Moreover, Amsab mentioned the use of ISBD cataloguing rules as a reference for their idiosyncratic element set and UPIP(MSH) uses EAD as an encoding format. These metadata can relatively easily be mapped to MARC21 Bibliographic metadata elements.

Also the DC Simple encoded metadata from the UPIP(BDIC) can easily be mapped to MARC21, in spite of the limited specificity of the metadata

For a listing of all HOPE Collections for which a mapping using a library profile is considered most appropriate, see Appendix A: HOPE Library Collections.

T2.3 decided to use MARC21 Bibliographic as a source for specifying the subset of library metadata elements. Since HOPE uses XML for encoding metadata, MARCXML Schema will be used for encoding metadata mapped to the library profile.

Since a number of Content Providers are using the ISBD cataloguing rules for creation of library records, the Library Profile will include references to ISBD in the logical units of the profile. The ISBD “areas of description” will be used for grouping the MARC based elements in logical units.

Encoding standard

The encoding standard used to define the HOPE library profile is MARCXML Schema. The MARCXML Schema supports XML markup of full MARC21 records, featuring lossless conversion to and from MARC21 records, a conversion toolkit and style sheets. It simply

duplicates the MARC content designation structure in a native XML encoding format.

For the specification of the subset of library elements, HOPE borrows element names and semantics from MARC21 Bibliographic.

MARC 21 Bibliographic is designed to be a standard for the representation and communication (exchange) of bibliographic and related information in machine-readable form (MARC is an acronym for MACHine Readable Cataloging). Metadata records are encoded into electronic form by assigning tags, numbers, letters, or words (i.e. codes) to discrete pieces of information in the description. MARC 21 Bibliographic is used to create and maintain bibliographic records concerning printed and manuscript textual materials, computer files, maps, music, continuing resources, visual materials, and mixed materials. Bibliographic data commonly includes titles, names, subjects, notes, publication data, and information about the physical description of an item.

MARCXML is maintained by the Library of Congress' Network Development and MARC Standards Office. The current version, used by T2.3 for specification of the library profile is MARC 21 XML Schema (version 1.2) 2009 (revised; version 1.0: 2002).

Reference documents:

[Official MARC21 website](#)¹⁵

[Official MARCXML website](#)¹⁶

[MARCXML: The MARC 21 XML Schema](#)¹⁷

Cataloguing standard

The logical framework of the HOPE library profile is provided by the International Standard Bibliographic Description consolidated edition (ISBD). ISBD is a set of rules to describe a wide range of library materials within the context of a catalogue. They also organise the display of bibliographic description of an item in a catalogue. The consolidated edition supersedes earlier separate ISBDs that were published for monographs, older monographic publications, cartographic materials, serials and other continuing resources, electronic resources, non-book materials, and printed music. The ISBD prescribes eight areas of description and their order. Each area is composed of multiple elements. Elements and areas that do not apply to a particular resource are left out of the description. Standardised punctuation is used to identify and separate the areas and the elements.

The logical structure is as follows:

- Title and statement of responsibility area (author, editor, artist)

¹⁵Official MARC21 website, <http://www.loc.gov/marc/bibliographic/>

¹⁶Official MARCXML website, <http://www.loc.gov/standards/marcxml/>

¹⁷MARCXML: The MARC 21 XML Schema, <http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>

- Edition area
- Material or type of resource specific information
- Publication and distribution area (“impressum”)
- Physical description area (“collation”)
- Series area
- Notes field
- Resource identifier (standard number) and terms of availability area

The Library profile has the following areas

- Identification Area (cf. Resource identifier area)
- Title Area
- Statement of Responsibility Area
- Edition Area
- Publication Area
- Physical Description Area
- Subject Area (added because ISBD does not include a defined area for subject headings)

For a listing of all domain specific elements contained by the HOPE Library Profile, see Appendix B: Library Profile

ISBD is currently maintained by the ISBD Review Group of the International Federation of Library Associations and Institutions (IFLA).

Reference documents:

[ISBD review group](#)¹⁸

[ISBD consolidated \(2007\)](#)¹⁹

[Draft review ISBD consolidated \(10-05-2010\)](#)²⁰

Europeana best practice

In the Europeana community, The European Library (TEL) is the domain-specific network for

¹⁸ISBD review group , <http://www.ifla.org/isbd-rg>

¹⁹ISBD consolidated (2007),
http://archive.ifla.org/VII/s13/pubs/ISBD_consolidated_2007.pdf (PDF, 1,73 MB)

²⁰Draft review ISBD consolidated (10-05-2010),
http://www.ifla.org/files/cataloguing/isbd/isbd_wwr_20100510_clean.pdf (PDF, 1,5 MB)



the library community. Moreover, TEL provides the organisational structure for the creation of Europeana, i.e. the cross-domain portal, aggregating library as well as archive and museum metadata. Although the creation of the Europeana has somehow evolved from The European Library, the initial idea however was to create an on-line portal providing access to the electronic resources of the National Libraries of Europa.

To this goal, TEL created a TEL Application Profile, based on the Dublin Core Library Application Profile. The intention was to have a clear data model, particularly supporting the functionalities of the TEL portal, in spite of the fact that Dublin Core is not able recording Library metadata in its full specificity. This TEL Application Profile has been the source for harvesting library metadata by the Rhine release of the Europeana Portal. TEL is currently working on an application profile that accommodates 'rich' library metadata, anticipating the forthcoming (EDM based) Danube release.

Regarding the compliancy between HOPE and TEL encoded library resources, HOPE is using the same metadata standards as TEL, i.e. MARC21 Bibliographic and DC. It is important to note that the current TEL Handbook includes a default mapping of MARC21 records to the TEL/DC Application Profile.

Further, HOPE should be aware of the research TELplus has done on the implementation of FRBR categories in the TEL Application Profile. TELplus was a follow-up project of the original TEL project, aiming at further strenghtening, expanding and improving The European Library. Within this scope, TELplus has designed and built a prototype of a web-service supporting searching and browsing according to the FRBR paradigm. The Europeana v.1.0 project is considering the integration of FRBR categories in a future release of the EDM. The HOPE project should follow further developments in this field, in order to benefit from tools developed to transform current DC-based TEL application profile to a FRBR compliant EDM profile.

2.4.2.3 HOPE visual profile

This chapter describes the rationale used for the specification of the HOPE Visual Profile.

Survey results

The content provider survey asked specifically to identify museum collections, instead of visual collections. However, a couple of collections defined as archive and library collections by the content providers might be considered visual collections as well.

As regards the collections identified by Content Providers as museum collections, the content provider survey revealed that:

- 5 CPs (26,8% of the HOPE metadata) supply one or more visual collections
- IISG uses MARC21 for encoding its visual metadata (59,9% of the museum metadata)
- SSA and UPIP(BDIC) use Dublin Core Simple for encoding visual metadata (11,9% or the museum metadata)
- Amsab and TA do not use a metadata standard for encoding this visual metadata, but one or more "idiosyncratic" sets of metadata elements (28,2% or the museum metadata)

When adding the potential visual collections from the archive and library domain, the numbers are as follows:

- 11 CPs (52% of the HOPE metadata) supply one or more collections which can be identified as visual collections
- 2 additional CPs (CGIL and Génériques) use EAD for encoding museum metadata (6,1% of the HOPE metadata)
- 3 CPs (FES archive, FMS and OSA) do not use a metadata standard for encoding their metadata, but one or more “idiosyncratic” sets of metadata elements (together with Amsab and TA 53,6% of the visual metadata)
- MARC21 and Dublin Core respectively represent 31% and 6,1% of the visual metadata.

In both cases, the variety in metadata standards and idiosyncratic element sets is too large to identify a 'commonly used' metadata standard. As for now, it can be concluded that several CPs use metadata standards borrowed from other 'domains' to encode rich item descriptions of visual resources, i.e. IISG (MARC21), CGIL and Génériques (EAD). None of the HOPE CPs use an encoding format from the museum community, such as CDWA lite, VRA core, museumdat. Several CPs seem to use idiosyncratic element sets, which are often extended archive or library formats.

For a listing of all HOPE Collections for which a mapping using a visual profile is considered most appropriate, see Appendix A: HOPE Visual Collections.

Regarding the encoding standard for the visual profile, T2.3 decided to deliberately choose the Europeana best practice metadata standard for the museum domain, i.e. LIDO, although none of the Content Providers is currently using this standard. The arguments are that LIDO is compliant with the best-practice encoding and cataloguing standards in the museum domain (i.e. CIDOC-CRM, museumdat, CDWA and Spectrum), the encoding standard is endorsed by the ATHENA related museum network in the Europeana Community, and mapping to LIDO will enable the integration of rich HOPE visual metadata in the forthcoming Danube release of the Europeana Portal.

In order to provide guidance for mapping the variety of metadata structures for visual resources into the LIDO format, the visual profile will refer to Spectrum Information Groups for grouping the LIDO based elements into logical units.

Encoding standard

The encoding standard for the HOPE visual profile originates from the museum domain: i.e. Lightweight Information Describing Objects (LIDO). LIDO is a metadata format for contributing museum information for resource discovery, developed by the ATHENA project, which enables to supply rich metadata about museum objects to the Europeana portal. Athena introduced this new standard (which is heavily based on CDWA, museumdat, CIDOC-

CRM and SPECTRUM) to overcome the shortcomings of Dublin Core (especially in describing museum objects), and to be able to anticipate future changes of the ESE set. LIDO, specified as XML Schema, is the result of a joint effort of the CDWA Lite, museumdat, SPECTRUM and CIDOC CRM communities. The schema combines the CDWA Lite and museumdat schemas and is informed by SPECTRUM. Being CIDOC-CRM compliant, it aims at contributing information of all kinds of museum objects for resource discovery. LIDO is currently maintained by The CIDOC Data Harvesting and Interchange Working Group, which provides information about, and supports the development and application of, formats and techniques for harvesting and interchanging information of relevance to the museum community. The Data Harvesting and Interchange Working Group aims to "centralize information and activities that are currently taking place with Lightweight Information Describing Objects (LIDO) under a committee that can sustain its understanding, advancement, exposure, and applicability." ([CIDOC: Data Harvesting and Interchange Working Group](#))

Reference documents:

- [CIDOC Data Harvesting and Interchange Working Group](#)²¹
- [ATHENA project](#)²²
- [ATHENA D3.3 Specification for conversion tools](#)²³
- [Appendix ATHENA D3.3](#)²⁴
- [LIDO Schema](#)²⁵

Cataloguing standard

As logical framework serves the SPECTRUM Standard for Collections Management (v3.1).

SPECTRUM is a collection management standard. It catches the complete spectrum of activities that are performed to document and manage a museum collection.

SPECTRUM architecture consists of two components:

- procedures: The SPECTRUM standard defines 20 procedures. Each of these procedures describes the steps that need to be carried out to perform a specific task in a museum. Examples of such tasks are: the entry of objects in the museum, loans

²¹CIDOC Data Harvesting and Interchange Working Group, http://cidoc.icom.museum/WG_Data_Harvesting%28en%29%28E1%29.xml

²²ATHENA project, <http://www.athenaeurope.org/index.php?en/91/information-on-the-project>

²³ATHENA D3.3 Specification for conversion tools, <http://www.athenaeurope.org/getFile.php?id=539> (PDF, 153 kb)

²⁴Appendix ATHENA D3.3, <http://www.athenaeurope.org/getFile.php?id=535> (261 kb)

²⁵LIDO Schema, <http://www.athenaeurope.org/getFile.php?id=536> (XSD, 85 kb)

of objects (in and out) and risk assessment. A separate chapter is devoted to each of these procedures. The 20 procedures form the first half of the SPECTRUM publication.

- units of information: At the end of each chapter a list of data elements is specified, in the SPECTRUM jargon called 'units of information'. The second half of the publication is an overview of all these units of information, their definition and usage and how they are linked to other units of information.

SPECTRUM is an open standard, maintained by The Collection Trust²⁶ on behalf of UK museums. The Collections Trust is funded in England by the Museums, Libraries and Archives Council, in Scotland by the Scottish Museums Council and in Wales by CyMAL to make SPECTRUM available free of charge for non-commercial use by museums. SPECTRUM is available for download as an electronic document. To obtain a copy you will first need to read and agree to the terms of the SPECTRUM licence.

The SPECTRUM standard is widely used in the UK and is adopted in quite a few other countries (especially in Europe, e.g. Belgium, Germany and the Netherlands). SPECTRUM is available from the Collections Trust in printed form or in electronic form. The electronic version is called [SPECTRUM interactive](#)²⁷.

The Visual Profile will group the Spectrum Information Units into following areas:

- Identification area
- Creation area
- Physical Description area
- Content Area
- Association Area

For a listing of all domain-specific elements contained by the HOPE Visual Profile, see Appendix B: Visual Profile.

Reference documents:

- [registration form SPECTRUM v3.1 \(UK\)](#)(March 2007)²⁸
- [SPECTRUM XML schema v3.1](#) (March 2007)²⁹

²⁶The Collections Trust, <http://www.collectionstrust.org.uk/>

²⁷SPECTRUM interactive, <http://www.collectionstrust.org.uk/spectrum>

²⁸registration form SPECTRUM v3.1 (UK)(March 2007), <http://www.collectionstrust.org.uk/spectrum>

²⁹SPECTRUM XML schema v3.1, <http://www.collectionstrust.org.uk/spectrum-schema-3-1.zip>

Europeana best practice

The subset of metadata elements for the HOPE visual profile will be borrowed from the LIDO XML Schema, which has been created by the ATHENA project.

ATHENA is a project that deliberately aims for providing museum metadata to Europeana and developed technical tools and standards, to facilitate the inclusion of digital content belonging to European museums into Europeana.

For this purpose ATHENA developed LIDO as a common metadata format for bringing rich metadata and digital content to Europeana.

ATHENA enables Content Providers mapping their metadata to the LIDO XML Schema, from where the metadata is mapped to the Europeana format. This process is analogous to the way LIDO acts as a domain profile in the HOPE System.

The idea for LIDO originated from the common view within the museum community that a DC derived metadata schemas (cf. TEL Application Profile) do not deliver a rich enough view of museum content. DC-based systems 'flatten out' museum metadata, with most of the data going into limited subset of elements. Which led the ATHENA Project to seek to use, or develop, a metadata schema suitable for harvesting museum data into the service environment. The end result of the process was the decision to create a schema based on the existing museumdat schema – LIDO (Light Information Describing Objects). LIDO is not really a new schema. It builds on existing standards and best practise from a number of different countries in Europe and the rest of the world. LIDO is the result of a joint effort of the CDWA Lite, museumdat, SPECTRUM and CIDOC CRM communities. The schema combines the CDWA Lite and museumdat schemas and is informed by SPECTRUM concepts. It is CIDOC CRM compliant and can be used to submit information about of all kinds of cultural heritage objects.

Reference documents:

- [CIDOC Data Harvesting and Interchange Working Group](#)³⁰
- [ATHENA project](#)³¹

2.4.2.4 HOPE audio-visual profile

³⁰CIDOC Data Harvesting and Interchange Working Group, http://cidoc.icom.museum/WG_Data_Harvesting%28en%29%28E1%29.xml

³¹ATHENA project, <http://www.athenaeurope.org/index.php?en/91/information-on-the-project>

This chapter describes the rationale used for the specification of the HOPE Visual Profile.

Survey results

Audio-visual collections have not been addressed specifically in the Content Providers Survey. The collection type has been introduced after the analysis of the survey results, as a result of CPs indicating the need for a specific profile for audio-visual resources. In the Survey, audio-visual collections have been denoted as museum, as well as library, archive and 'other' collections. However, audio-visual collections could be easily identified by the indicated material type. The content provider survey revealed that that:

- 6 CPs supply one or more sub-collections which can be identified as audio-visual sub-collections (26,8% of the HOPE metadata)
- Amsab, OSA and TA do not use a metadata standard for encoding audio-visual metadata, but one or more "idiosyncratic" sets of metadata elements (63,4% of the audio-visual museum metadata)
 - Amsab uses the same idiosyncratic element set as for its museum metadata.
 - OSA uses two different idiosyncratic element sets: one with specific elements for audio-visual media, and another that is partly based on Dublin Core Simple
 - TA also uses an idiosyncratic set that is partly based on Dublin Core Simple.
- SSA uses Dublin Core Simple for encoding audio-visual metadata (27,4% of the museum metadata)
- IISG uses MARC21 for encoding its audio-visual metadata (9% of the audio-visual metadata)
- We did not receive an metadata element set from VGA yet.

In view of the large variety of element sets among these 6 CPs, used to describe audio-visual objects, it is not possible to define a common metadata standard. Although 3 CPs use elements that are valid or based on Dublin Core Simple, 2 CPs borrow element sets from other communities in order to record rich metadata on audio-visual resources.

In order to ensure interoperability of the audio-visual metadata, the namespace for the HOPE audio-visual profile will depend on the best practice within the audio-visual community.

T2.3 has considered following encoding standards, which have been developed by Europeana related project, specifically dealing with audio-visual resources:

- EBUcore metadata set, the metadata format used by the EUscreen project for the supply of audio-visual metadata to Europeana
- EFG Schema, the XML Schema used by the European Film Gateway project as 'common interoperability framework' for servicing the EFG portal.

Both Schemas are implementations of a metadata standard for cinematographic works produced by the European Committee for Standardization (CEN). This Cinematographic Works Standard is being prepared in two parts:

- EN15744 specifies a minimum set of information elements for the unambiguous identification of film works. This element is an implementation of the DC application profile.
- EN15907 is a specification for structuring machine-processable metadata about cinematographic works. This part is presented as a data model, intended to serve as a guideline for information professionals seeking to build interoperable systems that carry information about moving images. EN15907 draws upon several existing standards, recommendations and reference models.

Faced with the conflicting needs for rich audio-visual profile by CPs such as IISG, AMSAB and OSA on the one hand, and CPs supplying audio-visual metadata encoded using DC such as SSA, OSA and TA on the other hand, T2.3 opted for the following solution.

The HOPE Audio-Visual Profile will be using metadata elements borrowed from EN15907 specification. For encoding the HOPE Content Providers metadata on audiovisual resources, the HOPE System will use modified EFG Schema that only includes the subset of EN15907 elements specified by HOPE.

For a listing of all HOPE Collections for which a mapping using a audio-visual profile is considered most appropriate, see Appendix A: HOPE Audio-Visual Collections.

Encoding standard

The EN15907 data model has been used by EFG as well as EUscreen for specifying a audio-visual metadata profile, tailored to the needs of their content providers. T2.3 also specified a HOPE specific set of metadata elements, that should ensure accomodating the rich audio-visual metadata supplied by IISG, OSA and AMSAB.

WP3 tends to re-use the EFG Schema, without the EFG specific elements and including all EN15907 specified by HOPE.

Reference documents:

- [EFG project home page](#)
- [EFG Guidelines and Standards](#)
- [EFG Schema](#)

Cataloguing standard

As for grouping of EN15907 compliant elements in logical units, there was no obvious cataloguing standard available. For pragmatic reasons, T2.3 decided to use the areas already specified in the Visual and Archive profiles. The Audio-Visual metadata elements are grouped into following areas:

- Identification area
- Creation area
- Physical Description area

- Content area
- Association area
- Conditions of Access and Use area

For a listing of all domain-specific elements contained by the HOPE audio-visual Profile, see Appendix B: Audio-Visual Profile.

Europeana best practice

The EN15907 Data Model receives acknowledgement for best practice as to recording audio-visual metadata.

Reference Documents:

- filmstandards.org
- [EN15907: Film identification - Enhancing interoperability of metadata - Element sets and structures](#)

2.4.2.5 HOPE Dublin Core profile

This chapter describes the rationale used for the specification of the HOPE Dublin Core Profile.

Survey results

After an analysis of the results from the HOPE Content Providers Survey, and the feedback from the first test mappings to the domain profiles in December 2010, it was decided to add a fifth 'generic' profile to the four domain-specific profiles. This profile particularly supports Content Providers that supply rich, domain-specific metadata transformed into Dublin Core Simple. During the first test mappings, mapping this 'flattened' data back into a domain-specific metadata structure appeared to be not feasible.

This profile may accommodate cross-domain metadata, supplied by:

- some of the archive collections supplied by OSA
- archive, library and museum collection supplied by UPIP BDIC
- archive and audio-visual collections supplied by TA
- audio-visual collections supplied by SSA.

Encoding standard

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio; "core" because its elements are broad and generic, usable for describing a wide range of resources.

The fifteen elements which are described in this standard are part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI).

All changes made to terms of the Dublin Core Metadata Element Set since 2001 have been reviewed by a DCMI Usage Board in the context of a DCMI Namespace Policy. The namespace policy describes how DCMI terms are assigned Uniform Resource Identifiers (URIs) and sets limits on the range of editorial changes that may allowably be made to the labels, definitions, and usage comments associated with existing DCMI terms.

The implementation of DC element set in the HOPE DC profile is analogous with the ESE implementation (see below), including following additions of DCMI terms elements :

- the dc:format element has been replaced by is dcterms sub-properties dcterms:extent and dcterms:format.
- the dc:coverage element has been replaced by is dcterms sub-properties dcterms:spatial and dcterms:temporal.

reference documents:

- [Dublin Core Metadata Initiative](#)
- [DC Simple and Qualified Schemas](#)

Cataloguing standard

As for grouping of the Dublin Core elements in logical units, T2.3 decided to use the areas already specified in the Audio-Visual profile, i.e. a selection of area from the Visual and Archive profiles. The Dublin Core metadata elements are grouped into following logical units:

- Identification area
- Creation area
- Physical Description area
- Content area
- Association area
- Conditions of Access and Use area

For a listing of all 'generic' elements contained by the HOPE Dublin Core Profile, see Appendix B: Dublin Core Profile.

Europeana best practice

As already discussed above, Dublin Core Application Profiles have been used within the scope of The European Library and in the first release of the Europeana Portal. For the Rhine release, the EDL foundation specified an application profile including Dublin Core elements for recording cross-domain descriptive metadata, and additional administrative elements (cf. isShownBy and isShownAt) for supporting the functionality of the Europeana Portal. This application profile was named Europeana Semantic Elements (ESE) and has been widely implemented as an exchange format by institutions that participate in the Europeana Community.

ESE will be succeeded by the EDM, supporting the Danube release of the Europeana Portal, but the application profile will be integrated in the EDM and data supply using the ESE format will be further sustained.

Reference Documents

- [Europeana Semantic Elements v3.3.1](#)
- [ESE v3.3 Schema](#)

2.5 HOPE XML Schema

Authors:	Joris Janssens – Amsab-ISG Bert Lemmens – Amsab-ISG Alessia Bardi – CNR-ISTI Paolo Manghi – CNR-ISTI
Internal Reviewers:	Armin Straube – FES Afelonne Doek – KNAW-IISG Alessia Bardi – CNR-ISTI

This chapter describes the rationale for the specification of the HOPE XML Schema, i.e. the formal specification of the HOPE Data Model using the XML Schema Language.

2.5.1 Introduction

The HOPE Data Model has been implemented by means of XML technology with an XML Schema, called the HOPE Schema. The HOPE Schema is an XML file specified using the XML Schema Language which describes the structure, content and semantics of XML documents. The HOPE Schema describes HOPE metadata records, that is XML documents containing information about materials belonging to a CP's collection.

The HOPE Aggregator stores XML documents in a database and indexes their content to provide search and browsing functionalities.

The main reasons for choosing this solution over other technology (e.g a relational database) are the following:

- Metadata records are harvested from the Content Providers and exported to the Discovery Services in XML format. In order to avoid a double transformation from XML to, for example, records in a relational database, and back to XML again for dissemination purpose, HOPE metadata records are stored as XML Documents.
- The design of a relational database implementing the HOPE data model would have to include many relationships between the different entities of the model. In such cases the database schema may be hard to implement and search performances may be affected in a negative manner when the number of stored records increase.

For more technical details about the database in use with the Aggregator and the rationale for choosing this database solution please refer to the Aggregator Technical Support Wiki³².

³²The release of the Technical Support Wiki is planned for M20 (December 2011)

The HOPE Schema³³ implements the HOPE Data Model by representing HOPE entities and their properties as XML elements and attributes. It is supposed to be as self-explanatory as possible, however an understanding of the general structure of the data model is necessary.

For a printed version of the HOPE XML Schema, see Appendix C: HOPE XML Schema. For the detailed specifications of the included metadata elements, see chapter “2.3 HOPE Data Model”. Section 2.5.3 of this chapter contains a description of the most interesting elements and implementation details of the HOPE XML Schema.

2.5.2 XML Schema

This paragraph describes the rationale used designing the HOPE XML Schema. The Schema is compliant with the W3C XML Schema Language. The reader should be familiar with some concepts of the XML schema to fully understand this documentation. Documentation about XML schema's can be found on the website of the W3C³⁴.

The HOPE XML schema is the implementation of the data model, in the paragraphs below we have elicited the detailed specifications, changes or additions which were needed for the transformation from the data model to the schema.

2.5.2.1 Enumerations

Enumerations are exact listings of all values that are allowed in a particular metadata element. Although many elements in the Data Model can only contain constant or controlled values (which are listed in the data model), these enumerations are not made explicit for all elements in the XML Schema. Adding all enumerations would cause the Schema to become too heavy, and it would lose its clarity. Moreover the value of these elements is defined in the (static) mappings from the different profiles to the HOPE data model, and is not susceptible for the input of incorrect values by the content providers.

For these elements and attributes the enumerations are not present in the schema, but enforced by the mapping:

- encoding attribute
- cataloguing attribute
- label attribute
- etc.

³³ [HOPE SCHEMA v1.3, ADD LINK](#)

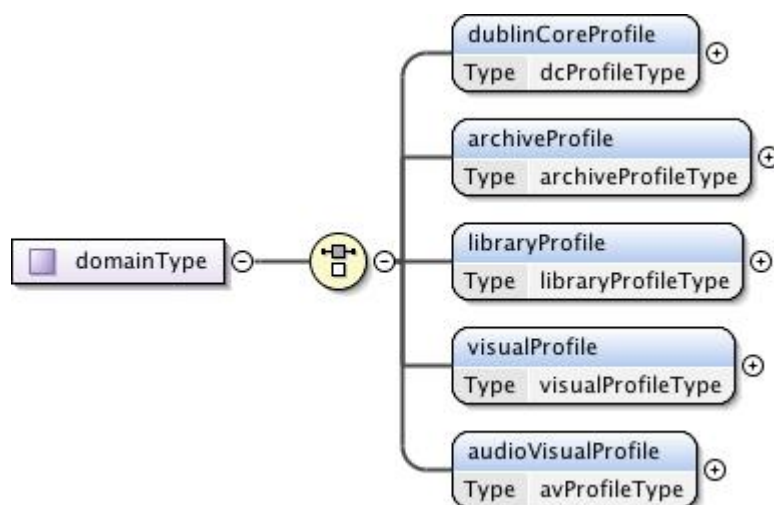
³⁴ <http://www.w3.org/TR/xmlschema-0/>

However, for some elements enumerations have been specified. These elements are listed below:

- description level/@normalised
- derivative/@derivativeType
- europeanaType
- access
- agent/type
- place/type

2.5.2.2 Domain

The element **domain** (/hopeEntity/descriptiveUnit/domain) is not a property of the descriptive unit in the data model. This element marks the metadata profile of the items described by this descriptive unit, in the data model this is the link from the descriptive unit to the different profiles. The element is necessary in the schema for defining profile specific elements. It is defined by the complexType **domainType**.



2.5.2.3 Bi-/one-directional relationships

In the HOPE Data Model, relationships between the different entities are defined in a bidirectional manner. In the HOPE Schema, most of the relationships are only defined in one direction (some exceptions apply, see below). For example the relationship **Associated agent** (from entity descriptive unit to the entity agent) in the descriptive unit entity is present in both the data model and the schema, while the inverse relationship **inverse associated agent** (from the entity agent to the entity descriptive unit) is only defined in the data model. Not only is it not always necessary to record these relationships in two

directions to comply with the functional requirements of the HOPE system, but also adding all inverse relationships to an XML record (e.g. all the records supplied by a specific content provider to the XML file about this record) would cause the XML files to become too heavy to handle.

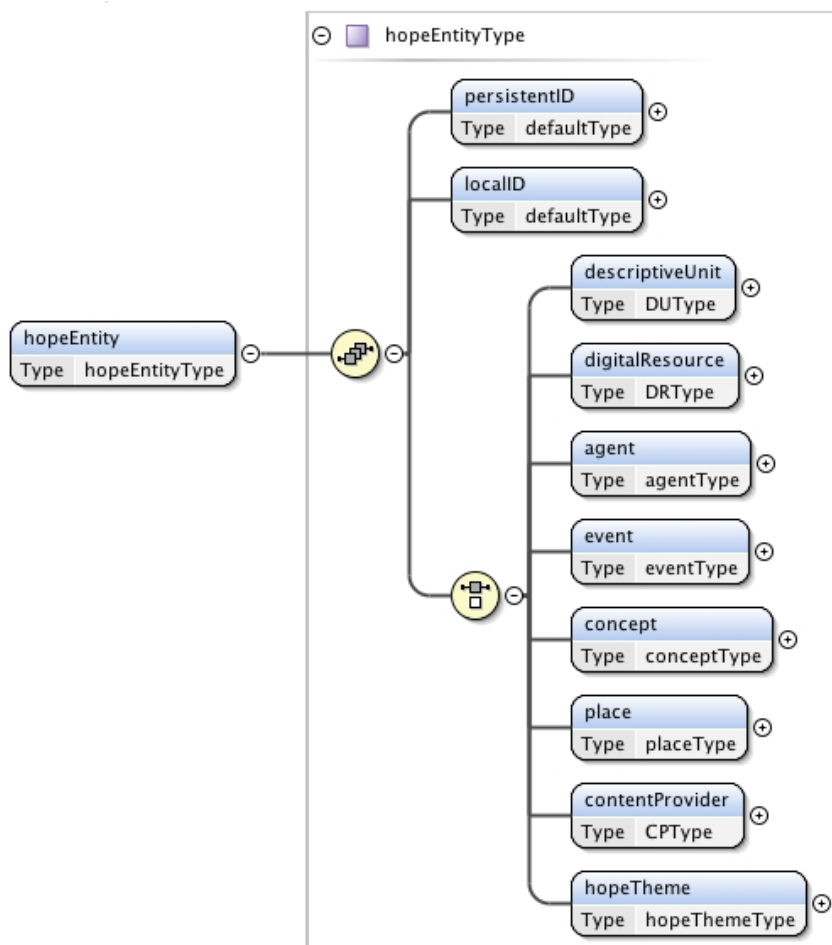
There is one exception, the element **represents** in the entity digitalResource and the element **isRepresentedBy** in the entity descriptiveUnit.

2.5.3 Structure of the XML Schema

This paragraph describes the structure of the HOPE XML Schema.

2.5.3.1 Entity level

The root element (hopeEntity) contains two elements which hold the identifiers (local and persistent IDs) and an element which defines the entity which is described in the XML file, this can be: descriptiveUnit, digitalResource, agent, event, concept, place, contentProvider or hopeTheme.



Every entity corresponds to one XML file, these XML files have only two elements in common. Every entity must have a persistent identifier (/hopeEntity/persistentID) to uniquely identify it in the HOPE information space. Another element all the entities share is the local identifier (/hopeEntity/localID). When this local identifier does not exist (this is the case for the entity contentProvider, hopeTheme and can be the case for agent, event, place and concept) the aggregator will create this value.

2.5.3.2 Complex types

Most of the elements do not simply contain a value but can have multiple attributes and subelements, these attributes are defined in XML Schema Complex Types³⁵.

Default Type

The most common type is **defaultType**. Elements defined by this type can have the following attributes of which only the label attribute is required for all elements.

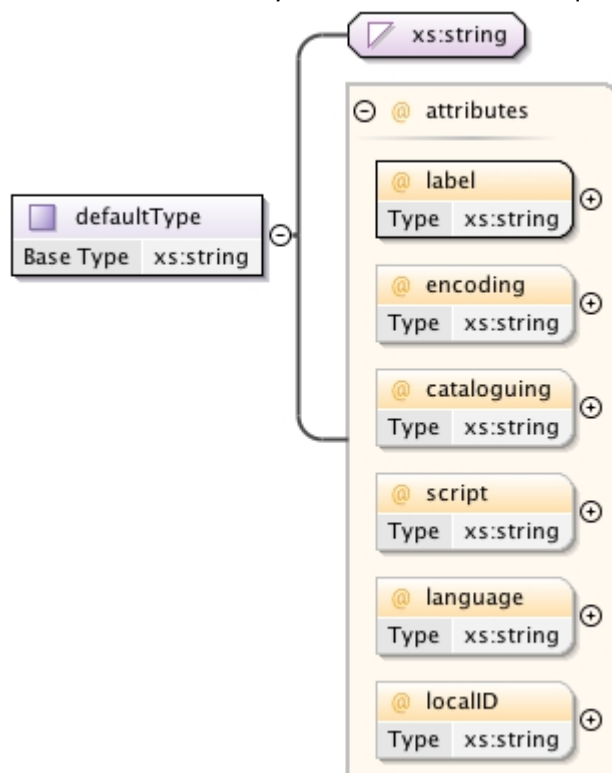


Table of attributes which can be used in defaultType

³⁵XML Schema, <http://www.w3.org/TR/xmlschema-0/#DefnDecls>

Name	Use	Definition
cataloguing	optional	Contains a reference to a metadata data standard that represents the cataloguing rules for creating the recorded metadata. This attribute contains the name of the cataloguing rules and the identification number of the element. The used cataloguing rules include ISAD(g) and ISBD 2007 Consolidated.
encoding	optional	Contains a reference to a metadata standard that is the source from which the HOPE element has been derived. This attribute contains a reference to an element name, including the namespace. The metadata standards include EAD, MARC21 Bibliographic, LIDO, EN15907, DC Qualified, EDM v5.2.1, PREMIS. Elements that are specific for HOPE refer to the HOPE namespace.
label	required	Contains a domain-specific label for displaying the value in the IALHI Portal.
language	optional	Contains a reference to the language used for a translated value.
localID	optional	Contains the ID, used for identifying the term in the local system of the CP.
script	optional	Contains a reference to the script for a transliterated value.

The majority of the elements used in the schema have almost the same structure. We use the complexType **defaultType** for these elements. When these attributes do not suffice one of the other complexTypes listed below are used.

PID Type

Those elements which express a relationship in the data model are of the type **PIDtype** in the schema. Since every entity needs to have a PID (delivered by the CP or generated by the HOPE Aggregator through the HOPE PID Service), these elements contain a required attribute **persistentID** which contains the PID of the related entity.

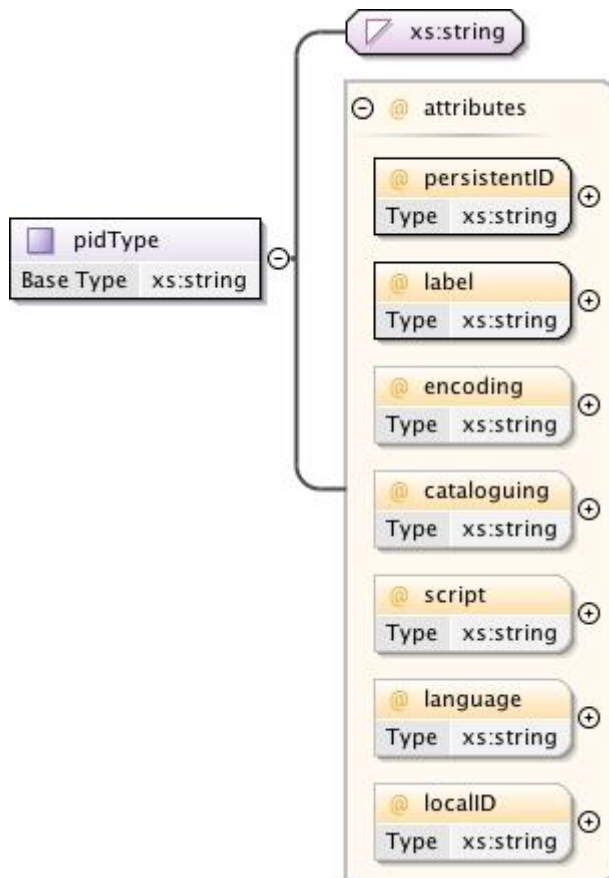


Table of attribute(s) which can be used in pidType (only those are listed which are not present in defaultType)

Name	Use	Definition
PersistentID	required	Contains a resolvable identifier used for identifying in and outside the HOPE system.

pidType is used by elements:

- /hopeEntity/digitalResource/contentProvider,
- /hopeEntity/digitalResource/represents,
- /hopeEntity/descriptiveUnit/associatedAgent,

- /hopeEntity/descriptiveUnit/associatedConcept,
- /hopeEntity/descriptiveUnit/associatedEvent,
- /hopeEntity/descriptiveUnit/associatedHopeTheme,
- /hopeEntity/descriptiveUnit/associatedPlace,
- /hopeEntity/descriptiveUnit/depictedAgent,
- /hopeEntity/descriptiveUnit/depictedConcept,
- /hopeEntity/descriptiveUnit/depictedEvent,
- /hopeEntity/descriptiveUnit/depictedPlace,
- /hopeEntity/descriptiveUnit/hasContributionsBy,
- /hopeEntity/descriptiveUnit/isCreatedBy,
- /hopeEntity/descriptiveUnit/isCreatedIn,
- /hopeEntity/descriptiveUnit/isPublishedBy,
- /hopeEntity/descriptiveUnit/isPublishedIn,
- /hopeEntity/descriptiveUnit/isSuppliedBy,
- /hopeEntity/event/spatialCoverage,
- /hopeEntity/hopeTheme/broaderTerm

Normalised Type

Some elements are of the type **normalisedType**, they have the same attributes as **defaultType** and an extra attribute **normalised**. The normalised value is put in this field, the original value as the value in the element. Most of these values must comply with a certain ISO standard (see: data model)

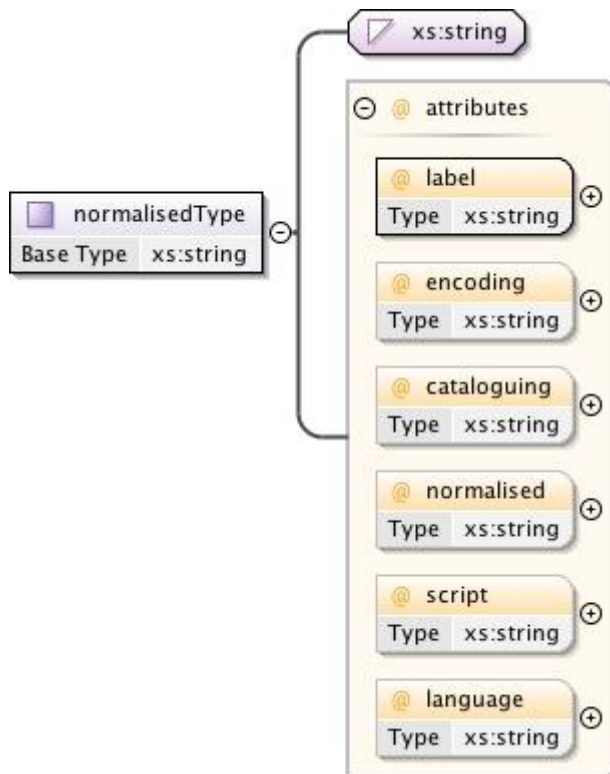


Table of attribute(s) which can be used in normalisedType (only those are listed which are not present in defaultType)

Name	Use	Annotation
Normalised	optional	Contains a normalised version of the term recorded in the value space.

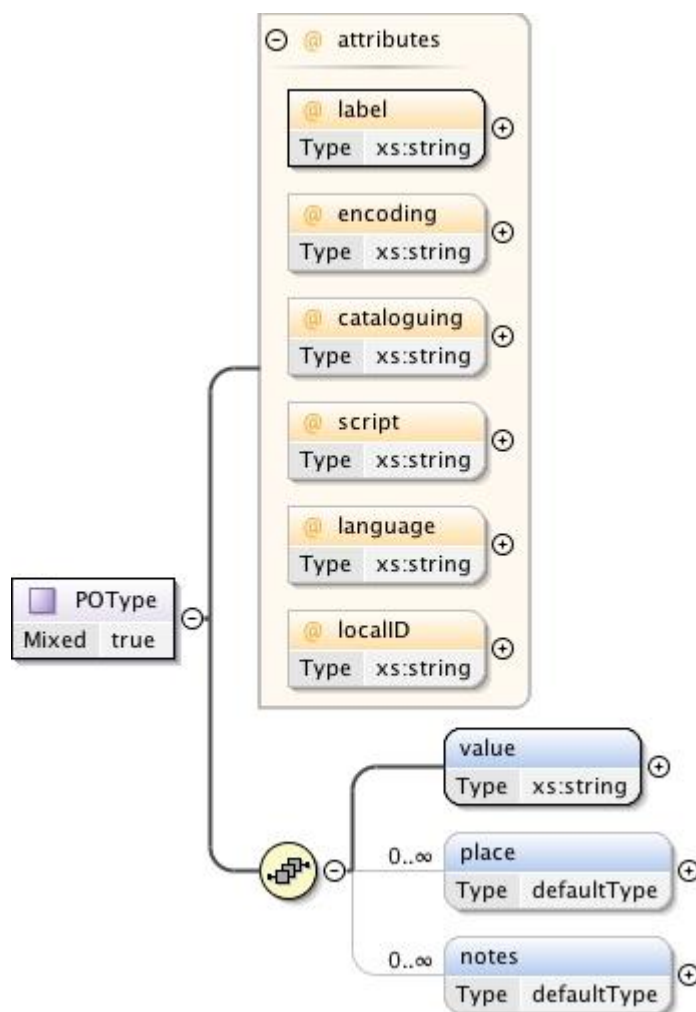
normalisedType is used by elements:

- /hopeEntity/contentProvider/country,
- /hopeEntity/contentProvider/language,
- /hopeEntity/digitalResource/copyright,
- /hopeEntity/digitalResource/language,
- /hopeEntity/descriptiveUnit/isRepresentedBy/language,
- /hopeEntity/descriptiveUnit/date,
- /hopeEntity/descriptiveUnit/language,
- /hopeEntity/descriptiveUnit/metadataLanguage,

- /hopeEntity/agent/date

PO Type

Elements which relate in some way to an agent are of the type **POType** (Persons and Organisation Type). It holds the same attributes as the defaultType, in addition this type defines three subelements. One for the value of the element and a notes and place subelement, which gives extra information.

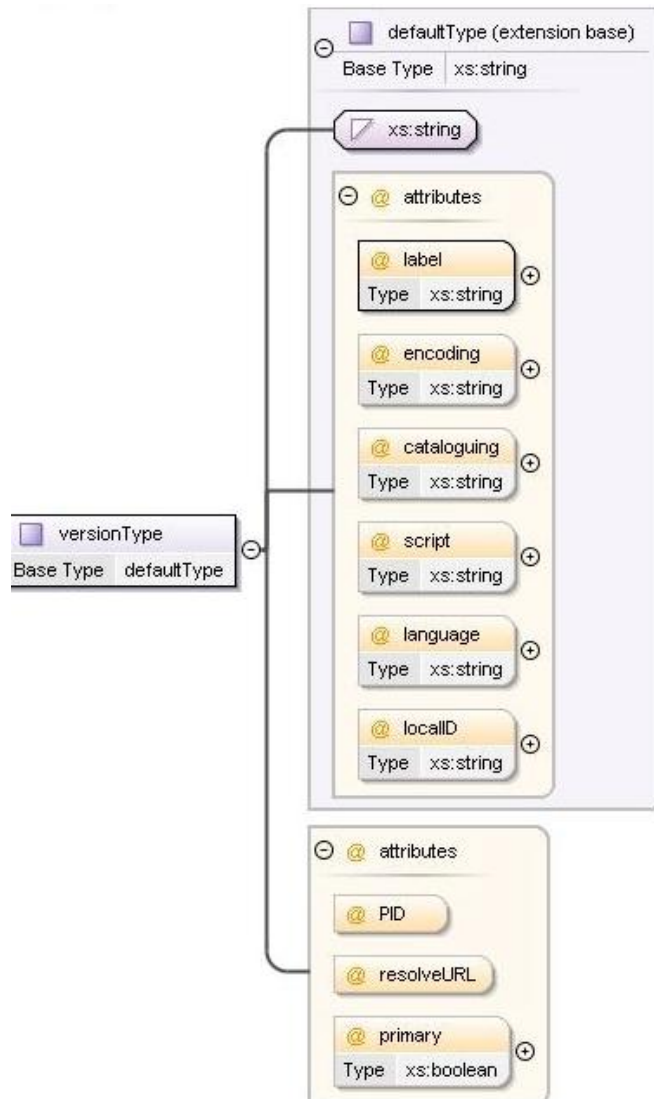


poType is used by elements:

- /hopeEntity/descriptiveUnit/contributor,
- /hopeEntity/descriptiveUnit/creator,
- /hopeEntity/descriptiveUnit/publisher

Version Type

The complex type `versionType` is used by these elements in the schema which record data on the different quality derivatives and the textual version of the master file described by the digital Resource Entity in which they reside. This type is an extension of the `defaultType` and thus shares its attributes. In addition `versionType` defines three extra attributes. The attribute `resolveURL` holds a URL that provides direct access to a derivative file stored in a local or shared object repository. The attribute `PID` holds the persistent identifier of the derivative file. The primary attribute indicates whether the derivative file is the preferred to be displayed with the corresponding descriptive unit.



versionType is used by elements:

- /hopeEntity/digitalResource/derivative2
- /hopeEntity/digitalResource/derivative3
- /hopeEntity/digitalResource/transcription

Those are the most commonly used types used in the Schema, however the Schema uses many others. Please refer to the schema **(LINK)**, and to the full documentation **(LINK: documentation³⁶)** for an exhaustive listing.

³⁶HOPE schema v. 1.0 documentation
<http://hopewiki.amsab.be/uploads/hopeSchema/hopeSchema1.0.html> (final location to be changed)

3. Harmonisation Requirements

This chapter contains a more detailed discussion on how the Common HOPE Metadata Structure deals with following technical issues:

- recording local and persistent identifiers
- recording hierarchical descriptions and compound objects (granularity)
- recording descriptive metadata about digital content and descriptive units

This chapter will also list the related requirements for making Local Metadata HOPE compliant.

3.1 Local and Persistent Identifiers

Authors:	Eric Beving – UPIP Bert Lemmens – Amsab-ISG
Internal reviewers:	Titia van der Werf– KNAW-IISG Alessia Bardi – CNR-ISTI

This chapter contains a detailed discussion about how the Common HOPE Metadata Structure deals with the creation, supply, storage and update of local and persistent identifiers.

3.1.1 Introduction

3.1.1.1 Objective

The chapter 3.1 is the output of T2.3 – sub-task 7 on the requirements for persistent resolvable identifiers. This sub-task specified scenarios and requirements for the creation, supply and update of “local identifiers” and “persistent resolvable identifiers (PIDs)” in the HOPE System, with the final objective of providing the common HOPE metadata structure with PIDs for each of the HOPE entities.

This chapter consists of 3 main sections:

- Section 2 gives general requirements about persistent unique identifiers and related URLs.
- Section 3 explains the purpose and practical use of PID services and resolver services in the HOPE System.

- Section 4 contains the requirements that CPs have to comply with when supplying identifiers and related URLs to the HOPE System.

3.1.1.2 Survey Results

Before starting with sub-task 7, the HOPE Content Providers Survey has sounded the current experience of the HOPE content providers with the use of resolvable identifiers and the uniqueness and persistence of the identifiers for the metadata records and digital content they will supply to HOPE. The survey came up with following conclusions:

- Only about half of the HOPE digital objects currently have a unique URL for each of the digital files they contain.
- 5 CPs, representing more than two thirds of the HOPE digital objects, are using PID technology. 2 CPs will use Handles for their HOPE digital content, which represents more than one third of the digital objects.
However, one CP currently only uses Handle for its non-HOPE content and the other one so far has only installed the Handle resolver.
- A third of the content providers who do not use PID technology today are intending its use in the future, another third is discussing it.
- Although currently only 1 CP has (fully) operational PID technology for its HOPE content, there is know-how on/experience with PID technology with 4 other CPs.

3.1.2. General requirements

This section deals with the general requirements for persistent unique identifiers and related URLs as far as the general objective of the HOPE project is concerned, i.e. ensuring a trusted identification of HOPE resources in- and outside the HOPE System, while allowing for flexible management of the related metadata or digital content.

Important: The requirements listed in this section are very general and do not make any difference between “local identifiers” (not managed through a PID service and thus not resolvable) and “persistent identifiers” managed through a PID service.

3.1.2.1 General requirement 01: identifiers and URLs

Data entities managed in the HOPE System **MUST** have a unique name, i.e. an identifier. Entities include for example:

- Descriptions of Collection Items, supplied by content providers
- Digital files supplied by HOPE compliant repositories
- Vocabulary terms, such as agents, places, concepts and events

A number of Data entities require URLs in order to allow Internet end users to display,

download, print or get information about the entity:

- Digital files:
 - URLs must allow end users to view and possibly download or print each digital derivative file. For example:
 - A URL to download a high-resolution derivative of a digital representation.
 - A URL to view a low-resolution derivative of a digital representation.
- Descriptions of Collection Items:
 - A URL to view a description of a collection and its digital representation on the Content Providers' website.

3.1.2.2. General requirement 02 : persistence of identifiers and URLs

In aggregator services like HOPE, the persistent identification of entities is one of the cornerstones to the quality of the overall service. The long time-persistence of identifiers and URLs is fundamental:

- Content providers supply metadata records and digital files to HOPE. These records and files are in turn supplied to other aggregators (e.g. Europeana) and websites (e.g. social websites). Data supply between all these systems can occur regularly to manage creations, updates and deletions. Therefore, stable identifiers allowing aggregators and websites to determine if a metadata record or a digital file must be created or just updated is very important.
- Internet end users looking for documents jump from one website to another website in order to launch a search, view a digital file, get contextual information about a collection, order a high-resolution copy. Therefore, URL links used for these jumps must remain stable and never be "broken". Moreover, once descriptive metadata is disseminated in several aggregators or websites, it is very difficult to ensure the propagation of URL updates in all these target systems.

Identifiers and URLs must be independent of any cataloguing or repository software in order to remain stable in case this software changes. Many systems use some kind of incremental key to uniquely identify a record in their database. Often, this key cannot be exported or externally set and its value changes during the migration to a new system. Thus, identifiers must be managed in a "user" specific data field that can be migrated and must be recognized by the new system. Content providers must particularly take care about this issue.

3.1.2.3. General requirement 03: uniqueness of identifiers

Each identifier used in the HOPE System must be globally unique and not need additional information to allow the retrieval of the corresponding entity. In particular, it means that the entity type (digital file, agent, concept, etc.) is not required to retrieve an entity with its identifier (which could have been a possible option). Moreover, identifiers must never be reused. When an entity is deleted, its identifier becomes useless. But even more important is that this identifier must not be used for a new entity to avoid any confusion between the old and the new entity.

However there are some exceptions to this requirement (not applicable to PIDs, identifiers managed through a PID service; see section 3):

- Since none of current HOPE content providers can ensure the global uniqueness of its identifiers (cf. Content Providers survey), HOPE allows an exception to this rule keeping the HOPE entry barrier for content providers as low as possible. Identifiers supplied by a content provider must be unique only within the local information system of the content provider. The Aggregator and the SOR will create globally unique identifiers for the CP's corresponding entities when ingesting them.
- Two digital derivative files may have the same identifier, provided that this identifier is always supplied with the derivative level (1, 2, 3 or OCRed text).

3.1.2.4. General requirement 04: identifiers (except PIDs) should not be URLs

When possible, identifiers must not be URLs and therefore, when an entity requires a URL, this URL must be stored independently from the identifier:

- URLs are more prone to persistence problems since they include a domain name and possibly folders, parameters and so on. This can cause problems when digital files are moved and change location for storage maintenance purposes. This is especially true for URLs provided by content providers that do not use a PID service.
- A URL is a network address that should allow retrieving "something". For many entities in the HOPE System, there is nothing to retrieve.

Nevertheless, there are some exceptions to this requirement:

- Some content providers are only able to supply URLs (without non-URL identifiers) for their descriptive metadata and/or their digital files. Again, in order to make the HOPE entry barrier as low as possible for content providers, HOPE allows these URL identifiers.
- PIDs (identifiers managed through a PID service; see section 3) can be URLs since the domain part of these PID URLs is very persistent; in particular, PIDs generated by the HOPE System will be URLs. However a PID URL must never lead to a "HTTP 404 - File not found" error. If necessary (no resolve URL), the PID service must display a default web page.

3.1.3. PID services and resolver services

3.1.3.1. What is a PID service?

A persistent identifier (PID) service has two main goals:

- Provide a name scheme for persistent unique identifiers that, in particular, makes these identifiers globally unique,

and/or

- Provide a resolver service to improve the persistence of URLs.

There are several established PID systems: Handle, Handle/DOI, ARK, PURL, N2T, URN... All systems do not implement both goals.

Name scheme for persistent identifiers

A common name scheme for persistent identifiers is based on a three parts structure:

- An “identifier type” (i.e. a URI scheme)
- A unique identifier attributed to each institution using the PID system (institution ID)
- An identifier generated by the institution for each entity

Examples:

- ARK (Archival Resource Key):
 - Example of persistent identifier: ark:/13030/tf5p30086k
 - 13030 is the institution ID which is called NAAN (Name Assigning Authority Number)
 - NB: “ark:” is not registered with IANA as a URI scheme or as a URN namespace.
- Handle:
 - Example of a persistent identifier: hdl:1066/1112
 - 1066 is the institution ID which is called a prefix or naming authority number
 - To get a Handle prefix, an institution needs to pay an administrative fee (ca. 50 dollars annually).
 - NB: “hdl:” is not registered with IANA as a URI scheme or as a URN namespace.

Resolver service

The use of a PID resolver service, implemented as a layer between the URLs embedded in an entity and the actual location of the entity on the web, improves the persistence of URLs and avoids broken links.

The URL related to the unique identifier and embedded in the entity (entity-URL) is not the actual location of the entity on the web (actual-URL) but redirects Internet end users towards a resolver server. The persistent unique identifier (PID) of the entity is part of the entity-URL. The resolver server has a database where PIDs are stored together with actual-URLs. The resolver service extracts the unique identifier from the entity-URL, searches its database for the actual-URL, and redirects the end user to this actual-URL called the resolve URL.

This mechanism, by providing a level of indirection, disconnects the actual web location of an entity (actual-URL = resolve URL) from the URL stored in the entity (entity-URL). The entity-URL is more stable than the actual-URL if the resolver service used is one of the established resolver services since it means the domain name used in the entity-URL is also persistent.

Example of entity-URLs:

- Handle: <http://hdl.handle.net/1066/1112>
- ARK: <http://anyresolver.com/ark:/13030/tf5p30086k>

If the actual location (actual-URL) of an entity changes, there is no need to update the entity-URL spread in multiple aggregators, discovery services or social sites. There is only one update in one location: the resolver database.

The relationship between the entity-URL and the actual-URL is called the binding.

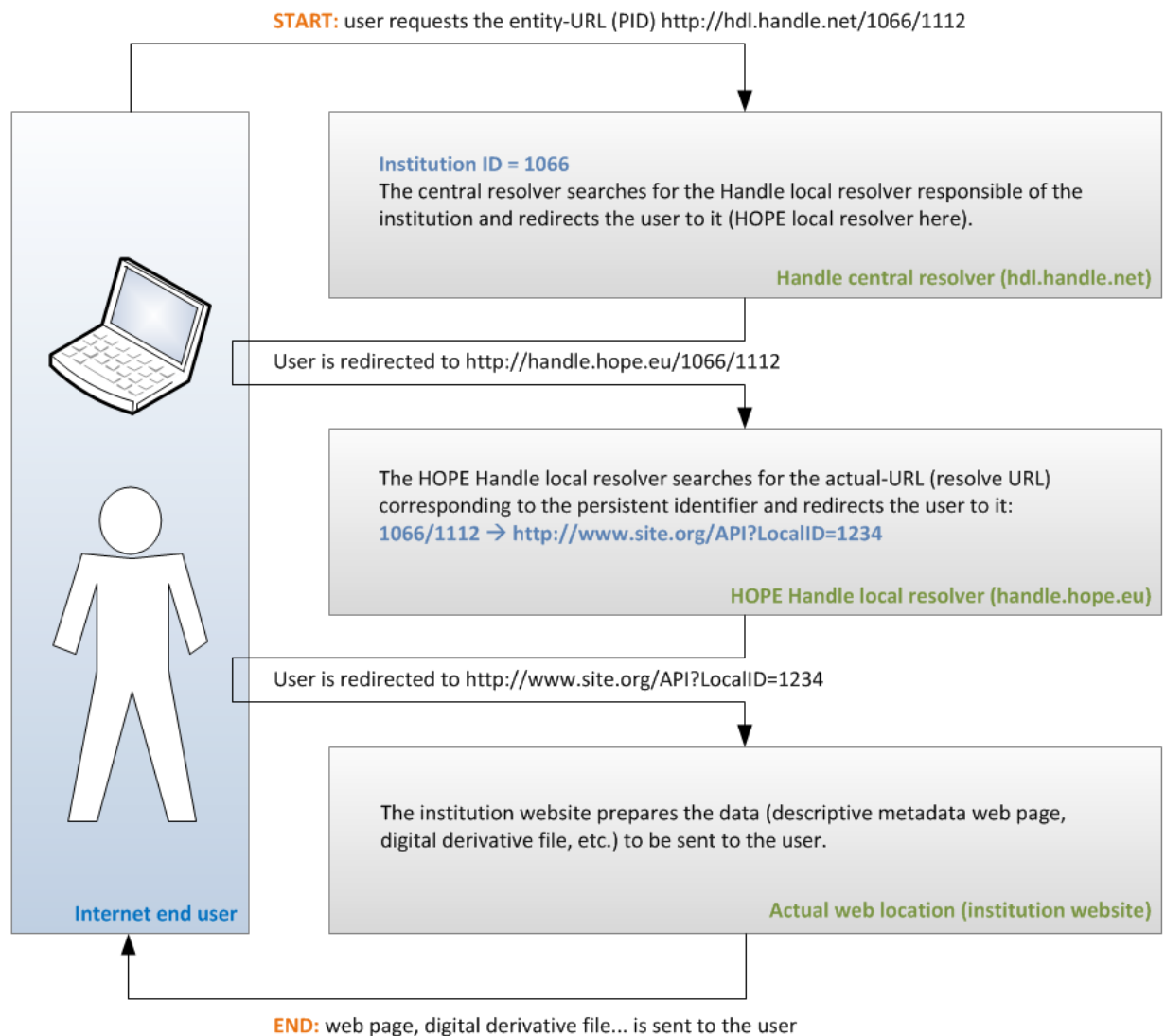
Example of the Handle system

Handle is taken as an example here since it implements both goals of PID systems (name scheme and resolver service) and HOPE is currently implementing it for its PID Service.

The diagram below represents the resolution process of the Handle system.

The Handle system has a central resolver (<http://hdl.handle.net>) that manages the institution ID (1066 in this example).

When an Internet end user clicks on the entity-URL (<http://hdl.handle.net/1066/1112>), his request is directed to the central resolver. This resolver extracts the institution ID and redirects the request to the local resolver that manages the institution. HOPE will have its own local resolver (<http://handle.hope.eu>). The local resolver extracts the persistent identifier (1066/1112), searches its database for the corresponding actual-URL (resolve URL) and redirects the request to this actual URL.



3.1.3.2 Requirements for the usage of PIDs in the HOPE System -

Glossary

As described above, a PID service makes identifiers globally unique (name scheme) and highly improves the persistence of identifiers if it includes a resolver service.

Using a resolver-based approach to deal with identification of HOPE entities is important to streamline the aggregation workflow, ease the identification of duplicates, reduce broken links with external partners, and to become a trusted knowledge hub. It is also a good preparation for upcoming usage in the Linked Open Data cloud (<http://linkeddata.org>), and increased usage of Semantic Web technologies.

Therefore, the HOPE System should manage all the persistent unique identifiers and URLs of its entities through a PID Service that includes a resolver service for the following entities:

- Aggregator entities
- SOR entities
- Entities supplied by Content Providers to HOPE
 - Metadata records
 - Digital files supplied to the SOR or digital files stored on LORs.

However, for content providers using proprietary cataloguing software or without technical skill, the implementation of a PID Service can be burdensome and expensive. There are two main actions a CP needs to perform when it chooses to use a PID service:

- First, the institution may have to modify its existing unique identifiers:
 - Create a new field in its database(s)
 - Generate the 'new' persistent identifiers in it or adapt existing identifiers if they can be made compliant with the PID service name scheme and characters authorised on URLs (encoding of characters on URLs (e.g. %20 for space) should be avoided because it can create problems)...
- Secondly, the institution has to manage the PID resolver database: when the institution creates a new entity, it has to create the corresponding entry in the resolver database. When the resolve URL of the entity is modified, the database entry has to be updated. When the entity is archived (no more available on-line) or deleted, the entry must also be updated/deleted.

A bad implementation of a PID system can be worse than no implementation. For example, if the PID resolver database update is ran asynchronously by batch (i.e. not managed by the cataloguing software) and the batch fails, some entities can be created in the local cataloguing system but without the corresponding entries in the resolver database. This creates broken links for Internet users.

PID usage requirement 01: the Aggregator and the SOR MUST manage the identifiers and URLs of their entities through a HOPE PID Service

The Aggregator and the SOR must manage the persistent unique identifiers and URLs of their entities through a PID service that includes a resolver service. The section 3.3 gives the requirements for this "HOPE PID Service".

PID usage requirement 02: contents providers are strongly recommended to implement PIDs.

In order to keep the HOPE entry barrier for content providers as low as possible, HOPE has decided that content providers are free to use or not a PID service to manage the identifiers and URLs of the entities they supply to HOPE.

If a CP decides to use a PID service, it can choose any well-known PID service that includes a resolver service. The implementation can be directly made by the CP ("local PID service") or by a third party (regional, national, international or commercial service; "third-party PID service"). CPs can also decide to use the "HOPE PID Service". The HOPE BPN Wiki gives examples of PID Services a CP can use or implement.

PID usage requirement 03: identifiers managed through a PID service MUST be URLs

Since URLs managed through a PID Service (e.g. <http://hdl.handle.net/1066/1112>) are as persistent as non-URL identifiers ((e.g. hdl:1066/1112), HOPE has decided that identifiers managed through a PID service must be URLs.

This requirement applies to identifiers of Aggregator and SOR entities but also of entities supplied by CPs to HOPE.

PID usage requirement 04: the Aggregator and the SOR MUST provide PIDs for CP entities not managed through a PID service

The Aggregator and the SOR must provide PIDs for entities supplied by CPs that do not use a PID service. This is important for the HOPE internal management of entities and, of course, the dissemination or export of these entities on the Internet.

PID usage requirement 05: the HOPE System MUST preserve PIDs supplied by CPs

The HOPE System must preserve PIDs provided by content providers. In particular, data disseminated or exported on the Internet must retain CP original PIDs without any HOPE replacement.

3.1.3.3 Requirements for the HOPE PID Service

HOPE PID Service requirement 01: the HOPE System MUST host a PID service.

The HOPE System must host a PID service which is called the “HOPE PID Service”. It must include a resolver service and must be implemented on the basis of an operational, global and distributed (as opposed to central) PID service.

HOPE has decided to implement its own instance of the Handle System.

HOPE PID Service requirement 02: the HOPE PID Service MUST provide a reverse look up functionality.

For each PID supplied for a CP entity not managed through a PID service, the HOPE PID Service must record :

- the corresponding local identifier
- the corresponding local URL (resolve URL) if it exists.

The HOPE PID Service must provide a “reverse look up” functionality that will allow the Aggregator and the SOR to retrieve the PID corresponding to a local identifier.

Thus, two PIDs can't be bound to the same local identifier.

3.1.4. Scenarios and requirements for creation, supply and update of identifiers by content providers

3.1.4.1 Structure of scenarios

Scenarios only takes into account entities supplied by CPs to the Aggregator and the SOR. They are organised according to three variables:

- Type of entity:
 - Metadata records (i.e. descriptive units) and landing pages
 - Digital representations and digital files
 - Authority records (agent, place, concept, event)
- Type of digital object repository:
 - HOPE Shared Object Repository (SOR)
 - Content provider local object repository (LOR)
- Use or not of a local PID service by the content provider:
 - Direct use of a local PID service by the CP. The term “local PID service” includes the use of a third-party PID service or the direct use of the HOPE PID Service. HOPE must not replace the PIDs supplied by the CP.
 - No use of a local PID service. HOPE must create PIDs for CP entities.

3.1.4.2 Requirements for metadata records (i.e. descriptive units) and landing pages

Each metadata record supplied by a content provider is mapped into one or more (e.g. in the case of hierarchical archival records) Descriptive Unit Entities in the HOPE data model. A descriptive unit is a set of metadata elements describing zero, one or several collection items. In the HOPE information space, a descriptive unit can describe:

- A single cultural heritage object, such as an archival document, a publication, a photograph, a movie.
- A series of cultural heritage objects, such as an archival fonds, series or file, series of monographs, a periodical, collection of photographs, a series of TV broadcasts.
- A collection of cultural heritage objects, for which the metadata is submitted as one or more data sets to the aggregator.

Each descriptive unit may contain as a property a Landing Page element, which is a web page provided by the Content Provider. This web page displays the descriptive unit metadata as well as embedded digital files(s). Landing pages are optional but their supply is recommended to Content Providers.

In the Europeana data model:

- PIDs for descriptive units are accommodated by the ore:proxy element.
- Landing pages are accommodated by the ens:landingPage property or its sub-property ese:isShownAt.

The workflows for creation, supply and update of PIDs for descriptive units depends whether the CP uses or not a local PID service (2 scenarios).

3.1.4.2.1 CP does not use a local PID service

Procedure:

- The CP records a local identifier and an optional landing page in each metadata record it supplies to the Aggregator. For hierarchical descriptions such as EAD, the term “metadata record” refers also to each nested descriptive unit (<c> element) and not only to the whole file.
- The Aggregator stores the local identifier and the landing page permanently.
- If the Aggregator does not find an existing PID bound to the local identifier (via the HOPE PID Service reverse look up functionality), it creates a PID for the descriptive unit. It binds the PID to the local identifier and the landing page (as a resolve URL).
- The Aggregator provides the CP with the PID of the descriptive unit for an optional storage in the CP information system.

Binding updates:

- When changing the local identifier, the CP must
 - First, update the binding recorded by the HOPE PID Service, using the HOPE PID Service administrative interface.
 - Secondly, supply the updated metadata record to the Aggregator.
- When changing the landing page URL, if this URL is not the local identifier, the CP must
 - Supply the updated metadata record to the Aggregator. The Aggregator will update the resolve URL binding recorded by the HOPE PID Service.

Requirements:

- CPs:
 - Must provide a local identifier for each metadata record (i.e. descriptive unit) it supplies to the Aggregator.
 - Should supply a landing page for each descriptive unit.
 - May record the PID created for the descriptive unit in its local information system.

3.1.4.2.2 CP uses a local PID service

Procedure:

- In each metadata record it supplies to the Aggregator, the CP inserts a PID identifying the record as well as a landing page URL that must be also a PID (the same as the metadata record identifier or another one). The PID management workflow in the CP information system is beyond the scope of this document. For

hierarchical descriptions such as EAD, the term “metadata record” refers also to each nested descriptive unit (<c> element) and not only to the whole file.

- The Aggregator stores the PID and the landing page permanently.

Binding updates:

- The CP cannot change the PID identifying the metadata record. If really necessary, the CP must delete the record in the Aggregator and then ingest the updated one. However, it may create broken links.
- When changing the landing page URL, if this URL is not the PID identifying the record, the CP must supply the updated record to the Aggregator.

Requirements:

- CPs
 - Must provide a PID for each metadata record (i.e. descriptive unit) it supplies to the Aggregator.
 - Should supply a landing page (which is also a PID) for each descriptive unit.

3.1.4.3 Requirements for digital files

This section describes the workflow for the creation, supply, recording and update of persistent identifiers for digital derivatives files and/or digital masters files.

In the HOPE Data Model, a collection item may be represented by one or more digital representations, i.e. one or more single, unique renditions of a Collection Item, such as a digital image, an audio-visual recording or a sound recording. The HOPE Data Model can accommodate multiple versions for each unique rendition, including high-resolution derivatives, low-resolution derivatives, preview derivatives, thumbnails and OCRred text derivatives.

The Digital Resource entity contains information about a single digital representation, including persistent identifiers for all available derivatives.

The HOPE data model does not model the ‘digital object’, i.e. all Digital Resources associated with a single Descriptive unit, as a separate entity. Digital resources are directly linked to a descriptive unit, optionally with some “next in sequence” information. Therefore, strictly speaking, the HOPE data model does not allow describing several digital objects for one descriptive unit; it allows describing an ordered list of digital resources without any “object grouping” information. If necessary, CPs must refine a too “global” descriptive unit by creating new ones from the global one in order to have a consistent and coherent list of resources bound to each new descriptive unit.

In the HOPE System, a Digital Resource entity MUST always contain a low-resolution

derivative. Other derivative types are optional.

In the HOPE data model, each digital derivative file MUST have a PID and a corresponding resolve URL.

In the Europeana data model, PIDs for digital derivative files are accommodated by the `ens:hasView` property and its sub-property `ese:isShownBy`, enabling direct access to the digital file. Thumbnails are managed through specific properties.

The workflow for the creation, supply, recording and update of PIDs for digital derivative files depends primarily on the repository where the digital file is stored, i.e. a the HOPE Shared Object Repository (SOR) or a local object repository (LOR). It depends also whether the CP uses or not a local PID service (2x2 = 4 scenarios).

3.1.4.3.1 CP uses the SOR and no local PID service

This paragraph describes the workflow for CPs storing their digital files in the SOR and having PIDs for digital files created and updated by the SOR using the HOPE PID Service.

Creation:

- The CP first registers with the Handle System and receives a naming authority number that uniquely identifies the CP. The CP then supplies the authority number to the HOPE PID Service via its administrative interface.
- The CP records a local identifier for each digital master file in its local information system, in relation with the identifier (PID or local identifier) of the corresponding metadata record.
- When providing a digital master file to the SOR, the CP also supplies the local identifier for the digital master file, using the “SOR processing instruction”.
- The SOR creates a resolve URL and a PID for the digital master file and each of its digital derivative files. The master and its derivatives have the same PID root with an additional parameter defining the level of derivative. The PID without parameter is the one of the level 2 derivative, i.e. the “on-line consultation” resolution. Since a local identifier can be bound to only one PID, the SOR binds each derivative PID to the master local identifier suffixed with a string indicating the derivative level. Thus, the master and its derivatives have the same local identifier root.
- The SOR updates the SOR processing instruction with the PIDs of master files and stores it in its staging area where it is available for download. This updated SOR processing instruction enables the CP to incorporate the PIDs in its local information system.

Supply:

- The CP inserts the local identifiers or the PIDs of the derivative files created by the SOR in the metadata records it supplies to the Aggregator. Since the derivative files of the same master have the same identifier root, CPs may insert only this root for each master file; the mapping worksheet will allow the definition of the derivatives since suffixes to add to the root are constant for each derivative level.
- When the CP only provides a local identifier, the Aggregator searches for the corresponding PID using the reverse look up functionality of the HOPE PID Service.

Binding updates:

- When the local identifier of a digital master file changes, the CP
 - Must update the PID binding for this master and for all its derivative using the administrative interface of the HOPE PID Service.or
 - Must update the SOR processing instruction, and notify the SOR that it has to update the local identifiers recorded in the HOPE PID Service for the master and its derivatives.
- When the SOR changes the resolve URLs of the digital files, the SOR updates the bindings in the HOPE PID service.
- When the resolve URLs of digital files changes because the CP moves from the SOR to another object repository, the CP must update the resolve URLs using the administrative interface of the HOPE PID Service.

Requirements:

- **General requirements:**
 - Must record a local identifier that uniquely identifies each digital master file in the CP's local information system.
 - Must record the binding between the local identifier of the digital master file and the descriptive unit (metadata record) in its local information system.
 - Should not use local URLs as local identifiers.
- **SOR specific requirements** :
(additional requirements for SOR users)
 - Must supply the local identifier of each digital master file to the SOR, using the SOR processing instruction.
 - Must retain local identifiers when switching from one local system to another, in order to ensure the correct identification of digital files by the

Aggregator and the SOR.

- **HOPE PID Service specific requirements:**
(additional requirements for SOR users related to the HOPE PID Service)
 - Must register with the Handle System and supply the Handle prefix (the naming authority) that identifies the CP to the HOPE PID Service via its administrative interface.
 - May record in their local system the PIDs generated for each digital master file and supplied via the 'SOR processing instruction'; the PIDs for derivative files can be automatically built from the corresponding master file PID.
 - Must supply the local identifier or the PID generated by the SOR for each digital derivative file to the Aggregator.
 - When the local identifiers of digital master files change because the CP moves from one information system to another (which is not recommended), the CP
 - Must update the binding recorded in the HOPE PID Service for the masters and their derivatives, using the HOPE PID Service administrative interface
 - or
 - Must update the SOR processing instruction and notify the SOR it must update the binding recorded in the HOPE PID Service for the masters and their derivatives.
 - When the resolve URLs of the digital files change because the CP moves from the SOR to another object repository, the CP must update the binding recorded in the HOPE PID Service for the masters and their derivatives, using the HOPE PID Service administrative interface.

3.1.4.3.2 CP uses the SOR and a local PID service

This paragraph describes the workflow for CPs storing their digital files in the SOR and having PIDs for digital files created and updated by a local PID service. The internal management of the PID service by the CP is beyond the scope of this document.

Creation:

- The CP creates a (local) PID for each digital master file and each derivative the SOR will create, using its local PID service. A master and its derivatives may have the same PID root with an additional parameter defining the level of derivative. The PID without parameter must be the one of the level 2 derivative, i.e. the "on-line consultation" resolution.
- When providing a digital master file to the SOR, the CP also supplies the PID for the digital master file, using the "SOR processing instruction".

- The SOR creates a resolve URL for the digital master file and each digital derivative file.
- The SOR updates the SOR processing instruction with the resolve URLs and stores it in its staging area where it is available for download. This updated SOR processing instruction enables the CP to retrieve the resolve URLs and to bind them to the corresponding PIDs in its local PID service.

Supply:

- The CP inserts the PIDs of the different digital derivative files in the metadata records it supplies to the Aggregator.

Binding updates:

- The CP must not change the PID identifying a digital derivative file.
- When the resolve URL of a digital file changes, the SOR administrator will notify the CP and the CP will have to update the resolve URL recorded by its local PID service.
- When the resolve URLs of digital files change, because the CP moves from the SOR to another object repository, the CP must update the resolve URLs recorded by its local PID service.

Requirements:

- **General requirements:**
 - Must create a local PID that uniquely identifies each digital master file and each digital derivative file created by the SOR using its local PID service and store the PIDs in its local information system.
 - Must record the binding between the PIDs of the digital files and the descriptive units in its local information system.
- **SOR specific requirements:**
(additional requirements for SOR users)
 - Must supply the PID of each digital master file to the SOR, using the SOR processing instruction.
- **Local PID specific requirements:**
(additional requirements for SOR users related to the local PID service)
 - Must not change the PID identifying a digital derivative file.
 - Must bind its local PIDs with the resolve URL created by the SOR.

- Must supply the PID for the digital master file and for each digital derivative file to the Aggregator.
- When receiving a notification from the SOR administrator, the CP must update the binding between SOR resolve URLs and PIDs using its local PID service.
- When resolve URLs of digital files change because the CP moves from the SOR to another object repository, the CP must update the binding between resolve URLs and PIDs using its local PID service.

3.1.4.3.3 CP uses a LOR and no local PID service

This paragraph describes the workflow for CPs storing their digital files in a LOR without using a local PID service for the management of identifiers for their digital files. The Aggregator will create PIDs for the CPs digital files using the HOPE PID Service.

Creation:

- The CP first registers with the Handle System and receives a naming authority number that uniquely identifies the CP. The CP then supplies the authority number to the HOPE PID Service via its administrative interface.
- The CP records a local identifier for each digital representation it is going to supply to the Aggregator in relation with the identifier (PID or local identifier) of the corresponding metadata record.
- The CP creates a resolve URL for each of the digital derivative file it is going to supply to the Aggregator. Each resolve URL must allow retrieving the corresponding digital derivative file from the CPs' LOR.
- The CP inserts the binding between the resolve URL and the local identifier of each digital derivative file in the corresponding metadata record.

Supply:

- The CP supplies local identifier and resolve URLs to the Aggregator via its metadata records.
- The Aggregator creates a PID using the HOPE PID service for each digital derivative file and binds it to the corresponding resolve URL.
- The Aggregator makes the PIDs available for the CP; the CP may insert these PIDs in its local information system.

Binding updates:

- When the local identifier of a digital representation changes, the CP must update the binding with the corresponding PID using the HOPE PID Service administrative interface.

- When the resolve URL of a digital derivative file changes, if this resolve URL is not the local identifier, the CP must
 - Update the resolve URL recorded in the HOPE PID Service using the HOPE PID service administrative interfaceor
 - Provide the Aggregator with the new resolve URL (submission of the updated metadata record). The Aggregator will automatically update the binding recorded in the HOPE PID Service.

Requirements:

- **General requirements:**
 - Must record a local identifier that uniquely identifies each digital representation in its local information system.
 - Must record the binding between the local identifier, the digital file and the descriptive unit in its local information system.
 - Should not use local URLs as local identifiers.
- **LOR specific requirements:**
(additional requirements for LOR users)
 - Must create a resolve URL for each digital derivative file stored by the LOR.
 - Must record the binding between the resolve URL and the local identifier in its local information system.
 - When the resolve URL of the digital file changes, the CP must update the binding with the PID using the administrative interface of the HOPE PID Service.
- **HOPE PID Service specific requirements:**
(additional requirements for LOR users related to the HOPE PID Service)
 - Must register with the Handle System and supply the handle prefix (the naming authority) that identifies the CP to the HOPE PID Service using its administrative interface.
 - Must provide the Aggregator with the local identifier and the resolve URL of each digital file.
 - May record the PIDs, made available by the Aggregator, in its information system.

3.1.4.3.4 CP uses a LOR and a local PID service

CP stores digital files in a LOR. PIDs for digital files are created and maintained by a local PID service.

This scenario may apply to CPs who already have a complete content management and digital repository system that already uses PIDs internally. The internal management of the local PID service by the CP is beyond the scope of this document.

Creation:

- The CP creates a (local) PID for each digital file it is going to supply to the Aggregator in relation with the identifier (PID or local identifier) of the corresponding metadata record. Derivatives of the same master file may have the same PID root with an additional parameter defining the level of derivative. The PID without parameter must be the one of the level 2 derivative, i.e. the “on-line consultation” resolution.
- The CP creates a resolve URL for each of the digital derivative file it is going to supply to the Aggregator. Each resolve URL must allow retrieving the corresponding digital derivative file from the CP LOR.
- The CP records the binding between the resolve URL and the local identifier of each digital derivative file using its local PID service.

Supply:

- The CP inserts the PIDs of the different digital derivative files in the metadata records it supplies to the Aggregator.

Binding updates:

- The CP must not change the PID identifying a digital derivative file.
- When the resolve URL of a digital derivative file changes, the CP MUST updates the binding between the resolve URL and the PID using its local PID service.

Requirements:

- **General requirements:**
 - Must record a local PID that uniquely identifies the digital derivative file in the CPs local information system.
 - Must record the binding between the PID of each digital derivative file and the corresponding descriptive unit in its local information system.
- **LOR specific requirements:**
(additional requirements for LOR users)
 - Must create a resolve URL each digital derivative file supplied to HOPE.

- Must record the binding between a resolve URL and the corresponding PID in its local PID service.
- When the resolve URL of a digital derivative file changes, the CP must update the binding using its local PID service.
- **Local PID service specific requirements:**
(additional requirements for LOR users related to the local PID service)
 - Must not change the PID identifying a digital derivative file.
 - Must create a PID for each digital derivative file, using its local PID service and bind it with the corresponding resolve URL.
 - Must record the PIDs generated by the local PID service in its local information system.
 - Must provide the aggregator with PIDs of digital derivative file.

3.1.4.4 Requirements for CP authority records

Content providers may supply to the Aggregator authority records about agents, places, concepts or events in order to be able to use the Aggregator's Authority File Manager Tool that allows the management of authority list by merging duplicate authority records coming from the same CP or from different CPs, thus allowing to create common HOPE authority lists. The use of the Authority File Manager Tool requires that

- Each CP authority record has an identifier.
- Each metadata record contains the identifiers of the CP authority records it is linked to.

The Aggregator may also build automatically the CP authority records provided that the CP metadata records contain authority identifiers and enough regularly structured information (for example, for agents: name, surname[,date of birth[-date of death]]).

As other entities, each CP authority records must have a unique persistent identifier. However, these identifiers will be used only internally by the HOPE System and never be published on the Internet. Consequently, the resolver service is not important for CP authority records (which is not the case for common HOPE authority records).

For the authority records identifiers, CP may use:

- Local identifiers
- PIDs managed in a local PID service
- Third-party identifiers (for example, identifiers created by the organisation managing the authority list, such as a national library). These identifiers are not PIDs (no resolver service), but since the resolver service is not important for CP authority

records and since these identifiers are unique and persistent, they can be considered as local PIDs.

The workflows for creation, supply and update of PIDs for CP authority records is split in two scenarios:

- The CP uses local identifiers for the records
- The CP uses PIDs or third-party identifiers for the records

3.1.4.4.1 CP uses local identifiers

Procedure:

- The CP inserts a local identifier in each authority record it supplies to the Aggregator.
- In each metadata record, the CP inserts the local identifiers of the authority records used by the metadata record.
- The Aggregator creates a PID using the HOPE PID Service for each CP authority record. When the CP merges its authority list with the common HOPE authority list, the Aggregator registers the CP authority PID in the corresponding HOPE authority record.

Binding updates:

- When changing a local identifier, the CP must update the binding with the PID recorded by the HOPE PID Service, using the HOPE PID Service administrative interface.

Requirements:

- See Procedure and Binding updates.

3.1.4.4.2 CP uses a local PID service (or third-party identifiers)

Procedure:

- The CP inserts a PID (or third-party identifier) in each authority record it supplies to the Aggregator.
- In each metadata record, the CP inserts the PIDs of the authority records used by the metadata record.
- When the CP merges its authority list with the common HOPE authority list, the Aggregator registers the CP authority PID in the corresponding HOPE authority record.

Binding updates:

- The CP cannot change the PID identifying its authority records.

Requirements:

- See Procedure and Binding updates.

3.2. Granularity

Authors:	Kathryn Máthé – KEE-OSA Bert Lemmens – Amsab-ISG
Internal reviewers:	Joris Janssens – Amsab-ISG Marien van der Heijden – KNAW-IISG

This chapter deals with the detailed specifications for recording hierarchical descriptions and compound digital objects in the HOPE Data Model, and the corresponding requirements for supply of identifiers to the HOPE Aggregator.

3.2.1 Survey Results

This section provides a summary of the information gathered through the HOPE Survey for Content Providers about the granularity of the digital objects and metadata records supplied by the Content Providers (CPs).

The HOPE Survey included questions on two particular aspects of the HOPE Metadata:

- **Can CPs ensure their metadata contains a one-to-one relationship between metadata records and digital objects?**

This Survey question was raised when considering the ESE specification that allows CPs to provide **only one** URL for providing direct access to the digital file representing the described item. This implies that ideally, CPs supply exactly one digital file for each metadata record.

The general conclusion was that 57,4% of the total number of collections, representing 42,9% of the metadata records, do **not** have a one-to-one relationship between descriptions and digital objects.

Moreover, 5 CPs, representing more than half of the digital objects (57,7%) claim that they are not able to change the granularity of the digital objects, because they lack the time and the people.

- **Do CPs provide collections that contain hierarchically structured metadata records?**

A major drawback of the ESE element set was the way it dealt with hierarchically structured metadata structures, particularly metadata on archival records. This problem has been addressed at length by the APEnet Project. This question tried to find out whether the hierarchical description issue also applied to the metadata supplied by HOPE CPs.

The general conclusion was that, based on the Content Providers Survey, the

current metadata supplied to the Aggregator contains only 14 collections with hierarchical descriptions, including 11 archive collections and 3 library collections. However, several CPs have indicated they would like to make hierarchically structured collections available via the HOPE Aggregator in the near future. Only 4 CPs answered affirmatively to the question whether they have the capacity (time, people, tools, know-how) to change the granularity of the metadata records in order to comply with Europeana requirements.

The following conclusions can be drawn from the survey results:

- If we consider a one-to-one relation between a metadata record and a digital file as the best practice to deliver to Europeana (conform to the ESE-requirements), then only a bigger half of the HOPE metadata records conforms to this.
- A substantial number of CPs however informed us that they lack the capacity (resources) to change the granularity of metadata.
- If alternatively we would consider changing the digital files in order to conform to the number of metadata objects, than no less than potentially 80% of the digital objects would have to be altered.
- The issue of hierarchically structured descriptions currently only applies to a minor part of the metadata, but HOPE will have to deal with this problem in the future.

3.2.2 Europeana and Granularity

However, in August 2010 Europeana published the technical specifications for the [European Data Model \(EDM\)](#), which will be implemented in the Danube release of Europeana (expected May 2011). These specifications open new perspectives as to mapping compound objects and hierarchical descriptions.

- The EDM describes *aggregations* as groupings of multiple web resources and by default a single proxy (the placeholder for the descriptive unit). The EDM specifies the `ens:hasView` property, allowing “an aggregation to have 0 to many resources as views of the described cultural heritage object”. This implies that relating zero to many digital files to one metadata record may be considered a technical possibility in EDM. [[EDMv5.2](#): 23-24]
- Following the test mappings by the APEnet project, the EDM allows for recording hierarchical descriptions using the `dc:isPartOf` property. This means that EDM is recording hierarchies of Descriptive Units recording the ID of the ‘parent record’ as a property of the ‘child’ record.

As to the display of multiple digital objects for one metadata record in the Europeana portal, the Europeana v1.0 project is currently working on the implementation of displaying multi-level descriptions in the Europeana Portal. This

feature is expected to be included in the Danube release (see [presentation David Haskiya](#), d.d. 16-09-2010).

However, it is important to note that currently (March 2011) only the technical specifications for the EDM have been published. The Mapping and Normalisation Guidelines, which are expected to include concrete guidelines for recording granularity, have not been published yet.

Allowing one-to-many relationships between metadata descriptions and digital files would considerably reduce the harmonisation work regarding the granularity level of the digital files to a minimum and retains the projected number of digital objects to be delivered to Europeana.

3.2.3 Recording Granularity in the Common HOPE Metadata Structure

This section describes how the hierarchical descriptions and compound objects are recorded in the Common HOPE Metadata Structure. First of all, a clear distinction should be made between (1) the way granularity of descriptive units and digital files is recorded in the domain profiles, and (2) the way it is recorded in the HOPE Data Model.

As a general rule:

- The domain profiles record granularity in accordance with the specifications of the corresponding encoding standard. Which means:
 - the archive profile will record hierarchical descriptions as nested <c> level records, as specified in EAD .
 - the library profile will record references to the parent record using the marc:774 Constituent Unit Entry metadata element, as specified in MARCXML. (Note that HOPE will not map metadata of the 'parent' record into the 80X-83X Series Added Entry Fields)
- The HOPE Schema will record metadata about hierarchical descriptions and compound objects in a way that is analogous with the Europeana Data Model, in order to ensure a simple mapping from the HOPE Schema to the forthcoming EDM format.
- The CPs map their information on hierarchies and compound objects to the domain profiles, after which the HOPE Aggregator will deal with the transformation of hierarchies and compound objects from domain profile to HOPE Schema profile by profile.

In the section below, we will discuss the metadata elements used by the HOPE Data Model for recording hierarchical descriptions in the Descriptive Unit entity and compound objects in the Digital Resource Entity.

3.2.4 Hierarchical Descriptions in the Descriptive Unit Entity

Hierarchical descriptions, as referred to in this section, may include:

- multi-level archival fonds,
- monograph series and issues of periodicals.
- audio-visual series and programs
- visual series and

The first section lists all metadata elements the HOPE Data Model uses for recording hierarchical descriptions, as well as the requirements for the supply by the CP of these elements to the aggregator.

The variety of hierarchical structures accommodated by the HOPE Data Model will be discussed in the second section.

3.2.4.1 Elements

Local Identifiers and PIDs

In the HOPE Data Model hierarchies are recorded by including cross references to the PID of a related descriptive unit. Therefore it is a HOPE System requirement that each Descriptive Unit in the HOPE System **MUST** be identified by a PID.

PIDs can be supplied by the CP as part of the metadata. If the CP is cannot supply PIDs, the aggregator will create them for the CP. In that case, the CP must provide the aggregator with a local identifier.

- **local identifier (1,1)**
an unambiguous reference to a description of one or more collection items in the local system of the Content Provider.

Requirements:

- CP **MUST** provide the Aggregator with a local ID that unambiguously identifies the described resource in its Local system. This local ID **MUST** be (convertable to) a string or a resolvable identifier (cf PID)

- **persistent identifier (1,1)**
a globally unique, resolvable identifier, identifying the Descriptive Unit entity in and outside the HOPE System.

Requirements:

- The CP CAN provide the Aggregator with a persistent ID that unambiguously identifies the described resource in the HOPE System

In the EDM, the persistent identifier will be the URI for the ore:proxy representing the Descriptive Unit.

Parent Identifiers

Cross referencing between Descriptive Units is performed in only one direction: from child record to parent record. Therefore it is a system requirement that each descriptive unit describing an entity that is 'contained by' another entity, MUST include a reference to the persistent identifier of the Descriptive Unit describing that 'parent' entity.

If the CP cannot provide the Aggregator with a persistent identifier for the Descriptive Unit of the parent entity, the CP must provide the corresponding local ID.

- **local identifier parent descriptive unit (0,1)**
local identifier of the parent metadata record, as it is identified in the local information system of the content provider.

requirements:

- If the Descriptive Unit is part of hierarchical description, the CP MUST provide the Aggregator with a local identifier for the parent descriptive unit.

- **is contained by (0,1)**
This relationship links a descriptive unit with another constituent descriptive unit that 'contains' the descriptive unit.

requirements:

- if the Descriptive Unit is part of hierarchical description, the CP CAN provide the Aggregator with a persistent identifier for the parent descriptive unit

In the EDM, the hope:isContainedBy relationship will be mapped as the dcterms:isPartOf property, relating two hierarchically structured proxies.

As for the mapping procedure, when the CP does not record references to parent records in its local data, the Aggregator provides the opportunity for the CP to describe in 'plain language' how hierarchies are recorded in its data set. For instance, a data set may record hierarchies using references to child records, or use nested elements in the XML structure

(as is the case with EAD). The explanation should allow the Aggregator to reconstruct the hierarchies in the HOPE System and create the values to be stored in the is contained by element.

Sequence

The HOPE Data Model provides the opportunity to arrange a series of child Descriptive Units that are contained by the same parent Descriptive Unit in a sequence. This may particularly apply to archival series.

In order to support this feature, the HOPE Data model specifies the Is Next In Sequence element, which contains the persistent identifier of the next Descriptive Unit in the ranking.

If the CP cannot provide the Aggregator with a persistent identifier for the next Descriptive Unit in sequence, the CP can also provide the corresponding local ID.

- **local identifier next descriptive unit in sequence (0,1)**
local identifier that relates two descriptive units that are 'ordered' parts of the same parent descriptive unit, such that one descriptive unit comes immediately after the other descriptive unit.
Requirement:
 - The CP CAN supply the local identifier for the next descriptive unit in sequence.

- **is next in sequence (0,1)**
This relationship links two descriptive units that are 'ordered' parts of the same constituent descriptive unit, such that one descriptive unit comes immediately after the other descriptive unit.
Requirements:
 - The CP CAN supply the persistent identifier for the next descriptive unit in sequence.

The hope:isNextInSequence relationship will be mapped to EDM as the ens:isNextInSequence property, relating two proxies that belong to a sequence of proxies.

As for the mapping procedure, when the CP records sequence in another way, the Aggregator provides the opportunity for the CP to describe in plain English how sequences are recorded in its data set. For instance, a particular data set may record a ranking number for each record in the sequence. The explanation should allow the Aggregator to reconstruct the ranking in the HOPE System and create the values to be stored in the "is next in sequence" element.

Level of Description

The HOPE Data Model allows for each Descriptive Unit to specify a 'description level', i.e. the level of arrangement of the descriptive unit.

The element Level of Description is used in archival descriptive practice to define the scope of the description (cf. the ISAD(G) compliant levels: fonds, series, file, item). The HOPE Data Model makes this element available also for library, visual, audio-visual and Dublin Core encoded materials. There are two main objectives for making this element available for all descriptive units.

- **Recording descriptive metadata on the level of arrangement, supplied by the CP.**
This particularly applies to archival metadata, but it might also be useful to record the distinction between series descriptions and issue/episode descriptions for library and audio-visual materials.
The harmonisation and recording of this metadata will provide a valuable search entry for the HOPE system. Within the scope of the HOPE project, it was not feasible to establish agreement on a controlled list for the description level element. The Data Model however has already specified domain-specific elements for each of the domain profiles. It is therefore proposed to consider the specification of vocabularies for Description Level as a priority in the next stage of the HOPE Project and to include this element in the normalisation workflow.
- **Allow the Aggregator to distinguish between top-level collection descriptions and bottom-level item descriptions, linked with digital content.**
As an alternative for the lack of domain-specific vocabularies for this element, the Data Model will specify a normalised value for each descriptive unit. This value is based on the way the CP specifies the mapping for his collection. The HOPE mapping sheet allows the specification of different mappings for top-level collection records, mid-level records and item records.
 - **Collection record:**
This is the top-level record for each collection supplied by the CP. Collection records are not part of the data set, but the metadata is created during the mapping by completing the collection form.
 - **Mid-level records:**
Include descriptions of groupings of item records or other mid-level records. Mid-level records have no related digital content.
 - **Item level records:**
include descriptions of single items. If the collection has been digitised, the digital content is linked at item level.

The table below contains examples of description levels for each specific domain and the normalised value it should be associated with in the mapping.

	Archives	Library	Audio-Visual	Visual
Collection	Collection (created at ingest time)	Collection (created at ingest time)	Collection (created at ingest time)	Collection (created at ingest time)
Mid-Level	Fonds > Series	Periodical, Monograph Series	Series	Series
Item	Folder, Document	Monograph, Issue	Film, Sound Recording, Program	Image, Object

The specifications for the Level of Description element in the HOPE Data Model are as follows:

- **description level (1,1)**
the level of arrangement of the descriptive unit.
Requirements:
 - The CP SHOULD provide a controlled term for the description level of the Descriptive Unit.

In order to show how this element looks in an HOPE XML instance, including the normalised value added via the mapping sheet, we have listed some examples:

```
<descriptionLevel normalised="collection"></descriptionLevel>
<descriptionLevel normalised="series">fonds</descriptionLevel>
<descriptionLevel normalised="item">folder</descriptionLevel>
<descriptionLevel normalised="series">subseries</descriptionLevel>
<descriptionLevel normalised="series">periodical</descriptionLevel>
<descriptionLevel normalised="item">issue</descriptionLevel>
<descriptionLevel normalised="series">series</descriptionLevel>
<descriptionLevel normalised="item">program</descriptionLevel>
<descriptionLevel normalised="item">film</descriptionLevel>
```

3.2.4.2 Hierarchical structures

This section describes the different hierarchical structures the HOPE data model supports, and the subsequent requirements for hierarchical description in HOPE collections. Note that at the top of each multi-level description in every domain, there is a Collection description which is not part of the local metadata, but which is created at ingest time. The metadata of

the lower-level descriptions are all recorded in the data set and should be mapped to the domain profiles, using the mapping sheets.

The sections below are presented by domain type and are further broken down into requirements for digital, non-digital, and so-called “mixed” collections. It should be underscored that while mixed collections have not been part of the discussion thus far, local realities will likely give rise to this scenario as the project progresses. There are, in fact, many reasons why a content provider might submit a mixture of digitised and non-digitised materials, ranging from the administrative infeasibility of digitising every object to restrictions over selected content.

Archival Collections

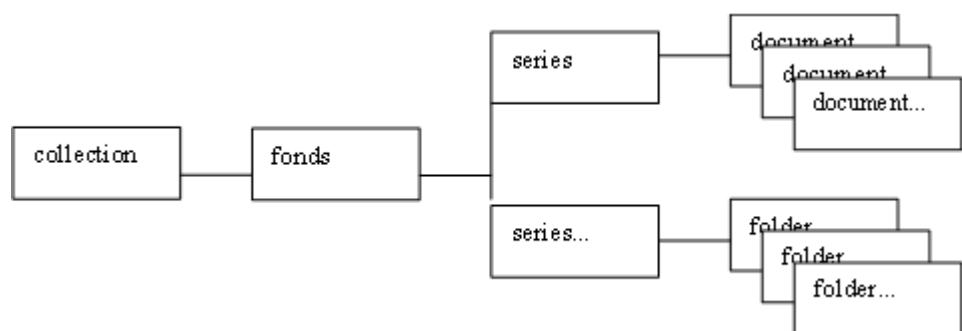
In the HOPE data model, the most granular level of an archival collection, the so-called “item”, can be either a document or a folder (i.e. group of documents). Above each document or folder the levels of description can be arrayed as follows:

- Document / Folder < Series* < Fonds* < Collection
- Document / Folder < Collection

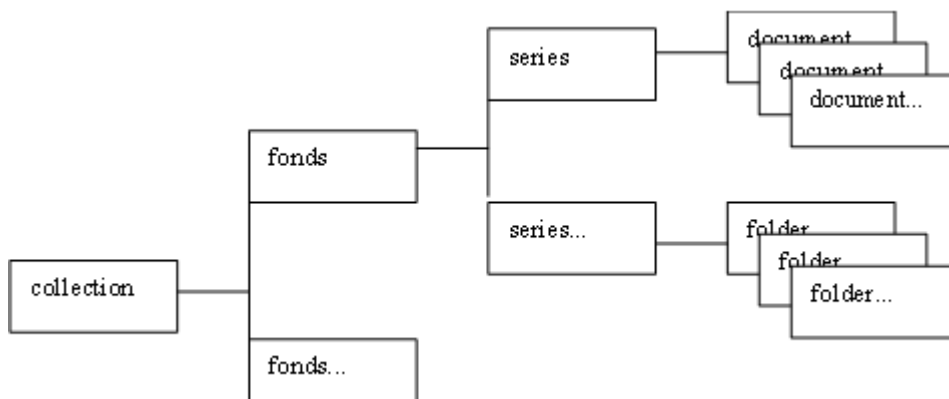
(*subfonds and subseries can also be added along with fonds or series throughout this section)

A single collection can be modelled as one of the following:

ARCHIVAL MODEL 1



ARCHIVAL MODEL 2



ARCHIVAL MODEL 3



Requirements:

- **Digital Archive Collections**

- **The lowest level of description (i.e. the “item”) MUST be a folder or a document and MUST correspond to the digital objects supplied for the collection.**

All digital archival collections must include items that are either folders or documents with digital objects of corresponding granularity.

Rationale: For the good of an appropriate display of the archival unit in the IALHI and Europeana Portal, each digital object should be accompanied with specific metadata providing the semantic context for that digital object. This unit should be the “lowest level granularity that is both intellectually meaningful and administratively practical” for a particular group of items.

- **A submitted fonds description SHOULD include all units present in the existing description, i.e. all subfonds, series, subseries, items, etc.**

A fonds and each of its child units should be submitted in their completeness.

Rationale: For the appropriate display of an archival unit in the IALHI and Europeana Portals and because of archival inheritance rules, the hierarchical

finding aid structure should not be violated if possible.

- **Items sharing a single parent description SHOULD be of a uniform type.**
In other words, a single series, subseries, or collection should have EITHER folders OR documents.
Rationale: The inclusion of mixed lowest level description types for the same set of material may disorient end users.
- **A selection of items from a single fonds SHOULD be treated as a separate collection.**
This might include a selection of multiple items from several series or a selection of a single or multiple series from the same fonds. If such a selection would result in the submission of a partial finding aid, then it is recommended to use Model 3 rather than to violate of the hierarchical finding aid structure.
Rationale: For the appropriate display of an archival unit to end users and because of complex inheritance rules, the hierarchical finding aid structure should not be violated if possible.
Note: *In this case it is recommended to provide links back to original description(s) as part of the item and/or collection descriptions.*
- **A selection of items from multiple fonds SHOULD be treated as a separate collection.**
This might include any selection of items or series from multiple fonds. If such a selection would result in the submission of one or more partial finding aids, then it is recommended to use Model 3 rather than to violate of the hierarchical finding aid structure.
Rationale: For the appropriate display of an archival unit to end users and because of complex inheritance rules, the hierarchical finding aid structure should not be violated if possible.
Note: *In this case it is recommended to provide links back to original description(s) as part of the item and/or collection descriptions.*
- **A selection of archival material treated as a separate collection (see above) SHOULD include links back to the formal archival finding aid.**
In the case that a archive type collection is composed of a selection from one or more fonds, a link from the collection and/or item descriptions to the appropriate local descriptive units should be made.

- **Non-Digitised Archive Collections**

The requirements for non-digitised archive type collections correspond to the above with the exceptions that:

- no correspondence between descriptive units and digital objects is required and
- non-digitised archive type collections MAY be submitted without lower-level descriptions.

Thus, a **non-digitised archive type collection may fit archival models 1-3 with the exception that only the level directly beneath the collection level is a requirement.**

- **Mixed Digital and Non-Digitised Archive Collections**

The requirements for mixed archival collections correspond to those of both sections above with the following provision:

- **Items sharing a single parent description MUST be either *all digital* or *all non-digital*.**

In other words, a single series, subseries, or collection must have either digital or non-digital items.

Rationale: The inclusion of mixed digital and non-digital items for the same set of material may disorient end users.

Note: *If the items are only partially digitised, then the unit is treated as a non-digital unit; the digital objects should not be submitted until the entire set has been digitised.*

Exception: In the case that digital objects from a single unit are only partially available based on externally imposed restrictions (e.g. copyright, 3rd party privacy, donor restrictions, etc.), then the inclusion of both digital and non-digital (or open and restricted) items can be considered.

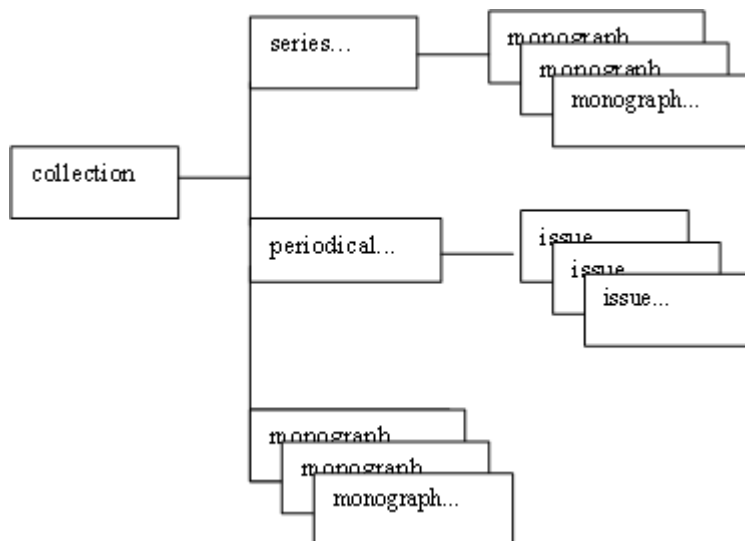
Library Collections:

In the HOPE data model, the most granular level of a library collection, the so-called “item”, can be either a monograph or an issue. Above each monograph or issue, the levels of description can be arrayed as follows:

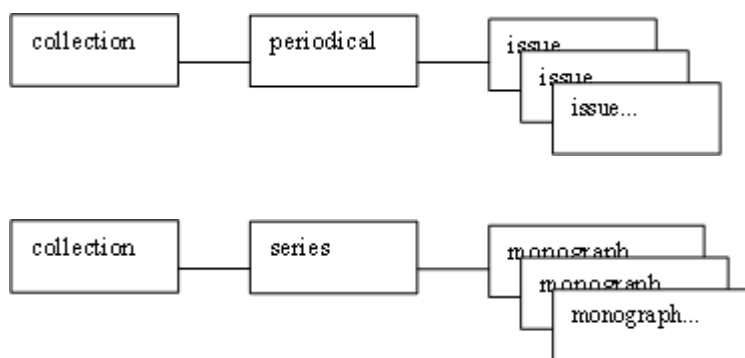
- Issue < Periodical < Collection
- Monograph < Series < Collection
- Monograph < Collection

A single collection can be modelled as one of the following:

LIBRARY MODEL 1



LIBRARY MODEL 2



Requirements:

- **Digital Library Collections**

- **The lowest level of description (i.e. the “item”) MUST be a monograph or an issue and MUST correspond to the digital objects supplied for the collection.**

All digital library collections must include items that are either monographs or periodical issues with digital objects of corresponding granularity.

Descriptive units on articles or other content-based sub-units of published

items should not be submitted as separate records.

Rationale: For the good of an appropriate display of the library unit in the IALHI and Europeana Portal, each digital object should be accompanied with specific metadata providing the semantic context for that digital object. This unit should be the “lowest level granularity that is both intellectually meaningful and administratively practical” for a particular group of items. For library items, this should correspond to published units and not content-based units.

- **A submitted periodical or monograph series SHOULD include all units or sub-units present in the existing description.**

A periodical or monograph series should be submitted with all issues or monographs on which the series description was based.

Rationale: For the appropriate display of an library unit in the Labour History and Europeana Portals, periodicals and series should be represented and described in a complete and consistent manner.

- **The units and levels of description for a given collection MAY be of different types.**

A single collection may have a mixture of monographs, monograph series, and periodicals.

Rationale: It is accepted practice in libraries to allow the search and display of series and periodical descriptions along with those of single monographs.

- **A multi-volume monograph MAY be treated either as a single monograph, as multiple monographs, or as a monograph series.**

In the case of a multi-volume work, it is at the discretion of CPs to determine which model of description is most appropriate. Digital objects will be stored with the lowest level of description. In the case that a multi-volume work is described as a single monograph, multiple volumes must be submitted as a single digital object.

- **A collection MAY be composed of a single monograph series or periodical.**

For a significant or substantial periodical or series, it may be appropriate to create a single collection for a single publication.

- **Archival material described as library items SHOULD include links back to the formal archival finding aid.** In the case that a library collection is composed of archival materials, a link from the collection and/or item descriptions to the appropriate local descriptive units should be made.

- **non-digitised library collections**

The requirements for non-digitised library collections correspond to the above with the exceptions that:

- no correspondence between descriptive units and digital objects is required and
- periodical and monograph series descriptions MAY be submitted without accompanying issue or monograph descriptions.

Thus, a **non-digitised library collection may fit library models 1-2 with the exception that only the level directly beneath the collection level is a requirement.**

- **mixed digital and non-digitised library collections**

- **Items from the same collection MUST be either *all digital* or *all non-digital*.** A library collection (both models 1 and 2) must have either digital or non-digital series and items.

Rationale: The inclusion of mixed digital and non-digital items for the same set of material may disorient end users.

Note: *If the items of a collection are only partially digitised, then the collection is treated as a non-digital collection; the digital objects should not be submitted until the entire collection has been digitised.*

Exception: In the case that digital objects from a collection are only partially available based on externally imposed restrictions (e.g. copyright, 3rd party privacy, donor restrictions, etc.), then the inclusion of both digital and non-digital (or open and restricted) items can be considered.

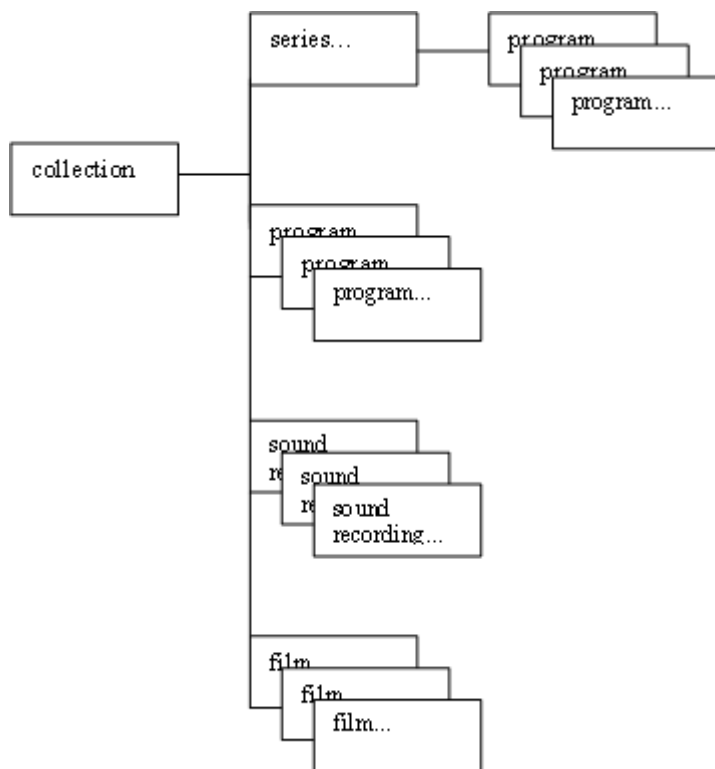
Audio-Visual Collections:

In the HOPE data model, the most granular level of an AV collection, the so-called “item”, can be either a film, a sound recording, or a program. Above each film, sound recording, or program, the levels of description can be arrayed as follows:

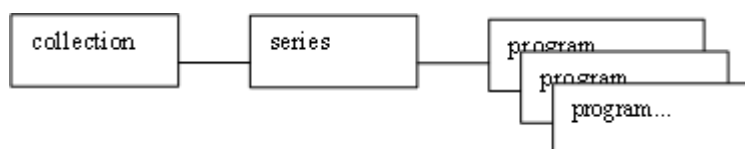
- Program < Series / Serial < Collection
- Film / Sound Recording / Program < Collection

A single collection can be modelled as one of the following:

AV MODEL 1



AV MODEL 2



Requirements:

- **Digital Audio-Visual Collections**
 - **The lowest level of description (i.e. the “item”) MUST be a program, film, or a sound recording and MUST correspond to the digital objects supplied for the collection.**
All digital av collections must include items that are either programs, films, or sound recordings with digital objects of corresponding granularity.
Rationale: For the good of an appropriate display of the av unit in the IALHI and Europeana Portal, each digital object should be accompanied with specific metadata providing the semantic context for that digital object. This

unit should be the “lowest level granularity that is both intellectually meaningful and administratively practical” for a particular group of items. For av items, this is generally a unit of production or distribution.

- **A submitted series SHOULD include units or sub-units present in the existing description.**
A series should be submitted with all programs on which the series description was based.
Rationale: For the appropriate display of an av unit in the IALHI and Europeana Portals, series should be represented and described in a complete and consistent manner.
- **The units and levels of description for a given collection MAY be of different types.**
A single collection may have a mixture of films, sound recordings, programs, and series.
Rationale: It is accepted practice in libraries and repositories to allow the search and display of series along with those of items.
- **A multi-part program MAY be treated either as a single program, as multiple programs, or as a series.**
In the case of a multi-part program, it is at the discretion of CPs to determine which model of description is most appropriate. Digital objects will be stored with the lowest level of description. In the case that a multi-part program is described as a single program, multiple parts must be submitted as a single digital object.
- **A collection MAY be composed of a single series.**
For a significant or substantial series, it may be appropriate to create a single collection for a single production.
- **Archival materials described as av items SHOULD include links back to the formal archival finding aid.**
In the case that an av collection is composed of archival materials, a link from the collection and/or item descriptions to the appropriate local descriptive units should be made.

- **Non-Digitised Audio-Visual Collections**

The requirements for non-digitised av collections correspond to the above with the exceptions that:

- no correspondence between descriptive units and digital objects is required and
- series descriptions MAY be submitted without accompanying program descriptions.

Thus, a non-digitised av collection may fit av models 1-2 with the exception that only the level directly beneath the collection level is a requirement.

- **Mixed Digital and Non-Digitised Audio-Visual Collections**

- **Items from the same collection MUST be either *all digital* or *all non-digital*.**
An av collection (both models 1 and 2) must have either digital or non-digital series and items.

Rationale: The inclusion of mixed digital and non-digital items for the same set of material may disorient end users.

Note: *If the items of a collection are only partially digitised, then the collection is treated as a non-digital collection; the digital objects should not be submitted until the entire collection has been digitised.*

Exception: In the case that digital objects from a collection are only partially available based on externally imposed restrictions (e.g. copyright, 3rd party privacy, donor restrictions, etc.), then the inclusion of both digital and non-digital (or open and restricted) items can be considered.

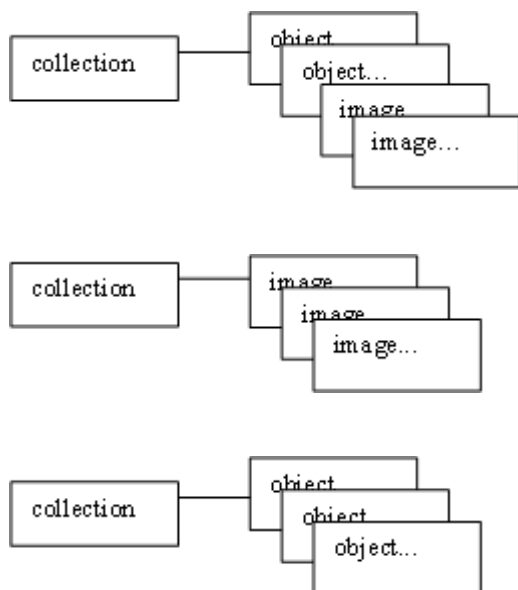
Visual Collections:

Visual collections basically consist of a set of single cultural heritage objects, i.e. photographs, prints, posters, badges, etc. In the HOPE data model, the most granular level of a visual collection, the so-called “item”, is an image or object. Above each image or object, there is a collection, as shown below:

- Image / Object < Collection

A single collection can be modelled in the following manner:

VISUAL MODEL



Requirements:

- **Digital Visual Collections:**

- **The lowest level of description MUST be an image or an object and MUST correspond to the digital objects supplied for the collection.**

All digital visual collections must include items that are either images or objects with digital objects of corresponding granularity. When an object consists of multiple components, the object as a whole must be treated as a single descriptive unit and submitted as a single digital object.

Rationale: For the good of an appropriate display of the visual unit in the IALHI and Europeana Portal, each digital object should be accompanied with specific metadata providing the semantic context for that digital object. This unit should be the “lowest level granularity that is both intellectually meaningful and administratively practical” for a particular group of items.

Note: *The visual profile has been developed for the item-level description of visual materials. For visual items published as a formal series, it is recommended to use the library profile. For visual items created and stored in formal or artificial collections, it is recommended to use the archival profile.*

- **The units of descriptions for a given collection MAY be of different types.**
A single collection may have a mixture of images and objects.

Rationale: It is accepted practice in museums to allow the common search

and display of image and object descriptions.

- **Archival materials described as visual items SHOULD include links back to the formal archival finding aid.**

In the case that a visual collection is composed of archival materials, a link from the collection and/or item descriptions to the appropriate local descriptive units should be made.

- **Non-Digitised Visual Collections**

The requirements for non-digitised visual collections correspond to the above with the exception that no correspondence between descriptive units and digital objects is required.

Thus, a **non-digitised visual collection MUST also fit the visual model.**

- **Mixed Digital and Non-Digitised Visual Collections**

- **Items from the same collection MUST be either *all digital* or *all non-digital*.**

A visual collection must have either digital or non-digital items.

Rationale: The inclusion of mixed digital and non-digital items for the same set of material may disorient end users.

Exception: In the case that digital objects from a collection are only partially available based on externally imposed restrictions (e.g. copyright, 3rd party privacy, donor restrictions, etc.), then the inclusion of both digital and non-digital (or open and restricted) items can be considered.

Note: *If the items of a collection are only partially digitised, then the collection is treated as a non-digital collection; the digital objects should not be submitted until the entire collection has been digitised.*

3.2.5 Compound Digital Objects in the Digital Resource Entity

‘Compound Digital Object’ usually refers to one or more *Digital Files* that, as whole, constitute the *Digital Representation* of a Collection Item. According to the findings of the HOPE Content Providers Survey (see 3.2.1 Survey Results), more than half of the collections supplied to the HOPE Aggregator contain compound digital objects. Therefore the HOPE Data Model should allow associating multiple digital files to one Descriptive Unit entity.

In the HOPE System, there is an additional level of complexity, since the functional requirements for the Common HOPE Metadata Structure require the HOPE Data Model to record multiple versions of a single digital file. These versions may include:

- A low-resolution derivative file
- A thumbnail file
- A transcription text file.

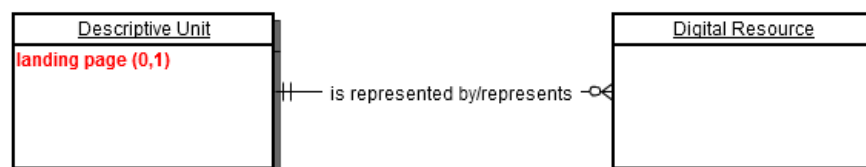
The means that the HOPE Data Model should distinguish between three levels of granularity for digital content: compound object, digital file and version of a file.

The following paragraph describes how the HOPE Data Model has modeled this three level hierarchical structure and how the related terminology has been defined.

(Compound) Digital Object

The (Compound) Digital Object has not been modeled in the HOPE Data Model as an entity in its own right. In order to record structural and administrative metadata about Simple as well as Compound Digital Objects, that HOPE Data Model considers the Digital Representation as the main unit of information for the exchange of metadata about digital content in the HOPE System. Hence, the digital object should be considered as a discrete unit of information, which consists of all instances of the Digital Resource entity that are associated with an instance of the Descriptive Unit entity (i.e. the description of the collection item). It should be noted that also the Europeana Data Model follows a similar approach, i.e. the Digital Object is a discrete unit of information that consists of all instances of `ens:webResource` that are gathered by an `ore:aggregation`.

So, from the perspective of the HOPE Data Model, a compound digital object is the name for one or more Digital Resource entities that are associated with one single Descriptive Unit. A Digital Object represents one single Collection Item, described by the Descriptive Unit entity, even if that item is represented by multiple images, i.e. the associated Digital Resource



entities.

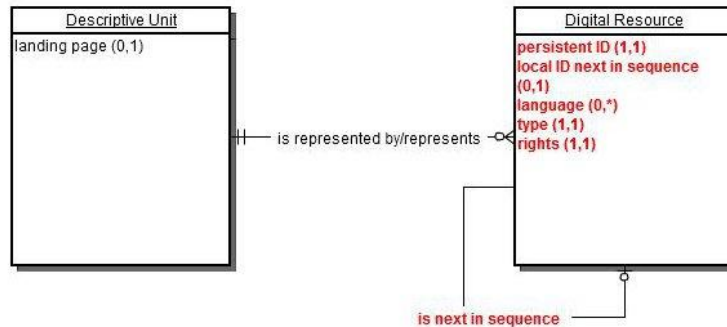
The only information, supplied by Content Providers, about Digital Object as such, may consist of a URL for a web page on Content Provider's website, which displays the digital object (i.e. a set of digital image files) as well as the description of the corresponding collection item.

The HOPE Data Model records this URLs using the Landing Page element, and requires the use of PIDs for this element. Since Digital Objects have a one-to-one relationship with the Descriptive Unit entity, and since the Landing Page also represents the metadata record, the Landing Page is recorded as a property of the Descriptive Unit entity.

Digital Representation

In order to accommodate multiple images creating one Digital Object, and at the same time accommodate multiple versions of one image, the HOPE Data Model specifies an

intermediary level that is denoted the Digital Representation. In the HOPE System, a Digital Representation is a single digital image or audio-visual / sound recording, that provides a single, unique rendition of a Collection Item, such as a single digital image, an audio-visual recording or a sound recording. In the HOPE Data Model, information about Digital Representations are accommodated by the Digital Resource entity.



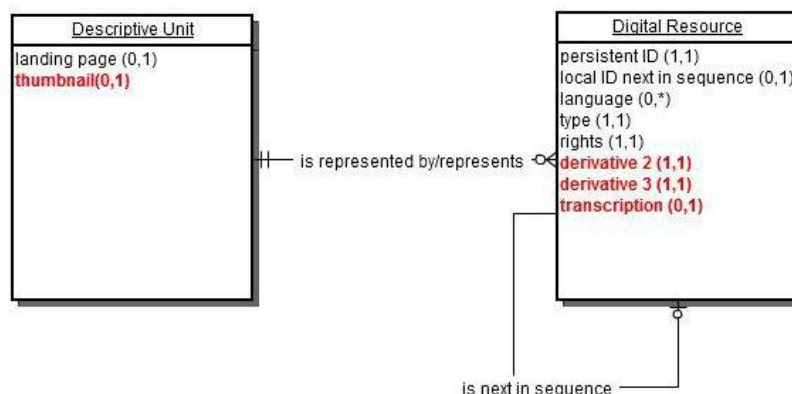
The Digital Resource entity records a persistent identifier for each Digital Representation, as well as structural (i.e. associated descriptive units, sequence information) and administrative (i.e. language, type, rights) metadata that apply to the image or recording. Detailed specifications for all these metadata elements can be found in section 2.3.2.

Digital file

In the HOPE System, a digital representation may be available through multiple versions of the same image. The HOPE Data Model accommodates metadata about three types of versions:

- low-resolution derivatives (cf. derivative 2)
- thumbnail derivatives (cf. derivative 3)
- transcriptions

These versions are stored as Digital Derivative Files in a Local or Shared Object Repository or by the Aggregator. A Digital Derivative File is an image, text, film or sound recording, encoded as a binary computer file, created for a specific use, such as printing, displaying or previewing of the digital representation. Metadata about these digital files is recorded as properties of the Digital Resource entity, using the Derivative 2, Derivative 3 and Transcription element.



These three elements allow recording PIDs as well as local IDs resolve URLs for each type of digital file, expressing the Digital Representation.

The supply of a Derivative 2 element is required by HOPE in order to supply digital content to Europeana. The supply of a Derivative 3 element is optional, but when the CP does not supply a thumbnail for the Digital Representation, the Aggregator will automatically derive one from the Derivative 2 element. So this is why the Data Model considers the derivative 3 element mandatory. The supply of a manually or automatically (i.e. OCRed) transcription of the digital representation is optional.

The Derivative 3 element of the first instance of the Digital Resource entity associated with a Descriptive Unit entity, will be mapped into the Thumbnail element of the Descriptive Unit. This image file will be used for display with metadata records in result lists.

Example

In order to make this three levels of granularity for digital content more concrete, we have included a practical example on how to record the metadata and digital content about a coin in the HOPE Data Model.

The Content Provider provides the aggregator with information about a coin, which belongs to a HOPE Collection. The coin itself is the **collection item**. The Content Provider provides the Aggregator with a **metadata record** that includes the description of the coin and information about the digital image of the coin.

The information that describes the coin is mapped into the **Descriptive Unit** entity. The metadata record contains information about two different **Digital Representations** of the coin, an image of the front side of the coin, and an image of the back side of the coin. This information is mapped into two different **Digital Resource** entities, one for each image. In the HOPE Data Model, these two digital resource entities are associated with the Descriptive Unit entity using the is represented by relationship. Together, these two instances of the Digital Resource entity constitute a **Digital Object**, representing the collection item.

Each digital resource contains information about the different **versions** of the digital representation, i.e. a **low-resolution derivative** of the front or back side image, a **thumbnail** of the front or back side image. The Derivative 2 and Derivative 3 elements contain a PIDs for each digital file, redirecting to the LOR where these digital files have been stored.

The following section lists all structural metadata elements the HOPE Data Model uses for recording Digital Resources, as well as the requirements for the supply by the CP of these elements to the aggregator. Please note that administrative metadata elements for Digital

Resources are discussed in detail in chapter 2.3.2.

The last section lists the requirements for the supply of structural metadata to the aggregator.

3.2.5.1 Elements

PIDs

In the HOPE Data Model the associations between Descriptive Units and Digital Resources are recorded by making cross/references between the persistent identifier of the associated entities.

Therefore it is a HOPE system requirement that also each Digital Resource in the HOPE System MUST be identified by a persistent identifier. The persistent identifier for the Digital Resource entity is created by the Aggregator.

- **persistent identifier (1,1)**
the globally unique, resolvable URL, identifying a digital representation in- and outside the HOPE information system.

Relationship Descriptive Unit – Digital Resource

In the HOPE Data Model, the concrete association between the Descriptive Unit and Digital Resource entities is recorded by the hope:isRepresentedBy and hope:represents relationships. Note this is the only relationship in the HOPE Data Model that is bi-directional.

The Aggregator creates this relationship based on the information provided by the Content Provider in the mapping worksheet.

- **is represented by (0,*)**
This relationship links the Descriptive Unit entity with one or more Digital Resource entities representing the described Collection Item.
- **represents (1,1)**
This relationship links the Digital Resource with the Descriptive Unit entity, describing the represented Collection Item
- .

requirements:

- The CP MUST record the binding between the descriptive metadata record and the corresponding digital file(s) in its local system.
- The CP MUST provide bindings between metadata record and the corresponding digital file(s) to the Aggregator via the data sets or via an

Aggregator Processing Instruction.

In the EDM, the relationship between Descriptive Unit and Digital Resource will be recorded by the `ens:hasView` property, which associates an `ore:aggregation` with zero or more `ens:webresources`.

Sequence

Analogous with the Descriptive Unit entity, the HOPE Data Model provides the opportunity to organise a series of Digital Resources that are associated with the same Descriptive Unit in a particular order or “sequence”. This may particularly apply to Compound Digital Objects where the series of images represents the logical order of the original resource (e.g. the separate digitised pages of a book). In order to support this feature, the HOPE Data model specifies the `Is Next In Sequence` element, which contains the persistent identifier of the next Digital Resource in the ranking. If the CP cannot provide the Aggregator with a persistent identifier for the next digital resource in sequence, the CP can also provide the corresponding local ID.

- **local identifier next digital resource in sequence (0,1)**
local identifier that relates two Digital Resources that are ‘ordered’ parts of the same compound digital object, such that one Digital Resource comes immediately after the other Digital Resource.

Requirement:

- The CP CAN supply the local identifier for the next digital resource in sequence.

- **is next in sequence (0,1)**
This relationship links two Digital Resources that are ‘ordered’ parts of the same compound digital object, such that one Digital Resource comes immediately after the other Digital Resource.

Requirement:

In the EDM, the `hope:isNextInSequence` relationship will be mapped as the `ens:isNextInSequence` property, relating two web resources that belong to a sequence.

As for the mapping procedure, when the CP does record rankings in another way, the Aggregator provides the opportunity for the CP to describe in ‘plain language’ how rankings are recorded in its data set. For instance, data sets may record ranking numbers for each digital file in the sequence. The explanation should allow the Aggregator to reconstruct the ranking in the HOPE System and create the values to be stored in the `is next in sequence` element.

Digital Files

The HOPE Data Model records information about Digital Files as properties of the Digital Resource entity. The HOPE Data Model specifies three metadata elements for this purpose:

The Derivative element records information about , low-resolution derivatives, thumbnails and transcriptions. This information is provided by the Content Provider via the mapping sheets. If no thumbnail is supplied, the aggregator will create one, derived from the low-resolution derivative.

- **Derivative 2 (1,1):**
globally unique, resolvable URL which allows retrieving a low-resolution derivative of the digital representation.
- **Derivative 3 (1,1):**
globally unique, resolvable URL which allows retrieving a thumbnail derivative of the digital representation.
 - occurrences: 1,1
 - value: record(value:URI; localID:literal; resolveURL:URI; label:constant value; encoding:constant value; primary:Boolean)
- **Transcription (0,1):**
Transcription of the textual content of a digital representation.

The value space of these three elements contains a PID that redirects to the corresponding low-resolution, thumbnail or transcription file. In addition, these three elements can have the following attributes:

- **local identifier**, which may contain a local identifier that identifies the digital file in the local system of the Content Provider.
- **resolve URL**, which may contain a URL that provides direct access to the digital file, stored at the Local or Shared Object Repository, or by the Aggregator.
- **label**, which contains an appropriate label for display with the digital file, i.e. low-resolution derivative, thumbnail or transcription.
- **encoding**, contains a reference to a metadata standard that is the source from which the metadata element has been derived.
- **primary**, is used to indicate whether the thumbnail is the preferred thumbnail that is to be mapped to the element Thumbnail in the corresponding Descriptive Unit. Only one thumbnail should be primary. If no thumbnail is primary, then the system will pick one to use with the descriptive unit. If there are several primary thumbnails, then the system will pick one of those.

3.2.5.2 Acceptance Levels

The actual requirements for the supply of structural metadata about Digital Resources to the HOPE Aggregator strongly depend on the acceptance level of the discovery services, targeted by the HOPE Aggregator. The requirements for the supply of digital derivative files and landing pages may be different for each of the discovery service. T2.3 identified the specific criteria for following discovery services:

HOPE Aggregator

These requirements are especially relevant to the Aggregator module in the HOPE System. Basically, they ensure the unambiguous identification of the Descriptive Unit and Digital Resource entities, as well as the relationship between both entities.

As a result, the supply of following elements by the CP to the HOPE Aggregator is **mandatory**:

- local ID for each Descriptive Unit
- binding between Descriptive Unit and Digital resource
- a resolve URL or PID for a Derivative 2 file

Europeana Portal

The requirements set by the Europeana Portal will be published in the forthcoming Europeana Mapping and Harmonisation Guidelines. The requirements listed below are based on the technical specifications of the EDM v5.2.1. They include acceptance criteria for the elements `ens:isShownBy`, `ens:isShownAt` and `ens:object`.

The supply of following elements by the Aggregator to Europeana are **mandatory**:

- `isShownBy` **or** `isShownAt`, i.e. a direct link to a digital file or a link to a webpage with the digital file embedded and displayed in its full information context.

These Europeana elements correspond with following elements in the HOPE Data Model:

- `digitalResource/derivative2`, **or**
- `descriptiveUnit/landingPage`

The supply of following elements by the Aggregator to Europeana are **recommended**:

- `ens:object`, i.e. a source image from which Europeana may create a thumbnail for display of the resource in result lists. If available, the Aggregator may supply the `ens:isShownBy`. If not the Aggregator **SHOULD** provide a link to an image that meets the technical requirements for the Europeana Portal Image.

This Europeana element corresponds with following element in the HOPE Data Model:

- `descriptiveUnit/thumbnail`

HOPE requirements for Europeana

One of the main objectives of the HOPE Project is supplying content to Europeana Portal. An additional objective is to deliberately improve the quality of the content supplied to the Europeana Portal. To this aim, the HOPE project seeks to raise the acceptance criteria for supply of digital content to the Europeana Portal.

These acceptance criteria MAY include:

The supply of following elements by the Aggregator to Europeana are **mandatory**:

- isShownBy **and** isShownAt, i.e. a direct link to a digital file or a link to a webpage with the digital file embedded and displayed in its full information context.
These Europeana elements correspond with following elements in the HOPE Data Model:
 - digitalResource/derivative2, **and**
 - descriptiveUnit/landingPage

- ens:object, i.e. a source image from which Europeana may create a thumbnail for display of the resource in result lists.
This Europeana element corresponds with following element in the HOPE Data Model:
 - descriptiveUnit/thumbnail

IALHI Portal

The functional requirements for the update of the IALHI Portal have already been drafted in D1.2. But in order to identify the acceptance criteria for digital content, they should be further specified by T3.7.

However, the HOPE Data Model provides the IALHI Portal with the following elements to allow a differentiated presentation of digital content in the IALHI Portal:

- digitalResource/derivative2, i.e. a low-resolution derivative of the digital file
- digitalResource/derivative3, i.e. a thumbnail derivative of the digital file
- digitalResource/transcription, i.e. a manually or automatically generated transcription of the text contained in the digital file.

For a detailed discussion of acceptance criteria for structural, as well as administrative and descriptive metadata, we refer to chapter 3.3.

3.3. Descriptive and Administrative Metadata

Authors:	Kathryn Máthé – KEE-OSA Bert Lemmens – Amsab-ISG
Internal reviewers:	Marien van der Heyden – KNAW-IISG Alessia Bardi – CNR-ISTI

This chapter describes the requirements for supply of descriptive and administrative metadata to the HOPE Aggregator, as well as the normalisation that will be performed after ingest.

3.3.1 Mandatory, Recommended & Optional Elements

This section describes the requirements for the supply of administrative and descriptive metadata to the HOPE Aggregator. These requirements provide the CP with guidelines as to which elements MUST/SHOULD/MAY be supplied to the Aggregator when preparing the mapping worksheet, Aggregator Processing Instruction and Dissemination Profile.

The actual specification of the ‘cardinality’ (i.e. if an element is mandatory (if), recommended or optional) of each element, will be indicated in the mapping worksheets. This section will provide you with the general principles for specifying these requirements.

These requirements heavily depend on the acceptance criteria for the discovery services that are part of the HOPE System. In other words, cardinality of the metadata elements depends on the acceptance criteria for services such as Europeana, IALHI and other Social Sites. As a result, there are different requirement levels for the HOPE Metadata, which are each related to a particular service. T2.3 identified four different requirement levels for the acceptance of Content Providers metadata to the HOPE Aggregator:

- **HOPE Aggregator:**

This acceptance level defines the system requirements, ensuring the basic functionality of the HOPE Aggregator. These requirements are independent of the targeted discovery services. These requirements are incorporated in the HOPE Data Model and have been specified by CNR-ISTI in the process of developing the Data Model.

- **Europeana:**
This acceptance level applies to metadata that is to be supplied to the Europeana Portal. The requirements for supply of metadata to the EDM based Danube release have not been published yet. But they are expected to resemble the acceptance level of the ESE based Rhine release, except for some specific elements. The requirements listed below are based on the ESE v3.3 specifications.
- **HOPE@Europeana:**
Regarding the supply of HOPE Metadata and Content to Europeana, it appeared during the HOPE BPN Workshop in Lisbon (February 2011) that HOPE aims for an acceptance level that is more ambitious than the specifications by ESE v3.3. T2.3 made a first attempt to describe this 'improved' acceptance criteria, in order to identify the stricter cardinality levels for the metadata elements involved. It is important to note that these requirements are a first draft. Agreement on these HOPE requirements for the supply of metadata to Europeana is still to be established.
- **IALHI Portal:**
Regarding the supply of metadata to the IALHI Portal, the acceptance criteria listed below are based on the preliminary functional requirements specified by T1.2. These requirements are to be further specified by T3.7. Hence the requirements listed below are only a first orientation, based on the possibilities offered by the HOPE Data Model.

3.3.1.1 HOPE Aggregator

These are the elements required by the aggregator infrastructure that ensure the basic functionalities of the HOPE Aggregator. These functionalities include:

- the unambiguous identification of all entities in the common HOPE metadata structure
- the unambiguous identification of hierarchical relations between entities.
- Identification of the language and script of the metadata, including translated and transliterated elements.

The supply of following metadata elements to the Aggregator is **mandatory**:

- Local ID Descriptive Unit (1,*)
- Local ID Parent Descriptive Unit (if applicable) (0,1)
- Metadata Language (1,1)

- Local ID, Resolve URL or PID for a Derivative 2 file (1,1)

Following metadata elements are required by the system, but will be created by the Aggregator

- Description Level@normalised (1,1)
- Persistent ID Digital Resource (1,1)
- Persistent ID or Agents, Concepts, Places and Events) (1,1)
- Persistent ID Content Provider (1,1)
- Persistent ID HOPE Theme (1,1)

The supply of following metadata element to the Aggregator is **recommended**:

- @language for each translated title or description (0,1) (if applicable)
- @script for each transliterated title or creator (0,1) (if applicable)
- @Local ID for Agents, Concepts, Places, and Events (0,*)
- Persistent ID Descriptive Unit (1,1)
- Persistent ID Parent Descriptive Unit (if applicable) (1,1)
- Description Level (0,1)

3.3.1.2 Europeana

These are the requirements for supply of metadata and digital content to Europeana. These requirements have been based on the specifications of ESE v3.3 and basically apply only to item level descriptions.

These requirements are to ensure the basic functionalities of Europeana, namely to support end user discovery, identification, and selection of material. The emphasis is on elements that facilitate rudimentary search, display, and retrieval and that provide basic information on ownership, content provision, and rights.

In the list below, we first indicate the EDM element, and then the corresponding element in the HOPE Data Model.

The supply of following metadata elements to the Aggregator is **mandatory**:

- ens:type
 - **(Europeana) Type** (1,1) (Digital Resource)
- dc:title **AND/OR** dc:description
 - **Title** (0,*) (Descriptive Unit) **AND/OR**
 - **Description** (0,*) (Descriptive Unit)
- ens:IsShownBy **AND/OR** ens:isShownAt
 - **Derivative2** (0,1) (Digital Resource), **AND/OR**

- **Landing Page** (0,1) (Descriptive Unit)

The following metadata elements are required by the system, but will be created by the Aggregator:

- ens:provider
 - **Aggregator** (1,1) (Descriptive Unit)
- ens:country
 - **Country** (1,*) (Content Provider)
- ens:language
 - **Language** (1,*) (Content Provider)

The supply of following metadata elements to the Aggregator is **recommended**:

- ens:rights
 - **Rights** (0,1) (Digital Resource)
- ens:object
 - **Thumbnail** (0,1) (Descriptive Unit)
- ens:dataProvider
 - **Name** (1,1) (Content Provider)
- dc:title
 - **Title@label="title"** (0,*) (Descriptive Unit)
 - **Title@label="title proper"** (0,*) (Descriptive Unit)
- dcterms:alternative
 - **Title@label="subtitle"** (0,*) (library)
 - **Title@label="translated title"** (0,*) (library)
 - **Title@label="parallel title"** (0,*) (library)
- dc:creator
 - **Creator** (0,*) (Descriptive Unit)
- dc:contributor
 - **Contributor** (0,*) (Descriptive Unit)
- dc:date

- **Date** (0,*) (Descriptive Unit)
- dcterms:created
 - **Date@label="object production date"** (0,*) (visual)
 - **Date@label="date of printing, manufacture or engraving"** (0,*) (library)
- dcterms:issued
 - **Date@label="date of publication"** (0,*) (library)
 - **Date@label="publication date"** (0,*) (audio-visual)

3.3.1.3 HOPE@Europeana:

In order to optimise Europeana's current search, browse, and display functionalities and to ensure a higher quality end user experience, HOPE aims for an acceptance level that is more ambitious than the specifications by ESE v3.3. This paragraph is a first attempt to identify the stricter requirements as to the supply of metadata and digital content supplied to Europeana. Agreement on these requirements within the HOPE network is still to be established.

The HOPE requirements for supply of metadata to Europeana include all requirements listed under 3.4.1.2, and following additional requirements:

The supply of following metadata elements to the Aggregator is **mandatory**:

- ens:isShownBy **AND** ens:isShownAt
 - **Derivative2** (0,1) (Digital Resource), **AND**
 - **Landing Page** (0,1) (Descriptive Unit)
- ens:object
 - **Thumbnail** (0,1) (Descriptive Unit)

3.3.1.4 IALHI Portal

These are the proposed requirements for the supply of metadata and digital content to the IALHI Portal. They have been proposed based on the preliminary functional requirements specified by T1.2—functional requirements which are to be further specified by T3.7. These requirements are intended to support a high-level discovery to delivery experience for the target user groups. (As a rule, Aggregator requirements are not repeated in these lists.)

This section specifies three sets of requirements, one for each level of description identified by the Aggregator. Requirements are separated and discussed by level of description for the following reasons:

- each descriptive level bears a different relationship to the description managed locally by CPs;
- each descriptive level will serve a different purpose on the IALHI Portal (as currently envisioned).

This reasoning will be further developed in the following section.

Collection Descriptive Units

The Collection Description is a unit of description based on the [Dublin Core Collection Description Profile](#)³⁷ and was created specifically for use in the IALHI Portal. Each collection will be defined by CPs according to guidelines provided by HOPE. Metadata on collections will be entered by the Content Provider into a form during the mapping phase of content submission. And, as the total number of collections for each CP is relatively small and as this metadata is created specifically for HOPE (and is not expected to be stored and managed in local systems), the requirements for collection description are relatively stringent. Collections are connected to series and/or to items but not to digital content.

Purpose of Collection Description: The collection record was created to provide a uniform access point to HOPE's wide ranging social history content. The purpose of the collection record is fourfold:

- To provide an overview of content available on the IALHI Portal. Collections are of a granularity and number which allow them to be showcased on maps, arrayed on a timeline or presented topically, alphabetically on a flat list, or by institutional clusters. Table 0 currently lists 130 digital collections.
- To help orient users by providing a relatively uniform set of basic information about each set of material submitted by a CP. Basic information includes the nature (both physical and intellectual) and structure of the underlying content and description. As such collection description is intended to sit atop the domain specific mid- and item-level descriptive units, which are steeped in their respective professional practice and often disorienting for end users to navigate between. Importantly, while in most cases it is infeasible to produce item-level descriptions in multiple languages, collection descriptions can easily be translated according to a defined language policy.
- To serve as a navigation device allowing users to jump between highly granular item-level results and broader sets of related material—and between content and context. Queries on collections can likewise provide an alternative to “noisy” item-level results set.

³⁷Dublin Core Collection Description Profile,
<http://dublincore.org/groups/collections/collection-application-profile/>

- To capture, present, and explain overarching administrative, access/use, and delivery policies. Collection description can help illuminate content as an administrative as well as an intellectual grouping.

Mandatory and Recommended Collection Description Elements: Mandatory (if) and recommended elements for collections have been set based on the 4 purposes set out above.

Purposes 1 and 3. To facilitate the browsing and searching of collections within the IALHI Portal, mandatory elements include:

- Collection Title
- Description
- Date Created
- Language (of items, if applicable)
- Item Type
- Data Provider
- Country of Data Provider
- HOPE Themes

Recommended elements include:

- Subject
- Spatial Coverage

Purpose 2. To help users identify and select materials, mandatory elements additionally include:

- Size (in either meters or number of items)

Recommended elements include:

- Item Format
- Temporal coverage
- Non-normalized values which accompany normalized values for dates and physical description fields

Purpose 4. To set out overarching policies, mandatory elements additionally include:

- Use Rights
- Access Rights

Recommended elements include:

- Custodial History
- Finding Aid

Mid-Level Descriptive Units

“Series”, as defined by HOPE in the most generic sense, are a set of items grouped together based on their creation/production or collection history. Series are represented by one or more units of description in several domains, including archival hierarchical fonds descriptions and library series/serial records. It is expected that HOPE CPs currently have mid-level “series” type records in their local collection management systems and that these have been created following professional content standards. Thus as a rule, few requirements are imposed on series description. Series are connected to collections, other series, and items but not to digital content.

Purpose of Mid-Level Description: Mid-level records are collected to provide the creation, production, and collecting context for HOPE content. They also provide an important level of granularity for the search and browse interfaces. As mid-level descriptive units are highly domain specific, it is best to discuss them accordingly:

Archival Fonds Description: Multi-level description is not one but several hierarchically structured mid-level descriptive units: Fonds, Subfonds, Series, Subseries. multi-level description is written as a single interconnected document with inheritance rules that pull description to the higher levels. As such, for IALHI Portal end users it is important that the structure is not violated—that existing units are kept and presented together and that the fonds unit is present. The requirements for submission of hierarchical units can be found under 3.2.3 Granularity. Hierarchical Descriptions in the Descriptive Unit Entity.

Because these are units of description that exist for purposes other than to facilitate the presentation of materials, it is not recommended to alter multi-level descriptions. Instead, the Collection Description has been developed to assume higher-level search and browsing roles. Hierarchical description still serves several important functions on the IALHI portal very similar to those listed above under Collection Description: to provide context about the creation and collection of materials, to provide administrative and access information, to serve as an important level of granularity which allows users to move between general and specific—content and context.

Element requirements have been set in accordance with ISAD(G) specifications, and not specifically with reference to IALHI functionality. The mandatory elements for fonds units include:

- Reference Code
- Level of Description
- Title
- Creator
- Dates
- Extent of the Unit of Description

Recommended elements for fonds units include:

- Administrative/Biographical History

- Archival History
- Language of the Described Materials (if applicable)
- Content Notes
- System of Arrangement
- Conditions Governing Access and Use
- and various Controlled Access Headings

As a result of inheritance, only Title, Level of Description, and Reference Code are required for other hierarchical descriptive units (i.e Subfonds, Series, Subseries). If Extent, Dates, or Creator are of a different value than those for fonds, they should of course be present. Recommended elements are Content Notes and System of Arrangement. Language of the Material and Conditions Governing Access and Use should also be present if they are different from the fonds.

Library and AV Series Description: Library and AV series are sets of items that are produced as a formal coherent unit and generally share a uniform title and often include a uniform description. In practice, series items are often described using a minimum of elements while more extensive description is pulled into the series description. Though they are also treated as mid-level descriptive units by the Aggregator, they are generally more granular than fonds-level description. As a result, library and AV series can be used in place of items to group and present search results, thus preventing the “noise” that would result from the presentation of all items bearing a similar name and description. The IALHI specifications from T1.2 make no reference to this issue. Here we have proposed mandatory elements which would allow library and AV series records to be searched together with other item records, thus providing a potentially valuable option for the IALHI Portal search interface.

Mandatory elements for Library and AV series include:

- Title
- Language (if applicable)
- Type

Dates and extent (in terms of number of items) are also highly recommended, but these can be tricky as the dates of publication/production and extent of a whole series are often broader than the dates and extent of the actual holdings of a single institution and can thus be difficult to establish. Other recommended elements for Library and AV series include:

- Authors, Directors, Editors, and Contributors
- Publisher and Place of Publication
- Producer and Place of Production
- Edition Statements
- Abstract or Synopsis
- Frequency
- Material Designation or Format

- Any Alternative or Translated Titles
- Subject Areas
- Rights and Access Conditions

Item-Level Descriptive Units

Items are the finest level of granularity in the context of the creation, production, or collection of the original materials. In HOPE, items are represented by units of description in each of the domains:

- Archival: documents, folders
- Library: monographs, periodical issues
- AV: films, sound recordings, programs
- Visual: images, objects.

Items are generally but not necessarily managed in local collection management systems. Archives in particular follow idiosyncratic practice in the creation and management of item-level descriptions. It is expected that HOPE CPs currently have some item-level description, especially for library and AV material, but that they will also dedicate time and resources to enhance and harmonize metadata to meet HOPE requirements. Thus, requirements have been set to maximize IALHI interface potential. In HOPE items are connected to collections, series, and digital content. The requirements for digital content are specifically related to items.

Purpose of Item-Level Description: Item-level records support users in the discovery, identification, selection, and retrieval of particular material. The purpose of item records in HOPE is fourfold:

- To support simple and advanced search and filter functionalities which lead to the discovery of specific material.
- To provide information on the identity, form, content, creation, and source of particular material.
- To serve to help users navigate between granular and broad results sets—and between content and context.
- To give practical information related to the source of, access to, and use of materials with direct links to digitized copies of the material when available or further information on access and delivery options when not.

Mandatory and Recommended Item Description Elements: Mandatory (if) and recommended elements for items have been set based on the 4 purposes set out above.

Purposes 1 and 3. To facilitate the searching and navigation of collections within the IALHI

Portal, mandatory elements across all domains include:

- Title
- Type
- Date
- Language (if applicable)
- Data Provider
- Thumbnail (for digital content and if applicable)

Recommended elements include:

- Alternative Titles
- Creators and Contributors
- Description, Abstract, Scope and Content, Synopsis
- Place of Publication, Production, Creation
- Subject Areas, Associations, Depictions
- Spatial Coverage

Purpose 2. To help users identify and select materials, recommended elements additionally include:

- Publisher or Producer
- Provenance, Administrative/Biographical History, Archival History
- Form, Genre, Object Type, Format
- Appearance of Material, Medium, Physical Description
- Extent, Dimensions, Duration, Length
- Temporal Coverage

Purpose 4. To provide practical information and links to content, mandatory elements additionally include:

- Copyright, Rights
- Access Conditions (for digital content)

Recommended elements additionally include:

- Local IDs, Call Numbers, or Reference Numbers
- Finding Aids
- Digital Resource PIDs and Sequence

3.3.2. Metadata Normalisation

In order to enable controlled access to a sub-set of HOPE elements, the Aggregator will perform a series of normalisation actions after ingest of the Content Providers' metadata.

These normalisations are semi-automated process, i.e:

- The Aggregator checks whether the values in a particular set of HOPE elements is compliant with their corresponding data value standard, i.e. a controlled syntax, list or vocabulary.
- The CP should check which terms are not compliant with the corresponding data value standard, and has to specify a compliant value. More information on this procedure can be found in the implementation guide https://wiki.peopleseurope.nl/index.php/Ingesting_your_data_sets_into_the_HOPE_Aggregator
- The CP provides the Aggregator with a matching table, including non-compliant values and their corresponding compliant value.
- The Aggregator records these compliant values as normalized values in the related HOPE elements.

The HOPE Data Model maintains the 'original' value, supplied by the Content Provider in the value space. The normalized value is recorded in the normalized attribute.

The section below lists all elements that will be considered in this normalization procedure. For each element, this section specifies:

- The objective of the normalization, i.e. which search access will be enabled.
- Which data value standard will be used as a reference, i.e. controlled syntax, controlled vocabularies, controlled lists
- Which elements in the HOPE Data Model will be checked.
- What are the subsequent requirements for supply of metadata to the Aggregator.

3.3.2.1 Dates

Objective:

The HOPE Data Model normalises metadata about dates in order to:

- Make Descriptive Units, HOPE Themes, Agents and Events searchable by date.
- Displaying Collections in the IALHI Portal on a Timeline.

Metadata standard:

ISO 8601:2004_ Data elements and interchange formats -- Information interchange -- Representation of dates and times

ISO 8601:2004 is a standard covering the exchange of date and time-related data. It was issued by the International Organization for Standardization (ISO) and was first published in 1988. The purpose of this standard is to provide an unambiguous and well-defined method of representing dates and times, so as to avoid misinterpretation of numeric representations

of dates and times, particularly when data is transferred between countries with different conventions for writing numeric dates and times.

The standard organizes the data so the largest temporal term (the year) appears first in the data string and progresses to the smallest term (the second). It also provides for a standardized method of communicating time-based information across time zones by attaching an offset to Coordinated Universal Time(UTC). [Wikipedia: http://en.wikipedia.org/wiki/ISO_8601]

ISO 8601:2004 is applicable whenever representation of dates in the Gregorian calendar, times in the 24-hour timekeeping system, time intervals and recurring time intervals or of the formats of these representations are included in information interchange.

Examples include:

- 2011 (calendar dates expressed in terms of calendar year)
- 2011-03 (calendar dates expressed in terms of calendar year and calendar)
- 2011-03-29 (calendar dates expressed in terms of calendar year, calendar month and calendar day of the month)
- 2010-05-21/2-11-03-29 (time intervals, using solidus or forward slash “/” as an interval designator)

ISO 8601:2004 does not cover dates and times where words are used in the representation and dates and times where characters are not used in the representation.

ISO 8601:2004 does not assign any particular meaning or interpretation to any data element that uses representations in accordance with ISO 8601:2004. Such meaning will be determined by the context of the application.

Elements:

The HOPE System normalizes dates recorded by following metadata elements:

- **Date** (Descriptive Unit) (0,*)
A date associated with an event in the life cycle of the described item.
Including:
 - Date of creation (archive)
 - Date of publication (library)
 - Date of printing or manufacture of engraving (library)
 - Object production date (visual)
 - Publication date (audio-visual)
 - Broadcast date (audio-visual)
 - Date (Dublin core)
- **Date** (HOPE Theme) (1,1).
Date associated with the HOPE Theme.

Remark: This date is specified as part of the HOPE List of Historical Themes and uses normalized dates by default.

- **Date (Agent) (0,*)**
Dates of birth and/or death of the person, or start and/or end date of activity of the organization.
Remark: The Agent/Date element does not record time intervals. Date of birth and date of death are recorded as separate occurrences, with type attribute indicating it is a start or end date.
- **Date (Event) (0,*)**
Start and/or end date of the Event.
Remark: The Event/Date element does not record time intervals. Start and end date are recorded as separate occurrences, with type attribute indicating it is a start of end date.

Requirements:

- The CP **MUST** provide a date for the collection and item-level Descriptive Units.
- The CP **SHOULD** provide a date for each mid-level Descriptive Unit.

3.3.2.2 Languages

Objective:

The HOPE Data Model normalizes metadata about languages in order to:

- Make Descriptive Units and Digital Resources searchable on language
- To identify translated values of titles and descriptions
- To enable adapting the user interface to the language of the Descriptive Unit.
- Provide Europeana with a value for the `ens:language` element.

Metadata Standard:

ISO 639-3:2007, Codes for the representation of names of languages — Part 3: Alpha-3 code for comprehensive coverage of languages

ISO 639-3:2007, is an international standard for language codes in the ISO 639 series. The standard describes three-letter codes for identifying languages. It extends the ISO 639-2 alpha-3 codes with an aim to cover all known natural languages.

It is intended for use in a wide range of applications, in particular computer systems where many languages need to be supported. It provides an enumeration of languages as complete

as possible, including living and extinct, ancient and constructed, major and minor, written and unwritten. However, it does not include reconstructed languages such as Proto-Indo-European. [wiki: <http://en.wikipedia.org/wiki/ISO-639-3>]

Examples:

- English: eng
- German: deu
- Dutch: ndl

Elements:

The HOPE Data Model records metadata about language in the following metadata elements.

- **Language (Content Provider) (1,*)**
Official language(s) of the country where the content provider is located.
- **Language metadata (Descriptive Unit) (1,1)**
Primary language of the metadata describing the item.
- **Language (Digital Resource) (0,*)**
Language(s) of the intellectual content of the digital resource.
- **Language attribute**
The language attribute contains a reference to the language, used for a translated value in the value space.
The attribute is applied for translations of titles and descriptions, including:
 - Translated title (library)
 - Translated subtitle (library)
 - Translated title (audio-visual)
 - Translated synopsis (audio-visual)

Note: The element Language (Descriptive Unit) is **not** normalized. This element contains free-text information about the language of the 'original' item, as part of the metadata description. In order to enable searching the digitized version of the 'original' item, the HOPE System uses the language digital content element.

Requirements:

- The CP **MUST** provide a value for the language metadata element.
- The CP **SHOULD** provide a language attribute for each translated title or description.

- The CP MUST provide a value for the language of all text type digital resources.

3.3.2.3 Scripts

Objective:

The HOPE Data Model normalizes metadata about scripts in order to:

- Allow the Aggregator to create indexes for values that are originally encoded using non-latin scripts.

Metadata Standard:

ISO 15924 - Codes for the representation of names of scripts

ISO 15924 defines two sets of codes for a number of writing systems (scripts). Each script is given both a four-letter code and a numeric one. Script is defined as "set of graphic characters used for the written form of one or more languages". So scripts that are not (yet) "used for the written form of one language" are explicitly excluded.

Where possible the codes are derived from ISO 639-2 where the name of a script and the name of a language using the script are. Preference is given to the 639-2 Bibliographical codes, which is different from the otherwise often favored use of the Terminological codes.

4-letter ISO 15924 codes are incorporated into the Language Subtag Registry for IETF language tags and so can be used in file formats that make use of such language tags. For example, they can be used in HTML and XML to help Web browsers determine which typeface to use for foreign text. This way one could differentiate, for example, between Serbian written in the Cyrillic (sr-Cyrl) or Latin (sr-Latn) script, or mark romanized text as such. [wiki: http://en.wikipedia.org/wiki/ISO_15924]

Examples:

- Latn : latin
- Arab : arabe
- Cyrl : cyrillique
- Yiii : yi

Elements:

The HOPE Data Model records metadata about scripts in the following attribute.

- **Script attribute**
The Script Attribute contains a reference to the script, used for a creator or title

which is not specified using Latin script.

This attribute is used to indicate creators and titles supplied by the CP using non-latin script. These values require a transliteration to Latin script in order to be indexed by the HOPE Aggregator.

Requirements:

- CPs **MUST** provide a value for the script used for each creator or title that is specified using non-Latin script, i.e. Cyrillic, Arabic, Hebrew.
- CPs **SHOULD** provide the Aggregator with a Latin transliterated value for each creator or title which is specified using non-Latin script.

3.3.2.4 Description Level

Objective:

The HOPE Data Model normalizes metadata about description levels in order to:

- Enable the Aggregator to distinguish between collection level, mid-level and item level Descriptive Units.

Controlled

T2.3 specified a Description Level elements in each sub-entity, enabling domain-specific vocabularies for each Archival, Library, Visual, Audio-Visual and Dublin Core Units. However, T2.3 was not able to establish agreement on these vocabularies.

Vocabulary:

Alternatively, each Description Level element currently has a 'generic' controlled list, which allows the Aggregator to distinguish between following description levels:

- **Collection level:**
This is the top-level record for each collection supplied by the CP. Collection records are not part of the data set, but the metadata is created during the mapping by completing the collection form.
- **Mid-level:**
include descriptions of groupings of item records or other mid-level records. Mid-level records have no related digital content.
- **Item level:**
include descriptions of single items. If the collection has been digitized, the digital content is linked at item level.

This normalized value is assigned when completing the mapping sheets for the data sets.

Depending on the mapping worksheet used, a normalized value will be assigned to a set of Descriptive Units.

Elements:

The HOPE Data Model records metadata about description levels in the following metadata elements.

- **description level (1,1)**
The level of arrangement of the descriptive unit.

Requirements:

- The CP SHOULD provide a controlled term for the description level of the Descriptive Unit.

3.3.2.5 Rights

Objective:

The HOPE Data Model normalizes metadata about Copyrights in order to:

- provide IALHI Portal users with information on use restrictions that apply to the displayed digital content.
- Provide Europeana with a statement about the rights status of the digital objects described in the metadata submitted to Europeana. (cf. ens:rights)

Controlled Vocabulary:

T1.4 will specify the copyright statements between which the IALHI portal will distinguish.

Regarding the ens:rights element, the forthcoming Mapping and Harmonisation Guidelines for the Europeana Danube Release are expected to include a list of preferred copyright statements for the ens:copyright element.

In the ESEv3.3, the inclusion of the ens:rights was recommended. In the forthcoming guidelines for EDM, this element is expected to become mandatory. The value ens:rights should be the URL of the appropriate (1) Creative Commons or (2) Europeana rights statement. Tools will be provided to assist the process of license.

The first set is from the Creative Commons (CC) domain and encompasses

- the Public Domain mark for objects that are not in copyright
- seven possible CC licenses for objects that are in copyright.

The second set comprises four possible statements defined by Europeana for use with objects that have reserved rights and may have on-line access restrictions.

The list includes following statements:

- CC License statements:
 - Public Domain Mark
(<http://creativecommons.org/publicdomain/mark/1.0/>)
 - CC – Zero (<http://creativecommons.org/publicdomain/zero/1.0/>)
 - CC BY (<http://creativecommons.org/licenses/by/3.0/>)
 - CC BY-SA (<http://creativecommons.org/licenses/by-sa/3.0/nl/>)
 - CC BY-NC (<http://creativecommons.org/licenses/by-nc/3.0/de/>)
 - CC BY-NC-SA (<http://creativecommons.org/licenses/by-nc-sa/2.0/fr/>)
 - CC BY-ND (<http://creativecommons.org/licenses/by-nd/2.0/es/>)
 - CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/1.0/fi/>)

- Europeana rights statements:
 - Rights Reserved - Free Access
(<http://www.europeana.eu/rights/rr-f/>)
 - Rights Reserved - Paid Access
(<http://www.europeana.eu/rights/rr-p/>)
 - Rights Reserved - Restricted Access
(<http://www.europeana.eu/rights/rr-r/>)
 - Unknown
(<http://www.europeana.eu/rights/unknown/>)

Changes in Europeana's Copyright policy have been anticipated in following document:

- Guidelines for the europeana:rights metadata element v4.0 (30/11/2010)
http://version1.europeana.eu/c/document_library/get_file?uuid=06e63d96-0358-4be8-9422-d63df3218510&groupId=10602

T2.3 did not establish agreement as to the Copyright Policy for Digital Resources.

Elements:

- **Rights** (Digital Resource) (0,1)
Information about rights held in and over the digital resource.

Requirements:

to be specified

3.3.2.6 Type

Objective:

The HOPE Data Model normalizes metadata about Europeana Types in order to:



- Provide Europeana with a value for the `ens:type` element, used to enable faceted browsing in the Europeana Portal.

Controlled list:

Europeana specifies following controlled list for the `ens:type` element:

- TEXT
- IMAGE
- AUDIO
- VIDEO

Inclusion of this element in the data supplied to Europeana is mandatory

Elements:

- **Type (1,1)**
The Europeana material type of the digital resource, which Europeana uses to categorize search results by text, image, sound or video..

Requirements:

- The CP **MUST** provide for a value for the Europeana Type element for each supplied Digital Resource.

4. Case Studies Semantic Enrichment

This chapter contains two cases studies regarding the possibilities for enriching the HOPE Metadata. This chapter reports on the findings of T2.4 as to the feasibility of creating common HOPE authorities and enabling multi-lingual access to the HOPE metadata.

4.1. Common HOPE Authorities

Authors:	Bert Lemmens – Amsab-ISG
Internal reviewers:	Ruth van Dijck – Amsab-ISG Alessia Bardi – CNR-ISTI

4.1.1. Survey results

This Sections includes the findings of the HOPE Content Survey and the Addition Survey regarding the use of controlled vocabularies by the HOPE content providers.

4.1.1.1 Content Providers Survey

The HOPE Content Provides Survey included a section on the use of controlled vocabularies by the content providers. The survey seeked information about the use of controlled vocabularies for the following entities: persons and organisations, geographic names, subjects, dates, time periods and co-ordinates.

The following table contains the percentage of metadata records that uses a controlled vocabulary for these entities. The survey made a distinction between controlled vocabularies created in-house by the content provider, and published data value standards adopted by the content provider.

Type of terminology	Use of a controlled vocabulary (% of metadata records)	Use of a published standard (% of metadata records)
geographic names	96,0%	18,2 %
subjects	95,4%	21,0 %
persons and organisations	78,5%	23,1 %

dates	48,0%	26,5 %
time periods	31,3%	0,6 %
co-ordinates	3,5%	3,5 %

Considering geographic names, subject, persons and organisations, a vast majority of the metadata has been validated using controlled vocabularies, which makes this metadata eligible for creating common HOPE authority lists. At the same time, only around 1 out of 5 of this metadata has been validated using a published data value standard (see list of standards below). Moreover, most of these Content Providers use national data value standards.

As regards the type of controlled vocabulary that is used: most of them are developed and maintained by the Content Providers. 36 Times a content provider mentioned that it used a self-developed vocabulary while only 19 times a content provider responded that the vocabulary it uses is a published standard:

Terminologies types	Published standard	
<i>Geographic names</i>	<ul style="list-style-type: none"> ● <i>ISO 3316-1 – Codes for the representation of names of countries and their subdivisions – Part 1: Country code (international)</i> ● <i>Merriam-Webster's Geographical Dictionary (international)</i> ● <i>MOTPRO (Finnish)</i> ● <i>RAMEAU (French)</i> 	
<i>Geographic co-ordinates</i>	<ul style="list-style-type: none"> ● <i>latitude/longitude referential system</i> 	
<i>Dates</i>	<ul style="list-style-type: none"> ● <i>ISO 8601:2000 (international)</i> ● <i>EN 28601 (European)</i> ● <i>ONKI Finnish Ontology Library Service: Lusto Aikaontologia (Lusto time ontology)</i> 	
<i>Time periods</i>	<ul style="list-style-type: none"> ● <i>RAMEAU (French)</i> 	
<i>Subjects</i>	<ul style="list-style-type: none"> ● <i>LCSH (international)</i> ● <i>Helvetosaurus (Swiss)</i> ● <i>ONKI Finnish Ontology Library Service: YSO (Finnish General Upper Ontology)</i> ● <i>TYPO (Finnish labour history thesaurus)</i> ● <i>RAMEAU (French)</i> 	

Persons and organisations	<ul style="list-style-type: none"> ● <i>Personennamdatei (PND) (German)</i> ● <i>Gemeinsame Körperschaftsdatei (GKD) (German)</i> ● <i>RAMEAU (French)</i> 	
---------------------------	---	--

7 Content providers (FES Library, KEE-OSA, KNAW-IISG, SSA, TA, UPIP-BDIC and UPIP-MSH) use at least one published data value standard. The other 6 content providers (Amsab-ISG, CGIL, FES Archive, FMS, Génériques and VGA) only use terminologies they developed themselves.

4.1.1.2 Additional Survey

After the HOPE Content Providers Survey, T2.4.1 decided to ask some additional questions regarding the use of controlled vocabularies for agents, places and concepts. The objective of this survey was to obtain a better view on the semantic quality and consistency of the controlled data for these entities, which is required to assess the feasibility of creating common HOPE authorities for these entities. The survey basically gathered information about the following aspects:

- the syntax of terms accommodated by these metadata elements,
- whether these metadata elements are either hard validated or also contain non-validated terms,
- the use of published standard vocabularies.

This additional survey provided T2.4.1 the following results³⁸:

Persons

- The survey asked for the syntax used for recording names of persons and organisations in the content providers' metadata.
 - 8 CPs indicated that they use a standard syntax for recording names of persons and organisations (CGIL, FES-A, FMS, Génériques, KNAW-IISG, UPIP(BDIC), VGA, SSA).
 - These CPs use the following components for building a standard syntax: name, surname, title, date of birth and death.
- Only 2 CPs (FES-A, SSA) indicated they ensure that all person and organisation names in their metadata are validated by a controlled vocabulary.
- Only 2 CPs use published standard vocabularies for recording person and organisation names.
 - FES-Library: Personennamdatei (PND), Gemeinsame Körperschaftsdatei (GKD)

³⁸We did not indicate any figures because we did not ask for any in-depth information. The results thus reflect general tendencies. Some CPs moreover did not answer (all) the survey (questions).

- UPIP(BDIC): RAMEAU

Geographical names

- 7 Content providers (CGIL, AMSAB, FES-A, FES-L, Génériques, SSA, VGA) use an idiosyncratic controlled vocabulary for geographical names, i.e. a controlled vocabulary that is created and maintained within the institution.
- 4 Content providers adopted a published controlled vocabulary for recording geographical names:
 - KNAW-IISG: ISO 3166-1 (*Codes for the representation of names of countries and their subdivisions – Part 1: Country codes*), but this will be replaced by MARC Code List of Countries
 - KEE-OSA: ISO 3316-1, Merriam-Webster's Geographical Dictionary
 - TA: Finnish web thesaurus MOTPRO
 - UPIP (BDIC): Rameau
- Most controlled vocabularies for geographical names are flat lists. However, 2 content providers use a hierarchically structured vocabulary for geographical names:
 - SSA: idiosyncratic
 - UPIP (BDIC): RAMEAU
- The survey asked for the geographical scope of the terms of the controlled vocabulary, in particular whether a single metadata element includes geographical names that apply to multiple levels.
 - Most controlled vocabularies are multilevel, including local, regional, national and supra-national terms.
 - 3 Content providers record geographical terms that have a single level.
KNAW-IISG: national terms
CGIL: national terms
TA: local terms
- 2 CPs (KNAW-IISG, CGIL) indicated not all geographical names in their metadata are all validated by a controlled vocabulary.

Subjects

- 5 Content providers (AMSAB-ISG, CGIL, FES-A, FES-L, KNAW-IISG) indicated they use an idiosyncratic controlled vocabulary for recording subjects.
- 4 Content providers adopted a published controlled vocabulary for recording subjects:
 - OSA: OSA Subject Headings
 - SSA: Helvetosaurus
 - TA: YSO (Finnish General Upper Ontology) and TYPO (labour history thesaurus)
 - UPIP(BDIC): RAMEAU

- The survey asked which kind of terms are recorded as subjects, in particular to identify which *type* of subject terms are recorded in one single metadata element: topical terms, personal or family names, corporate names, names of events, geographical names, chronological terms, genre/form terms.
6 Content providers indicated they record multiple subject types in one metadata element:
 - AMSAB: topical and genre/form terms
 - CGIL: topical names and event names; or
topical terms, names of events, personal/family names
 - FES-A: topical terms, names of events, chronological terms
 - TA: topical terms, names of events, geographical terms
 - KEE-OSA: geographic names with topical terms or
chronological terms; or
topical terms, names of events, corporate names,
geographic
names, chronological terms
 - UPIP(BDIC): geographical names and dates; or
topical terms, personal/family names, corporate names,
names
of events, geographical names, chronological terms,
genre/form
terms
- The survey also specifically asked how subject terms are expressed in the content providers' metadata:
 - 6 CPs (AMSAB-ISG, CGIL, FES-Archive, KNAW-IISG, KEE-OSA, TA) record one or more terms,
 - 2 CPs (KEE-OSA, UPIP(BDIC)) records articulated terms, i.e. a term consisting of multiple parts, organised into a meaningful order,
 - 2 CPs (KNAW-IISG, SSA) adopted a classification system for recording subjects.
- 3 CPs (KEE-OSA, CGIL, KNAW-IISG) indicated that not all subject terms in their metadata are validated by a controlled vocabulary.

4.1.2. Mapping and merging procedure

This section contains the conclusions of T2.4.1 as to the feasibility of creating common HOPE authorities for persons and organisations, geographical locations and subject.

4.1.2.1 HOPE Authority File Manager Tool

The HOPE Aggregator contains a tool for creating and merging authority records, which thus

enables the creation of common vocabularies for a defined set of entities. One of the objectives of T2.4.1 is to identify the entities in the content providers' metadata, for which it is feasible to use this Authority File Manager Tool.

Feasibility

Based on the survey results discussed above, the following conclusions can be made:

- Metadata on **persons** and **organisations** is validated by CPs representing 78,5% of the supplied metadata records. Content Providers indicate that the syntax of this metadata is fairly consistent.

T2.4.1 proposes to use the Authority File Manager Tool in order to create a common authority list for persons and organisations.

The following issues should be considered when specifying the mapping and merging procedure for this entity:

- not all terms in the metadata recording person and organisation names are hard validated by a controlled vocabulary,
- metadata about person and organisation names, often also include descriptive metadata about the title of the person and the dates of birth and death.

- Metadata on **geographical locations** is validated CPs representing 96% of the supplied metadata records. This metadata includes local, regional, national as well as supra-national terms.

T2.4.1 proposes to use the Authority File Manager Tool in order to create a common authority list for persons and organisations.

The following issues should be considered when specifying the mapping and merging procedures for this entity:

- not all terms in the metadata recording geographical names are hard validated by a controlled vocabulary,
- the common vocabulary should distinguish between geographical names with a different scope: i.e. local, regional, national and supra-national terms.

- Metadata on **subject terms** is validated by CPs representing 95,4% of the supplied metadata records. This metadata includes topical terms, as well as personal or family names, corporate names, names of events, geographical names, chronological terms and genre/form terms.

T2.4.1 proposes to look into the possibilities for using named entity extraction tools in order to distinguish between the different subject types, before implementing the Authority File Manager Tool for creating a common subject authority list.

The following issues should be considered when specifying the mapping and merging procedure for this entity:

- Not all terms in the metadata recording geographical names are hard validated by a controlled vocabulary.

- Some content providers provide articulated terms for topical terms and genre/form terms, i.e. a term consisting of multiple parts, organised into a meaningful order.

In the course of research performed by subT10 on feasibility of creating Common HOPE Authorities, an important technical drawback appeared for the implementation of the Authority File Manager Tool. The tool requires content providers supply controlled vocabularies as separate authority files, including a local identifier for each vocabulary term. These identifiers are required in order to record the provenance of the term in the HOPE system.

This technical requirement has not been surveyed by T2.4.1, which renders it impossible to draw final conclusions as to the feasibility of implementing the Authority File Manager Tool for creating common HOPE Authorities for person and organisation names, geographical locations and subjects.

The conclusions drafted above should be considered the basis for developing a business case.

Published Controlled vocabularies

Regarding the use of published controlled vocabularies by the HOPE content providers, none of the standards used was considered to be widespread enough to propose it as a best practice.

However, T2.4.1 proposes to use a published controlled vocabulary as a reference vocabulary in the merging process. That is to say, the Authority File Manager Tool should merge 'local' vocabularies supplied by the HOPE content providers. But the Authority File Manager Tool should also provide the possibility to merge terms with a published controlled vocabulary, to be specified by the HOPE content providers.

Following the involvement of Europeana with Linked Open Data, T2.4.1 will propose the adoption of the following published vocabularies:

- VIAF for person names
- geonames.org for geographical locations

Data Model

In order to allow the implementation of the Authority File Manager Tool, the HOPE Data Model has specified following elements for recording vocabulary terms for persons, organisations, places and events. (We included events, since we borrowed this entity for EDM. The practical use of this entity in the short run will be limited.)

Descriptive Unit Entity

Vocabulary terms are provided to the aggregator as part of the metadata record. In first instance the vocabulary term is mapped as a property of the Descriptive Unit. A vocabulary term will be mapped in two distinctive elements:

- One element (e.g. Creator) will record the vocabulary term supplied by the content provider as part of the metadata record. The term is kept as a back-up of the original value.
This element also records the local identifier of the vocabulary term in an attribute.
- Another element (e.g. Is Created by) will record the vocabulary term after being merged with other vocabulary terms. This term is the 'unified' term, which is also recorded in the corresponding entity (e.g. Agent)
This element records the local identifier and the persistent identifier of the term, referring to the Agent Entity.

Agent/Place/Concept/Event Entity

In a second stage, all vocabulary terms are also mapped to a corresponding entity (e.g. person and organisation terms are mapped to the Agent Entity.)

These entities by default record the following metadata elements:

- The persistent identifier of the vocabulary term. This identifier is created by the Aggregator and unambiguously identifies the term in the HOPE System.
- The local identifier of the vocabulary term. This identifier refers to the local system of the content provider who supplied the term.

In addition, each entity also records a set of descriptive metadata for each vocabulary term, such as dates, description, spatial coverage and source. These elements are specified in the 2.2 Data Model Specifications.

4.1.2.2 Agents

Objective:

The HOPE Data Model enables to merge controlled vocabularies for persons and organisations in order to:

- enable unified search results for persons and organisations in the IALHI Portal,
- allow to link HOPE metadata about persons and organisations with a published controlled vocabulary, such as VIAF.

Metadata Standard:

T2.4.1 proposes to consider associating person and organisations names with the following published vocabularies:

- VIAF
- ULAN

Elements:

The HOPE Data Model specifies the following elements for recording vocabulary terms of

agents.

The following metadata elements record the person or organisation name supplied by the content provider:

- **Creator** (Descriptive Unit) (0,*)
- **Contributor** (Descriptive Unit) (0,*)
- **Publisher** (Descriptive Unit) (0,*)
- **Subject** (Descriptive Unit) (0,*)

The following metadata elements record the 'unified' term, created by using the Authority File Manager Tool:

- **IsCreatedBy** (Descriptive Unit) (0,*)
- **Has Contributions By** (Descriptive Unit) (0,*)
- **Is Published By** (Descriptive Unit) (0,*)
- **Associated Agent** (Descriptive Unit) (0,*)
- **Depicted Agent** (Descriptive Unit) (0,*)

4.1.2.3 Places

Objective:

The HOPE Data Model enables to merge controlled vocabularies for geographical locations in order to:

- enable unified search results for geographical locations in the IALHI Portal,
- allow to link HOPE metadata about geographical locations with a published controlled vocabulary, such as geonames.org.

Standard

Vocabulary:

T2.4.1 proposes to consider to associate geographical locations with the following published vocabularies:

- geonames.org

Elements:

The HOPE Data Model specifies the following elements for recording vocabulary terms of geographical locations.

The following metadata elements record the geographical term supplied by the content provider.

- **Creator/place** (Descriptive Unit) (0,*)
- **Publication/place** (Descriptive Unit) (0,*)
- **Spatial coverage** (Descriptive Unit) (0,*)

The following metadata elements record the 'unified' term, created by using the Authority File Manager Tool.

- **Is created In** (Descriptive Unit) (0,*)
- **Is Published In** (Descriptive Unit) (0,*)
- **Associated Place** (Descriptive Unit) (0,*)
- **Depicted Place** (Descriptive Unit) (0,*)

4.1.2.4 Concepts

Please note that merging concepts is possible only after extracting terms such as topical terms, genre/form terms from the content providers' metadata. The specifications below only apply to these kinds of subject headings.

Objective:

The HOPE Data Model enables to merge controlled vocabularies for concepts in order to:

- enable unified search results for concepts in the IALHI Portal,
- allow to link HOPE metadata about concepts with a published controlled vocabulary, such as LCSH, AAT.

Metadata Standard:

T2.4.1 proposes to consider to associate concepts with the following published vocabularies:

- LCSH
- AAT

Elements:

The HOPE Data Model specifies the following elements for recording vocabulary terms of concepts.

The following metadata elements record the concept term supplied by the content provider:

- **Subject** (Descriptive Unit) (0,*)

The following metadata elements record the 'unified' term, created by using the Authority File Manager Tool:

- **Associated Concept** (Descriptive Unit) (0,*)
- **Depicted Concept** (Descriptive Unit) (0,*)

4.1.2.5 Events

Please note that merging events is currently a theoretical option, since there is not much

HOPE metadata about events.

Objective:

The HOPE Data Model enables to merge controlled vocabularies for events in order to:

- enable unified search results for events in the IALHI Portal.

Metadata Standard:

T2.4.1 was, within the scope of the HOPE project, not able to identify a published controlled vocabulary for events.

Elements:

The HOPE Data Model specifies the following elements for recording vocabulary terms of events.

The following metadata elements record the event name supplied by the content provider:

- **Temporal Coverage** (Descriptive Unit) (0,*)

The following metadata elements record the 'unified' term, created by using the Authority File Manager Tool.

- **Associated Events** (Descriptive Unit) (0,*)
- **Depicted events** (Descriptive Unit) (0,*)

4.1.3 HOPE Themes

In addition to the Agent, Place, Concept and Event entities, the HOPE Data Model has specified a HOPE-specific entity for HOPE Themes. The HOPE Theme Entity accommodates thematic headings, specific to the fields of social and labour history.

HOPE Themes are not assigned via mappings or vocabulary mergings. HOPE themes are assigned in two different ways:

- HOPE Themes are assigned to Collections, via completing the Collections form that is part of the mapping worksheet. Based on this information, the Aggregator will create relationships between Descriptive units and HOPE Themes.
- HOPE Themes are assigned to mid-level and item level records, using the HOPE tagging tool. The functionality of this tool is still to be developed but should allow to create sub-sets of Descriptive Units for which a HOPE Theme can be assigned.

A list of 47 HOPE Themes to be accommodated by the HOPE Theme Entity has been specified by T1.6³⁹. This Task also listed following requirements³⁹ as to the specification of the HOPE Theme Entity:

³⁹ [Available separately](#)

- the HOPE Theme Entity should allow hierarchical relations between a set of broader terms, i.e. cluster titles, and a set of narrower terms, i.e. the actual HOPE Themes to be assigned to the metadata.

Please note that the HOPE system will at first stage provide management of a **flat** list of themes and, only in a later stage a hierarchical list (if it is feasible and usable). In the linked document themes are grouped under the same letter (e.g., A POLITICAL PARTIES AND MOVEMENTS) for sake of comprehension, but in the first stage only those having a number will be in the theme list (e.g., A.01 Anarchist movements, C.03 Jobless)

- the HOPE Theme Entity should allow the possibility for recording translations for each HOPE Theme,
- the HOPE Theme Entity should allow the following descriptive metadata about each HOPE Theme:
 - a date or date range for the HOPE theme,
 - a description for the HOPE theme.

The HOPE Data Model specified the following elements for the HOPE Entity:

- **persistent identifier (1,1)**
unique, resolvable identifier, identifying the HOPE Theme in the HOPE information system.
- **preferred term (1,1)**
The accepted or most desirable term for the theme used in the HOPE system.
- **alternative term (0,*)**
A translation of the preferred term for the HOPE Theme, used in the HOPE system.
- **date (1,1)**
date associated with the HOPE Theme
- **description (0,*)**
An account of the HOPE Theme

The HOPE Data Model specified the following relationships for the HOPE Entity:

- **associated HOPE Theme (0,*)**
This relationship links an object with a HOPE theme that is the topic of the object.
- **broader term (1,1)**
This relationship links a historical theme with another higher-level historical theme that includes the historical theme.

4.2. Multilingual Access

Authors:	Ruth Van Dyck – Amsab-ISG Bert Lemmens – Amsab-ISG Kathryn Máthé – KEE-OSA Katalin Dobó – KEE-OSA Afelonne Doek – KNAW-IISG
Internal reviewers:	Sjoerd Siebinga – KNAW-IISG Alessia Bardi – CNR-ISTI

4.2.1 Introduction

4.2.1.1 Sources of information

- HOPE content providers survey
- Additional survey information from IISG
- Additional survey information from KEE-OSA
- Additional survey information from Génériques
- Presentation on Europeana language resources at Europeana plenary
- Skype call d.d. 21-12-2010 with Franca Debole (CNR-ISTI) on MultiMatch tools
- Skype call d.d. 22-12-2010 with and review paper v.08 d.d. 14.01.2011 by Sjoerd Siebinga (EuropeanaConnect) on Europeana language resources
- Website MultiMatch2
- Website Europeana language resources
- Further bibliographical references

4.2.1.2 Purpose of this document

This document reports on the findings of Task 2.4.2 “Specification for harmonisation requirements” regarding the harmonisation and enrichment of multilingual metadata by the HOPE aggregator.

The DoW spells out two implementation issues as to multilingualism on which we have to reach an agreement:

1) the “recording [of] metadata in foreign languages” (DoW, Part A, p. 10/36)

This may involve agreement on how the HOPE aggregator accommodates information on the language of the HOPE metadata, as well as original and translated values of (a selection

of) metadata elements

2) “how to tackle local practices of transliterations and translations” (DoW, Part A, p. 10/36)

This involves agreement on how the HOPE aggregator accommodates information on the scripts and alphabets used by in the HOPE metadata, as well as original and transliterated values of (a selection of) metadata elements.

As for the enrichment of the HOPE metadata, this document will look into the available tools for establishing multilingual access to the HOPE metadata. We will particularly focus on the feasibility of implementing multilingual controlled vocabularies for subjects and translation services for unstructured text.

First, this document briefly summarises the findings of the HOPE content providers survey on the use of different languages and scripts in the HOPE metadata, completed with the additional information we gathered from CPs (see chapter 2.).

The results of Task 1.2 – Creation of use cases – have some bearing on the multilingualism issue, the sense that language has been explicitly mentioned as a parameter for sorting and filtering search and browsing result sets in the IALHI Portal.

Since HOPE wants to adhere to the global project philosophy “not to develop or devise new technologies and solutions, but to adhere to best practices and to use existing tools and methods” (DoW, Part B, p. 13/76), we shall look into the results of a number of European(a) projects about multilingualism issues. Task 2.4.2 explicitly mentions the resources and tools developed by the European MultiMatch project⁴⁰ but the DoW also states that the work of “Europeana Connect-WP1 (...) in the area of multi-linguality can be shared with HOPE.” (Part B, p. 41/76) We shall thus report on the results that have come out of both MultiMatch (see section 3.3.) and EuropeanaConnect (see section 3.4.) projects and assess whether and if so, in what way, we could make use of them.

The DoW separates the multilingualism issues (T2.4.2.) from the issues that relate to the use of “metadata encoding schemes” (and metadata cleansing practices), about which equally agreement has to be established (T2.4.1). In the framework of this task 2.4 the latter relates to metadata *value* encoding schemes: established lists of normalised/controlled terms or headings used to ensure uniformity, i.e. controlled vocabularies/authorities.⁴¹ Although these issues – multilingualism and the sharing/harmonising of controlled vocabularies – are mentioned separately, they are related. For controlled vocabularies can be used to create “translation dictionaries” i.e. mappings (matches) between terms in different languages. In this paper we shall look into a number of multilingual vocabularies/authorities (see section 3.6) in order to assess the feasibility and usefulness to use (some of) these lists to create translation dictionaries.

⁴⁰ “This task will work together with CNR-ISTI (WP4 leader) and look into the effects of applying tools and translation resources developed by MultiMATCH to reach agreement on the practicalities of implementing such solutions.” The DoW also mentions that “[t]he Aggregator services will implement crosslanguage searching by re-using available multi-lingual resources and tools developed in the context of the European Project MultiMatch”. (DoW, part B, p. 14/76)

⁴¹ The DoW specifies that “[t]he schemes of particular importance to HOPE concern relevant entities to social history such as events, persons, organisations, places and time expressions/periodisation” and that, unfortunately, there “are as yet no authoritative semantic schemes for such entities”. (DoW, Part A, p. 10/36)

The conclusions of this document contain a proposal on how to deal with respectively translation issues (section 4.1), transliteration issues (section 4.2), and the use of special characteristics and diacritics (section 4.3) within the scope of the HOPE project. These paragraphs cover the criteria for recording multilingual metadata in the HOPE Aggregator, and make suggestions for improving multilingual access to the HOPE metadata.

The final chapter contains a more high level discussion of the multilingual issues the HOPE Aggregator is facing and a list of the issues HOPE has to deal with on short, mid and long term. This should help HOPE to catch up with the current and future development of tools enabling multilingual access to digital resources.

4.2.1.3 Building a multilingual controlled vocabulary

The basic problem

In an ideal situation a user would be able to perform searches in a language of his or her choice, the system would then retrieve all relevant results (and only these) in all the languages of the metadata and/or the digital objects, and the user would (automatically) receive a translation of the results in the language(s) desired. Although multilingual access to digital resources is a “hot” actual research topic, in the current stage of development this is still utopia, as we shall demonstrate in this paper.

The key problem in the automation of multilingual access to digital resources are the semantic specificities of each language and the specificity of the – in our case: social-history – knowledge domain in which content providers and users operate when creating or searching for information. When a user is entering a particular search term “myriad semantic difficulties can arise. Virtually every single word (...) has more than one meaning or sense, and many of those senses have more than one nuance; many words can be used as nouns, verbs, adjectives and/or adverbs. Search systems that purport to allow the use of natural language cannot yet successfully distinguish among different meanings or various parts of speech in very large general systems (...). In addition there is evidence that people writing about the same concepts often do not use the same words to express them, and people searching for the same concept do not think of the same words to search for it.” (Taylor & Joudrey, 2008: 333) Since these semantic difficulties are specific for each language, they are multiplied in case of a search system that deals with multilingual searching (metadata or full text digital objects in multiple languages).

Addressing the problem: approaches at different levels

Addressing multilingualism issues in practice means addressing semantic problems (subjects) and dealing with the problem of different names or variant forms of the same name. This can be done at different levels: user interface level, metadata level, domain knowledge representation level (ontologies/thesauri/controlled vocabularies) (DoW, part B, p. 13-14). Within T2.4 this means first and foremost addressing them at the metadata level. We shall deal with the metadata level in chapter 4: sections 4.1 and 4.2. However, we cannot leave out the other two levels, since they are part of the problem and/or the solution. We shall deal with them in chapter 3 (resources, tools and services) and chapter 4: section 4.3

(presence of special characteristics and diacritics).⁴² In any case, given the current state of development, we shall be forced to adopt a pragmatic approach.

We now first define the scope of the multilingual issues for the HOPE metadata by presenting a short analysis of the languages and scripts used in the HOPE metadata, and the translations and transliterations (already or still to be) performed.

4.2.2 Hope Metadata: Language and Scripts, Translations and Transliterations

4.2.2.1 Languages / translations

A random check of the element sets gathered by the HOPE content providers survey showed that most CPs by default provide dedicated metadata elements recording the language of the metadata and or the language of the described material/digital content. When considering multilingual access, this information is important for linking translation vocabularies, or for adapting the user interface to the metadata displayed.

The vast majority of content providers (11) has described their sub-collections in one (primary) language only: 83,5% of the HOPE metadata are unilingual descriptions.

Almost half (47,5%) of all the unilingual metadata records of the content providers is in German, at a large distance followed by English (14,2%) and Dutch (9,6%). Italian, Portuguese and Finnish each count for about 3,5% of the metadata records. French and Hungarian are the primary languages of respectively 1,6% and 0,4% of the unilingual descriptions. The KEE-OSA metadata records count almost entirely for the (unilingual) metadata records in English (KEE-OSA: 14%, KNAW-IISG: 0,2%).

Only 16,5% of the descriptions are bilingual, the vast majority being in Dutch/English (16,4%, KNAW-IISG). French/German (KNAW-IISG) and Hungarian/English (KEE-OSA) each count for less than 1% of the bilingual descriptions. KNAW-IISG bilingual descriptions are not fully bilingual though. Most of the time it concerns a translation of the (transliterated) title only, which is entered in a note field and which can be searched.

Analysis of the element sets provided by the content providers supplying bilingual metadata showed that only KEE-OSA currently indicated dedicated fields for translated values. These fields accommodate 'translated titles' and 'translated descriptions' of archival, library as well as audio-visual metadata.

Requirements from CPs:

- CPs have metadata specifying the primary language of metadata and or the digital content. The metadata should be accommodated by the HOPE aggregator.

⁴² Problems due to the presence of special characteristics and diacritics in languages are partly caused by variations in indexing, i.e. metadata entry, but some problems are related to the display of them. We therefore treat this issue in a separate section.

- Two CPs supply metadata in more than one language. KEE-OSA specifies dedicated fields for translated titles and descriptions. The aggregator should enable recording translations of (a selection of) metadata elements.

4.2.2.2 Scripts / transliterations

Only (an estimated) 1,5% of the HOPE metadata records contains metadata in a script other than Latin. KEE-OSA and UPIP-BDIC have a number of such descriptions in Russian Cyrillic. However, most of the metadata records that UPIP-BDIC will provide to HOPE, including those in Cyrillic, are not yet in its OAI repository. UPIP-BDIC thus still has to create these Cyrillic HOPE metadata records. Moreover, in its current configuration, the OAI repository does not allow for more than one instance of the title element. UPIP-BDIC intends to add the transliterated metadata, but since the OAI repository must be significantly reconfigured first no precise date has been set yet. The KEE-OSA metadata in Russian Cyrillic for HOPE, at this moment still a modest number, is expected to grow substantially.

KNAW-IISG also transliterates from Russian Cyrillic, plus from Chinese, Hebrew, Bengalese and Urdu scripts. Russian Cyrillic and Chinese are the most prominent scripts. Currently all the metadata in these non-Latin scripts are transliterated. The proportion of transliterated metadata amounts to an estimated 15-20%, but KNAW-IISG only registers the transliterated metadata. A working group will however shortly start studying the possibility to add some original elements to the transliterated ones, e.g. title and creator.

When Génériques captures the metadata that was originally in Arabic script, these have already been transliterated by the source from which the metadata is copied. Génériques does not enter the original metadata.

At the time of the survey CPs involved with transliteration of metadata only have metadata available that already has been transliterated to Latin script. However, they all consider adapting their practice providing the transliterated and non-transliterated values side by side.

As regards the use of transliteration schemes, for Russian Cyrillic the CPs concerned currently use different schemes. UPIP-BDIC and KNAW-IISG both use ISO schemes, albeit different ones: UPIP-BDIC copies the transliterated metadata from the French national catalogue SUDOC, which uses the national (French) ISO 9: 1995 standard⁴³, while KNAW-IISG manually performs all the transliterations itself thereby using the older ISO/R 9:1968⁴⁴ scheme. KNAW-IISG leaves out the diacritics on the “e” though, but keeps them on the other letters.⁴⁵ KEE-OSA uses the Russian transliteration scheme of the Library of Congress (ALA-LC Romanisation Tables: Russian), entirely without diacritics. We deal with the differences

⁴³ NF ISO 9: 1995 Information and documentation – Transliteration of Cyrillic characters into Latin characters – Slavic and non-Slavic languages.

⁴⁴ ISO/R 9:1968 International system for the transliteration of Slavic Cyrillic characters (which was revised by ISO/R 9:1986 and ISO 9: 1995).

⁴⁵ This is done because transliterating e.g. the hacek (ч) on the ‘c’ as either c with hacek (č) or c without hacek (c) results in different words (meanings). Adding the hacek thus enables a correct reverse transliteration” and translation of these words. Other diacritics by contrast, e.g. the diaeresis (¨), only effect the pronunciation, not the semantics, and therefore can be omitted in the transliteration.

between these schemes and the CPs' reasons for using them in section 4.2.

HOPE should specify best practices as to the transliteration of Russian Cyrillic, Arabic and Chinese, ensuring consistent transliteration of metadata on Russian Cyrillic, Arabic and Chinese resources.

4.2.3 Resources, Tools and Services

4.2.3.1 Introduction

A knowledge domain that is reflected in a representational vocabulary enhances the probability that the information system that has applied this vocabulary captures the right connotation when a user performs a search.

The development of a multilingual knowledge domain-specific controlled vocabulary can be approached in three different ways⁴⁶:

- 1. Starting from scratch**

This can be done either by starting with one language and adding another language or languages, or by starting with more than one language simultaneously.

- 2. Starting from/combining existing vocabularies.**

This can be done by **merging** two or more existing vocabularies into one new information retrieval language that is to be used in indexing and retrieval, or by **linking** existing controlled vocabulary languages to each other thereby using the existing controlled vocabulary languages both in indexing and retrieval.

- 3. Translating a controlled vocabulary into one or more other languages.**

Since there is no authoritative social history vocabulary we can immediately exclude the third approach. Building a new vocabulary from scratch (bottom-up) “is only viable in cases where a new information language is envisaged. Although “[t]he main advantage is that the languages involved can be treated equally” (IFLA, 2006: 1), this is –bearing in mind what is stipulated in the DoW – out of the HOPE scope.

This leaves us to the second approach, which itself consists of two sub-approaches. In any case however “in the creation of a controlled vocabulary it is necessary to link all the synonymous and nearly synonymously terms that should be brought together under a single authorized term.” (Taylor & Joudrey, 2008: 337) When a vocabulary is built by merging several existing lists in different languages into one, on the one hand, this leads to a new vocabulary that is based on the subject headings of the existing ones. “In the end, both indexers and searchers, have to use the new thesaurus resulting from the merging.”

When subject heading lists are linked, on the other hand, “the users, both indexers and searchers, can continue to use the same subject heading language as before. However, through the linking it becomes possible to search in collection A, indexed with subject heading language X, using subject heading language Y, used to index collection B. Linking can be done with monolingual subject heading languages in the same language (...) with multilingual subject heading languages, and with monolingual subject heading languages in different languages, or with any combination of subject heading languages. Linking is done

⁴⁶ Adapted from ‘Introduction’ IFLA, [Guidelines for Multilingual Thesauri](#), 2005: 1.

by mapping headings of subject heading languages. The result can be displayed as a table with as many columns as subject heading languages involved. In each row of the table the “equivalent” terms of the different subject heading languages are given. A cell of the table can contain zero, one or more terms.” (IFLA, 2006: 13)

Since HOPE does not intend to build a new vocabulary that in the end should be used by the Content Providers to index the (new) digital objects, it is the linking or mapping technique that we would use to build a social-history vocabulary. If we would proceed with this we should not necessarily restrict this mapping to the CPs local lists. Mapping the CPs’ subject headings to an existing (multilingual) controlled vocabularies would be an aid to produce these mappings (matches).

Creating mappings between subject headings requires the creation of translation dictionaries first since subject headings necessarily must be mapped on the basis of what they mean (connotation, semantics), not the way they are written (morphology). It is only when terms have been translated that equivalences between them can be established. Eight (primary HOPE) languages would produce 56 unique pairs of languages, hence an equal number of translation lexicons.

In case of proper names entered as subjects we are not dealing with synonyms or near synonyms. But names can have different forms or one and the same entity can have different names. A mapping (matching) technique would be used to collocate these different forms or names.

So far we have been talking about controlled subject headings or proper names lists (authorities). Although a large majority of the HOPE metadata for subjects and proper names uses a controlled vocabulary list, there is a substantial part that uses uncontrolled (free) proper names and subject text metadata which may contain other information within field in which the subject or proper name is entered. Subjects and proper names from these unstructured text fields should be extracted first before they could be mapped to another controlled vocabulary, be it an international one or the local CPs’ lists.

Which resources and tools/services are currently available for subjects and proper names? These would include:

- resources for structured text (metadata): (international) multilingual controlled vocabularies,
- tools/services for extracting structured data from unstructured text (metadata),
- tools/services for matching structured text to a controlled vocabulary.

For a discussion on the issue of extracting different proper names out of subject fields, we refer to the discussion paper on 2.4.1. The following sections and chapters thus only deal with subject headings/topical terms (and genre/form terms if they are not indexed in a separate field).

We shall now provide an overview of projects which have resulted in resources and tools/services which HOPE could consider making use of in the future.

4.2.3.2 Availability of controlled vocabularies and tools

From the Survey of Content Providers we learnt that for three quarters of the HOPE metadata records a published standard (controlled vocabulary) is used for subjects. That is to say: is *potentially* used, since we did not ask whether a subject field is mandatory and/or whether the information is always available.

As regards existing tools developed by other projects, unfortunately most of them are in an early stage of development. In a recent state-of-the-art overview of problems of 'multi-linguality', G. Clavel-Merrin⁴⁷ concludes that [a]lthough many exploratory projects exist and elements of solutions are available in the research and commercial sector, to our knowledge no fully operational system covering all aspects of multi-linguality is available for the digital library."

(Clavel-Merrin, 2010: 144).

Moreover, many of these projects, especially those dealing with mappings (matches) of authority files in different languages, are situated in the realm of the (digital) *library* domain. This is of course no surprise since the idea of authority control was born in the library world. Yet meanwhile the "rest of the metadata world has also begun to see a need for access points either with consistent form from record to record, or with links to resources that identify a person, corporate body, concept, and so forth, uniquely." (Taylor & Joudrey, 2008: 246).

4.2.3.3 MultiMatch project

MultiMatch ("Multilingual/Multimedia access to cultural heritage") was a research project under the Sixth Framework Programme supported by the European Commission. It started on 1 May 2006 and lasted 30 months, thus ending on 1 November 2008. Its overall objective was to "enable users to explore and interact with online C[ultural] H[eritage] content, across media types and language boundaries (...)." ([MultiMatch website](#)) It sought to do this by developing a multilingual search engine, "designed specifically for the access, organisation and personalized presentation of cultural heritage information" ([Amato et. al. 2006](#)), i.e. "a smart search system that is multilingual, multimedia and optimised for cultural heritage." ([Cordis, ICT Results](#)). CNR-ISTI was the technical partner of the project, but it was different team/research group than the one working for HOPE.

The MultiMatch translation resources are based on the WorldLingo machine translation system augmented with domain-specific dictionary resources gathered automatically from the WWW. ([Jones et. al. 2007](#)) CNR-ISTI constructed "translation lexicons" and applies them for query translation in multilingual metadata search. Unfortunately the multilingual cultural heritage vocabulary that was built is currently not being maintained let alone further developed. Since the application for the follow-up project was not approved, there is not money to do this. (Call with F. Debole, 22-12-2010)

The final version of the search engine was expected for September 2008 ([Amato et. al. 2006](#)), but currently there is only a Second Prototype available online. This prototype is "fluent in six languages": Dutch, English, German, Italian, Polish and Spanish, meaning that e.g. "a search entered in Polish can be targeted to look for results in Spanish, or English, Italian, Dutch and German" ([Cordis, ICT Results](#)). The website has not been updated since

⁴⁷ G. Clavel-Merrin, "Problems of multi-linguality", in M. Collier (ed.), [Business Planning for Digital Libraries. International Approaches](#), 2010, Leuven University Press, pp.137-146



2008 and several links in the menu are dead. The MultiMatch project does not figure on Europeana's ThoughtLab, where the results of the work of Europeana's partners is gathered.

4.2.3.4 EuropeanaConnect

EuropeanaConnect is a Best Practice Network funded by the European Commission within the area of Digital Libraries of the eContent*plus* Programme. It aims to deliver core components which are essential for the realisation of Europeana as an interoperable, multilingual and user-oriented service. EuropeanaConnect started in May 2009 and will last for 30 months. ([Website EuropeanaConnect](#)) Although several results have come out of EuropeanaConnect until now, this is still work in progress, and all software developed as part of the research project is still a prototype. Implementation by Europeana partners should have integrated into the upcoming Danube release but currently very little information is made public about this.

Language Resource Repository

EuropeanaConnect has built a "Language Resource Repository", which contains both linguistic resources (data), necessary for query translation and vocabulary mapping, and software.

One task [within its WP2] is "to establish mappings between controlled vocabularies in different languages (...); another is to set up translation modules or services for cross-lingual user queries. Such services rely on language resources: software and information for processing digital representations of human language in various ways. It is not in the scope of EuropeanaConnect to build such resources, but to collect and assess available resources and to adapt and maintain them as necessary.

The set of 10 languages that should be supported within EuropeanaConnect is divided into two groups:

- Core Languages: English, French, German, Italian, Polish, Spanish
 - Secondary Languages: Dutch, Hungarian, Portuguese, Swedish
- (...)

The repository currently contains the following types of resources which are necessary for query translation and vocabulary mapping

- Stop word lists: lists of "non-content" words, such as articles, conjunctions, prepositions, which can be ignored for specific tasks of processing, especially for indexing, retrieval or even translation. This resource will be mainly used by the indexer.
- Language identifiers: a tool that is necessary whenever the language of the query is not explicitly known. It will be used by the indexer and by the query translation module.
- Morphological analyzers: software modules that perform tokenization and lemmatization, but also decompounding, multi-word detection and part of speech tagging

- Named entity recognizers: software modules that identify named entities, such as person names, geographic names, organisation names, etc.
- Translation dictionaries: mappings between terms in different languages.

(...)

More categories can be added, for example, if at a later stage of the project other types of language resources turn out to be necessary or useful for new functionalities.”

Because of IPR issues access to the [repository](#) is currently restricted: it is “a password-protected subversion repository that can only be accessed by Europeana partners and Europeana developers (EuropeanaLabs). “Once these (...) issues are resolved, the repository will be split into an open-source section that is publicly accessible, and a proprietary section that is accessible only to Europeana developers.”

The interface for using resources from the Europeana Linguistic Resource Repository consists of the following software components:

“LanguageIdentifier	LanguageIdentifier interface is the core interface for the language identification
MorphoAnalyzer	A service that performs morphological analysis, which may cover any of: segmentation, stemming/lemmatization, and part of speech tagging.
MorphoAndNamedEntity	A component that does the jobs of a MorphoAnalyzer and a NamedEntityRecognizer simultaneously.
NamedEntityRecognizer	A service that performs named entity extraction from text.
TranslationDictionary	Provides lookup in one or more bilingual dictionaries.”

A full list of resources with detailed information is available to the public via the [resource register](#) (but a password is required to download them). “The initial motivation for collecting and adapting linguistic resources was to support query translation and vocabulary mapping work within the EuropeanaConnect project, but the resources are potentially of interest to developers outside of Europeana as well. We therefore provide a web interface that outsiders can use to see what resources have been selected for use in Europeana and adapted to the standard Europeana APIs, the licensing conditions under which the resources are available, and contact addresses for the rights holders.”

(Source: EuropeanaConnect, D2.2.1 – [Europeana Language Resources Repository](#), v1.0, 31-07-2010, pp. 5-7, 19, 23)

Semantic Data Layer

EuropeanaConnect has equally built a “**First Operational Version**” of the “Semantic Data Layer” which “aims at providing a uniform, machine-actionable, web-enabled access to the reference knowledge capitalized by the various stakeholders of Europeana.eu. This

knowledge mostly comes in the form of controlled vocabularies: thesauri, subject heading lists, classification schemes, authority lists for person names and place names, etc. Those are consistently used in the metadata describing the objects ingested in Europeana.eu. The objective is to make this knowledge available for enabling the Europeana.eu users to benefit from semantics-intensive functions, as will be specified in other WP1 deliverables. (...) To match the aforementioned objectives, the controlled vocabularies that form the Semantic Layer have been converted to the RDF format, using the SKOS model. This allows a uniform representation of the concepts present in the vocabulary; it also paves the way for semantically aligning those concepts, as will be done in WP1.2.” (EuropeanaConnect, D1.1.2 – [Semantic Data Layer First Operational Version](#), v1.0, 16-08-2010, p.5)

When this report was published (August 2010), the Semantic Layer contained SKOS/RDF data for 19 vocabularies used by (national) libraries, archives, audio-visual archives and museums in Europe, e.g. RAMEAU (French National Library; includes common nouns, names of organizations, persons and places, time periods, and titles of works), the OSZK thesauri (National Library of Hungary; includes names of places, and subjects) and the thesaurus used at the Austrian media library, the Getty AAT (Art and Architecture Thesaurus) and the Dutch version thereof (Dutch AAT), the vocabularies created by the Italian Ministry for Cultural Heritage (ICCD) used in the catalogue of the *Fondazione Federico Zeri* (includes subjects, materials and genres of objects, a classification for photographic techniques, and a basic classification for photographs). For the complete list of vocabularies, see D1.1.2 – [Semantic Data Layer First Operational Version](#), pp.9-10.)

Unfortunately, the link to the thesaurus mentioned in the deliverable is not working anymore.

The second version of the Semantic Data Layer should be released in 2011.

EuropeanaConnect Gazetteer and EuropeanaConnect Geoparser

EuropeanaConnect has developed two tools (prototypes) for enriching metadata: the “Gazetteer” and the “Geoparser” (see [Europeana ThoughtLab: Enriching Metadata](#)).

The EuropeanaConnect **Gazetteer** is a Geographical Information Service (GIS) which “will give service providers access to over 9 million geographic names, co-ordinates, and boundaries. By enriching Europeana's metadata with these geographic references, it will be possible to identify features such as continents, countries, cities, monuments and rivers contained in the objects on Europeana. (...)The information in the EuropeanaConnect Gazetteer has been collected from free data sources, which means there are no legal constraints to its use and re-use. Information from additional data sources is also continuously integrated, ensuring the service is kept up to date.”

The EuropeanaConnect **Geoparser**, which works with the EuropeanaConnect Gazetteer, automatically extracts structured information about places and historical periods from unstructured and semi-structured text. Users can provide textual sentences or metadata records, and the geoparsed results can be returned as XML or HTML.

See the [EuropeanaConnect Gazetteer Prototype](#) and the [EuropeanaConnect Geoparser Prototype](#).

4.2.3.5 Hitime

The DoW specifies that (...) the “KNAW-IISG is currently involved in an ICT-research project with the University of Tilburg (CATCH project HITIME) which aims to create an open source toolkit for historical text mining. Through mining the domain semantics from digital text sources, a historical web of entities (such as names of persons, organisations, events, occupations, etc.) could be developing and would create interesting synergies. (...)HOPE (...) will not invest resources in research activities in the framework of this project, but liaison between Hitime, Europeana and CNR-ISTI will be encouraged and might lead to interesting practical implementation opportunities.”

4.2.3.6 Multilingual vocabularies for subjects / topical terms

“Digital library resources and their metadata will increasingly co-exist with traditional resources and associated metadata”, S. Gradmann (Europeana WP3 Leader) remarks. Therefore “when referring to authority data for named entities such as persons, corporate bodies or for concept resources digital library metadata should point to the same resources as do traditional cataloguing data.” For, if this is not the case this could lead to “inconsistent usage scenarios in the future”. (Gradmann, 2010: 182) He then suggests that these authority files could be VIAF (Virtual International Authority File) for personal names and LCSH (Library of Congress Subject Headings) for concepts. MACS, TELplus and CrissCross and are (were) project which link the topical terms (headings) of LCSH, RAMEAU (*Répertoire d'autorité-matière encyclopédique et alphabétique unifié*) and SWD (*Schlagwortnormdatei*). As regards geographical names the ISO 3166-1 standard and the MARC Code List for Countries are international standards that are widely used for traditional resources. Merriam-Webster's Geographical Dictionary and GeoNames are more recent sources (databases) for place names.

HOPE subject headings

If it would be unfeasible to create a HOPE controlled vocabulary for subjects by starting from existing (multilingual) vocabularies, we could still consider constructing these by merging the local Content Providers' lists into one common list for each of the entities.

In order to map (match) subject terms, however, translation dictionaries would need to be created. We could consider using Google Translate (see section 4.2.1. of this paper) as an aid for performing the basic translations and thus as a start for integrating the lists. Although names do not need to be ‘translated’ different forms and names of the same entity would need to be collocated, thus creating a true authority list. All this would still (and perhaps even more) be a very time-consuming enterprise.

Relating local lists to an existing controlled vocabulary/authority would thus be preferable, in any case as a first step.

Published vocabularies for subject headings

Library of Congress Subject Headings (LCSH)

Although the LCSH is a list of terms (controlled vocabulary for subjects) that is created and used by the Library of Congress, the national library of the United States, in the meantime it has become a de facto standard that is used worldwide. According to Taylor and Joudrey it can be used by any agency that wants to provide controlled subject access to metadata records. (2008: 462) The latest (32nd) edition (2010) contains more than 317,000 total

subject headings and references.

It is however not made available free of charge. A subscription to the paper version of the latest edition (currently 6 volumes) costs \$ 345 (€ 260,50) outside North America. A [subscription to the electronic version](#) is possible through the LC Classification Web (annual subscription solo user: \$ 325 (€ 245,40), 1 to 4 concurrent users: \$ 525 (€ 396,50), etc.)

The vocabulary has been published as linked data (serialized as RDF/XML or N-triples) and is since 5 January 2011 available free of charge for bulk [download](#) from the Library of Congress Authorities and Vocabularies service.

MACS

[MACS](#) or “Enabling large-scale Multilingual Access to Subjects” established equivalence links between the topical terms of the LCSH and the subject headings authority files of the German ([SWD](#)) and French (RAMEAU) national libraries. The end result is reportedly is “neither a translation nor a new thesaurus but a mapping of existing and widely used indexing languages”. ([About MACS](#)) According to G. Clavel-Merrin (member of MACS team) “the MACS project concentrated on mapping until 2007 when work began anew on a search interface” (2010: 141). To our knowledge there are two projects which have used the MACS data to create links between the concepts of the aforementioned authorities: TELplus and CrissCross.

TELplus

TELplus was a project funded by the European Commission under the eContent*plus* Programme, within the domain of ‘Digital Libraries’. The overall goal of TELplus was to strengthen, expand and improve The European Library. (...) [which] has been officially acknowledged by the European Commission as the organisational structure for the creation of Europeana”. ([About TELplus](#))

TELplus tested mapping and alignment of vocabularies in SKOS using MACS data. A service has been set up that uses SKOS to provide [RAMEAU subject headings as SKOS linked data](#). The RAMEAU conversion to SKOS was produced by Antoine Isaac, currently Scientific Coordinator at Europeana, and who is also involved in the development of the Europeana Data Model (EDM) and related semantic issues (Europeana WP3). The Library of Congress subject headings include links to associated RAMEAU subject headings that are available through this service. However the available concepts are only a small part of the RAMEAU vocabulary and the URIs may not always resolve. Phrased differently: also this service is still in an experimental stage of development. ([About Library of Congress Authorities and Vocabularies](#))

An updated version of the MACS data with links to the German SWD published as linked data is also available. This leads us to the CrissCross project of the German national library, and its Linked Data Service.

CrissCross

[CrissCross](#) was a joint project German National Library and the Cologne University of Applied Sciences, which ran from 2006 to (June) 2010. The topical terms of the German subject headings authority file (SWD) were mapped to the German Dewey Decimal Classification (“DDC Deutsch”). In addition equivalent subject headings in the SWD, the LCSH and RAMEAU were linked, based on the approaches of the MACS project. The result is a thesaurus-based

research vocabulary published as [linked data](#) which can be used (for further research) free of charge.

The current data pool currently not only holds (about 188,000) subject headings from the SWD but also (about 1.8 million) differentiated persons from the German personal name authority file (PND) and (about 1.3 million) names of corporate bodies from the German corporate body authority File(GKD). In addition 51,748 classes and 110 subject headings of the German DDC are comprised. It not only includes links from the SWD subject headings to the corresponding equivalents in the LCSH, RAMEAU and German DDC but also links from the PND to the Virtual International Authority File (VIAF) (see below). ([DNB](#), 2010) An [FTP download](#) of the entire data pool (455 MB ZIP file, date 28-04-2010) is available [online](#).

4.2.3.7 Conclusion

In this chapter we discussed the following controlled vocabularies and tools/services:

	Resources	Extracting and other tools
Subjects/ topical terms	Europeana Language Resources Repository (EuropeanaConnect) Semantic Data Layer (EuropeanaConnect) TranslationDictionary (EuropeanaConnect) Library of Congress Subject Headings MACS/TELplus/CrissCross	EuropeanaConnect MorphoAnalyzer
Proper names		EuropeanaConnect NamedEntityRecognizer Europeana Geoparser EuropeanaConnect Geoparser

As regards these controlled vocabularies, each has as specific constellation of advantages and disadvantages:

Subjects/topical terms	Advantages	Disadvantages
Europeana Language Resources Repository (EuropeanaConnect)	<ul style="list-style-type: none"> ▪ meant for enabling cross-lingual searching ▪ multilingual data (core languages: en, fr, de, lt, pl, es; secondary languages: nl, hu, pt, sv) ▪ ... 	<ul style="list-style-type: none"> ▪ early development stage ▪ <i>access?</i> ▪ <i>not used in traditional (cataloguing) data</i> ▪ ...

Semantic Data Layer (EuropeanaConnect)		
	<ul style="list-style-type: none"> ▪ meant for enabling cross-lingual searching ▪ broad scope: includes 19 vocabularies used by (national) libraries, archives, audio-visual archives and museums in Europe ▪ multilingual data (<i>which languages?</i>) ▪ SKOS/RDF data ▪ ... 	<ul style="list-style-type: none"> ▪ early development stage (“first operational version”) although second version should be released soon ▪ not many vocabularies included are currently used by HOPE CPs ▪ <i>possibility to use?</i> ▪ <i>If yes: conditions?</i> ▪ ...
Library of Congress Subject Headings (LCSH)	<ul style="list-style-type: none"> ▪ well established international standard ▪ also used for traditional (library) resources ▪ available in electronic form ▪ available in RDF/XML or N-triples ▪ ... 	<ul style="list-style-type: none"> ▪ CH domain-wise: limited scope: mainly used for library resources ▪ <i>subject-wise: broad but perhaps too general for HOPE?</i> ▪ unilingual (en) and single script ▪ not available free of charge (annual licence fee) → <i>still true?</i> ▪ ...
MACS/TELPlus/ CrissCross	<ul style="list-style-type: none"> ▪ links subject headings of three important authority files (LCHS, SWD, RAMEAU) ▪ multilingual (en, fr, de) ▪ entire data pool (ZIP file) freely available online (CrissCross) ▪ ... 	<ul style="list-style-type: none"> ▪ (very) early / experimental development stage ▪ not very CH domain-specific: built on authority data for library resources ▪ ...

A first conclusion after studying the available multilingual tools and resources, is that none of the projects that have developed multilingual resources and tools so far has implemented something that is fully operable. Developing applications that enable true multilingual searching requires a huge human effort. Although costs are difficult to estimate, it is clear that manual mapping is very time-consuming and costly. E.g. For the CrissCross project 4 person years were spent to link the headings of the German SWD the 50,000 LCSH headings, and 3 additional person years to link 70,500 DDC notations. (Clavel-Merrin, 2010: 140) Once resources and tools have been built, the tools are made available free of charge but the resources often are not. Yet

another major problem seems to be the difficulty to continue to maintain and develop these applications and resources (because of lack of money?, because the last development stage is the hardest?). EuropeanaConnect looks the most promising project in that it has developed the most advanced applications in 10 languages that cover 7 of the current HOPE primary languages (Finnish being the only exception). On the other hand not many of the resources included in EuropeanaConnect's Semantic Data Layer are currently used by the HOPE CPs.

4.2.4 Recording Multilingual HOPE Metadata

This chapter deals with the requirements for the HOPE aggregator for recording multilingual metadata, as well as the criteria for multilingual metadata to be accepted by the HOPE aggregator.

4.2.4.1 Translation issues

This paragraph deals with translations of descriptive metadata, and metadata about the language of the metadata and the digital content.

Language Elements

The HOPE Data Model contains dedicated metadata elements for recording the language of the metadata, as well as the digital content. These elements record the information supplied by the Content Provider as a string, but records a normalised language code in the normalised attribute of the element. HOPE uses the ISO-639-3 data value standard for language codes.

The language elements specified in the HOPE Data Model are:

- **Europeana language (1,1)**

This element indicates one or more official languages, associated with the country the content provider is located. The element is mandatory, since it is required by Europeana.

- **Language Metadata (1,1)**

This element indicates the primary language, in which the metadata describing the original resources has been drafted. This element is mandatory and can have only one occurrence. Secondary languages, used for translation of specific values, are recorded using the language attribute.

- **Language Digital Content (0,*)**

This element records one or more languages used in or by the original resource, represented by the digital object. This element is optional, since not all HOPE resources use language.

- **Language (0,*)**

This element records metadata on languages, as part of the descriptive metadata about the original resource. The semantics of this element correspond with the Language Digital Content element, providing it is a free-text field that enables recording any legacy data about the language of the original resource, supplied by the content provider.

Depending on the domain profile, used by the content provider mapping metadata to the HOPE aggregator, the labels for the language element may change.

- Archival metadata: language of the described material
- Audio-visual metadata: original language, language used, sub-title language

- **Language Attribute (0,1)**

The language attribute belongs to a set of general attributes that are by default available for each metadata element in the HOPE data model. This attribute is used to indicate the language of a value which is not drafted in the primary language specified in the Language Metadata element. According to the HOPE Content Providers Survey, there are currently two content providers supplying translations of titles and translations of descriptions (see 4.1.3). Anticipating the supply of translations for other elements, the language attribute will be available for all descriptive metadata elements.

Please note, that some domain profiles already contain particular elements for translations of descriptive metadata, since they are part of the encoding and cataloguing standards used by the domain profiles

- Library metadata: parallel title, translated title, translated sub-title
- Audio-visual metadata: translated title, translated synopsis

HOPE domain profiles: metadata labels

Given the overwhelming presence of unilingual descriptions, we asked ourselves whether it would be useful to have translations for the labels of the HOPE metadata. For this might enable synchronising the language of the labels with the language of the metadata, which greatly enhances the consistency of the display of the HOPE metadata.

Content Providers map their metadata to one of the five domain profiles, which contain about 250 terms, currently only available in English terms. Eight primary non-English languages are currently used to record the HOPE metadata. Translating the metadata labels of the source language (English) to all the target languages would thus produce seven bilingual lexicons of about 250 terms. This seems feasible, all the more because these labels are taken from international standards of which (albeit not always official) translations in various languages exist.

It might be useful to have them also translated to other important languages (i.e. spoken by a large number of people) which are currently not present (yet), e.g. Spanish or Russian.

HOPE metadata

In addition to the labels, it may be useful and/or feasible to provide translations for the values contained in a selection of metadata elements, such as those taken from controlled lists and some free-text elements for which we noted that CPs provide translated values.

We know that at least content providers (KEE-OSA, KNAW-IISG, Génériques) translate the original metadata that are not in their national language into the language of the catalogue or into English. Some of them enter the translated and the original text, others only enter the translated text. Within the scope of the Content Providers Survey, it was not feasible to look systematically into the actual translation practices: how are (which?) metadata fields in

other languages translated in the (primary) language of the catalogue? Some CPs have given us an idea of what they do, but we have no systematic overview.

In the HOPE aggregator translated values are recorded as additional occurrences, next to the original value. The distinction between translated and original value is made by the domain-specific label for the value. The language attribute indicates the language of the translated and the original values.

Based on a survey of the element sets used by the content providers, HOPE will enable mapping translated values for following elements:

- **Titles**

Translations of titles are often parallel or translated titles of library or audio-visual resources, mostly translation of titles in the language of the catalogue/institution. These translations are done manually and supplied next to the title proper in the original language.

- **Description, abstract, summary, notes fields (free text fields)**

Given that automatic translation of controlled text (e.g. thesaurus terms) is problematic, this is even more true for uncontrolled (free-text) metadata. This would put a heavy burden on the CPs, something we cannot ask, at least not in the short run.

We could however consider the suggestion made Clavel-Merrin (2010: 144) to have a part of the 'results' translated when it is displayed, i.e. on the fly. In case of HOPE this would not be the (textual) digital object but some metadata, e.g. a description, abstract, summary, or notes field, if there would be any. If the goal would only be to offer the user the means to decide whether the object retrieved corresponds to his or her wishes, offering an online translation service/tool such as Google Translate to the user might be helpful. Since the Google translation software does not apply grammar rules nor human-like skills but a statistical analysis, translations are not perfect and even not always completely accurate. But it would be a pragmatic solution for a complex issue for which, given the current state of information-technological development, there exists no satisfactory solution yet.

- **HOPE Historical Theme**

Controlled vocabulary, crafted by HOPE T1.6 with a list of historical themes that are relevant to the field of social history. This vocabulary will be multilingual.

4.2.4.2 Transliteration issues

This paragraph deals with transliteration of descriptive metadata, and metadata about the scripts and alphabets used to record the metadata.

Apart from enabling multilingual access to the HOPE resource, transliteration is required to enable indexing the metadata. For indexed fields Content Providers should perform transliterations of metadata values which are in a non-Latin script/alphabet to the Latin script.

The transliterated value is recorded as an additional occurrence of the metadata element.

The HOPE data model does not provide specific labels to distinguish between the proper and transliterated element, but allows an attribute indicating the script of alphabet of the proper and/or transliterated value.

It is however not necessary to have all the metadata transliterated. We would suggest as a best practice to have the name(s) of the creator(s) and the (main) title transliterated.

Whether the transliteration is done manually or (semi-)automatically is up to the Content Provider but it is Best Practice to use a standard transliteration scheme.

Russian Cyrillic

Russian Cyrillic is the main non-primary language/script that is transliterated to Latin script or captured as transliterated metadata by KNAW-IISG, KEE-OSA, UPIP-BDIC and Génériques. Among these Content Providers, three transliteration schemes are currently used for transliterating Russian Cyrillic to Latin script:

- **ISO-R9:1968** permits some deviations from the main standard (in countries where tradition favours it, a set of alternative transliterations, but only as a group).
KNAW-IISG chose the ISO 9:1968 transliteration scheme a long time ago because this was then used by the major Dutch libraries and universities. KNAW-IISG thus wanted to adhere to the same standard.
- **ISO-R9:1995** The edition of 1995 cancels and replaces the edition of 1986. The latest version emphasizes the unambiguity of the transliteration instead of the phonemic representation. This version ensures:
 - A universal 1:1 mapping of letters. Univocal, as one character is represented by one equivalent character.
 - Reliance on multiple diacritics.
 - Represents the original spelling and allows for reverse transliteration (or retroconversion).
 - Language-independent.

Since UPIP-BDIC (and Génériques) copy the transliterated metadata from the French Union Catalogue, they do not make an active choice for the ISO scheme that SUDOC is using. However, they will have other metadata to transliterate for which a scheme still must be chosen.

- **ALA-LC** Used by North American libraries and the British Library, as well as in publications throughout the English-speaking world. Often used without the additional diacritics (the scheme used by OSA) – Reverse transliteration is not considered.

The latest version was published by the American Library Association & Library of

Congress in 1997. ALA-LC Romanization tables:

<http://www.loc.gov/catdir/cpsol/roman.html>

KEE-OSA opted for the Library of Congress scheme.

T2.4 did not establish agreement on which transliteration scheme for Russian-Cyrillic is best practice for the HOPE BPN. However, there is an 'window of opportunity' for establishing agreement, since the CPs involved are currently reconsidering their local practice as to transliteration.

T2.4 suggests considering following issues when selecting a transliteration scheme.

- Possibility for retro-conversion
- Lack of ambiguity
- Reliance on diacritics
- Widespread use

Following table includes a comparison of the three transliteration schema used by HOPE Content Providers when transliteration Russian-Cyrillic to Latin.

Cyrillic Upper case/Lower case letters		ISO-R9:1968		ISO-R9:1995		ALA-LC with diacritics		ALA-LC without diacritics	
А	а	A	a	A	a	A	a	A	a
Б	б	B	b	B	b	B	b	B	b
В	в	V	v	V	v	V	v	V	v
Г	г	G	g	G	g	G	g	G	g
Д	д	D	d	D	d	D	d	D	d
Е	е	E	e	E	e	E	e	E	e
Ё	ё	Ё	ё	Ё	ё	Ё	ё	E	e
Ж	ж	Zh/Ž	zh/ž	Ž	ž	Zh	zh	Zh	zh
З	з	Z	z	Z	z	Z	z	Z	z
И	и	I	i	I	i	I	i	I	i
Й	й	J	j	J	j	ĭ	ĭ	I	i
К	к	K	k	K	k	K	k	K	k
Л	л	L	l	L	l	L	l	L	l
М	м	M	m	M	m	M	m	M	m
Н	н	N	n	N	n	N	n	N	n
О	о	O	o	O	o	O	o	O	o
П	п	P	p	P	p	P	p	P	p

Р	р	Р	р	R	r	R	r	R	r
С	с	С	с	S	s	S	s	S	s
Т	т	Т	т	T	t	T	t	T	t
У	у	У	у	U	u	U	u	U	u
Ф	ф	Ф	ф	F	f	F	f	F	f
Х	х	Ch/H	ch/h	H	h	Kh	kh	Kh	kh
Ц	ц	Ts/C	ts/c	C	c	TS with ligature	ts with ligature	TS	ts
Ч	ч	Č	č	Č	č	Ch	ch	Ch	ch
Ш	ш	Sh/Š	sh/š	Š	š	Sh	sh	Sh	sh
Щ	щ	Shch/Ṣ̌	shch/ṣ̌	Ṣ̌	ṣ̌	Shch	shch	Shch	shch
Ъ	ъ		“		“		“		“
Ы	ы	Y	y	Y	y	Y	y	Y	y
Ь	ь		’		’		’		’
Э	э	È	è	È	è	É	é	E	e
	Ю	Û	û	Û	û	IU with ligature	iu with ligature	IU	iu
Я	я	Â	â	Â	â	IA with ligature	ia with ligature	IA	ia

**Modern alphabet. Archaic (pre-revolutionary) letters not included.*

When setting best practice, the issue should be treated more broadly. Beyond an analysis of the merits of various transliteration schemes, the practicality / benefits of transliteration v. original alphabet cataloguing should be considered in the context of institutional policies, professional practice, and intended end use. It is important to keep in mind the following issues:

- Internal context:
 - in-house language / alphabet policies,
 - professional standards followed (specifically, library and archival content and structural standards and authorities),
 - copy cataloguing practices,
 - technical environment (e.g. support of character sets)
- External context:
 - language / transliteration practices for external material potentially presented and searched with the material

- language / transliteration practice followed by national institutions or similar profile institutions
- union catalogues, aggregator, portal rules and guidelines
- End user context:
 - typical research and search patterns
 - technical environment

Arabic

As regards the **Arabic** script Génériques enters transliterated Arabic metadata, which were romanised according to the ISO 233-2:1993 schema.⁴⁸ We would therefore suggest this scheme to be used as best practice.

Chinese

KNAW-IISG is transliterating Chinese characters to Latin script using the Pinyin Scheme, i.e. ISO 7098:1982 – Documentation – Romanisation of Chinese. We would therefore suggest this scheme to be used as best practice.

Best Practice as to recording original metadata

Next there is the question of original metadata entry. Currently KEE-OSA sometimes enters the original metadata, next to the transliterated one, and KNAW-IISG will shortly be discussing the pros and cons about adding the original metadata wholly or partially.

We would suggest as a best practice to index, next to the metadata that is transliterated, at least the name(s) of the creator(s) and the (main) titles that were originally in a non-Latin script. And as good practice we would suggest to enter other original metadata, if they would be present, which relate to the content of the object, notably the data that are registered in the description, summary, abstract, note, ... fields (the name varies according to the domain profile). If this metadata has already been transliterated. If not, we would leave it to the CP to decide whether to perform these additional transliterations and register the transliterated values.

4.2.4.3 Issues related to the use of special characteristics and diacritics

The HOPE content providers must be able to export all descriptive metadata records encoded in UNICODE UTF-8. "In UTF-8, each character is encoded on 1 to 4 bytes according a specific algorithm. ASCII characters need 1 byte, all other European characters (including non-Latin scripts) and Middle-Eastern scripts need 2 bytes, and 3-4 bytes are needed for Asian scripts and some special characters such as mathematical symbols. This encoding scheme is preferred where space or transmission speed is an issue and is therefore widespread on the web." (Clavel, 2006)

At first sight this does not seem to be a problem for the content providers. 8 CPs, representing 70,1% of the metadata, already use UNICODE UTF-8 only, and the other CPs either use UTF-8 and ISO 8859-1 (2 CPs, 26% metadata) or ISO 8859-1 only (3 CPs, 3,8%

⁴⁸ ISO 233-2:1993 Information and documentation – Transliteration of Arabic characters into Latin characters – Part 2: Arabic language – Simplified transliteration



metadata). ISO 8859-1 can easily be converted to UTF-8 since ISO 8859-1 only contains characters from Western European languages. Nearly all Western European languages have special characters or use diacritics. English is the only exception.

However, even in an "all-Unicode world" there are still obstacles related to the use of diacritics. All languages except English have special characteristics or use diacritics. This might hinder the interoperability of metadata and thus the accessibility of resources due to the application of different indexing rules and the use of different keyboards (data input), and divergences in the alphabetic sorting of search results (display). Clavel (2006) and Clavel-Merrin (2010) both suggest some solutions.

Problems

Data input

Difficulties may arise because of different indexing rules when the user cross-searches multiple collections containing objects that were indexed in different languages/scripts (metadata) i.e. according to different indexing rules, or text objects which themselves contain various languages/scripts. And difficulties may come about because of the use of different keyboards, i.e. when the user uses a different keyboard than the one that was used for indexing the resources. For, keyboards tend to be "localised" according to the language/script of the country in which they are used, to make them more user-friendly, which is however not necessarily the language/script of the material indexed.

Display

Alphabetical sorting of search results is also linked to the use of a language/script. In some languages/scripts special characteristics are placed after the 'z', not before. The user who is not accustomed to this then risks to miss resources because he or she does expect to find them somewhere else in the set of results.

Solutions

Clavel (2006) sees three solutions for the problems caused by variations in indexing:

Provide users with means to type terms as precisely as possible:

- increase users' awareness that special characters may be important by offering them a (simple) list of the special characters that are individually indexed,
- add a virtual keyboard on the search page, so that users can type special characters that are not present on their actual keyboard.

Use enriched authority records: "[i]f the most important names and terms containing a special character individually indexed according to one's locale but not elsewhere have see-references with an 'unlocalized' form, all users will have a better chance to reach relevant records, i.e. even those unaware of a locale particularity."

Duplicate indexes and have one set localised and one set unlocalised, i.e. "where all diacritics are ignored and all special characters converted to ordinary letters. A different Z39.50 port destined to international access and portals would then point to the unlocalized indexes. Notice that this would exchange silence against noise, since names or words

differentiated within some locales would then be indexed the same way. For this reason, it should be made optional, allowing users to select the access best suited to their linguistic knowledge and the characteristics of their search."

When speaking for The European Library (TEL) Clavel (2006) added that providing users with means to type terms as precisely as possible (the first solution mentioned) "is technically bit feasible today" but he is pleased to see "that several TEL-partners apply the second solution [use enriched authority records] and complete a selection of their authority records with unlocalized forms."

As regards the **problems related to the alphabetical sorting of search results**, "these can be solved by making [the sorting] centrally on the portal, on which it is not difficult to have sorting modules for all European locales. One could let the user choose the locale she wants to apply or, more simply, derive it from the chosen interface language. It would be necessary to sort again ordered sets returned by some systems, in order to apply the expected locale." (Clavel, 2006)

Allocate time to test data access and display

According to Clavel-Merrin "[w]hen planning for access to a multi-lingual digital collection, it is essential to allocate time to test data access and display". The time required for testing a system in a multilingual environment "will depend on the complexity of languages and scripts present". Although the author is talking about digital objects ("a collection with documents in multiple scripts will require more testing, and staff expertise"), this is also true for metadata in multiple scripts. (Clavel-Merrin, 2010: 138)

HOPE

Clavel's and Clavel-Merrin's analyses are partly relevant for HOPE. The only thing HOPE could do is signal the potential display problem (alphabetical sorting) discussed by Clavel to the developers of the Europeana and the IALHI portals. If there would be a problem at all. This would have to be studied first.

The problems that are related to variations in indexing are real potential problems for HOPE but the solutions suggested by Clavel would put a heavy burden on the Content Providers. In the short run it does not seem feasible to impose such solutions as best practices for all CPs.

Taking the time for testing a system is something we should in any case do, not only with regards to these issues but in general.

4.2.5 Enrichment: Building Multilingual HOPE Metadata

Given the exploratory nature of most projects, if one is planning, Clavel-Merrin advises organisations which are planning to “to introduce elements of multi-linguality into a system” to adopt a pragmatic approach and to build incrementally (2010: 144). This is exactly what we are proposing for HOPE.

Labels

In the very short run is feasible and very useful to translate the English metadata labels of the five domain profiles in the seven other primary HOPE languages (see section 4.1) so that, even if the retrieved metadata values are not translated, the user understands what type of information the field contains. This is the easiest thing to do in the short term.

Linked Data

It is also important not to miss the “Linked Data” boat.⁴⁹ Although results in this area are still very preliminary it should be a best practice to prepare the HOPE metadata in such a way that it is compatible with this “recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF” (linkeddata.org). This might involve recording persistent identifiers borrowed from published standard vocabularies for agents, places and concepts, and considering the implementation of SKOS elements to record metadata on concepts. Vocabularies endorsed within the Europeana Community include mainly VIAF and geonames.org.

Online Translation Tools

As regards the display of results on the IALHI Labour History Portal we should explore applying a tool which enables searching in one language and presenting a (on-the-fly) translation of the results in (all the) other (HOPE) languages. This is far from ideal, but for the time being is as far as we can get. Google Translate currently is the most developed online translation tool.

Transliterations

As regards transliterations we propose to apply the following schemes as best practice:

⁴⁹ Wikipedia defines cites Tim Berners-Lee who outlined four principles of Linked Data:

1. Use URIs to identify things.
2. Use HTTP URIs so that these things can be referred to and looked up (“dereferenced”) by people and user agents.
3. Provide useful information about the thing when its URI is dereferenced, using standard formats such as RDF/XML.
4. Include links to other, related URIs in the exposed data to improve discovery of other related information on the Web.”

(...) Note that although the second rule mentions “standard formats”, it does not require any specific standard, such as RDF/XML.” http://en.wikipedia.org/wiki/Linked_Data; accessed at 06-01-2011

- for Russian Cyrillic: to be agreed on: LC-ALA; ISO-R9:1968; ISO-R9:1995
- for Arabic: ISO 233-2:1993 Information and documentation – Transliteration of Arabic characters into Latin characters – Part 2: Arabic language – Simplified transliteration
- for Chinese: ISO 7098:1982

Multilingual vocabularies

Unfortunately there is currently at best only a beginning of a solution for HOPE in terms of multilingual controlled vocabularies and tools that could be used as a starting point for mapping the CPs' local lists of subject headings and making them searchable in multiple languages. We cannot but conclude that, given the current stage of knowledge and experience (and our limited knowledge about it) it is not feasible to create a multilingual social history controlled vocabulary within the scope of the current HOPE project. Whether or not to initiate such a project requires the development of a business case both assessing (1) the technical feasibility and (2) the time and money it would cost.

This business case should include following components:

- create a additional module in the HOPE System for developing multilingual vocabularies:
gathering the expertise and software components and ensuring the long-term sustainability of the software used.
- develop a user-friendly shell:
develop a user-friendly interface for the creation and maintenance of multilingual vocabularies by the HOPE Content Providers
- ensure the 'semantic' maintenance of the multilingual vocabulary:
ensure the commitment of the HOPE/IALHI network in sustaining the quality of the social-history vocabulary in the long run, including data maintenance, cleaning, software updates, availability of time and human resources.

Persons & Places

In addition we propose to have a number of CPs to try to match a part of their proper names metadata (values) with the standard controlled vocabularies. We should however bear in mind that there shall never be a 100% match. The proportion of matches is likely to be (much) lower. The proper names that do not match will thus have to be collocated in another way. This test will allow us to assess the added value of associating HOPE terms with Linked Data resources.

As regards producing mapping for persons and organisations we tend to propose VIAF (and ULAN for visual collections?). For geographic places we propose geonames.org. It is expected that both vocabularies will also be associated with the Europeana Portal.

HOPE Historical Themes

Some start could also be made by T1.6, which will define a list of historical themes in the social history domain. It is proposed to have the list (translated in) all the current eight primary HOPE languages. Although these lists will be no authorities using “keywords” or “terms” but classification lists containing maximum 60 terms, they do cover the content of the HOPE collections. The Content Providers will be able to link these themes to the (higher level and the lower level) descriptions of the digital objects, in a way that they will be searchable by the user.

Named Entity Extration

Regarding the creation of multilingual access to 'named entities', such as persons and places, it is proposed to look into the possibilities for implementing 'named entity extraction' tools. In the short time, the implementation of tools for extracting named entities from full-text resources, and linking them with multilingual vocabularies, is considered the easiest way of creating multilingual access to the HOPE metadata. However, also here a business case should clarify whether or not to initiate such a project within the scope of HOPE. This thus equally requires the development of a business case both assessing (1) the technical feasibility and (2) the time and money it would cost.

BIBLIOGRAPHY

All online publications were accessed in December 2010 and January 2011

Amato, G., Cigarrán, J., Gonzalo, J., Peters, C., and Savino, P., [MultiMatch – Multilingual Multilingual/Multimedia Access to Cultural Heritage](#), Fifth International Workshop on Content-Based Multimedia Indexing (CBMI-2007), 25-27 June 2007, Bordeaux, France.

Clavel, P., [How localization challenges international portals : character sets and international access](#), World Library and Information Congress: 72nd IFLA General Conference and Congress, 20-24 February 2006, Seoul, Korea.

Clavel-Merrin, G. ,[Problems of multi-linguality](#), in M. Collier (ed.), [Business Planning for Digital Libraries. International Approaches](#), Leuven, Leuven University Press, 2010, pp. 137-146.

EuropeanaConnect, D2.2.1 – [Europeana Language Resources Repository](#), v1.0, 31-07-2010.

EuropeanaConnect, D1.1.2 – [Semantic Data Layer First Operational Version](#), v1.0, 16-08-2010.

Gradman, S., Digital Library Metadata, in M. Collier (ed.), [Business Planning for Digital Libraries. International Approaches](#), Leuven, Leuven University Press, 2010, pp. 157-164.

IFLA, [Guidelines for Multilingual Thesauri](#), April 2005.

Jones, G.J.F., Zhang, Y., Newman, E., Fantino, F. and Debole, F.(2007) [Multilingual search for cultural heritage archives via combining multiple translation resource](#), in LaTeCH 2007/ACL Workshop on Language Technology for Cultural Heritage Data, 28 June 2007, Prague, Czech Republic.

Taylor, A.G. and Joudrey, D.N., The organization of information, Westport, Connecticut/London, Libraries Unlimited, 2008 (3rd ed.).

APPENDIX A, B, C

APPENDIX A

HOPE Archive Collections

This table holds an overview of all collections that HOPE CPs have been identified as **archive collections**. The metadata of these collections could be mapped to the Common HOPE Metadata Structure using a HOPE archive profile. The table respectively lists the name of the content provider (A), the name of the collection (B), the number of digital files per collection (C), the number of metadata records per collection (D), the metadata standard or element set used to encode the metadata (E), the collection type as specified by the content provider (F), the element set type (G) and the Europeana classification type (H). The last column (I) contains a suggestion for the most appropriate HOPE domain profile, derived from the information in the preceding columns. Collections for which the CP might consider using another profile are put in color: red for visual profiles, green for audio-visual profiles.

archive collections: 413562 records of which 872 encoded using Dublin Core Simple.
 consider an archive profile: 188977 records 45,69 % of the metadata identified as archive sub-collections
 consider an audio-visual profile: 717 records 0,17 % of the metadata identified as archive sub-collections
 consider a visual profile: 223868 records 54,13 % of the metadata identified as archive sub-collections
 add from library sub-collections: 248 records

Content Provider	Name Sub-collection	# digital files	# metadata records	metadata standards (Table 2)	sub-collection type (Q.101)	element set type (Table 2)	Europeana classification type (Q.153)	HOPE domain profile
Amsab-ISG	Socialist Party Meeting Reports	1700	1700	ISAD(G) v.2 / idiosyncr	archive	archival	Text	archive
Amsab-ISG	Trade Union Meeting Reports	4800	4800	ISAD(G) v.2 / idiosyncr	archive	archival	Text	archive
CGIL	1. CGIL Executive bodies' minutes (1	368	408	EAD 2002	archive	archival	Text	archive
CGIL	3. CGIL Executive bodies' minutes (1	1131	1230	EAD 2002	archive	archival	Text	archive
CGIL	4. CGIL Bureau's circular letters (194	4057	4058	EAD 2002	archive	archival	Text	archive
CGIL	2. CGIL Photo Archive (1945-1980)	25000	25000	EAD 2002	archive	audio-visual	Image	visual
FES Archive	Allgemeiner Deutscher Arbeitervere	257	1	idiosyncratic FES-A	archive	archival	Text	archive
FES Archive	Leaflets	182017	6564	idiosyncratic FES-A	archive	archival	Text	archive
FES Archive	Banners	159	159	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Postcards	630	630	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Stickers	873	873	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Posters	47683	47683	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Photographs	127402	127402	idiosyncratic FES-A	archive	archival	Image	visual
FMS	11. Cultural and Political magazine fi	2301	54	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	12. Cultural and political magazine fi	8252	56	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	7. Personal archives on support of pi	1000	300	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	13. Personal archive of a Portuguese	1123	384	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	5. Portuguese Socialist Party archive	5000	900	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	4. Students Movements archives and	10000	1700	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	6. Cultural and political opposition a	8500	2000	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	10. Labour Movement archives	10000	2500	idiosyncratic FMS (set	archive	archival	Image	archive
FMS	3. African Independence Movement	38100	3600	idiosyncratic FMS (set	archive	archival	Image and Audio	archive

FMS	8. Political opposition against Portug	30000	5000 idiosyncratic FMS (set archive	archival	Image	archive
FMS	9. Republican Movement archives ar	45000	10700 idiosyncratic FMS (set archive	archival	Image	archive
FMS	2. FMS Photo Archive	36700	4200 idiosyncratic FMS (set archive	archival	Image	visual
Génériques	Books of songs of "French-Arabic Mi	1423	189 EAD 2002 + admin. mc archive	archival		archive
Génériques	Newspapers of migrants in France, p	913	913 EAD 2002 + admin. mc archive	archival		archive
Génériques	Newspapers of migrants in France, p	917	917 EAD 2002 + admin. mc archive	archival		archive
Génériques	Photographs of Arabic artists in Fran	60	1 EAD 2002 + admin. mc archive	archival	Image	visual
Génériques	Posters collection Migration in Franc	611	611 EAD 2002 + admin. mc archive	archival	Image	visual
Génériques	Posters collection Migration in Franc	1118	1118 EAD 2002 + admin. mc archive	archival	Image	visual
Génériques	Posters collection Migration in Franc	1575	1575 EAD 2002 + admin. mc archive	archival	Image	visual
KEE/OSA	12. HU OSA 102-02-01 Supplementa	1280	289 Dublin Core archive	archival	Text	archive
KEE/OSA	10. 3380003 Transcripts of Viewers (1500	39 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	11. HU OSA 386-02 Records of the P	2443	138 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	13. HU OSA 300-01-02 Information I	225000	68500 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	2. Office of the High Representative	128000	16000 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	3. HU OSA (3)59, Records the UN Spi	32700	8220 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	4. HU OSA 300-08-03 RFE/RL RI: Pub	174000	31000 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	5. HU OSA 300-08-47 RFE/RL RI: Situ	66000	212 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	8. 1956 Digital Archive: Donald and \	17000	365 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	9. 300-040-08 RFE/RL RI: Media Mo	29236	349 idiosyncratic OSA (set archive	archival	Text	archive
KEE/OSA	1. HU OSA 300-08-109 Video Record	646	646 idiosyncratic OSA (set archive	audio-visual	Video	audio-visual
KEE/OSA	6. Virtual Museum of Film slides, Prc	18000	248 idiosyncratic OSA (set library	audio-visual	Image	archive
KEE/OSA	7. HU OSA 358 Records of the Civil D	66	66 idiosyncratic OSA (set archive	audio-visual	Video	audio-visual
KNAW-IISG	6. Ferdinand Domela Nieuwenhuis A	40000	1022 EAD 2002 archive	archival	no answer	archive
KNAW-IISG	1. Archives Akseľ'rod, Bakunin, Bern.	155864	3270 EAD 2002 archive	archival	no answer	archive
TA	Oral history, topics of interviews or i	2227	2227 idiosyncratic (set 2) (p archive	archival	Text, image, video, sc	archive
TA	Minutes of Social Democratic Party's	130000	8500 idiosyncratic (set 2) (p archive	archival	Text, image, video, sc	archive
UPIP(BDIC)	13. Notebooks of Antonio Blanca	123	2 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	5. Archives : Victor Bulin	192	4 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	3. Archives: revision of the 'Bonnet	855	9 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	7. Archives : correspondence Union	2587	11 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	16. Archives : Société (Association) c	509	16 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	15. Archives : Association Russe pou	554	18 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	4. Preparatory texts Convention de (1100	53 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	2. Archives : Académie de Lille	5673	250 Dublin Core Simple archive	bibliographic	Image	archive
UPIP(BDIC)	6. Jules Prudhommeaux collection o	19844	509 Dublin Core Simple archive	bibliographic	Image	archive
VGA	Films: Mayday-Demostration, SPÖ, €	5	5 ? (no answer) archive	audio-visual (addc	audio-visual (added	audio-visual
VGA	AZ-Fotoarchiv (available at the VGA	5000	5000 ? (no answer) archive	audio-visual (addc	audio-visual (added	visual
VGA	AZ-Fotoarchiv (online available)	9616	9616 ? (no answer) archive	audio-visual (addc	audio-visual (added	visueal

Commonly used Metadata Standards for encoding Archive Collections

Sum # metadata records			
standard	Content Provider	Totaal	
? (no answer)	VGA	14621	
Total ? (no answer)		14621	3,53 %
Dublin Core	KEE/OSA	289	
	UPIP(BDIC)	872	
Total Dublin Core		1161	0,28 %
EAD 2002	CGIL	30696	
	Génériques	5324	
	KNAW-IISG	4292	
Total EAD 2002		40312	9,74 %
idiosyncratic	Amsab-ISG	6500	
	FES Archive	183312	
	FMS	31394	
	KEE/OSA	125783	
	TA	10727	
Total idiosyncratic		357716	86,44 %
Total:		413810	archive' metadata records

HOPE Visual Collections

This table holds an overview of all collections that HOPE CPs have been identified as **museum or 'other: visual resource' collections**. The metadata of these collections could be mapped to the Common HOPE Metadata Structure using a HOPE Visual Profile. The table respectively lists the name of the content provider (A), the name of the collection (B), the number of digital files per collection (C), the number of metadata records per collection (D), the metadata standard or element set used to encode the metadata (E), the collection type as specified by the content provider (F), the element set type (G) and the Europeana classification type (H). The last column (I) contains a suggestion for the most appropriate HOPE domain profile, derived from the information in the preceding columns. Hence, several archive and library might be considered as visual collections as well. These collections are put in color: red for native archive collections, green for native library collections.

visual collections 239654 records, of which 28484 encoded using Dublin Core Simple.
 add from archive collections 223868 records
 add from library collections 200 records

Content Provider	Name Collection	# digital files	# metadata records	metadata standards (Table 2)	sub-collection type (Q.101)	element set type (Table 2)	Europeana classification type (Q.153)	HOPE domain profile
CGIL	2. CGIL Photo Archive (1945-	25000	25000	EAD 2002	archive	audio-visual	Image	visual
FES Archive	Banners	159	159	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Postcards	630	630	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Stickers	873	873	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Posters	47683	47683	idiosyncratic FES-A	archive	archival	Image	visual
FES Archive	Photographs	127402	127402	idiosyncratic FES-A	archive	archival	Image	visual
FMS	2. FMS Photo Archive	36700	4200	idiosyncratic FMS (set 2)	archive	archival	Image	visual
Génériques	Photographs of Arabic artist:	60	1	EAD 2002 + admin. md	archive	archival	Image	visual
Génériques	Posters collection Migration	611	611	EAD 2002 + admin. md	archive	archival	Image	visual
Génériques	Posters collection Migration	1118	1118	EAD 2002 + admin. md	archive	archival	Image	visual
Génériques	Posters collection Migration	1575	1575	EAD 2002 + admin. md	archive	archival	Image	visual
VGA	AZ-Fotoarchiv (available at tl	5000	5000	? (no answer)	archive	audio-visual (ad	audio-visual (added t	visual
VGA	AZ-Fotoarchiv (online availa	9616	9616	? (no answer)	archive	audio-visual (ad	audio-visual (added t	visual
KNAW-IISG	7. The Dutch Labour Movem	6000	200	MARC21	library	audio-visual	no answer	visual
Amsab-ISG	Pictures from De Morgen Ne	20000	20000	SPECTRUM v3.2 / idiosyncratic A	museum	museum	Image	visual
Amsab-ISG	Visual collection: photograpl	27570	27570	SPECTRUM v3.2 / idiosyncratic A	museum	museum	Image	visual
UPIP(BDIC)	1. Posters Commune de Pari	1170	1170	Dublin Core Simple	museum	bibliographic	Image	visual
UPIP(BDIC)	10. Photographs of the Gula	145	145	Dublin Core Simple	museum	bibliographic	Image	visual
UPIP(BDIC)	11. Photographs of the Span	220	220	Dublin Core Simple	museum	bibliographic	Image	visual
UPIP(BDIC)	12. Photographs by Elie Kag	2534	2543	Dublin Core Simple	museum	bibliographic	Image	visual

SSA	1. Social issues and social mc	18857	18857 Dublin Core Simple	other	audio-visual	Image	visual
SSA	2. Political propaganda (post	5549	5549 Dublin Core Simple	other	audio-visual	Image	visual
KNAW-IISG	3. Image collection Descaves	1700	1000 MARC21	other (visual)	audio-visual	no answer	visual
KNAW-IISG	4. Image and Sound collectic	105000	105000 MARC21	other (visual)	audio-visual	no answer	visual
KNAW-IISG	8. The Dutch Labour Movem	5700	5700 MARC21	other (visual)	audio-visual	no answer	visual
KNAW-IISG	9. Algemeen Hollands Fotop	1900	1900 MARC21	other (visual)	audio-visual	no answer	visual
KNAW-IISG	11. Posters Paradiso and Me	5000	5000 MARC21	other (visual)	audio-visual	no answer	visual
KNAW-IISG	12. Posters Dutch political ar	25000	25000 MARC21	other (visual)	audio-visual	no answer	visual
TA	Posters and photographs Fin	20000	20000 idiosyncratic TA (set 1) (partly D	other (visual)	audio-visual	Text, image, video, so	visual

Commonly used Metadata Standards for encoding Visual Collections

Metadata Standard	Sum # metadata records	
? (no answer)	14616	3,15 %
VGA	14616	
Dublin Core Simple	28484	6,14 %
SSA	24406	
UPIP(BDIC)	4078	
EAD 2002	28305	6,10 %
CGIL	25000	
Génériques	3305	
idiosyncratic	248517	53,59 %
Amsab-ISG	47570	
FES Archive	176747	
FMS	4200	
TA	20000	
MARC21	143800	
KNAW-IISG	143800	31,01 %
Total:	463722	'visual' metadata records

HOPE Audio-Visual Collections

This table holds an overview of all collections that HOPE CPs have been identified as 'other: audio-visual resource' collections. The metadata of these collections could be mapped to the Common HOPE Metadata Structure using a HOPE audio-visual profile. The table respectively lists the name of the content provider (A), the name of the collection (B), the number of digital files per collection (C), the number of metadata records per collection (D), the metadata standard or element set used to encode the metadata (E), the collection type as specified by the content provider (F), the element set type (G) and the Europeana classification type (H). The last column (I) contains a suggestion for the most appropriate HOPE domain profile, derived from the information in the preceding columns. Collections for which CPs might consider an archive profile are put in red.

audio-visual collections 1510 records of which 610 encoded using Dublin Core Simple.
 add from archive collections: 717 records

Content Provider	Name Sub-collection	# digital files	# metadata records	metadata standards (Table 2)	sub-collection type (Q.101)	element set type (Table 2)	Europeana classification type (Q.153)	HOPE domain profile
Amsab-ISG	Social History Documentaries and Socialist Party Broadcasts	500	500	SPECTRUM v3.2 / idiosyncratic	other (audio-visual)	audio-visual	Video	audio-visual
KEE/OSA	1. HU OSA 300-08-109 Video Recordings of Central Television Pr	646	646	idiosyncratic OSA (set 1)	archive	audio-visual	Video	audio-visual
KEE/OSA	7. HU OSA 358 Records of the Civil Defense Alliance (ABC "parar	66	66	idiosyncratic OSA (set 7)	archive	audio-visual	Video	audio-visual
KNAW-IISG	5. Social movements 20th century; interviews, recordings of me	200	200	MARC21	other (audio-visual)	audio-visual	no answer	audio-visual
SSA	3. Films about various social movements	70	70	Dublin Core Simple	other	audio-visual	Video	audio-visual
SSA	4. Collections "Stadt in Bewegung", "SOVAZ", "Da und Fort": you	150	150	Dublin Core Simple	other	audio-visual	Video	audio-visual
SSA	5. Swiss trade union films: political propaganda and self manifes	30	30	Dublin Core Simple	other	audio-visual	Video	audio-visual
SSA	6. Collection "Voll-versammlungen": youth movement	10	10	Dublin Core Simple	other	audio-visual	Sound	audio-visual
SSA	7. Records of trade union sessions (metalworkers, transport wor	350	350	Dublin Core Simple	other	audio-visual	Sound	audio-visual
TA	Trade Union movement, SDP's organisations, socio-political film	200	200	idiosyncratic TA (set 1)	other (visual)	audio-visual	Text, image, video, s	audio-visual
VGA	Films: Mayday-Demonstration, SPÖ, elections, Karl Seitz	5	5	? (no answer)	archive	audio-visual	ad audio-visual (added	audio-visual

Commonly used Metadata Standards for encoding Audio-Visual Collections

Metadata Standards	Sum # metadata records	
? (no answer)	5	0,22 %
VGA	5	
Dublin Core Simple	610	
SSA	610	27,39 %
idiosyncratic	1412	63,40 %
Amsab-ISG	500	
KEE/OSA	712	
TA	200	
MARC21	200	8,98 %
KNAW-IISG	200	
Total:	2227	audio-visual' metadata records

HOPE Library Collections

This table holds an overview of all collections that HOPE CPs have been identified as **library collections**. The metadata of these collections could be mapped to the Common HOPE Metadata Structure using a HOPE Library Profile. The table respectively lists the name of the content provider (A), the name of the collection (B), the number of digital files per collection (C), the number of metadata records per collection (D), the metadata standard or element set used to encode the metadata (E), the collection type as specified by the content provider (F), the element set type (G) and the Europeana classification type (H). The last column (I) contains a suggestion for the most appropriate HOPE domain profile, derived from the information in the preceding columns. Collections for which the CP might consider using an archive profile are put in red.

library collections 238349 records, of which 209 encoded using Dublin Core Simple.
 consider an archive profile: 248 records 0,10 % of the metadata identified as library sub-collections

Content Provider	Name Sub-collection	# digital files	# metadata records	metadata standards (Table 2)	sub-collection type (Q.101)	element set type (Table 2)	Europeana classification type (Q.153)	HOPE domain profile
Amsab-ISG	AMSAB-ISG editions	300	300	ISBD 2007 / idiosyncr	library	bibliographic	Text	library
Amsab-ISG	Trade Union Periodicals	1480	1480	ISBD 2007 / idiosyncr	library	bibliographic	Text	library
Amsab-ISG	Trade Union Periodicals	11510	11510	ISBD 2007 / idiosyncr	library	bibliographic	Text	library
Amsab-ISG	De Vooruit Socialist Party Daily N	17000	17000	ISBD 2007 / idiosyncr	library	bibliographic	Text	library
CGIL	5. CGIL Library brochures (1945-:	925	925	MODS	library	bibliographic	Text	library
FES Library	Portal Breslauer Arbeiterbewegu	48	48	MAB (original format	library	bibliographic	Text	library
FES Library	Congress reports, Sozial-demokr:	578	83	MAB (original format	library	bibliographic	Text	library
FES Library	Publications of the Internationa	95	95	MAB (original format	library	bibliographic	Text	library
FES Library	Periodical: Sozialistische Mitteilu	343	151	MAB (original format	library	bibliographic	Text	library
FES Library	Biographical dictionary of the ÖT	213	213	MAB (original format	library	bibliographic	Text	library
FES Library	Sources on the development of t	360	360	MAB (original format	library	bibliographic	Text	library
FES Library	Periodicals of 'Die Naturfreunde'	381	381	MAB (original format	library	bibliographic	Text	library
FES Library	Monatsberichte über die Entwicl	405	405	MAB (original format	library	bibliographic	Text	library
FES Library	Collection: Arbeitersport	570	570	MAB (original format	library	bibliographic	Text	library
FES Library	Periodical: Sozial-demokratische	988	988	MAB (original format	library	bibliographic	Text	library
FES Library	Programmatic documents and st	27264	1383	MAB (original format	library	bibliographic	Text	library
FES Library	Periodical: Die Arbeit	7372	1436	MAB (original format	library	bibliographic	Text	library
FES Library	Periodical: Berliner Frauenzeitun	6459	1825	MAB (original format	library	bibliographic	Text	library

FES Library	'FES-Netz-Quelle' History and po	1834	1834 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Arbeiterwohlfahrt	5487	2630 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Der Funke	2246	3205 MAB (original format library	bibliographic	Text	library
FES Library	Newspaper: Volkswacht für Schl	30000	4500 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Gewerkschaftliche M	7180	4833 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Neue Zeit	61351	4963 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Sozialistische Monats	43500	5153 MAB (original format library	bibliographic	Text	library
FES Library	Chronik der deutschen Sozialden	5412	5412 MAB (original format library	bibliographic	Text	library
FES Library	Digital Library of the Friedrich-Et	12330	6000 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Archiv für Sozialgesch	35519	7218 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Neue Gesellschaft0Fr	49493	8619 MAB (original format library	bibliographic	Text	library
FES Library	Press releases SPD	52880	25510 MAB (original format library	bibliographic	Text	library
FES Library	40 Trade union newspapers befo	383000	50000 MAB (original format library	bibliographic	Text	library
FES Library	Periodical: Sozial0demokratische	101068	63104 MAB (original format library	bibliographic	Text	library
FMS	1. Historical and Political Science	16600	198 idiosyncratic FMS (sel library	bibliographic	Image	library
KEE/OSA	6. Virtual Museum of Film slides,	18000	248 idiosyncratic OSA (sel library	audio-visual	Image	archive
KNAW-IISG	10. International Review of Socia	22000	1528 idiosyncratic IISG (jou library	bibliographic	no answer	library
KNAW-IISG	2. Publications Fédération jurass	4136	22 MARC21 library	bibliographic	no answer	library
KNAW-IISG	7. The Dutch Labour Movement	6000	200 MARC21 library	audio-visual	no answer	library
UPIP(BDIC)	14. 58 Serial publications : Spani	30000	58 Dublin Core Simple library	bibliographic	text	library
UPIP(BDIC)	8. Serial : 'Journaux de tranchées	9571	102 Dublin Core Simple library	bibliographic	Image	library
UPIP(BDIC)	9. 49 Serial publications : Spanis	52056	49 Dublin Core Simple library	bibliographic	text	library
UPIP(MSH Dijon/C	1. Cahiers de l'Institut Maurice T	19000	127 EAD 2002 (original fo library	bibliographic	Texte, Image	library
UPIP(MSH Dijon/C	2. Revue Société française	3000	61 EAD 2002 (original fo library	bibliographic	Texte, Image	library
UPIP(MSH Dijon/C	3. Brochures Bibliothèque Marx:	2165	3581 EAD 2002 (original fo library	bibliographic	Texte, Image	library
VGA	Arbeiterzeitung (Frakturschrift)	365	1 ? (no answer) library	bibliographic (adr	? (no answer)	library
VGA	Glühlichter, Neue Glühlichter (Fr	1042	5 ? (no answer) library	bibliographic (adr	? (no answer)	library
VGA	Sozialdemokratische Parteitag	3882	35 ? (no answer) library	bibliographic (adr	? (no answer)	library

Commonly used Metadata Standards for encoding Library Collections

Metadata Standar	Sum # metadata records	
? (no answer)	41	0,02 %
VGA	41	
Dublin Core Simpl	209	0,09 %
UPIP(BDIC)	209	
EAD 2002	3769	1,58 %
UPIP(MSH Dijon, idiosyncratic)	3769	
Amsab-ISG	32264	13,54 %
Adlib / ISBD 20	30290	
FMS	198	
FMS (set 1)	198	
KEE/OSA	248	
OSA (set 6)	248	
KNAW-IISG	1528	
(journalpublish)	1528	
marc family	202066	84,78 %
CGIL	925	
MODS	925	
FES Library	200919	
MAB (original f	200919	
KNAW-IISG	222	
MARC21	222	
Total:	238349	'library' metadata records

APPENDIX B

HOPE Archive Profile: Collection Form			Dublin Core Collection Application Profile / Europeana Namespace				Local Element	Mapping Local-EAD
area	sub-area	collection element	definition	occurrences	obligation	conditions	EAD element	
Description Control Area		data provider	ens:dataProvider	Name of the institution supplying the metadata to the HOPE aggregator.	single value	mandatory	x	/ead/eadheader/profiledesc/creation[encodinganalogue="ens:d
		aggregator	ens:provider	Name of the aggregator, supplying metadata to the Europeana, Labour History Portal and other services	single value	mandatory	x	/ead/eadheader/eadid@mainagencycode
		language metadata	dc:language	Primary language of the metadata describing the resource.	single value	mandatory	x	ead/eadheader/eadid/profiledesc/language/language@langcode
Identification Area	Collection Identifier	local ID	dc:identifier	An unambiguous reference to an entity in the local system of the content provider describing the	single value	mandatory, if	x	/ead/archdesc/did/unitid
		persistent ID	dc:identifier	A persistent identifier, which allows identifying the collection in and outside the HOPE.	single value	mandatory, if	x	/ead/archdesc/did/unitid/@identifier
	Description Level	description level	hope:descriptionLevel	Identifying the hierarchical level of description, i.e. collection	single value	mandatory	x	/ead/archdesc/@level
		title	dc:title	The name of the collection.	single value	mandatory	x	/ead/archdesc/did/unittitle
		alternative title	dcterms:alternative	Any form of the name used as a substitute or alternative to the formal name of the collection.	repeatable	optional	x	/ead/archdesc/did/unittitle[encodinganalogue="dcterms:altern
	Collector	collector - person	dc:creator	A person who gathers (or gathered) the items in a collection together.	repeatable	optional	x	/ead/archdesc/did/origination/persname
		collector - organisation	dc:creator	An organisation who gathers (or gathered) the items in a collection together.	repeatable	optional	x	/ead/archdesc/did/origination/corporname
	Date	date created	cid:dateItemsCreated	A controlled date value over which the individual items within the collection were created.	single value	mandatory	x	/ead/archdesc/did/unitdate@normalised[encodinganalogue="ci
		date created statement	cid:dateItemsCreated	A statement on the dates over which the individual items within the collection were created.	single value	optional	x	/ead/archdesc/did/unitdate[encodinganalogue="cid:dateItemsC
date accumulated		dcterms:created	A controlled date value over which the individual items within the collection were collected.	single value	mandatory	x	/ead/archdesc/did/unitdate@normalised[encodinganalogue="d	
	date accumulated statement	dcterms:created	A statement on the dates over which the individual items within the collection were collected.	single value	optional	x	/ead/archdesc/did/unitdate[encodinganalogue="dcterms:creat	
Physical Description Area	Size	size in meters	dcterms:extent	A statement on the size of the collection in linear meters.	single value	mandatory, if	x	/ead/archdesc/did/physdesc/extent[unit="meters"]
		size in items	dcterms:extent	A statement on the number of items in the collection.	single value	mandatory, if	x	/ead/archdesc/did/physdesc/extent[unit="items"]
	Type	item type	cid:itemType	A term giving the nature or genre of one more items within the collection.	repeatable	mandatory	x	/ead/archdesc/did/physdesc/genreform@normalised[encoding
		item type statement	cid:itemType	A statement on the nature of genre of the items within the collection.	single value	optional	x	/ead/archdesc/did/physdesc/genreform[encodinganalogue="ci
	Format	item format	cid:itemFormat	A term giving a media type, physical or digital, of one or more items within the collection.	repeatable	optional	x	/ead/archdesc/did/physdesc/genreform@normalised[encoding
		item format statement	cid:itemFormat	A statement giving the media type, physical or digital, of one or more items within the collection.	single value	optional	x	/ead/archdesc/did/physdesc/genreform[encodinganalogue="ci
Content Area		language	dc:language	The language(s) of the items within the collection.	repeatable	mandatory, if	x	/ead/archdesc/did/langmaterial/language@langcode
		language statement	dc:language	A statement on the language of the items within the collection.	single value	optional	x	/ead/archdesc/did/langmaterial/language
		abstract	dcterms:abstract	A free text summary description of the collection.	single value	mandatory	x	/ead/archdesc/scopecontent
		custodial history	dcterms:provenance	A statement of any changes in ownership and custody of the collection that are significant for its authenticity, integrity and interpretation.	single value	optional	x	/ead/archdesc/custodhist
		accrual policy	dcterms:accrualPolicy	A statement on the policy governing the addition of items to the collection.	single value	optional	x	/ead/archdesc/accruals[encodinganalogue="dcterms:accrualPol
		super-collection	dcterms:isPartOf	A second collection that contains the current collection.	single value	optional	x	/ead/archdesc/arrangement[encodinganalogue="dcterms:isPart
		catalogue or index	cid:catalogueOrIndex	A catalogue for, or index of, the collection	single value	optional	x	/ead/archdesc/otherfindaid[encodinganalogue="cid:catalogueO
		associated publication	dcterms:isReferencedBy	A publication that is based on the use, study, or analysis of the collection.	repeatable	optional	x	/ead/archdesc/bibliography
Association Area		subject	dc:subject	A value for a subject or topic of the collection.	repeatable	optional	x	/ead/archdesc/controlaccess/subject[encodinganalogue="dc:su
		spatial coverage	dcterms:spatial	A statement on the spatial scope of the collection.	repeatable	optional	x	/ead/archdesc/controlaccess/subject[encodinganalogue="dcter
		temporal coverage	dcterms:temporal	A statement on the temporal scope of the collection.	repeatable	optional	x	/ead/archdesc/controlaccess/subject[encodinganalogue="dcter
Conditions of Access and Use Area		owner	marcrel:OWN	An entity who has legal possession of the collection.	single value	optional	x	/ead/archdesc/acqinfo[encodinganalogue="marcrel:OWN"]
		repository	cid:isLocatedAt	archival institution or agency providing access to the described materials.	repeatable	mandatory	x	/ead/archdesc/repository
		use rights	dc:rights	A statement of any rights held in/over the collection.	single value	mandatory	x	ead/accessrestrict/userrestrict
		access rights	dcterms:accessRights	A statement explicating the access restrictions placed over the analog and digital content, including allowed users, charges, formats, etc.	single value	mandatory	x	ead/archdesc/accessrestrict

HOPE Archive Profile: Mapping Worksheet

HOPE Archive Profile: Mapping Worksheet								Local Element	Mapping Local-EAD		
area	subarea	archive element	EAD 2002 / Premis / Europeana Namespace	definition	occurrences	cardinality	conditions		ead element		
Description Control Area		content provider aggregator	ens:dataProvider	Name of the institution supplying the metadata to the HOPE aggregator.	single value	mandatory		x	/ead/eadheader/profiledesc/creation[@encodinganalogue="ens:dataProvider"]		
		language metadata type	ens:provider	Name of the aggregator, supplying metadata to one or more discovery services	single value	mandatory		x	/ead/eadheader/eadid/mainagencycode		
		Digital Resource Area	Administrative Metadata	language	hope:language	Primary language of the metadata describing the collection item(s).	single value	mandatory		x	/ead/archdesc/dsc/c/did/daogrp/daodescnote/table/tgroup/tbody/row/entry[@colname="type"]
				rights	ens:type	The Europeana material type of the digital representation, which is used to create the Europeana Type facet search.	single value	mandatory		x	/ead/archdesc/dsc/c/did/daogrp/daodescnote/table/tgroup/tbody/row/entry[@colname="type"]
		Derivative 2		language	premis:significantProperties	The language(s) of the intellectual content of the digital representation.	repeatable	mandatory, if	mandatory if the digital representation contains text	x	/ead/archdesc/dsc/c/did/daogrp/daodescnote/table/tgroup/tbody/row/entry[@colname="language"]
				resolve URL	ens:rights	Statement about the rights held in or over the digital representation	single value	mandatory		x	/ead/archdesc/dsc/c/did/daogrp/daodescnote/table/tgroup/tbody/row/entry[@colname="rights"]
		Derivative 3		local ID	ens:isShownBy	An unambiguous URL which allows retrieving a low resolution derivative of the digital representation from a Shared or Local Object Repository.	single value	mandatory, if	mandatory if the CP does not create PIDs for low resolution derivative files	x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="derivative 2"]@href
				persistent ID	ens:isShownBy	An unambiguous reference to a low resolution derivative of the digital representation, identifying this derivative in the local system of the content provider.	single value	mandatory, if	mandatory if the SOR creates PIDS for low resolution derivative files	x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="derivative 2"]@id
		Derivative 3		sequence	ens:isShownBy	A globally unique, resolvable URL which allows identifying and retrieving a low-resolution derivative of the digital representation in- and outside the HOPE information system	single value	mandatory, if	mandatory if the CP creates PIDS for low resolution derivative files	x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="derivative 2"]@href
				resolve URL	ens:isNextInSequence	An identifier for the digital representation that comes immediately after the current digital representation, as part of an ordered series	single value	optional		x	sequence daogrp wrappers OR if sequence references are available in the CPs metadata:
		Transcription		local ID	ens:object	An unambiguous URL which allows retrieving a thumbnail of the digital representation from a Shared or Local Object Repository.	single value	optional		x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="derivative 3"]@href
				persistent ID	ens:object	An unambiguous reference to a thumbnail of the digital representation, identifying this thumbnail in the local system of the content provider.	single value	optional		x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="derivative 3"]@id
		Landing Page		transcription	premis:significantProperties	A manually or automatically generated transcription of the textual content of a digital representation.	single value	optional		x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="derivative 3"]@href
				resolve URL	premis:relatedObjectIdentification	An unambiguous URL reference to a text file containing the transcription.	single value	optional		x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="transcription"]@href
		Landing Page		local ID	premis:relatedObjectIdentification	An unambiguous reference to a text file containing the transcription, identifying this text file in the local system of the content provider.	single value	optional		x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="transcription"]@id
				persistent ID	premis:relatedObjectIdentification	A globally unique, resolvable URL which allows identifying and retrieving a text file containing the transcription, in- and outside the HOPE information system	single value	optional		x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="transcription"]@href
		Identity Statement Area	Reference Code (1.1)	resolve URL	ens:landingPage	An unambiguous URL reference to the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP does not create PIDs for landing pages	x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="landing page"]@href
				persistent ID	ens:landingPage	A globally unique, resolvable URL which allows identifying and retrieving the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP creates PIDS for landing pages	x	/ead/archdesc/dsc/c/did/daogrp/daoloc[@label="landing page"]@href
		Parent Record		call number	ead:unitid	A reference number or otherwise uniquely identifying number of the described materials [i.e. the An unambiguous reference to an entity in the local system of the content provider describing the collection item(s)].	single value	mandatory		x	/ead/archdesc/dsc/c/unitid@type="call number"
				local ID	ead:unitid	An unambiguous reference to an entity in the local system of the content provider describing the collection item(s).	single value	mandatory, if	mandatory if a CP does not create PIDs for descriptive units	x	/ead/archdesc/dsc/c/unitid@type="local id"
Sequence		persistent ID	ead:c@id	A persistent identifier for an entity in the HOPE system describing the collection item(s).	single value	mandatory, if	mandatory if a CP creates PIDS for descriptive units	x	/ead/archdesc/dsc/c@id		
		local ID	dcterms:isPartOf	A local identifier of a related entity describing a series or collection in which the described collection item(s) are physically or logically included.	single value	mandatory, if	mandatory if a CP does not create PIDs for descriptive units	x	Nest the c level of the current descriptive unit in the c level for which the local id (/ead/archdesc/dsc/c/unitid@type="local id") corresponds with the value provided by the CP.		
Level of Description (1.4)		persistent ID	dcterms:isPartOf	A persistent identifier of a related entity describing a series or collection in which the described collection item(s) are physically or logically included.	single value	mandatory, if	mandatory if a CP creates PIDs for descriptive units	x	Nest the c level of the current descriptive unit in the c level for which the persistent id (/ead/archdesc/dsc/c@id) corresponds with the value provided by the CP.		
		local ID	ens:isNextInSequence	A local identifier of another descriptive unit, as part of an ordered series of descriptive unit, that should be displayed immediately after the given descriptive unit.	single value	optional		x	Map the c level of the current descriptive unit immediately after the c level for which the local id (/ead/archdesc/dsc/c/unitid@type="local id") corresponds with the value provided by the CP.		
Title (1.2)		persistent ID	ens:isNextInSequence	A persistent identifier of another descriptive unit, as part of an ordered series of unit, that should be displayed immediately after the given unit.	single value	optional		x	Map the c level of the current descriptive unit immediately after the c level for which the persistent id (/ead/archdesc/dsc/c@id) corresponds with the value provided by the CP.		
		level of description	ead:c@level	A term identifying the hierarchical level of entity describing the collection item(s).	single value	mandatory		x	/ead/archdesc/dsc/c@level		
Date(s) (1.3)		title	ead:unittitle	The title or name of the described materials.	repeatable	mandatory		x	/ead/archdesc/dsc/c/did/unittitle		
		date of creation	ead:unitdate	A statement naming the year(s), month(s), or day(s) the described materials have been created.	repeatable	optional		x	/ead/archdesc/dsc/c/did/unitdate		
Extent and medium of the unit		appearance of the material	ead:physdesc	Information about the appearance, condition, and extent of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/physdesc		
		extent	ead:extent	Information about the extent of the described materials, specified by using the additional unit of measurement.	repeatable	optional		x	/ead/archdesc/dsc/c/did/physdesc/extent		
Text Area		genre of the fonds	ead:genreform	Information about specific types of archival materials, e.g. collections, or about specific physical characteristics.	repeatable	optional		x	/ead/archdesc/dsc/c/did/physdesc/genreform		
		origin	ead:origination	The institution or person responsible for the creation, accumulation, or assembly of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/origination		
Administrative / Biographical Archival History (2.3)		origin - name of person	ead:origination/persname	A person responsible for the creation, accumulation, or assembly of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/origination/persname		
		origin - name of corporation	ead:origination/corpname	An institution responsible for the creation, accumulation, or assembly of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/origination/corpname		
Immediate Source of Acquisition		administrative / biographical history	ead:biohist	Information about the creation and the creator(s) of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/biohist		
		archival history	ead:custodhist	Information about the custody and formation of the fonds.	repeatable	optional		x	/ead/archdesc/dsc/c/did/custodhist		
Content and Structure Area		immediate source of acquisition	ead:acqinfo	Information about the immediate source of the described materials and the circumstances under which they were received.	repeatable	optional		x	/ead/archdesc/dsc/c/did/acqinfo		
		scope and content	ead:scopecontent	A short abstract summarising the topical coverage of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/scopecontent		
Appraisal, destruction and sch Accruals (3.3)		appraisal information	ead:appraisal	Information about the process of determining the archival value of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/appraisal		
		accruals	ead:accruals	Information about anticipated additions to the described materials, that can indicate quantity and frequency.	repeatable	optional		x	/ead/archdesc/dsc/c/did/accruals		
System of Arrangement (3.4)		arrangement	ead:arrangement	Information about the principle characteristics of the internal structure, identifying the physical or logical groupings within the hierarchy of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/arrangement		
		controlled access headings	ead:controlaccess	A list of keywords associated with or covered by the described materials, including persons, corporations, materials, topical terms as well as places.	repeatable	optional		x	/ead/archdesc/dsc/c/did/controlaccess		
Conditions of Access and Use / Conditions Governing Access / Conditions Governing Reproduct		index of persons	ead:controlaccess/persname	A list of persons, including e.g. individuals' first and surnames, honorific titles, and added names.	repeatable	optional		x	/ead/archdesc/dsc/c/did/controlaccess/persname		
		index of corporations	ead:controlaccess/corpname	A list of corporations or groups, including names of associations, institutions, business firms, or	repeatable	optional		x	/ead/archdesc/dsc/c/did/controlaccess/corpname		
Conditions of Access and Use / Conditions Governing Reproduct		index of families	ead:controlaccess/famname	A list of family names; in contrast to the index of persons used when tagging the name(s) of a group of persons closely related to one another or persons who form a household.	repeatable	optional		x	/ead/archdesc/dsc/c/did/controlaccess/famname		
		index of materials	ead:controlaccess/materials	A list of materials, genre forms, or physical characteristics.	repeatable	optional		x	/ead/archdesc/dsc/c/did/controlaccess/materials		
Conditions of Access and Use / Conditions Governing Reproduct		index of subjects	ead:controlaccess/subject	A list of topical terms associated with or covered by the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/controlaccess/subject		
		index of places	ead:controlaccess/geogname	A list of geographic names, including names for places, natural features, or political jurisdictions.	repeatable	optional		x	/ead/archdesc/dsc/c/did/controlaccess/geogname		
Language / Scripts of Material		conditions governing access	ead:accessrestrict	Information about the conditions that affect the availability of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/accessrestrict		
		conditions governing use	ead:userrestrict	Information about limitations, regulations, or special procedures that affect the use of the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/userrestrict		
Physical Characteristics and Te		languages of the described materia	ead:langmaterial	A statement about the language(s) the described materials are written in.	repeatable	mandatory, if	mandatory if descriptive unit is an item and physical item contains text	x	/ead/archdesc/dsc/c/did/langmaterial		
		physical characteristics	ead:phystech	A description of physical conditions or characteristics that affect the storage, preservation, or use of the material.	repeatable	optional		x	/ead/archdesc/dsc/c/did/phystech		
Finding Aids (4.5)		other finding aids	ead:othfindaid	A reference to other finding aids or guides to the described materials, e.g. card files, inventories, or lists created by the originator.	repeatable	optional		x	/ead/archdesc/dsc/c/did/othfindaid		

Allied Materials Area	Repository	repository	ead:repository	An archival institution or agency providing access to the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/repository
	Existence and Location of Originals	location of originals	ead:originalsloc	Information about the existence, location, and availability of originals where the described materials contain copies.	repeatable	optional		x	/ead/archdesc/dsc/c/did/originalsloc
	Existence and Location of Copies	alternative form available	ead:altformavail	A reference to alternatively available forms of the described materials, e.g. microfilms or digital reproductions.	repeatable	optional		x	/ead/archdesc/dsc/c/did/altformavail
	Related Units of description (5)	related material	ead:relatedmaterial	Information about other archival materials that are related to the described ones in regards of content or origination, but not included.	repeatable	optional		x	/ead/archdesc/dsc/c/did/relatedmaterial
		separated material	ead:separatedmaterial	Information about materials that are associated by provenance to the described ones, but have been physically separated or removed.	repeatable	optional		x	/ead/archdesc/dsc/c/did/separatedmaterial
References (5.4)	bibliography	ead:bibliography	Citations to books, articles, web sites, or other forms of information that are based on, about, or of special value when using the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/bibliography	
Notes Area	Note (6.1)	general notes	ead:note	General remarks on the described materials.	repeatable	optional		x	/ead/archdesc/dsc/c/did/note
		other descriptive data	ead:odd	Further information concerning some special types of archival material such as historical public records or bulls, that cannot be tagged in any other element.	repeatable	optional		x	/ead/archdesc/dsc/c/did/odd

HOPE Library Profile: Collection Form							Local Element	Mapping Local-MARCXML	
area	sub-area	collection element	Dublin Core Collection Application Profile / Europeana Namespace	definition	occurrences	obligation	conditions	MARCXML element	
Description Control Area		data provider	ens:dataProvider	Name of the institution supplying the metadata to the HOPE aggregator.	single value	mandatory		x	/collection/record/datafield[@tag="040"]/subfield[@code="a"]
		language metadata	dc:language	Primary language of the metadata describing the resource.	single value	mandatory		x	/collection/record/datafield[@tag="040"]/subfield[@code="b"]
Collection Identifier	Collection Identifier	local ID	dc:identifier	An unambiguous reference to an entity in the local system of the content provider	single value	mandatory, if	mandatory if the CP does not create PIDs for	x	/collection/record/datafield[@tag="035"]/subfield[@code="a"]
		persistent ID	dc:identifier	A persistent identifier, which allows identifying the collection in and outside the HOPE.	single value	mandatory, if	mandatory if the CP creates PIDs for descriptive	x	/collection/record/datafield[@tag="001"]
	Description Level	description level	hope:descriptionLevel	Identifying the hierarchical level of description, i.e. collection	single value	mandatory		x	/collection/record/datafield[@tag="592"]
		title	dc:title	The name of the collection.	single value	mandatory		x	/collection/record/datafield[@tag="245"]/subfield[@code="a"]
	Collector	alternative title	dcterms:alternative	Any form of the name used as a substitute or alternative to the formal name of the collection.	repeatable	optional		x	/collection/record/datafield[@tag="245"]/subfield[@code="b"]
		collector - person	dc:creator	A person who gathers (or gathered) the items in a collection together.	repeatable	optional		x	/collection/record/datafield[@tag="100"]/subfield[@code="a"]
	Date	collector - organisation	dc:creator	An organisation who gathers (or gathered) the items in a collection together.	repeatable	optional		x	/collection/record/datafield[@tag="110"]/subfield[@code="a"]
		date created	clid:dateItemsCreated	A controlled date value over which the individual items within the collection were created.	single value	mandatory		x	/collection/record/datafield[@tag="362"]/subfield[@code="a"] (until separator ;)
	Date	date created statement	clid:dateItemsCreated	A statement on the dates over which the individual items within the collection were created.	single value	optional		x	/collection/record/datafield[@tag="362"]/subfield[@code="a"] (from separator ;)
		date accumulated	dcterms:created	A controlled date value over which the individual items within the collection were collected.	single value	mandatory		x	/collection/record/datafield[@tag="541"]/subfield[@code="d"] (until separator ;)
Date	date accumulated statement	dcterms:created	A statement on the dates over which the individual items within the collection were collected.	single value	optional		x	/collection/record/datafield[@tag="541"]/subfield[@code="d"] (from separator ;)	
	Physical Description Area; Size	size in items	dcterms:extent	A statement on the number of items in the collection.	single value	mandatory, if	mandatory if the size of your archive can be expressed in number of items	x	/collection/record/datafield[@tag="300"]/subfield[@code="a"]
Type	Type	item type	clid:itemType	A term giving the nature or genre of one more items within the collection.	repeatable	mandatory		x	/collection/record/datafield[@tag="336"]/subfield[@code="a"] (until separator ;)
		item type statement	clid:itemType	A statement on the nature of genre of the items within the collection.	single value	optional		x	/collection/record/datafield[@tag="336"]/subfield[@code="a"] (from separator ;)
	Format	item format	clid:itemFormat	A term giving a media type, physical or digital, of one or more items within the collection.	repeatable	optional		x	/collection/record/datafield[@tag="337"]/subfield[@code="a"] (until separator ;)
		item format statement	clid:itemFormat	A statement giving the media type, physical or digital, of one or more items within the collection.	single value	optional		x	/collection/record/datafield[@tag="337"]/subfield[@code="a"] (from separator ;)
Content Area		language	dc:language	The language(s) of the items within the collection.	repeatable	mandatory, if	mandatory if your archive includes textual resources	x	/collection/record/datafield[@tag="041"]/subfield[@code="a"]
		language statement	dc:language	A statement on the language of the items within the collection.	single value	optional		x	/collection/record/datafield[@tag="546"]/subfield[@code="a"]
		abstract	dcterms:abstract	A free text summary description of the collection.	single value	mandatory		x	/collection/record/datafield[@tag="520"]/subfield[@code="a"]
		custodial history	dcterms:provenance	A statement of any changes in ownership and custody of the collection that are significant for its authenticity, integrity and interpretation.	single value	optional		x	/collection/record/datafield[@tag="561"]/subfield[@code="a"]
		accrual policy	dcterms:accrualPolicy	A statement on the policy governing the addition of items to the collection.	single value	optional		x	/collection/record/datafield[@tag="584"]/subfield[@code="a"]
		super-collection	dcterms:isPartOf	A second collection that contains the current collection.	single value	optional		x	/collection/record/datafield[@tag="774"]/subfield[@code="n"]
		catalogue or index	clid:catalogueOrIndex	A catalogue for, or index of, the collection	single value	optional		x	/collection/record/datafield[@tag="555"]/subfield[@code="a"]
		associated publication	dcterms:isReferencedBy	A publication that is based on the use, study, or analysis of the collection.	repeatable	optional		x	/collection/record/datafield[@tag="504"]/subfield[@code="a"]
Association Area		subject	dc:subject	A value for a subject or topic of the collection.	repeatable	optional		x	/collection/record/datafield[@tag="650"]/subfield[@code="a"]
		spatial coverage	dcterms:spatial	A statement on the spatial scope of the collection.	repeatable	optional		x	/collection/record/datafield[@tag="651"]/subfield[@code="a"]
		temporal coverage	dcterms:temporal	A statement on the temporal scope of the collection.	repeatable	optional		x	/collection/record/datafield[@tag="648"]/subfield[@code="a"]
Conditions of Access and Use Area		owner	marcrel:OWN	An entity who has legal possession of the collection.	single value	optional		x	/collection/record/datafield[@tag="561"]/subfield[@code="a"]
		repository	clid:isLocatedAt	archival institution or agency providing access to the described materials.	repeatable	mandatory		x	/collection/record/datafield[@tag="850"]/subfield[@code="a"]
		use rights	dc:rights	A statement of any rights held in/over the collection.	single value	mandatory		x	/collection/record/datafield[@tag="540"]/subfield[@code="a"]
		access rights	dcterms:accessRights	A statement explicating the access restrictions placed over the analog and digital content, including allowed users, charges, formats, etc.	single value	mandatory		x	/collection/record/datafield[@tag="506"]/subfield[@code="a"]

Library Profile: Mapping Worksheet										Local Element	Mapping Local-MARCXML
area	sub-area	library element	MARCXML / Premis / Europeana Namespace	MODS	definition	occurrences	obligation	conditions		MARCXML element	
Digital Resource Area	Administrative Metadata	type	ens:type		The European material type of the digital representation, which is used to create the Europeana Type facet search.	single value	mandatory		x	/collection/record/datafield[@tag="843" and /subfield[@code="8"]="1"]/subfield[@code="a"] (subfield 8 = number of digital resource)	
		language	premis:significantProperties		The language(s) of the intellectual content of the digital representation.	repeatable	mandatory, if	mandatory if the digital representation contains text	x	/collection/record/datafield[@tag="843" and /subfield[@code="8"]="1"]/subfield[@code="n"]	
		rights	ens:rights		Statement about the rights held in or over the digital representation	single value	mandatory		x	/collection/record/datafield[@tag="845" and /subfield[@code="8"]="1"]/subfield[@code="a"]	
	Derivative 2	resolve URL	ens:isShownBy	identifier@type='url'@displayLabel='derivative 2'	An unambiguous URL which allows retrieving a low resolution derivative of the digital representation from a Shared or Local Object Repository.	single value	mandatory, if	mandatory if the CP does not create PIDs for low resolution derivative files	x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="z"]='derivative 2']/subfield[@code="d"]	
		local ID	ens:isShownBy	identifier@type='id'@displayLabel='derivative 2'	An unambiguous reference to a low resolution derivative of the digital representation, identifying this derivative in the local system of the content provider.	single value	mandatory, if	mandatory if the SOR creates PIDS for low resolution derivative files	x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="z"]='derivative 2']/subfield[@code="f"]	
		persistent ID	ens:isShownBy	identifier@type='pid'@displayLabel='derivative 2'	A globally unique, resolvable URL which allows identifying and retrieving a low-resolution derivative of the digital representation in- and outside the HOPE information system	single value	mandatory, if	mandatory if the CP creates PIDS for low resolution derivative files	x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="z"]='derivative 2']/subfield[@code="u"]	
		sequence	ens:isNextInSequence		An identifier for the digital representation that comes immediately after	single value	optional		x	/collection/record/datafield[@tag="856"/subfield[@code="8"]	
	Derivative 3	resolve URL	ens:object	identifier@type='url'@displayLabel='derivative 3'	An unambiguous URL which allows retrieving a thumbnail of the digital representation from a Shared or Local Object Repository.	single value	optional		x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="z"]='derivative 3']/subfield[@code="d"]	
		local ID	ens:object	identifier@type='id'@displayLabel='derivative 3'	An unambiguous reference to a thumbnail of the digital representation, identifying this thumbnail in the local system of the content provider.	single value	optional		x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="z"]='derivative 3']/subfield[@code="f"]	
		persistent ID	ens:object	identifier@type='pid'@displayLabel='derivative 3'	A globally unique, resolvable URL which allows identifying and retrieving a thumbnail of the digital representation in- and outside the HOPE information system	single value	optional		x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="z"]='derivative 3']/subfield[@code="u"]	
	Transcription	transcription	premis:significantProperties		A manually or automatically generated transcription of the textual content of a digital representation.	single value	optional		x	/collection/record/datafield[@tag="591"]	
		resolve URL	premis:relatedObjectIdentifier	identifier@type='url'@displayLabel='transcription'	An unambiguous URL reference to a text file containing the transcription.	single value	optional		x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="g"]='transcription']/subfield[@code="u"]	
		local ID	premis:relatedObjectIdentifier	identifier@type='id'@displayLabel='transcription'	An unambiguous reference to a text file containing the transcription, identifying this text file in the local system of the content provider.	single value	optional		x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="f"]='transcription']/subfield[@code="u"]	
		persistent ID	premis:relatedObjectIdentifier	identifier@type='pid'@displayLabel='transcription'	A globally unique, resolvable URL which allows identifying and retrieving a text file containing the transcription, in- and outside the HOPE information system	single value	optional		x	/collection/record/datafield[@tag="856" and /subfield[@code="8"]="1 and /subfield[@code="z"]='transcription']/subfield[@code="u"]	
	Landing Page	resolve URL	ens:landingPage	identifier@type='url'@displayLabel='landing page'	An unambiguous URL reference to the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP does not create PIDs for landing pages	x	/collection/record/datafield[@tag="856" and /subfield[@code="z"]='landing page']/subfield[@code="u"]	
		persistent ID	ens:landingPage	identifier@type='pid'@displayLabel='landing page'	A globally unique, resolvable URL which allows identifying and retrieving the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP creates PIDS for landing pages	x	/collection/record/datafield[@tag="856" and /subfield[@code="z"]='landing page']/subfield[@code="u"]	
	Resource Identifier Area	Resource Identifier	book number	marc:091	identifier@type='book number or alphanumeric designation of the original resource, assigned	single value	mandatory		x	/collection/record/datafield[@tag="091"]	
			ISBN (8.1)	marc:020 ## a	identifier@type='isbn'	International Standard Book Number denoting the original resource.	single value	optional		x	/collection/record/datafield[@tag="020"/subfield[@code="a"]
			ISSN (8.1)	marc:022 ## a	identifier@type='issn'	International Standard Serial Number denoting the original resource	single value	optional		x	/collection/record/datafield[@tag="022"/subfield[@code="a"]
			local identifier	marc:035 ## a	identifier@type='id'@displayLabel='local identifier'	unambiguous reference to an entity in the local system of the content provider	single value	conditional	mandatory if a CP does not create PIDs for descriptive units	x	/collection/record/datafield[@tag="035"/subfield[@code="a"]
Parent Record		persistent identifier	marc:001	identifier@type='URI'@displayLabel='persistent identifier'	persistent identifier for an entity in the HOPE system describing the local identifier of a related descriptive unit about a series or collection	single value	conditional	mandatory if a CP creates PIDs for descriptive units	x	/collection/record/controlfield[@tag="001"]	
		local identifier parent descriptive	marc:774 1# o	relatedItem	persistent identifier of a related descriptive unit about a series or collection in which the described collection item(s) are physically or logically included.	single value	conditional	mandatory if a CP creates PIDs for descriptive units	x	/collection/record/datafield[@tag="774"/subfield[@code="w"]	
Sequence		local identifier next descriptive	marc:785 10 w		original identifier of another descriptive unit, as part of an ordered series of descriptive unit, that should be displayed immediately after the given descriptive unit.	single value	optional		x	/collection/record/datafield[@tag="785"/subfield[@code="w"]	
		persistent identifier next descriptive	marc:785 10 o		persistent identifier of another descriptive unit, as part of an ordered series of unit, that should be displayed immediately after the given unit.	single value	optional		x	/collection/record/datafield[@tag="785"/subfield[@code="o"]	
Title Area		Level of Description (1.4)	level of description	marc:leader 07		Identifying the hierarchical level of entity describing the collection item(s).	single value	mandatory		x	/collection/record/datafield[@tag="592"]
		General Material Designation	general material designation	marc:leader 06	typeOfResource	The characteristics and general material type of content of the resource	repeatable	optional		x	/collection/record/datafield[@tag="245"/subfield[@code="h"]
	Title Proper	title	marc:245 00 a	titleInfo/title	General title field, if there is no title (proper)	single value	conditional	mandatory if there is no title proper	x	/collection/record/datafield[@tag="245"/subfield[@code="a"] (until separator ;)	
		title proper (1.1)	marc:245 00 a	titleInfo/title	Name given to the resource as it appears on the described resource	single value	conditional	mandatory if there is no title proper	x	/collection/record/datafield[@tag="245"/subfield[@code="a"] (from separator ;)	
	Parallel Title	parallel title (1.3)	marc:245 00 b	titleInfo/title/@lang='[iso639 2b code]'	Title as it appears on the described resource, but in another language and/or script than the title proper and presented as an equivalent of the prescribed source of information.	single value	optional		x	/collection/record/datafield[@tag="245"/subfield[@code="b"] (until separator ;)	
		translated title	marc:245 00 b	titleInfo/title/@type='translated'	Title translated into another language and/or script	single value	optional		x	/collection/record/datafield[@tag="242"/subfield[@code="a"]	
	Other Title Information	subtitle (1.4)	marc:245 00 b	titleInfo/subTitle	Word, phrase or group of characters, appearing in conjunction with and subordinate to the title proper, parallel title	single value	optional		x	/collection/record/datafield[@tag="245"/subfield[@code="b"] (from separator ;)	
		translated subtitle	marc:245 00 b	titleInfo/subTitle/@type='translated'	Subtitle translated into another language or script	single value	optional		x	/collection/record/datafield[@tag="242"/subfield[@code="b"]	

Statement of Responsibility Area		statement of responsibility (1.1)	marc:245 00 c	name	A statement of responsibility consists of name(s), phrase(s) or group of characters relating to the identification and/or function of any persons or corporate bodies responsible for or contributing to the creation or realisation of the intellectual or artistic content of a work contained in the resource described.	repeatable	optional		x	/collection/record/datafield[@tag="245"]/subfield[@code="c"]
	Person	author - person co-author or contributor - person	marc:100 1# a marc:700 10 a	name@type='personal'	An entity primarily responsible for making the content of the resource.	repeatable	optional		x	/collection/record/datafield[@tag="100"]/subfield[@code="a"]
	Corporate	author - corporate co-author or contributor - corporate	marc:110 2# a marc:710 2# a	name@type='corporate'	An entity primarily responsible for making the content of the resource.	repeatable	optional		x	/collection/record/datafield[@tag="110"]/subfield[@code="a"]
	Meeting	author - meeting	marc:111 2# a	name@type='conference' name/namePart/role@roleTerm='author'	An entity primarily responsible for making the content of the resource, in which the entity is a meeting name.	repeatable	optional		x	/collection/record/datafield[@tag="111"]/subfield[@code="a"]
		co-author or contributor - meeting	marc:711 2# a	name@type='conference' name/namePart/role@roleTerm='co-author or contributor'	An entity additionally responsible for making the content of the resource or for making contributions to the content of the resource, in which the entity is a meeting name.	repeatable	optional		x	/collection/record/datafield[@tag="711"]/subfield[@code="a"]
		date of meeting (as author)	marc:111 2# d	name@type='conference' name/namePart@type='date'	date of meeting (as author)	repeatable	optional		x	/collection/record/datafield[@tag="111"]/subfield[@code="d"]
		location of meeting (as author)	marc:111 2# c	name@type='conference' name/namePart@type='location'	location of meeting (as author)	repeatable	optional		x	/collection/record/datafield[@tag="111"]/subfield[@code="c"]
Edition Area	Edition Statement	edition statement (2.1)	marc:250 ## a	originInfo/edition	The edition statement consists of a term, phrase or group of characters relating to: (1) the copies of a resource formally identified as constituting a named and/or numbered edition, or (2) the copies of a resource in a particular form of presentation having significant differences from other copies in the same form of presentation.	repeatable	optional		x	/collection/record/datafield[@tag="250"]/subfield[@code="a"]
	Parallel Edition Statement	parallel edition statement (2.2)	marc:250 ## b	originInfo/edition/@lang='[iso6392b code]'	A parallel edition statement is an equivalent of the edition statement in another language and/or script.	repeatable	optional		x	/collection/record/datafield[@tag="250"]/subfield[@code="b"] (until separator ';')
	Statements of responsibility	statement of responsibility reference	marc:250 ## b	note/@type='edition statement'	Identifying the hierarchical level of entity describing the collection item(s).	single value	mandatory		x	/collection/record/datafield[@tag="250"]/subfield[@code="b"] (from separator ';')
Publication Area	Publication	publisher (4.2)	marc:260 ## b	originInfo/publisher	The name of the publisher, producer or distributor corresponds to the	repeatable	optional		x	/collection/record/datafield[@tag="260"]/subfield[@code="b"]
		place of publication (4.1) date of publication (4.4)	marc:260 ## a marc:260 ## c	originInfo/place@placeTerm/originInfo/dateIssued	The place of publication, production or distribution is the name of the The dates of publication, production or distribution of the resource.	repeatable	optional		x	/collection/record/datafield[@tag="260"]/subfield[@code="a"]
	Production	printer, manufacturer or engraver	marc:260 ## f	originInfo/publisher@type='c'	Name of printer, manufacturer or engraver and other printing	repeatable	optional		x	/collection/record/datafield[@tag="260"]/subfield[@code="c"]
		place of printing, manufacture or distribution	marc:260 ## e	originInfo/place@placeTerm/originInfo/dateCreated	The place of printing, manufacture or engraving and the name of the Date of printing or manufacture. When the date of printing or manufacture is given in place of an unknown date of publication, production or distribution, it is not repeated here.	repeatable	optional		x	/collection/record/datafield[@tag="260"]/subfield[@code="e"]
	specific material designation (5.1)	marc:260 ## g		physicalDescription/form	The form of communication through which a work is expressed. Used in conjunction with General , which indicates the general type of content of the resource. Field 336 information enables expression of more specific content types and content types from various lists.	repeatable	optional		x	/collection/record/datafield[@tag="260"]/subfield[@code="g"]
Physical Description Area		extent (5.1)	marc:336 ## a	physicalDescription/extent	Number of pages, volumes, etc. of the physical or digital unit or units constituting the resource, adding other measures of extent as appropriate.	repeatable	optional		x	/collection/record/datafield[@tag="336"]/subfield[@code="a"]
		publication frequency	marc:300 ## a	originInfo/frequency	Complete statement, exclusive of dates, of the current publication frequency	repeatable	optional		x	/collection/record/datafield[@tag="300"]/subfield[@code="a"]
		other physical details (5.2)	marc:310 ## a	physicalDescription/note@displayLabel="other physical details"	Other physical characteristics of the resource, such as method of production, colour, material from which the resource is made.	repeatable	optional		x	/collection/record/datafield[@tag="310"]/subfield[@code="a"]
		dimensions (5.3)	marc:300 ## b	physicalDescription/dimensions	Dimensions of the resource, such as size and duration.	repeatable	optional		x	/collection/record/datafield[@tag="300"]/subfield[@code="b"]
		abstract	marc:300 ## c	abstract	Summary of the content of the resource	repeatable	optional		x	/collection/record/datafield[@tag="300"]/subfield[@code="c"]
		table of contents	marc:505 0# a	tableOfContents	Table of contents of the resource	repeatable	optional		x	/collection/record/datafield[@tag="505"]/subfield[@code="a"]
Subject Area		keywords	marc:653 ## a	subject/topic (uncontrolled)	A list of keywords associated with or covered by the described materials, including persons, corporations, materials, topical terms as well as places.	repeatable	optional		x	/collection/record/datafield[@tag="653"]/subfield[@code="a"]
		subject - person	marc:600 14 a	subject/name@type='person'	Name(s) of one or more persons on which a work is focused.	repeatable	optional		x	/collection/record/datafield[@tag="600"]/subfield[@code="a"]
		subject - corporate	marc:610 24 a	subject/name@type='corporate'	Name(s) of one or more corporate bodies on which a work is focused.	repeatable	optional		x	/collection/record/datafield[@tag="610"]/subfield[@code="a"]
		subject - genre	marc:655 #4 a	subject/genre	The style or technique of the intellectual content of textual, graphic, ...	repeatable	optional		x	/collection/record/datafield[@tag="655"]/subfield[@code="a"]
		subject - topic	marc:650 #4 a	subject/topic	A term or phrase representing the primary topic(s) on which a work is focused.	repeatable	optional		x	/collection/record/datafield[@tag="650"]/subfield[@code="a"]
		subject - geographic	marc:651 #4 a	subject/geographic	Name(s) of one or more geographic entities on which a work is focused.	repeatable	optional		x	/collection/record/datafield[@tag="651"]/subfield[@code="a"]
		subject - temporal	marc:648 #4 a	subject/temporal	Temporal term(s), date(s) or time period(s) on which a work is focused	repeatable	optional		x	/collection/record/datafield[@tag="648"]/subfield[@code="a"]

HOPE Visual Profile: Mapping Worksheet									Local Element	Mapping Local-LIDO
area	sub-area	visual element	LIDO/ Premis / Europeana Namespace	definition	occurrences	obligation	conditions		LIDO element	
Digital Resource Area	administrative metadata	type	ens:type	The European material type of the digital representation, which is used to create the Europeana Type facet search.	single value	mandatory		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceDescription[@lido:type="europeana type"]	
		language	premis:significantProperties	The language(s) of the intellectual content of the digital representation.	repeatable	mandatory, if	mandatory if the digital representation contains text	x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceDescription[@lido:type="language"]	
		rights	ens:rights	Statement about the rights held in or over the digital representation	single value	mandatory		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:rightsResource/lido:creditLine	
	derivative 2	resolve URL	ens:isShownBy	An unambiguous URL which allows retrieving a low resolution derivative of the digital representation from a Shared or Local Object Repository.	single value	mandatory, if	mandatory if the CP does not create PIDs for low resolution derivative files	x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="derivative 2"]/lido:linkResource	
		local ID	ens:isShownBy	An unambiguous reference to a low resolution derivative of the digital representation, identifying this derivative in the local system of the content provider.	single value	mandatory, if	mandatory if the SOR creates PIDs for low resolution derivative files	x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="derivative 2"]/lido:linkResource	
		persistent ID	ens:isShownBy	A globally unique, resolvable URL which allows identifying and retrieving a low-resolution derivative of the digital representation in- and outside the HOPE information system	single value	mandatory, if	mandatory if the CP creates PIDs for low resolution derivative files	x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="derivative 2"]/lido:linkResource	
	derivative 3	sequence	ens:isNextInSequence	An identifier for the digital representation that comes immediately after the	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="derivative 3"]/lido:linkResource	
		resolve URL	ens:object	An unambiguous URL which allows retrieving a thumbnail of the digital representation from a Shared or Local Object Repository.	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="derivative 3"]/lido:linkResource	
		local ID	ens:object	An unambiguous reference to a thumbnail of the digital representation, identifying this thumbnail in the local system of the content provider.	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="derivative 3"]/lido:linkResource	
	transcription	persistent ID	ens:object	A globally unique, resolvable URL which allows identifying and retrieving a thumbnail of the digital representation in- and outside the HOPE information system	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="derivative 3"]/lido:linkResource	
		transcription	premis:significantProperties	A manually or automatically generated transcription of the textual content of a digital representation.	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceDescription[@lido:type="transcription"]	
		resolve URL	premis:relatedObjectIdentifi	An unambiguous URL reference to a text file containing the transcription.	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="transcription"]/lido:linkResource	
	landing page	local ID	premis:relatedObjectIdentifi	An unambiguous reference to a text file containing the transcription, identifying this text file in the local system of the content provider.	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="transcription"]/lido:linkResource	
		persistent ID	premis:relatedObjectIdentifi	A globally unique, resolvable URL which allows identifying and retrieving a text file containing the transcription, in- and outside the HOPE information system	single value	optional		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="transcription"]/lido:linkResource	
		resolve URL	ens:landingPage	An unambiguous URL reference to the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP does not create PIDs for landing pages	x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="landing page"]/lido:linkResource	
		persistent ID	ens:landingPage	A globally unique, resolvable URL which allows identifying and retrieving the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP creates PIDs for landing pages	x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:resourceWrap/lido:resourceSet/lido:resourceRepresentation[@lido:type="landing page"]/lido:linkResource	
		identification area	object number	lido:workID	A unique number identifying the collection item in the local information system	single value	mandatory		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectIdentificationWrap/lido:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/lido:object/lido:objectID[@lido:type="localID"]
			local ID	lido:lidoRecID	An unambiguous reference to an entity in the local system of the content provider.	single value	mandatory, if	mandatory if a CP does not create PIDs for descriptive	x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectIdentificationWrap/lido:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/lido:object/lido:objectID[@lido:type="localID"]
persistent ID	lido:lidoRecID		A globally unique, resolvable URL, identifying the described materials in the local system of the content provider.	single value	mandatory, if	mandatory if a CP creates PIDs for descriptive	x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectIdentificationWrap/lido:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/lido:object/lido:objectID[@lido:type="PID"]		
parent record	local ID	dcterms:isPartOf	A reference number or otherwise uniquely identifying number of a related resource.	single value	mandatory, if	mandatory if a CP does not create PIDs for descriptive	x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/lido:object/lido:objectID[@lido:type="PID"]		
	persistent identifier next	dcterms:isPartOf	A persistent identifier of a related resource in which the described materials are related to.	single value	mandatory, if	mandatory if a CP creates PIDs for descriptive	x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/lido:object/lido:objectID[@lido:type="PID"]		
	local ID	ens:isNextInSequence	A local identifier of another descriptive unit, as part of an ordered series of unit, that should be displayed immediately after the given unit.	single value	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/lido:object/lido:objectID[@lido:type="localID"]		
sequence	persistent ID	ens:isNextInSequence	A persistent identifier of another descriptive unit, as part of an ordered series of unit, that should be displayed immediately after the given unit.	single value	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:relatedWorksWrap/lido:relatedWorkSet/lido:relatedWork/lido:object/lido:objectID[@lido:type="persistentID"]		
	level of description	lido:recordType	A term identifying the hierarchical level of description	single value	mandatory		x	/lido:lidoWrap/lido:lido/administrativeMetadata/lido:recordWrap/lido:recordType		
	object name	lido:objectWorkType	A description of the form, function or type of visual resource.	single value	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
brief description	title	lido:titleSet	The name assigned to an object or group of objects by the artist/creator or publisher.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
	brief description	lido:objectDescriptionSet	A textual description of the object, used to identify the object. The description is a transcription or description of any distinguishing or identifying	single value	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
	inscription	lido:inscriptionTranscription	A transcription or description of any distinguishing or identifying	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
Creation Area	name	lido:objectProductionPerson	A person or organisation involved in the design, creation or manufacture of an object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
	person	lido:objectProductionPerson	A person or organisation involved in the design, creation or manufacture of an object. This may be a person or an organisation.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
	organisation	lido:objectProductionPerson	A person or organisation involved in the design, creation or manufacture of an object. This may be a person or an organisation.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
date	object production date	lido:eventSet > lido:eventDate	The date when a stage in the design, creation or manufacture of an object took place.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
	place	lido:eventSet > lido:eventPlace	A place where the design, creation or manufacture of an object took place.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
	technique	lido:eventSet > lido:eventMaterial	Processes, methods, techniques or tools used to fabricate or decorate an object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
Physical Description Area	material	lido:eventSet > lido:eventMaterial	The basic materials and media from which an object is constructed.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		
	dimension	lido:objectMeasurementsSet	The dimensions of the object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/descriptiveMetadata/lido:objectClassificationWrap/lido:objectTitle		

		technical attribute	lido:objectDescriptionSet	The name of a technical attribute possessed by an object which can be described and quantified.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@lido:type="technical attribute"]/lido:descriptiveNoteValue
Depiction Area	person	depicted person	lido:subjectActor	A person depicted in or described by an object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	organisation	depicted organisation	lido:subjectActor	The organisation depicted in or described in an object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	date	depicted date	lido:subjectDate	A date depicted in or described by an object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject[@lido:type="depiction"]/lido:subjectDate/lido:displayDate
	place	depicted place	lido:subjectPlace	A place depicted in or described by an object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	object name	depicted object	lido:subjectObject	An object depicted in or described by another object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject[@lido:type="depiction"]/lido:subjectObject/lido:displayObject
	concept	depicted concept	lido:subjectConcept	A concept depicted in or described by an	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
Association Area	event	depicted event	lido:subjectEvent	An event depicted in or described by an object.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	keywords	keywords	lido:displaySubject	A list of keywords associated with or covered by the described materials, including persons, organisations, materials, topical terms as well as places.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject/lido:displaySubject
	person	associated person	lido:subjectActor	A person associated with an object's or group of objects' history.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	organisation	associated organisation	lido:subjectActor	An organisation associated with an object's or group of objects' history.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	object	associated object	lido:subjectObject	An object associated with an object or group of objects.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject[@lido:type="association"]/lido:subjectObject/lido:displayObject
	concept	associated concept	lido:subjectConcept	A concept associated with an object or group of objects.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	place	associated place	lido:subjectPlace	A place associated with an object or group of objects.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	date	associated date	lido:subjectDate	A date associated with an object or group of objects.	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject[@lido:type="association"]/lido:subjectDate/lido:displayDate
	event	associated event name	lido:subjectEvent	An historical event associated with an object or group of objects, not including	repeatable	optional		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectRelationWrap/lido:subject
	Conditions of Access and Use A	owner	owner	lido:eventSet > lido:eventA	Details of a people, person or organisation who owned an object before title was transferred to the organisation.	repeatable	optional		x
		repository	lido:repositoryName	The name of the archive or other institution in possession of the collection item.	repeatable	mandatory		x	/lido:lidoWrap/lido:lido/lido:descriptiveMetadata/lido:objectIdentificationWrap/lido:repositoryWrap/lido:repositoryName/lido:legalBodyName
		credit line	lido:creditLine	Acknowledgement of the rights associated with the physical and/or digital object as requested.	repeatable	optional		x	/lido:lidoWrap/lido:lidoWrap/lido:lido/lido:administrativeMetadata/lido:rightsWorkWrap/lido:rightsWorkSet/lido:creditLine

HOPE Visual Profile: Collection Form							Local Element	Mapping Local-LIDO		
area	sub-area	collection element	Dublin Core Collection Application Profile / Europeana Namespace	definition	occurrences	obligation	conditions	LIDO element		
Identification Area	Description Control Area	data provider	ens:dataProvider	Name of the institution supplying the metadata to the HOPE	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:RecID[@lido:type="localID"]/lido:source	
		language metadata	dc:language	Primary language of the metadata describing the resource.	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/@xml:lang	
	Collection Identifier	local ID	dc:identifier	An unambiguous reference to an entity in the local system of the	single value	mandatory, if	mandatory if the CP does not	x	/lido:lidoWrap/lido:lido:lido:RecID[@lido:type="localID"]	
		persistent ID	dc:identifier	A persistent identifier, which allows identifying the collection in	single value	mandatory, if	mandatory if the CP creates	x	/lido:lidoWrap/lido:lido:lido:RecID[@lido:type="PID"]	
	Description Level	description level	hope:descriptionLevel	Identifying the hierarchical level of description, i.e. collection	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:administrativeMetadata/lido:recordWrap/lido:recordType	
		Title	title	dc:title	The name of the collection.	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectIdentificationWrap/lido:titleWrap/lido:titleSet/lido:appellationValue
	Collector	alternative title	dcterms:alternative	Any form of the name used as a substitute or alternative to the formal name of the collection.	repeatable	optional		x	x	
		collector - person	dc:creator	A person who gathers (or gathered) the items in a collection together.	repeatable	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType/lido:term="collecting"/lido:eventActor/lido:actor[@lido:type="person" and /lido:roleActor="owner"]/lido:nameActorSet/lido:appellationValue	
	collector - organisation	dc:creator	An organisation who gathers (or gathered) the items in a collection together.	repeatable	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType/lido:term="collecting"/lido:eventActor/lido:actor[@lido:type="organisation" and /lido:roleActor="owner"]/lido:nameActorSet/lido:appellationValue		
	Date	date created	cd:dateltmsCreated	A controlled date value over which the individual items within the collection were created.	single value	mandatory		x	/lido:lidoWrap/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType="creation"/lido:eventDate/lido:date/lido:earliestDate /lido:lidoWrap/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType="creation"/lido:eventDate/lido:date/lido:latestDate	
		date created statement	cd:dateltmsCreated	A statement on the dates over which the individual items within the collection were created.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType="creation"/lido:eventDate/lido:displayDate	
		date accumulated	dcterms:created	A controlled date value over which the individual items within the collection were collected.	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType="collecting"/lido:eventDate/lido:date/lido:earliestDate /lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType="collecting"/lido:eventDate/lido:date/lido:latestDate	
		date accumulated statement	dcterms:created	A statement on the dates over which the individual items within the collection were collected.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType="creation"/lido:eventDate/lido:displayDate	
	Physical Description Area:Size	Type	size in items	dcterms:extent	A statement on the number of items in the collection.	single value	mandatory, if	mandatory if the size of your archive can be expressed in number of items	x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="size in items"]/lido:descriptiveNoteValue
			item type	cd:itemType	A term giving the nature or genre of one more items within the collection.	repeatable	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectClassificationWrap/lido:objectWorkTypeWrap/lido:objectWorkType/lido:term
item type statement		cd:itemType	A statement on the nature or genre of the items within the collection.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="item type statement"]/lido:descriptiveNoteValue		
Format		item format	cd:itemFormat	A term giving a media type, physical or digital, of one or more items within the collection.	repeatable	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType="creation"/lido:eventMaterialsTech/lido:materialsTech/lido:termMaterialsTech[@lido:type="material"]/lido:term	
		item format statement	cd:itemFormat	A statement giving the media type, physical or digital, of one or more items within the collection.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="item format statement"]/lido:descriptiveNoteValue	
language		dc:language	The language(s) of the items within the collection.	repeatable	mandatory, if	mandatory if your archive includes textual resources	x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectClassificationWrap/lido:classification[@type="language"]/term		
Content Area	language statement	dc:language	A statement on the language of the items within the collection.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="language statement"]/lido:descriptiveNoteValue		
	abstract	dcterms:abstract	A free text summary description of the collection.	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="abstract"]/lido:descriptiveNoteValue		
	custodial history	dcterms:provenance	A statement of any changes in ownership and custody of the collection that are significant for its authenticity, integrity and interpretation.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="custodial history"]/lido:descriptiveNoteValue		
	accrual policy	dcterms:accrualPolicy	A statement on the policy governing the addition of items to the collection.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="accrual policy"]/lido:descriptiveNoteValue		
	super-collection	dcterms:isPartOf	A second collection that contains the current collection.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:relatedWorksWrap/lido:relatedWorksSet/lido:relatedWorkType="super collecting"/lido:relatedWork/lido:displayDate		
	catalogue or index	cd:catalogueOrIndex	A catalogue for, or index of, the collection	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="catalogue or index"]/lido:descriptiveNoteValue		
	associated publication	dcterms:isReferencedBy	A publication that is based on the use, study, or analysis of the collection.	repeatable	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectDescriptionWrap/lido:objectDescriptionSet[@type="associated publication"]/lido:descriptiveNoteValue		
	Association Area	subject	dc:subject	A value for a subject or topic of the collection.	repeatable	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject/lido:displaySubject	
		spatial coverage	dcterms:spatial	A statement on the spatial scope of the collection.	repeatable	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject[@type="association"]/lido:subjectPlace/lido:place/lido:namePlaceSet/lido:appellationValue	
		temporal coverage	dcterms:temporal	A statement on the temporal scope of the collection.	repeatable	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectRelationWrap/lido:subjectWrap/lido:subjectSet/lido:subject[@type="association"]/lido:subjectDate/lido:displayDate	
Conditions of Access and Use Area	owner	marcrel:OWN	An entity who has legal possession of the collection.	single value	optional		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:eventWrap/lido:eventSet/lido:event/lido:eventType/lido:term="ownership"/lido:eventActor/lido:actor/lido:roleActor="owner"/lido:nameActorSet/lido:appellationValue		
	repository	cd:isLocatedAt	archival institution or agency providing access to the described materials.	repeatable	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:objectIdentificationWrap/lido:repositoryWrap/lido:repositoryName/lido:legalBodyName		
	use rights	dc:rights	A statement of any rights held in/over the collection.	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:rightsWorkWrap/lido:rightsWorkSet/lido:rightsType="use rights"/lido:creditLine		
	access rights	dcterms:accessRights	A statement explicating the access restrictions placed over the analog and digital content, including allowed users, charges, formats, etc.	single value	mandatory		x	/lido:lidoWrap/lido:lido:lido:descriptiveMetadata/lido:rightsWorkWrap/lido:rightsWorkSet/lido:rightsType="access rights"/lido:creditLine		

HOPE Audio-Visual Profile

area	sub-area	audio-visual elements	definition	occurrences	cardinality	conditions	
description control area		content provider	Name of the institution supplying the metadata to the HOPE aggregator.	single value	mandatory	x	
		aggregator	Name of the aggregator, supplying metadata to one or more discovery services	single value	mandatory	x	
		country content provider	Name of the country in which the Content Provider is based.	single value	mandatory	x	
		language content provider	Official language of the country where the Content Provider is located.	repeatable	mandatory	x	
		language metadata domain	Primary language of the metadata describing the collection item(s). Type of the metadata profile, providing domain-specific context for the HOPE metadata	single value	mandatory	x	
digital resource area	identifier	resolve URL digital resource	URL directing to a webpage displaying the metadata about a digital resource, i.e. structural and administrative metadata and references to all available versions of the digital representation.	single value	conditional	mandatory if the CP stores its digital files in a LOR and the PIDs for the digital files are created by the Aggregator using the HOPE PID Service.	
		local identifier digital resource	unambiguous reference to the digital representation of a collection item, as it is identified in the local system of the Content Provider.	repeatable	conditional	mandatory if the CP does not create PIDs for digital resources	
		persistent identifier digital resource	PID, identifying the digital representation of a collection item in the HOPE System	repeatable	conditional	mandatory if the CP creates PIDs for digital resources	
	sequence	local identifier next digital resource in sequence	local identifier of another digital representation as part of an ordered series of representations, that should be displayed immediately after the given digital representation.				
		persistent identifier next digital resource in sequence	PID of another digital representation as part of an ordered series of representations, that should be displayed immediately after the given representation.	single value	optional	x	
	administrative metadata	europena type	The Europeana material type of the digital resource, which is used to create the Europeana Type facet search.	single value	mandatory	x	
		language digital content	Language(s) of the intellectual content of the digital resource.	repeatable	conditional	mandatory if the digital representation contains text	
		copyright	Information about copyright that apply to the digital resource	single value	mandatory	x	
		dissemination rights	information about the discovery services to which the digital resource is made available.	single value	mandatory	x	
	landing page	resolve URL landing page	An unambiguous URL reference to the digital object on the provider's OPAC in its full information context.	single value	conditional	mandatory if the CP does not create PIDs for landing pages	
		persistent identifier landing page	PID, identifying the landing page in the HOPE information system	single value	conditional	mandatory if the CP creates PIDs for landing pages	
	derivative	resolve URL thumbnail file	An unambiguous URL reference to a thumbnail file, stored in the Shared or Local Object Repository.	single value	optional	x	
		persistent identifier thumbnail file	PID, identifying the thumbnail file in the HOPE information system	single value	optional	x	
		resolve URL high resolution derivative file	An unambiguous URL reference to a high resolution derivative file, stored in the Shared or Local Object Repository.	single value	optional	x	
		persistent identifier high resolution derivative file	PID, identifying the high-resolution derivative file in the HOPE information system	single value	optional	x	
		resolve URL low resolution derivative file	An unambiguous URL reference to a low resolution derivative file, stored in the Shared or Local Object Repository.	single value	conditional	mandatory if the CP does not create PIDs for low resolution derivative files	
		persistent identifier low resolution derivative file	PID, identifying the low-resolution derivative file in the HOPE information system	single value	conditional	mandatory if the CP creates PIDs for low resolution derivative files	
		resolve URL preview derivative file	An unambiguous URL reference to a preview derivative file, stored in the Shared or Local Object Repository.	single value	optional	x	
		persistent identifier preview derivative file	PID, identifying the preview derivative file in the HOPE information system	single value	optional	x	
resolve URL OCRred text derivative file		An unambiguous URL reference to an OCRred text derivative file, stored in the Shared or Local Object Repository.	single value	optional	x		
persistent identifier OCRred text derivative file		PID, identifying the OCRred text derivative file in the HOPE information system	single value	optional	x		
Identification Area	identifier	inventory number	An identifier such as a call number, shelf mark or similar, required by the holding institution to uniquely identify the copy.	single value	mandatory	x	
		local identifier descriptive unit	An unambiguous reference to an entity in the local system of the content provider describing the collection item(s).	single value	conditional	mandatory if a CP does not create PIDs for descriptive units	
		persistent identifier descriptive unit	persistent identifier, identifying the described materials in the HOPE information system	single value	conditional	mandatory if a CP creates PIDs for descriptive units	

HOPE Audio-Visual Profile

area	sub-area	audio-visual elements	definition	occurrences	cardinality	conditions	
	parent record	original identifier parent descriptive unit	reference number or otherwise uniquely identifying number of a related resource in which the described materials are physically or logically included.	single value	conditional	mandatory if a CP does not create PIDs for descriptive units	
		persistent identifier parent descriptive unit	persistent identifier of a related resource in which the described materials are physically or logically included.	single value	conditional	mandatory if a CP creates PIDs for descriptive units	
	sequence	original identifier next descriptive unit in sequence	original identifier of another descriptive unit, as part of an ordered series of unit, that should be displayed immediately after the given unit.	single value	optional	x	
		persistent identifier next descriptive unit in sequence	persistent identifier of another descriptive unit, as part of an ordered series of unit, that should be displayed immediately after the given unit.	single value	optional	x	
	level of description	level of description	identifying the hierarchical level of description	single value	mandatory	x	
	instantiation type	instantiation type	A term or phrase describing the item type relative to the duplication process (e.g. "original negative", "dupnegative", "positive", "original positive (reversal film)", "dupe positive", "Lavender", "image negative", "sound negative", "non-film analogue carrier", "non-film digital carrier").	repeatable	optional	x	
	title	title	A word, phrase, character, or group of characters, naming the work, a group of works, a particular variant or manifestation, or an individual item. Depending on the cataloguing rules in use, titles may be determined from the item itself, or from any suitable secondary source.	repeatable	optional	x	
		translated title	Title translated into another language and/or script	repeatable	optional	x	
	Creation Area	production	Cast	Person acting or appearing in the cinematographic work	repeatable	optional	x
			Director - Person	Person responsible for directing the cinematographic work	repeatable	optional	x
Director - Organisation			Organisation responsible for directing the cinematographic work	repeatable	optional	x	
Producer			Person responsible for producing the cinematographic work	repeatable	optional	x	
Production Company			Organisation responsible for producing the cinematographic work	repeatable	optional	x	
Production Location			Any geographic name(s) or address(es) of the location(s) where the event took place.	repeatable	optional	x	
Production Date			The date or time-span during which the event took place.	repeatable	optional	x	
publication		Distributor	Organisation responsible for distribution of the cinematographic work	repeatable	optional	x	
		Publication Location	The name of the city or smaller geographic entity where the event took place.	repeatable	optional	x	
		Publication Date	The date on which the exhibition took place, or the date on which items of the manifestation became available for purchase or hire.	repeatable	optional	x	
Physical Description Area		Carrier Type	The type of physical carrier used for the manifestation. Digital manifestations may be bound to a physical carrier (such as DVD) or exist without a defined carrier (such as online streaming media, podcasts, etc.).	single value	optional	x	
		Gauge	The width of the film stock or other carrier (such as magnetic tape) used for the manifestation. Should include value and unit (e.g. "35 mm", "1/2 in.")	single value	optional	x	
		Original Duration	The total physical, logical or temporal extent of the manifestation of a cinematographic work, or of the item from a manifestation. Physical extent is the total length of the medium carrying the manifestation or item. Logical extent is the number of discrete units, e.g. image frames, bytes (octets), etc., of which the manifestation or item is composed. Temporal extent is the running time under normal or specific playback conditions (e.g. a specific frame rate).	repeatable	optional	x	
		Original Length	The total physical, logical or temporal extent of the manifestation of a cinematographic work, or of the item from a manifestation. Physical extent is the total length of the medium carrying the manifestation or item. Logical extent is the number of discrete units, e.g. image frames, bytes (octets), etc., of which the manifestation or item is composed. Temporal extent is the running time under normal or specific playback conditions (e.g. a specific frame rate).	repeatable	optional	x	

HOPE Audio-Visual Profile

area	sub-area	audio-visual elements	definition	occurrences	cardinality	conditions
		Colour System	A name for the colour system or process	single value	optional	x
		Sound System	The name of the system by which sound is recorded either on the carrier, on a separate medium, or as part of the digital encoding (e.g. Western Electric (Westrex) Movietone, RCA Photophone, Tobis, Dolby Digital, etc.).	single value	optional	x
		Aspect Ratio	The ratio between width and height of the image. For anamorphic formats, this element should contain the aspect ratio of the projected image rather than that on the carrier. Some aspect ratios are known not only by numeric values, but also by names (e.g. "full frame", "Cinemascope").	single value	optional	x
Content Area		Synopsis	Summary. A textual description of the contents of the cinematographic work.	repeatable	optional	x
		Translated Synopsis	Summary translated in another language	repeatable	optional	x
		Shotlist	Shotlist. A textual description of the contents of the cinematographic work.	repeatable	optional	x
		Language Used	Secondary languages, such as dubbed languages	repeatable	optional	x
		Subtitle Language	Language used for subtitles	single value	optional	x
		Original Language	The language or languages of the spoken, sung or written content of the variant or manifestation of a cinematographic work.	repeatable	optional	x
Association Area		Subject - Person	Person. A term or set of terms describing the content of the cinematographic work.	repeatable	optional	x
		Subject - Organisation	Organisation. A term or set of terms describing the content of the cinematographic work.	repeatable	optional	x
		Subject - Genre	Genre. A term or set of terms describing the content of the cinematographic work.	repeatable	optional	x
		Subject - Topic	Topic. A term or set of terms describing the content of the cinematographic work.	repeatable	optional	x
		Subject - Place	Place. A term or set of terms describing the content of the cinematographic work.	repeatable	optional	x
		Country of Reference	An element used for describing the geographic origin of a	repeatable	optional	x
		Year of Reference	work, typically associated with its creation, availability or registration (for	repeatable	optional	x
Conditions of Access and Use Area		Holding institution	The name of the archive or other institution possessing the copy or authorised to make it available. Alternatively, if available, a suitable URI (e.g. from the ISIL namespace according to ISO 15511), or a registered namespace identifier for the institution.	repeatable	optional	x
		IPR Registration	Organisation claiming copyright in the cinematographic work.	repeatable	optional	x
		Access Conditions	Specific restrictions for accessing the content of the manifestation.	repeatable	optional	x

HOPE Dublin Core Profile: Collection Form								Local Element	Mapping Local-Dublin Core Qualified	
area	sub-area	collection element	Dublin Core Collection Application Profile / Europeana Namespace	definition	occurrences	obligation	conditions		DC Qualified element	ESE 3.4 element
Description Control Area		data provider	ens:dataProvider	Name of the institution supplying the metadata to the HOPE	single value	mandatory		x	x	/europeana:metadata/europeana:record/europeana:dataPr
		language metadata	dc:language	Primary language of the metadata describing the resource.	single value	mandatory		x	/metadata/record/dc:language (until first separator ';')	/europeana:metadata/europeana:record/dc:language
Identification Area	Collection Identifier	local ID	dc:identifier	An unambiguous reference to an entity in the local system of the	single value	mandatory, if	mandatory if the CP does not create	x	/metadata/record/dc:identifier	/europeana:metadata/europeana:record/dc:identifier
		persistent ID	dc:identifier	A persistent identifier, which allows identifying the collection in	single value	mandatory, if	mandatory if the CP creates PIDs for	x	/metadata/record/dc:identifier[@type="URI"]	/europeana:metadata/europeana:record/dc:identifier[@typ
	Description Level	description level	hope:descriptionLevel	Identifying the hierarchical level of description, i.e. collection	single value	mandatory			/metadata/record/dc:type	/europeana:metadata/europeana:record/dc:type
		title	dc:title	The name of the collection.	single value	mandatory			/metadata/record/dc:title	/europeana:metadata/europeana:record/dc:title
	Title	alternative title	dcterms:alternative	Any form of the name used as a substitute or alternative to the formal name of the collection.	repeatable	optional			/metadata/record/dcterms:alternative	/europeana:metadata/europeana:record/dcterms:alternativ
		collector - person	dc:creator	A person who gathers (or gathered) the items in a collection together.	repeatable	optional			/metadata/record/dc:creator (until separator ';')	/europeana:metadata/europeana:record/dc:creator
	Collector	collector - organisation	dc:creator	An organisation who gathers (or gathered) the items in a collection together.	repeatable	optional			/metadata/record/dc:creator (from separator ';')	/europeana:metadata/europeana:record/dc:creator
		Date	date created	cid:dateItemsCreated	A controlled date value over which the individual items within the collection were created.	single value	mandatory			/metadata/record/cid:dateItemsCreated (until separator ';')
	date created statement		cid:dateItemsCreated	A statement on the dates over which the individual items within the collection were created.	single value	optional			/metadata/record/cid:dateItemsCreated (from separator ';')	/europeana:metadata/europeana:record/dcterms:created
	date accumulated		dcterms:created	A controlled date value over which the individual items within the collection were collected.	single value	mandatory			/metadata/record/dcterms:created (until separator ';')	/europeana:metadata/europeana:record/dcterms:created
date accumulated statement	dcterms:created		A statement on the dates over which the individual items within the collection were collected.	single value	optional			/metadata/record/dcterms:created (from separator ';')	/europeana:metadata/europeana:record/dcterms:created	
Physical Description Area:Size	Type	size in items	dcterms:extent	A statement on the number of items in the collection.	single value	mandatory, if	mandatory if the size of your archive can be expressed in number of items	x	/metadata/record/dcterms:extent	/europeana:metadata/europeana:record/dcterms:extent
		item type	cid:itemType	A term giving the nature or genre of one more items within the collection.	repeatable	mandatory			/metadata/record/cid:itemType (until separator ';')	/europeana:metadata/europeana:record/dcterms:medium
	Format	item type statement	cid:itemType	A statement on the nature or genre of the items within the collection.	single value	optional			/metadata/record/cid:itemType (from separator ';')	/europeana:metadata/europeana:record/dcterms:medium
		item format	cid:itemFormat	A term giving a media type, physical or digital, of one or more items within the collection.	repeatable	optional			/metadata/record/cid:itemFormat (until separator ';')	/europeana:metadata/europeana:record/dc:format
		item format statement	cid:itemFormat	A statement giving the media type, physical or digital, of one or more items within the collection.	single value	optional			/metadata/record/cid:itemFormat (from separator ';')	/europeana:metadata/europeana:record/dc:format
		language	dc:language	The language(s) of the items within the collection.	repeatable	mandatory, if	mandatory if your archive includes textual resources		x	/metadata/record/dc:language (until second separator ';')
Content Area	language statement	language statement	dc:language	A statement on the language of the items within the collection.	single value	optional			/metadata/record/dc:language (from second separator ';')	/europeana:metadata/europeana:record/dc:language
		abstract	dcterms:abstract	A free text summary description of the collection.	single value	mandatory			/metadata/record/dcterms:abstract	/europeana:metadata/europeana:record/dc:description
	custodial history	dcterms:provenance	A statement of any changes in ownership and custody of the collection that are significant for its authenticity, integrity and interpretation.	single value	optional			/metadata/record/dcterms:provenance	/europeana:metadata/europeana:record/dcterms:provenan	
	accrual policy	dcterms:accrualPolicy	A statement on the policy governing the addition of items to the collection.	single value	optional			/metadata/record/dcterms:accrualPolicy	/europeana:metadata/europeana:record/dcterms:provenan	
	super-collection	dcterms:isPartOf	A second collection that contains the current collection.	single value	optional			/metadata/record/dcterms:isPartOf	/europeana:metadata/europeana:record/dcterms:isPartOf	
	catalogue or index	cid:catalogueOrIndex	A catalogue for, or index of, the collection	single value	optional			/metadata/record/cid:catalogueOrIndex	/europeana:metadata/europeana:record/dcterms:tableOfCo	
	associated publication	dcterms:isReferencedBy	A publication that is based on the use, study, or analysis of the collection.	repeatable	optional			/metadata/record/dcterms:isReferencedBy	/europeana:metadata/europeana:record/dcterms:isReferen	
	subject	dc:subject	A value for a subject or topic of the collection.	repeatable	optional			/metadata/record/dc:subject	/europeana:metadata/europeana:record/dc:subject	
Association Area	spatial coverage	dcterms:spatial	A statement on the spatial scope of the collection.	repeatable	optional			/metadata/record/dcterms:spatial	/europeana:metadata/europeana:record/dcterms:spatial	
	temporal coverage	dcterms:temporal	A statement on the temporal scope of the collection.	repeatable	optional			/metadata/record/dcterms:temporal	/europeana:metadata/europeana:record/dcterms:temporal	
Conditions of Access and Use Area	owner	owner	marcrel:OWN	An entity who has legal possession of the collection.	single value	optional			/metadata/record/marcrel:OWN	/europeana:metadata/europeana:record/dcterms:provenan
		repository	cid:isLocatedAt	archival institution or agency providing access to the described materials.	repeatable	mandatory			/metadata/record/cid:isLocatedAt	/europeana:metadata/europeana:record/dcterms:provenan
	use rights	dc:rights	A statement of any rights held in/over the collection.	single value	mandatory			/metadata/record/dc:rights	/europeana:metadata/europeana:record/dc:rights	
	access rights	dcterms:accessRights	A statement explicating the access restrictions placed over the analog and digital content, including allowed users, charges, formats, etc.	single value	mandatory			/metadata/record/dcterms:accessRights	/europeana:metadata/europeana:record/dc:rights	

HOPE Dublin Core Profile: Mapping Worksheet								Local Element	Mapping Local-Dublin Core		
area	sub-area	dublin core element	Dublin Core Qualified/ Premis / Europeana Namespace	definition	occurrences	obligation	conditions		DC Qualified element	ESE 3.4	
Digital Resource Area	administrative metadata	type	ens:type	The Europeana material type of the digital representation, which is used to create the Europeana Type facet search.	single value	mandatory		x	/metadata/record/dc:type	/europeana:metadata/europeana:record/europeana:type	
		language	premis:significantProperties	The language(s) of the intellectual content of the digital representation.	repeatable	mandatory, if	mandatory if the digital representation contains text	x	/metadata/record/dc:language	/europeana:metadata/europeana:record/dc:language	
		rights	ens:rights	Statement about the rights held in or over the digital representation	single value	mandatory			x	/metadata/record/dcterms:license	/europeana:metadata/europeana:record/europeana:rights
	derivative 2	resolve URL	ens:isShownBy	An unambiguous URL which allows retrieving a low resolution derivative of the digital representation from a Shared or Local Object Repository.	single value	mandatory, if	mandatory if the CP does not create PIDs for low resolution derivative files	x	/metadata/record/dcterms:hasVersion	/europeana:metadata/europeana:record/europeana:isShownBy[@type="URI"] (from separator ';')	
		local ID	ens:isShownBy	An unambiguous reference to a low resolution derivative of the digital representation, identifying this derivative in the local system of the content provider.	single value	mandatory, if	mandatory if the SOR creates PIDs for low resolution derivative files	x	/metadata/record/dcterms:hasVersion	/europeana:metadata/europeana:record/europeana:isShownBy (until separator ';')	
		persistent ID	ens:isShownBy	A globally unique, resolvable URL which allows identifying and retrieving a low-resolution derivative of the digital representation in- and outside the HOPE information system	single value	mandatory, if	mandatory if the CP creates PIDs for low resolution derivative files	x	/metadata/record/dcterms:hasVersion	/europeana:metadata/europeana:record/europeana:isShownBy[@type="URI"] (from separator ';')	
	derivative 3	resolve URL	ens:object	An unambiguous URL which allows retrieving a thumbnail of the digital representation from a Shared or Local Object Repository.	single value	optional		x	/metadata/record/dcterms:hasFormat	/europeana:metadata/europeana:record/europeana:object[@type="URI"] (from separator ';')	
		local ID	ens:object	An unambiguous reference to a thumbnail of the digital representation, identifying this thumbnail in the local system of the content provider.	single value	optional		x	/metadata/record/dcterms:hasFormat	/europeana:metadata/europeana:record/europeana:object (until separator ';')	
		persistent ID	ens:object	A globally unique, resolvable URL which allows identifying and retrieving a thumbnail of the digital representation in- and outside the HOPE information system	single value	optional		x	/metadata/record/dcterms:hasFormat	/europeana:metadata/europeana:record/europeana:object[@type="URI"] (from separator ';')	
	landing page	resolve URL	ens:landingPage	An unambiguous URL reference to the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP does not create PIDs for landing pages	x	/metadata/record/dcterms:isReferencedBy	/europeana:metadata/europeana:record/europeana:isShownAt	
		persistent ID	ens:landingPage	A globally unique, resolvable URL which allows identifying and retrieving the digital representation on the provider's OPAC, in its full information context.	single value	mandatory, if	mandatory if the CP creates PIDs for landing pages	x	/metadata/record/dcterms:isReferencedBy	/europeana:metadata/europeana:record/europeana:isShownAt	
	Identification area	identifier	resource identifier	dc:identifier	A unique number identifying the collection item in the local information system	single value	mandatory		x	/metadata/record/dc:identifier[@type="URI"]	/europeana:metadata/europeana:record/dc:identifier
			local ID	dc:identifier	An unambiguous reference to an entity in the local system of the content provider	single value	mandatory, if	mandatory if a CP does not create PIDs for	x	/metadata/record/dc:identifier	/europeana:metadata/europeana:record/dc:identifier
			persistent ID	dc:identifier	A globally unique, resolvable URL, identifying the described materials in the local system	single value	mandatory, if	mandatory if a CP creates PIDs for	x	/metadata/record/dc:identifier[@type="URI"]	/europeana:metadata/europeana:record/dc:identifier
parent record		local ID	dcterms:isPartOf	A reference number or otherwise uniquely identifying number of a parent record	single value	mandatory, if	mandatory if a CP does not create PIDs for	x	/metadata/record/dcterms:isPartOf	/europeana:metadata/europeana:record/dcterms:isPartOf	
		persistent ID	dcterms:isPartOf	A persistent identifier of a related resource in which the described resource is part of	single value	mandatory, if	mandatory if a CP creates PIDs for	x	/metadata/record/dcterms:isPartOf	/europeana:metadata/europeana:record/dcterms:isPartOf	
		level of description	dc:type	A term identifying the hierarchical level of description	single value	mandatory		x	/metadata/record/dc:type	/europeana:metadata/europeana:record/dc:type	
description	title	dc:title	A name given to the resource.	single value	optional		x	/metadata/record/dc:title	/europeana:metadata/europeana:record/dc:title		
	description	dc:description	An account of the resource.	single value	optional		x	/metadata/record/dc:description	/europeana:metadata/europeana:record/dc:description		
Creation Area	creator	dc:creator	A person or organisation involved in the design, creation or manufacture of the resource.	repeatable	optional		x	/metadata/record/dc:creator	/europeana:metadata/europeana:record/dc:creator		
	contributor	dc:contributor	An entity responsible for making contributions to the resource.	repeatable	optional		x	/metadata/record/dc:contributor	/europeana:metadata/europeana:record/dc:contributor		
	publisher	dc:publisher	An entity responsible for making the resource available.	repeatable	optional		x	/metadata/record/dc:publisher	/europeana:metadata/europeana:record/dc:publisher		
	date created	dcterms:created	Date of creation of the resource.	repeatable	optional		x	/metadata/record/dcterms:created	/europeana:metadata/europeana:record/dcterms:created		
	date issued	dcterms:issued	Date of formal issuance (e.g., publication) of the resource.	repeatable	optional		x	/metadata/record/dcterms:issued	/europeana:metadata/europeana:record/dcterms:issued		
Physical Description Area	extent	dcterms:extent	The size or duration of the resource.	repeatable	optional		x	/metadata/record/dcterms:extent	/europeana:metadata/europeana:record/dcterms:extent		
	medium	dcterms:medium	The material or physical carrier of the resource.	repeatable	optional		x	/metadata/record/dcterms:medium	/europeana:metadata/europeana:record/dcterms:medium		
Association Area	subject	dc:subject	The topic of the resource.	repeatable	optional		x	/metadata/record/dc:subject	/europeana:metadata/europeana:record/dc:subject		
	spatial coverage	dcterms:spatial	Spatial characteristics of the resource.	repeatable	optional		x	/metadata/record/dcterms:spatial	/europeana:metadata/europeana:record/dcterms:spatial		
	temporal coverage	dcterms:temporal	Temporal characteristics of the resource.	repeatable	optional		x	/metadata/record/dcterms:temporal	/europeana:metadata/europeana:record/dcterms:temporal		
	relation	dc:relation	A related resource.	repeatable	optional		x	/metadata/record/dc:relation	/europeana:metadata/europeana:record/dc:relation		
Conditions of Access and Use Area	provenance	dcterms:provenance	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.	repeatable	optional		x	/metadata/record/dcterms:provenance	/europeana:metadata/europeana:record/dcterms:provenance		
	rights	dc:rights	Information about rights held in and over the resource.	repeatable	optional		x	/metadata/record/dc:rights	/europeana:metadata/europeana:record/dc:rights		

APPENDIX C

```

1: <?xml version="1.0" encoding="utf-8"?>
2: <!--
3: ? HOPE XML SCHEMA v1.1.3
4: -->
5: <xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
6:   xmlns:xs="http://www.w3.org/2001/XMLSchema">
7:   <xs:element xmlns:hope="http://www.peoplesheritage.eu/hope" name="hopeEntity"
8:     type="hopeEntityType">
9:     <xs:annotation>
10:      <xs:documentation> The overall HOPE entity; can contain the following sub-elements
11:        (sub-entities): descriptiveUnit, digitalResource, agent, event, concept, place,
12:        contentProvider, hopeTheme. </xs:documentation>
13:    </xs:annotation>
14:  </xs:element>
15:  <xs:complexType name="hopeEntityType">
16:    <xs:sequence>
17:      <xs:element maxOccurs="1" minOccurs="1" name="persistentID" type="defaultType">
18:        <xs:annotation>
19:          <xs:documentation>globally unique, resolvable identifier, identifying the entity
20:            in- and outside the HOPE system</xs:documentation>
21:        </xs:annotation>
22:      </xs:element>
23:
24:      <xs:element maxOccurs="1" minOccurs="1" name="localID" type="defaultType">
25:        <xs:annotation>
26:          <xs:documentation>an unambiguous reference to a description of one or more
27:            collection items in the local system of the content provider</xs:documentation>
28:        </xs:annotation>
29:      </xs:element>
30:      <xs:choice>
31:        <xs:annotation>
32:          <xs:documentation>only one entity can be chosen, every entity is defined by its
33:            own type</xs:documentation>
34:        </xs:annotation>
35:        <xs:element name="descriptiveUnit" type="DUType"/>
36:        <xs:element name="digitalResource" type="DRType"/>
37:        <xs:element name="agent" type="agentType"/>
38:        <xs:element name="event" type="eventType"/>
39:        <xs:element name="concept" type="conceptType"/>
40:        <xs:element name="place" type="placeType"/>
41:        <xs:element name="contentProvider" type="CPType"/>
42:        <xs:element name="hopeTheme" type="hopeThemeType"/>
43:      </xs:choice>
44:    </xs:sequence>
45:  </xs:complexType>
46:
47:  <xs:complexType name="defaultType">
48:    <xs:annotation>
49:      <xs:documentation> defaultType will be used for adding the default attributes to
50:        elements, only the label attribute is mandatory in all these elements
51:    </xs:documentation>
52:  </xs:annotation>
53:  <xs:simpleContent>
54:    <xs:extension base="xs:string">
55:      <xs:attribute name="label" type="xs:string" use="required">
56:        <xs:annotation>
57:          <xs:documentation>contains a domain-specific label for displaying the value
58:            in the IALHI Portal. </xs:documentation>
59:        </xs:annotation>
60:      </xs:attribute>
61:      <xs:attribute name="encoding" type="xs:string">
62:        <xs:annotation>
63:          <xs:documentation>contains a reference to a metadata standard that is the
64:            source from which the HOPE element has been derived. This attribute
65:            contains a reference to an element name, including the namespace. The
66:            metadata standards include EAD, MARC21 Bibliographic, LIDO, EN15907, DC
67:            Qualified, EDM v5.2.1, PREMIS. Elements that are specific for HOPE refer
68:            to the HOPE namespace. </xs:documentation>
69:          </xs:annotation>
70:        </xs:attribute>
71:        <xs:attribute name="cataloguing" type="xs:string">
72:          <xs:annotation>
73:            <xs:documentation>contains a reference to a metadata data standard that
74:              represents the cataloguing rules for creating the recorded metadata.
75:              This attribute contains the name of the cataloguing rules and the
76:              identification number of the element. The used cataloguing rules include

```

```

77:         ISAD(g) and ISBD 2007 Consolidated. </xs:documentation>
78:     </xs:annotation>
79: </xs:attribute>
80: <xs:attribute name="script" type="xs:string">
81:     <xs:annotation>
82:         <xs:documentation>contains a reference to the script for a transliterated
83:         value. </xs:documentation>
84:     </xs:annotation>
85: </xs:attribute>
86: <xs:attribute name="language" type="xs:string">
87:     <xs:annotation>
88:         <xs:documentation>contains a reference to the language used for a translated
89:         value</xs:documentation>
90:     </xs:annotation>
91: </xs:attribute>
92: <xs:attribute name="localID" type="xs:string">
93:     <xs:annotation>
94:         <xs:documentation>contains the ID, used for identifying the term in the
95:         local system of the CP.</xs:documentation>
96:     </xs:annotation>
97: </xs:attribute>
98: </xs:extension>
99: </xs:simpleContent>
100: </xs:complexType>
101:
102: <xs:complexType name="pidType">
103:     <xs:annotation>
104:         <xs:documentation> pidType adds the same attributes as defaultType with the addition
105:         of a PID attribute to mark the relationships between objects and other entities
106:     </xs:documentation>
107: </xs:annotation>
108: <xs:simpleContent>
109:     <xs:extension base="xs:string">
110:
111:         <xs:attribute name="persistentID" type="xs:string" use="required">
112:             <xs:annotation>
113:                 <xs:documentation>contains a resolvable identifier used for identifying in
114:                 the HOPE system.</xs:documentation>
115:             </xs:annotation>
116:         </xs:attribute>
117:         <xs:attribute name="label" type="xs:string" use="required"/>
118:         <xs:attribute name="encoding" type="xs:string"/>
119:         <xs:attribute name="cataloguing" type="xs:string"/>
120:         <xs:attribute name="script" type="xs:string"/>
121:         <xs:attribute name="language" type="xs:string"/>
122:         <xs:attribute name="localID" type="xs:string"/>
123:     </xs:extension>
124: </xs:simpleContent>
125: </xs:complexType>
126:
127:
128: <xs:complexType mixed="true" name="POType">
129:     <xs:annotation>
130:         <xs:documentation>POType is used for a specific set of elements (creator, contributor,
131:         publisher)</xs:documentation>
132:     </xs:annotation>
133:     <xs:sequence>
134:         <xs:element maxOccurs="1" minOccurs="1" name="value" type="xs:string"/>
135:         <xs:element maxOccurs="unbounded" minOccurs="0" name="place" type="defaultType">
136:             <xs:annotation>
137:                 <xs:documentation>location where the Cultural Heritage Object has been created
138:                 or published </xs:documentation>
139:             </xs:annotation>
140:         </xs:element>
141:         <xs:element maxOccurs="unbounded" minOccurs="0" name="notes" type="defaultType">
142:             <xs:annotation>
143:                 <xs:documentation>additional information on creator, publisher or contributor
144:             </xs:documentation>
145:             </xs:annotation>
146:         </xs:element>
147:     </xs:sequence>
148:     <xs:attribute name="label" type="xs:string" use="required"/>
149:     <xs:attribute name="encoding" type="xs:string"/>
150:     <xs:attribute name="cataloguing" type="xs:string"/>
151:     <xs:attribute name="script" type="xs:string"/>
152:     <xs:attribute name="language" type="xs:string"/>

```

```

153:     <xs:attribute name="localID" type="xs:string"/>
154: </xs:complexType>
155: <xs:complexType name="normalisedType">
156:   <xs:annotation>
157:     <xs:documentation> normalisedType is used for elements which contain both the
158:       original value and a normalised one </xs:documentation>
159:   </xs:annotation>
160:   <xs:simpleContent>
161:     <xs:extension base="xs:string">
162:       <xs:attribute name="label" type="xs:string" use="required"/>
163:       <xs:attribute name="encoding" type="xs:string"/>
164:       <xs:attribute name="cataloguing" type="xs:string"/>
165:       <xs:attribute name="normalised" type="xs:string">
166:         <xs:annotation>
167:           <xs:documentation>contains a normalised version of the term recorded in the
168:             value space.</xs:documentation>
169:         </xs:annotation>
170:       </xs:attribute>
171:       <xs:attribute name="script" type="xs:string"/>
172:       <xs:attribute name="language" type="xs:string"/>
173:     </xs:extension>
174:   </xs:simpleContent>
175: </xs:complexType>
176: <xs:complexType name="versionType">
177:   <xs:complexContent>
178:     <xs:extension base="defaultType">
179:       <xs:attribute name="PID"/>
180:       <xs:attribute name="resolveURL">
181:         <xs:annotation>
182:           <xs:documentation>URL that provides direct access to a derivative file
183:             stored in a local or shared object repository</xs:documentation>
184:         </xs:annotation>
185:       </xs:attribute>
186:       <xs:attribute name="primary" type="xs:boolean" use="optional">
187:         <xs:annotation>
188:           <xs:documentation>The primary attribute indicates whether the derivative
189:             file is the preferred to be displayed with the corresponding descriptive
190:             unit</xs:documentation>
191:         </xs:annotation>
192:       </xs:attribute>
193:     </xs:extension>
194:   </xs:complexContent>
195: </xs:complexType>
196: <xs:simpleType name="europeanaType">
197:   <xs:annotation>
198:     <xs:documentation> The Europeana material type of the digital resource, which
199:       Europeana uses to categorize search results by text, image, sound or video.
200:     </xs:documentation>
201:   </xs:annotation>
202:   <xs:restriction base="xs:string">
203:     <xs:enumeration value="TEXT"/>
204:     <xs:enumeration value="IMAGE"/>
205:     <xs:enumeration value="SOUND"/>
206:     <xs:enumeration value="VIDEO"/>
207:   </xs:restriction>
208: </xs:simpleType>
209:
210: <xs:complexType name="DUType">
211:   <xs:annotation>
212:     <xs:documentation> the Descriptive Unit entity represents one metadata record about
213:       one or more cultural heritage objects, which are instead represented in the HOPE
214:       data model by the digital resource entity. </xs:documentation>
215:   </xs:annotation>
216:   <xs:sequence>
217:     <xs:element maxOccurs="unbounded" minOccurs="0" name="localIDCollectionItem"
218:       type="defaultType">
219:       <xs:annotation>
220:         <xs:documentation> an unambiguous reference to a collection item preserved by
221:           the content provider
222:         </xs:documentation>
223:       </xs:annotation>
224:     </xs:element>
225:     <xs:element maxOccurs="1" minOccurs="0" name="localIDParent" type="defaultType">
226:       <xs:annotation>
227:         <xs:documentation> local identifier of the parent metadata record as it is
228:           identified in the local information system of the content provider

```

```

229:         </xs:documentation>
230:     </xs:annotation>
231: </xs:element>
232: <xs:element maxOccurs="1" minOccurs="0" name="isContainedBy" type="xs:string">
233:     <xs:annotation>
234:         <xs:documentation> this relationship links a descriptive unit with another
235:             constituent descriptive unit that contains the descriptive unit; the value
236:             must be a PID </xs:documentation>
237:     </xs:annotation>
238: </xs:element>
239: <xs:element maxOccurs="1" minOccurs="0" name="localIDNext" type="defaultType">
240:     <xs:annotation>
241:         <xs:documentation> local identifier that relates two descriptive units that are
242:             ordered parts of the same parent descriptive unit, such that one descriptive
243:             unit comes immediately after the other descriptive unit </xs:documentation>
244:     </xs:annotation>
245: </xs:element>
246:
247: <xs:element maxOccurs="1" minOccurs="0" name="isNextInSequence" type="xs:string">
248:     <xs:annotation>
249:         <xs:documentation> This relationship links two descriptive units that are
250:             'ordered' parts of the same constituent descriptive unit, such that one
251:             descriptive unit comes immediately after the other descriptive unit; the
252:             value must be a PID </xs:documentation>
253:     </xs:annotation>
254: </xs:element>
255:
256: <xs:element maxOccurs="unbounded" minOccurs="0" name="aggregator" type="defaultType">
257:     <xs:annotation>
258:         <xs:documentation>name of the organisation aggregating the
259:             collection</xs:documentation>
260:     </xs:annotation>
261: </xs:element>
262: <xs:element maxOccurs="1" minOccurs="1" name="isSuppliedBy" type="pidType">
263:     <xs:annotation>
264:         <xs:documentation>this relationship links the Descriptive Unit with the Content
265:             Provider entity </xs:documentation>
266:     </xs:annotation>
267: </xs:element>
268: <xs:element maxOccurs="unbounded" minOccurs="0" name="isRepresentedBy">
269:     <xs:annotation>
270:         <xs:documentation>this relationship links a descriptive unit with a digital
271:             representation of the described CHO </xs:documentation>
272:     </xs:annotation>
273:     <xs:complexType>
274:         <xs:annotation>
275:             <xs:documentation>these elements are copied from the represented digital
276:                 resource</xs:documentation>
277:         </xs:annotation>
278:         <xs:sequence>
279:             <xs:element name="rights" type="defaultType"/>
280:             <xs:element name="language" type="normalisedType"/>
281:             <xs:element name="persistentID" type="xs:string"/>
282:             <xs:element name="localID" type="xs:string"/>
283:         </xs:sequence>
284:         <xs:attribute name="label" type="xs:string" use="required"/>
285:     </xs:complexType>
286: </xs:element>
287: <xs:element maxOccurs="1" minOccurs="0" name="landingPage" type="defaultType">
288:     <xs:annotation>
289:         <xs:documentation>An unambiguous URL reference to the digital resource on the
290:             content provider's website in it's full information context
291:         </xs:documentation>
292:     </xs:annotation>
293: </xs:element>
294: <xs:element maxOccurs="1" minOccurs="0" name="thumbnail" type="defaultType">
295:     <xs:annotation>
296:         <xs:documentation> The persistent identifier of a thumbnail representing the
297:             described cultural heritage object in the targeted dicoverly service, by
298:             default the thumbnail is the derivative 3 of the first digital resource
299:             entity related with the descriptive unit entity. </xs:documentation>
300:     </xs:annotation>
301: </xs:element>
302: <xs:element maxOccurs="1" minOccurs="1" name="metadataLanguage" type="normalisedType">
303:     <xs:annotation>
304:         <xs:documentation> primary language of the metadata describing the cultural

```



```

305:         heritage object, normalised value must comply with the ISO639-3 standard
306:     </xs:documentation>
307: </xs:annotation>
308: </xs:element>
309:
310: <xs:element maxOccurs="unbounded" minOccurs="0" name="title" type="defaultType">
311:   <xs:annotation>
312:     <xs:documentation> name given to the CHO language attribute: the value must
313:       comply with the ISO639-3 standard script attribute: the value must comply
314:       with the ISO15924 standard </xs:documentation>
315:   </xs:annotation>
316: </xs:element>
317:
318: <xs:element maxOccurs="unbounded" minOccurs="0" name="creator" type="POType">
319:   <xs:annotation>
320:     <xs:documentation> creator, publisher and contributor are the non relational
321:       equivalents of the elements isCreatedBy, hasContributionsBy, isPublishedBy A
322:       person or organisation primarily responsible for making the cultural
323:       heritage object </xs:documentation>
324:   </xs:annotation>
325: </xs:element>
326:
327: <xs:element maxOccurs="unbounded" minOccurs="0" name="isCreatedBy" type="pidType">
328:   <xs:annotation>
329:     <xs:documentation>this relationship links a descriptive unit with an agent that
330:       created the cultural heritage object</xs:documentation>
331:   </xs:annotation>
332: </xs:element>
333: <xs:element maxOccurs="unbounded" minOccurs="0" name="isCreatedIn" type="pidType">
334:   <xs:annotation>
335:     <xs:documentation>this relationship links a descriptive unit with a place where
336:       the cultural heritage object was created</xs:documentation>
337:   </xs:annotation>
338: </xs:element>
339: <xs:element maxOccurs="unbounded" minOccurs="0" name="contributor" type="POType">
340:   <xs:annotation>
341:     <xs:documentation>a person or organisation primarily responsible for making
342:       contributions to the cultural heritage object</xs:documentation>
343:   </xs:annotation>
344: </xs:element>
345:
346: <xs:element maxOccurs="unbounded" minOccurs="0" name="hasContributionsBy" type="pidType">
347:   <xs:annotation>
348:     <xs:documentation>this relationship links a descriptive unit with an agent that
349:       contributed to the cultural heritage object</xs:documentation>
350:   </xs:annotation>
351: </xs:element>
352: <xs:element maxOccurs="unbounded" minOccurs="0" name="publisher" type="POType">
353:   <xs:annotation>
354:     <xs:documentation>a person or organisation primarily responsible for making the
355:       cultural heritage object available</xs:documentation>
356:   </xs:annotation>
357: </xs:element>
358: <xs:element maxOccurs="unbounded" minOccurs="0" name="isPublishedBy" type="pidType">
359:   <xs:annotation>
360:     <xs:documentation>this relationship links a descriptive unit with an agent that
361:       published a cultural heritage object</xs:documentation>
362:   </xs:annotation>
363: </xs:element>
364: <xs:element maxOccurs="unbounded" minOccurs="0" name="isPublishedIn" type="pidType">
365:   <xs:annotation>
366:     <xs:documentation>this relationship links a descriptive unit with a place where
367:       the cultural heritage object was published</xs:documentation>
368:   </xs:annotation>
369: </xs:element>
370: <xs:element maxOccurs="unbounded" minOccurs="0" name="date" type="normalisedType">
371:   <xs:annotation>
372:     <xs:documentation> a date associated with an event in the lifecycle of the
373:       cultural heritage object the normalised value must comply with the ISO3166-1
374:       standard ISO8601 </xs:documentation>
375:   </xs:annotation>
376: </xs:element>
377:
378: <xs:element maxOccurs="unbounded" minOccurs="0" name="description" type="defaultType">
379:   <xs:annotation>
380:     <xs:documentation> an account of the cultural heritage object language

```

```

381:         attribute: the value must comply with the ISO639-3 standard script
382:         attribute: the value must comply with the ISO15924 standard
383:     </xs:documentation>
384: </xs:annotation>
385: </xs:element>
386:
387: <xs:element maxOccurs="unbounded" minOccurs="0" name="spatialCoverage"
388:     type="defaultType">
389:     <xs:annotation>
390:         <xs:documentation>the special extent or scope of the cultural heritage
391:         object</xs:documentation>
392:     </xs:annotation>
393: </xs:element>
394: <xs:element maxOccurs="unbounded" minOccurs="0" name="associatedPlace" type="pidType">
395:     <xs:annotation>
396:         <xs:documentation>this relationship links a descriptive unit with a place that
397:         covers the spatial extent or scope of the cultural heritage object
398:     </xs:documentation>
399:     </xs:annotation>
400: </xs:element>
401: <xs:element maxOccurs="unbounded" minOccurs="0" name="depictedPlace" type="pidType">
402:     <xs:annotation>
403:         <xs:documentation>this relationship links a descriptive unit with a place that
404:         is depicted by the cultural heritage object</xs:documentation>
405:     </xs:annotation>
406: </xs:element>
407: <xs:element maxOccurs="unbounded" minOccurs="0" name="temporalCoverage"
408:     type="defaultType">
409:     <xs:annotation>
410:         <xs:documentation>the temporal extent or scope of the cultural heritage
411:         object</xs:documentation>
412:     </xs:annotation>
413: </xs:element>
414: <xs:element maxOccurs="unbounded" minOccurs="0" name="associatedEvent" type="pidType">
415:     <xs:annotation>
416:         <xs:documentation>this relationship links a descriptive unit with an event that
417:         is the topic of the cultural heritage object</xs:documentation>
418:     </xs:annotation>
419: </xs:element>
420: <xs:element maxOccurs="unbounded" minOccurs="0" name="depictedEvent" type="pidType">
421:     <xs:annotation>
422:         <xs:documentation>this relationship links a descriptive unit with an event that
423:         is depicted by the cultural heritage object</xs:documentation>
424:     </xs:annotation>
425: </xs:element>
426: <xs:element maxOccurs="unbounded" minOccurs="0" name="subject" type="defaultType">
427:     <xs:annotation>
428:         <xs:documentation>the topic of the cultural heritage object</xs:documentation>
429:     </xs:annotation>
430: </xs:element>
431: <xs:element maxOccurs="unbounded" minOccurs="0" name="associatedAgent" type="pidType">
432:     <xs:annotation>
433:         <xs:documentation>this relationship links a descriptive unit with an agent that
434:         is the topic of the cultural heritage object is about</xs:documentation>
435:     </xs:annotation>
436: </xs:element>
437: <xs:element maxOccurs="unbounded" minOccurs="0" name="depictedAgent" type="pidType">
438:     <xs:annotation>
439:         <xs:documentation>this relationship links a descriptive unit with an agent that
440:         is depicted by the cultural heritage object </xs:documentation>
441:     </xs:annotation>
442: </xs:element>
443: <xs:element maxOccurs="unbounded" minOccurs="0" name="associatedConcept" type="pidType">
444:     <xs:annotation>
445:         <xs:documentation>this relationship links a descriptive unit with a concept
446:         that is the topic of the cultural heritage object </xs:documentation>
447:     </xs:annotation>
448: </xs:element>
449: <xs:element maxOccurs="unbounded" minOccurs="0" name="depictedConcept" type="pidType">
450:     <xs:annotation>
451:         <xs:documentation>this relationship links a descriptive unit with a concept
452:         that is depicted by the cultural heritage object</xs:documentation>
453:     </xs:annotation>
454: </xs:element>
455: <xs:element maxOccurs="unbounded" minOccurs="0" name="language" type="normalisedType">
456:     <xs:annotation>

```

```

457:         <xs:documentation> Language of the cultural heritage object the normalised
458:             value must comply with the ISO639-3 standard </xs:documentation>
459:     </xs:annotation>
460: </xs:element>
461:
462: <xs:element maxOccurs="unbounded" minOccurs="0" name="rights" type="defaultType">
463:     <xs:annotation>
464:         <xs:documentation>information about rights held in and over the
465:             CHO</xs:documentation>
466:     </xs:annotation>
467: </xs:element>
468: <xs:element maxOccurs="unbounded" minOccurs="0" name="provenance" type="defaultType">
469:     <xs:annotation>
470:         <xs:documentation> statement of any changes in ownership and custody of the CHO
471:             since it's creation that are significant for it's authenticity, integrity
472:             and interpretation </xs:documentation>
473:     </xs:annotation>
474: </xs:element>
475: <xs:element maxOccurs="1" name="domain" type="domainType">
476:     <xs:annotation>
477:         <xs:documentation> domain is used to mark the profile of the descriptive unit,
478:             it adds some profile specific elements. This element is not available in the
479:             data model </xs:documentation>
480:     </xs:annotation>
481: </xs:element>
482: <xs:element maxOccurs="unbounded" minOccurs="0" name="associatedHopeTheme"
483:     type="pidType">
484:     <xs:annotation>
485:         <xs:documentation> this relationship links a descriptive unit with a HOPE theme
486:             that is the topic of a digital cultural heritage object </xs:documentation>
487:     </xs:annotation>
488: </xs:element>
489: <xs:element maxOccurs="unbounded" minOccurs="0" name="relation" type="defaultType">
490:     <xs:annotation>
491:         <xs:documentation>Description of a related resource outside the HOPE
492:             information system</xs:documentation>
493:     </xs:annotation>
494: </xs:element>
495: <xs:element maxOccurs="1" minOccurs="0" name="resolveURL" type="defaultType">
496:     <xs:annotation>
497:         <xs:documentation>URL, which provides direct access to the web resource
498:             representing the descriptive unit. This is the URL the persistent ID of the
499:             descriptive unit will redirect to. </xs:documentation>
500:     </xs:annotation>
501: </xs:element>
502: <xs:element maxOccurs="1" minOccurs="0" name="misc">
503:     <xs:complexType>
504:         <xs:sequence>
505:             <xs:element name="miscfield" maxOccurs="unbounded" minOccurs="1">
506:                 <xs:complexType>
507:                     <xs:sequence>
508:                         <xs:element name="key" maxOccurs="1" minOccurs="1"/>
509:                         <xs:element name="value" maxOccurs="1" minOccurs="1"/>
510:                     </xs:sequence>
511:                 </xs:complexType>
512:             </xs:element>
513:         </xs:sequence>
514:     </xs:complexType>
515: </xs:element>
516: </xs:sequence>
517: </xs:complexType>
518: <xs:complexType name="DRType">
519:     <xs:sequence>
520:         <xs:element maxOccurs="1" minOccurs="0" name="localIDNext" type="defaultType">
521:             <xs:annotation>
522:                 <xs:documentation> Local identifier of a digital representation that comes
523:                     immediately after the current digital representation, as part of an ordered
524:                     series.
525:                 </xs:documentation>
526:             </xs:annotation>
527:         </xs:element>
528:         <xs:element maxOccurs="1" minOccurs="1" name="derivative2" type="versionType">
529:             <xs:annotation>
530:                 <xs:documentation>Globally unique, resolvable URL which allows retrieving a
531:                     low-resolution derivative of the digital representation.
532:             </xs:documentation>

```

```

533:         </xs:annotation>
534:     </xs:element>
535: <xs:element maxOccurs="1" minOccurs="1" name="derivative3" type="versionType">
536:     <xs:annotation>
537:         <xs:documentation>Globally unique, resolvable URL which allows retrieving a
538:             thumbnail derivative of the digital representation.
539:         </xs:documentation>
540:     </xs:annotation>
541: </xs:element>
542: <xs:element maxOccurs="1" minOccurs="0" name="transcription" type="versionType">
543:     <xs:annotation>
544:         <xs:documentation>Globally unique, resolvable URL which allows retrieving a
545:             transcription of the textual content of the digital representation
546:         </xs:documentation>
547:     </xs:annotation>
548: </xs:element>
549: <xs:element maxOccurs="1" minOccurs="0" name="isNextInSequence" type="xs:string">
550:     <xs:annotation>
551:         <xs:documentation> This relationship links two digital resource entities that
552:             are part of an ordered series, such that one digital resource entity comes
553:             immediately after the other digital resource entity.</xs:documentation>
554:     </xs:annotation>
555: </xs:element>
556:
557: <xs:element maxOccurs="unbounded" minOccurs="0" name="language" type="normalisedType">
558:     <xs:annotation>
559:         <xs:documentation> language of the intellectual content of the digital
560:             representation; the normalised value must comply with the ISO639-3 standard.
561:         </xs:documentation>
562:     </xs:annotation>
563: </xs:element>
564: <xs:element maxOccurs="1" minOccurs="1" name="type">
565:     <xs:annotation>
566:         <xs:documentation> The Europeana material type of the digital representation,
567:             which Europeana uses to categorise search results by text, image, sound
568:             or video.
569:         </xs:documentation>
570:     </xs:annotation>
571:     <xs:complexType>
572:         <xs:simpleContent>
573:             <xs:extension base="europeanaType">
574:                 <xs:attribute name="label" type="xs:string" use="required" />
575:             </xs:extension>
576:         </xs:simpleContent>
577:     </xs:complexType>
578: </xs:element>
579: <xs:element maxOccurs="1" minOccurs="0" name="rights" type="normalisedType">
580:     <xs:annotation>
581:         <xs:documentation>Information about rights held in and over the digital
582:             representation.
583:         </xs:documentation>
584:     </xs:annotation>
585: </xs:element>
586: <xs:element maxOccurs="1" minOccurs="1" name="contentProvider" type="pidType">
587:     <xs:annotation>
588:         <xs:documentation> elements taken from the DU the Digital Resource represents
589:         </xs:documentation>
590:     </xs:annotation>
591: </xs:element>
592: <xs:element maxOccurs="1" minOccurs="1" name="represents" type="pidType">
593:     <xs:annotation>
594:         <xs:documentation> this relationship links the digital resource with the
595:             descriptive unit describing the represented cultural heritage object
596:         </xs:documentation>
597:     </xs:annotation>
598: </xs:element>
599: <xs:element maxOccurs="1" minOccurs="0" name="misc">
600:     <xs:complexType>
601:         <xs:sequence>
602:             <xs:element name="miscfield" maxOccurs="unbounded" minOccurs="1">
603:                 <xs:complexType>
604:                     <xs:sequence>
605:                         <xs:element name="key" maxOccurs="1" minOccurs="1" />
606:                         <xs:element name="value" maxOccurs="1" minOccurs="1" />
607:                     </xs:sequence>
608:                 </xs:complexType>

```

```

609:         </xs:element>
610:     </xs:sequence>
611: </xs:complexType>
612: </xs:element>
613: </xs:sequence>
614: </xs:complexType>
615: <xs:complexType name="agentType">
616:     <xs:annotation>
617:         <xs:documentation>the agent entities will be created during post-processing after
618:             aggregation </xs:documentation>
619:     </xs:annotation>
620:     <xs:sequence>
621:         <xs:element maxOccurs="unbounded" minOccurs="1" name="alternateID"
622:             type="defaultType"/>
623:         <xs:element maxOccurs="unbounded" minOccurs="1" name="preferredTerm"
624:             type="defaultType"/>
625:         <xs:element maxOccurs="unbounded" minOccurs="0" name="alternativeTerm"
626:             type="defaultType"/>
627:         <xs:element maxOccurs="unbounded" minOccurs="0" name="standardVocabulary">
628:             <xs:complexType>
629:                 <xs:simpleContent>
630:                     <xs:extension base="defaultType">
631:                         <xs:attribute name="source" type="xs:string" use="required"/>
632:                         <xs:attribute name="baseURL" type="xs:string"/>
633:                     </xs:extension>
634:                 </xs:simpleContent>
635:             </xs:complexType>
636:         </xs:element>
637:         <xs:element maxOccurs="1" minOccurs="1" name="type">
638:             <xs:complexType>
639:                 <xs:simpleContent>
640:                     <xs:restriction base="defaultType">
641:                         <xs:enumeration value="person"/>
642:                         <xs:enumeration value="organisation"/>
643:                     </xs:restriction>
644:                 </xs:simpleContent>
645:             </xs:complexType>
646:         </xs:element>
647:         <xs:element maxOccurs="1" minOccurs="1" name="description" type="defaultType"/>
648:         <xs:element maxOccurs="1" minOccurs="0" name="source"/>
649:         <xs:element maxOccurs="unbounded" minOccurs="0" name="date">
650:             <xs:complexType>
651:                 <xs:simpleContent>
652:                     <xs:extension base="normalisedType">
653:                         <xs:attribute name="rangeType" use="required">
654:                             <xs:annotation>
655:                                 <xs:documentation> Indicates whether the date is the start
656:                                     (from) or the end (to) of a date range. </xs:documentation>
657:                             </xs:annotation>
658:                         </xs:attribute>
659:                     </xs:extension>
660:                 </xs:simpleContent>
661:             </xs:complexType>
662:         </xs:element>
663:     </xs:sequence>
664: </xs:complexType>
665: <xs:complexType name="placeType">
666:     <xs:sequence>
667:         <xs:element maxOccurs="unbounded" minOccurs="1" name="alternateID"
668:             type="defaultType"/>
669:         <xs:element maxOccurs="unbounded" minOccurs="1" name="preferredTerm"
670:             type="defaultType"/>
671:         <xs:element maxOccurs="unbounded" minOccurs="0" name="alternativeTerm"
672:             type="defaultType"/>
673:         <xs:element maxOccurs="unbounded" minOccurs="0" name="standardVocabulary">
674:             <xs:complexType>
675:                 <xs:simpleContent>
676:                     <xs:extension base="defaultType">
677:                         <xs:attribute name="source" type="xs:string" use="required"/>
678:                         <xs:attribute name="baseURL" type="xs:string"/>
679:                     </xs:extension>
680:                 </xs:simpleContent>
681:             </xs:complexType>
682:         </xs:element>
683:         <xs:element maxOccurs="1" minOccurs="0" name="type">
684:             <xs:complexType>

```

```

685:         <xs:simpleContent>
686:             <xs:restriction base="defaultType">
687:                 <xs:enumeration value="local" />
688:                 <xs:enumeration value="regional" />
689:                 <xs:enumeration value="national" />
690:                 <xs:enumeration value="supra natural" />
691:             </xs:restriction>
692:         </xs:simpleContent>
693:     </xs:complexType>
694: </xs:element>
695: <xs:element maxOccurs="1" minOccurs="0" name="coordinates">
696:     <xs:annotation>
697:         <xs:documentation> the value must comply with EPSG4326 </xs:documentation>
698:     </xs:annotation>
699: </xs:element>
700: <xs:element maxOccurs="unbounded" minOccurs="0" name="description"
701:     type="defaultType" />
702: <xs:element maxOccurs="1" minOccurs="0" name="source" />
703: </xs:sequence>
704: </xs:complexType>
705: <xs:complexType name="CPType">
706:     <xs:annotation>
707:         <xs:documentation> The Content Provider entity represents all information about the
708:             institution that provides the HOPE aggregator with metadata and digital content
709:         </xs:documentation>
710:     </xs:annotation>
711:     <xs:sequence>
712:         <xs:element maxOccurs="1" minOccurs="1" name="name" type="defaultType">
713:             <xs:annotation>
714:                 <xs:documentation>name of the institution representing the content provider
715:             </xs:documentation>
716:             </xs:annotation>
717:         </xs:element>
718:
719:         <xs:element maxOccurs="1" minOccurs="1" name="country" type="normalisedType">
720:             <xs:annotation>
721:                 <xs:documentation>country where the content provider is located, the value must
722:                 comply with the ISO3166-1 standard </xs:documentation>
723:             </xs:annotation>
724:         </xs:element>
725:         <xs:element maxOccurs="unbounded" minOccurs="1" name="language" type="normalisedType">
726:             <xs:annotation>
727:                 <xs:documentation>the official language(s) of the country where the content
728:                 provider is located, value must comply with the ISO639-3 </xs:documentation>
729:             </xs:annotation>
730:         </xs:element>
731:     </xs:sequence>
732: </xs:complexType>
733: <xs:complexType name="eventType">
734:     <xs:sequence>
735:         <xs:element maxOccurs="unbounded" minOccurs="1" name="alternateID"
736:             type="defaultType" />
737:         <xs:element maxOccurs="unbounded" minOccurs="1" name="preferredTerm"
738:             type="defaultType" />
739:         <xs:element maxOccurs="unbounded" minOccurs="0" name="alternativeTerm"
740:             type="defaultType" />
741:         <xs:element maxOccurs="unbounded" minOccurs="0" name="standardVocabulary">
742:             <xs:complexType>
743:                 <xs:simpleContent>
744:                     <xs:extension base="defaultType">
745:                         <xs:attribute name="source" type="xs:string" use="required" />
746:                         <xs:attribute name="baseURL" type="xs:string" />
747:                     </xs:extension>
748:                 </xs:simpleContent>
749:             </xs:complexType>
750:         </xs:element>
751:         <xs:element maxOccurs="unbounded" minOccurs="0" name="description"
752:             type="defaultType" />
753:         <xs:element maxOccurs="1" minOccurs="0" name="source" />
754:         <xs:element maxOccurs="1" minOccurs="0" name="date">
755:             <xs:complexType>
756:                 <xs:simpleContent>
757:                     <xs:extension base="defaultType">
758:                         <xs:attribute name="rangeType" />
759:                     </xs:extension>
760:                 </xs:simpleContent>

```

```

761:         </xs:complexType>
762:     </xs:element>
763:     <xs:element maxOccurs="unbounded" minOccurs="0" name="spatialCoverage"
764:         type="pidType" />
765: </xs:sequence>
766: </xs:complexType>
767: <xs:complexType name="conceptType">
768:     <xs:annotation>
769:         <xs:documentation>within an object the concept can either be associated or depicted
770:     </xs:documentation>
771:     </xs:annotation>
772:     <xs:sequence>
773:         <xs:element maxOccurs="unbounded" minOccurs="1" name="alternateID" type="defaultType" />
774:         <xs:element maxOccurs="unbounded" minOccurs="1" name="preferredTerm"
775:             type="defaultType" />
776:         <xs:element maxOccurs="unbounded" minOccurs="0" name="alternativeTerm"
777:             type="defaultType" />
778:         <xs:element maxOccurs="unbounded" minOccurs="0" name="standardVocabulary">
779:             <xs:complexType>
780:                 <xs:simpleContent>
781:                     <xs:extension base="defaultType">
782:                         <xs:attribute name="source" type="xs:string" use="required" />
783:                         <xs:attribute name="baseURL" type="xs:string" />
784:                     </xs:extension>
785:                 </xs:simpleContent>
786:             </xs:complexType>
787:         </xs:element>
788:         <xs:element maxOccurs="unbounded" minOccurs="0" name="description" type="defaultType" />
789:         <xs:element maxOccurs="1" minOccurs="0" name="source" />
790:     </xs:sequence>
791: </xs:complexType>
792: <xs:complexType name="hopeThemeType">
793:     <xs:annotation>
794:         <xs:documentation>the historical theme entities will be created during post-processing
795:             after aggregation </xs:documentation>
796:     </xs:annotation>
797:     <xs:sequence>
798:
799:         <xs:element maxOccurs="unbounded" minOccurs="1" name="name" type="defaultType">
800:             <xs:annotation>
801:                 <xs:documentation>name is repeatable for different languages, list of languages
802:                     TBD </xs:documentation>
803:             </xs:annotation>
804:         </xs:element>
805:
806:         <xs:element maxOccurs="unbounded" minOccurs="0" name="broaderTerm" type="pidType" />
807:
808:     </xs:sequence>
809: </xs:complexType>
810: <xs:complexType name="domainType">
811:     <xs:annotation>
812:         <xs:documentation>sub-element of the descriptive unit</xs:documentation>
813:     </xs:annotation>
814:     <xs:choice>
815:         <xs:element name="dublinCoreUnit" type="dcProfileType" />
816:         <xs:element name="archiveUnit" type="archiveProfileType" />
817:         <xs:element name="libraryUnit" type="libraryProfileType" />
818:         <xs:element name="visualUnit" type="visualProfileType" />
819:         <xs:element name="audioVisualUnit" type="avProfileType" />
820:     </xs:choice>
821: </xs:complexType>
822: <xs:complexType name="archiveProfileType">
823:     <xs:sequence>
824:         <xs:element maxOccurs="1" minOccurs="1" name="descriptionLevel">
825:             <xs:annotation>
826:                 <xs:documentation>the level of arrangement of the descriptive unit
827:             </xs:documentation>
828:             </xs:annotation>
829:             <xs:complexType mixed="true">
830:                 <xs:attribute name="normalised" use="required">
831:                     <xs:simpleType>
832:                         <xs:restriction base="xs:string">
833:                             <xs:enumeration value="collection" />
834:                             <xs:enumeration value="mid-level" />
835:                             <xs:enumeration value="item" />
836:                         </xs:restriction>

```

```

837:         </xs:simpleType>
838:     </xs:attribute>
839:     <xs:attribute name="label" use="required" />
840:     <xs:attribute name="cataloguing" use="required" />
841:     <xs:attribute name="encoding" use="required" />
842: </xs:complexType>
843: </xs:element>
844:
845: <xs:element maxOccurs="unbounded" minOccurs="0" name="appearanceOfMaterial"
846:     type="defaultType">
847:     <xs:annotation>
848:         <xs:documentation>information about the appearance or construction of the
849:             described materials. </xs:documentation>
850:     </xs:annotation>
851: </xs:element>
852: <xs:element maxOccurs="unbounded" minOccurs="0" name="genreOfFonds" type="defaultType">
853:     <xs:annotation>
854:         <xs:documentation>information about specific types of archival materials, e.g.
855:             collections, or about specific physical characteristics. </xs:documentation>
856:     </xs:annotation>
857: </xs:element>
858: <xs:element maxOccurs="unbounded" minOccurs="0" name="extent" type="defaultType">
859:     <xs:annotation>
860:         <xs:documentation>information about the quantity of the materials being
861:             described or an expression of the physical space they occupy.
862:         </xs:documentation>
863:     </xs:annotation>
864: </xs:element>
865: <xs:element maxOccurs="unbounded" minOccurs="0" name="physicalCharacteristics"
866:     type="defaultType">
867:     <xs:annotation>
868:         <xs:documentation> A description of important physical conditions or
869:             characteristics that affect the storage, preservation, or use of the
870:             materials described. </xs:documentation>
871:     </xs:annotation>
872: </xs:element>
873: </xs:sequence>
874: </xs:complexType>
875: <xs:complexType name="libraryProfileType">
876:     <xs:sequence>
877:
878:     <xs:element maxOccurs="1" minOccurs="1" name="descriptionLevel">
879:         <xs:annotation>
880:             <xs:documentation> the level of arrangement of the descriptive unit.
881:             </xs:documentation>
882:         </xs:annotation>
883:         <xs:complexType mixed="true">
884:             <xs:attribute name="normalised" use="required">
885:                 <xs:simpleType>
886:                     <xs:restriction base="xs:string">
887:                         <xs:enumeration value="collection" />
888:                         <xs:enumeration value="mid-level" />
889:                         <xs:enumeration value="item" />
890:                     </xs:restriction>
891:                 </xs:simpleType>
892:             </xs:attribute>
893:             <xs:attribute name="label" use="required" />
894:             <xs:attribute name="encoding" use="required" />
895:         </xs:complexType>
896:
897:     </xs:element>
898:     <xs:element maxOccurs="unbounded" minOccurs="0" name="generalMaterialDesignation"
899:         type="defaultType">
900:         <xs:annotation>
901:             <xs:documentation> The characteristics and general material type of content of
902:                 the resource. </xs:documentation>
903:         </xs:annotation>
904:     </xs:element>
905:
906:     <xs:element maxOccurs="unbounded" minOccurs="0" name="specificMaterialDesignation"
907:         type="defaultType">
908:         <xs:annotation>
909:             <xs:documentation> Physical description of the described item, including its
910:                 extent, dimensions, and such other physical details as a description of any
911:                 accompanying materials and unit type and size. </xs:documentation>
912:         </xs:annotation>

```



```

913:     </xs:element>
914:   <xs:element maxOccurs="unbounded" minOccurs="0" name="extent" type="defaultType">
915:     <xs:annotation>
916:       <xs:documentation> Number of pages, volumes, etc. of the physical or digital
917:         unit or units constituting the resource, adding other measures of extent as
918:         appropriate. </xs:documentation>
919:     </xs:annotation>
920:   </xs:element>
921:   <xs:element maxOccurs="unbounded" minOccurs="0" name="publicationFrequency"
922:     type="defaultType">
923:     <xs:annotation>
924:       <xs:documentation>Complete statement, exclusive of dates, of the current
925:         publication frequency.</xs:documentation>
926:     </xs:annotation>
927:   </xs:element>
928:   <xs:element maxOccurs="unbounded" minOccurs="0" name="dimensions" type="defaultType">
929:     <xs:annotation>
930:       <xs:documentation>Dimensions of the resource, such as size and duration.
931:     </xs:documentation>
932:     </xs:annotation>
933:   </xs:element>
934:   <xs:element maxOccurs="unbounded" minOccurs="0" name="otherPhysicalDetails"
935:     type="defaultType">
936:     <xs:annotation>
937:       <xs:documentation> Other physical characteristics of the resource, such as
938:         method of production, colour, material from which the resource is made.
939:     </xs:documentation>
940:     </xs:annotation>
941:   </xs:element>
942: </xs:sequence>
943: </xs:complexType>
944: <xs:complexType name="visualProfileType">
945:   <xs:sequence>
946:     <xs:element maxOccurs="1" minOccurs="1" name="descriptionLevel">
947:       <xs:annotation>
948:         <xs:documentation>the level of arrangement of the descriptive unit
949:       </xs:documentation>
950:       </xs:annotation>
951:       <xs:complexType mixed="true">
952:         <xs:attribute name="normalised" use="required">
953:           <xs:simpleType>
954:             <xs:restriction base="xs:string">
955:               <xs:enumeration value="collection"/>
956:               <xs:enumeration value="mid-level"/>
957:               <xs:enumeration value="item"/>
958:             </xs:restriction>
959:           </xs:simpleType>
960:         </xs:attribute>
961:         <xs:attribute name="label" use="required"/>
962:         <xs:attribute name="encoding" use="required"/>
963:       </xs:complexType>
964:     </xs:element>
965:     <xs:element maxOccurs="unbounded" minOccurs="0" name="objectName" type="defaultType">
966:       <xs:annotation>
967:         <xs:documentation>A description of the form, function or type of object.
968:       </xs:documentation>
969:       </xs:annotation>
970:     </xs:element>
971:     <xs:element maxOccurs="unbounded" minOccurs="0" name="technique" type="defaultType">
972:       <xs:annotation>
973:         <xs:documentation>Processes, methods, techniques or tools used to fabricate or
974:           decorate an object.</xs:documentation>
975:       </xs:annotation>
976:     </xs:element>
977:     <xs:element maxOccurs="unbounded" minOccurs="0" name="material" type="defaultType">
978:       <xs:annotation>
979:         <xs:documentation>The basic materials and media from which an object is
980:           constructed. </xs:documentation>
981:       </xs:annotation>
982:     </xs:element>
983:     <xs:element maxOccurs="unbounded" minOccurs="0" name="dimensions" type="defaultType">
984:       <xs:annotation>
985:         <xs:documentation>The measurement of a dimension of the object.
986:       </xs:documentation>
987:       </xs:annotation>
988:     </xs:element>

```

```

989:     <xs:element maxOccurs="unbounded" minOccurs="0" name="technicalAttribute"
990:         type="defaultType">
991:         <xs:annotation>
992:             <xs:documentation> The name of a technical attribute possessed by an object
993:                 which can be described and quantified. </xs:documentation>
994:         </xs:annotation>
995:     </xs:element>
996: </xs:sequence>
997: </xs:complexType>
998: <xs:complexType name="avProfileType">
999:     <xs:sequence>
1000:         <xs:element maxOccurs="1" minOccurs="1" name="descriptionLevel">
1001:             <xs:annotation>
1002:                 <xs:documentation>the level of arrangement of the descriptive unit
1003:                 </xs:documentation>
1004:             </xs:annotation>
1005:             <xs:complexType mixed="true">
1006:                 <xs:attribute name="normalised" use="required">
1007:                     <xs:simpleType>
1008:                         <xs:restriction base="xs:string">
1009:                             <xs:enumeration value="collection"/>
1010:                             <xs:enumeration value="mid-level"/>
1011:                             <xs:enumeration value="item"/>
1012:                         </xs:restriction>
1013:                     </xs:simpleType>
1014:                 </xs:attribute>
1015:                 <xs:attribute name="label" use="required"/>
1016:                 <xs:attribute name="encoding" use="required"/>
1017:             </xs:complexType>
1018:         </xs:element>
1019:         <xs:element maxOccurs="unbounded" minOccurs="0" name="instantiationType"
1020:             type="defaultType">
1021:             <xs:annotation>
1022:                 <xs:documentation>A term or phrase describing the item type relative to the
1023:                 duplication process </xs:documentation>
1024:             </xs:annotation>
1025:         </xs:element>
1026:
1027:         <xs:element maxOccurs="unbounded" minOccurs="0" name="carrierType" type="defaultType">
1028:             <xs:annotation>
1029:                 <xs:documentation> The type of physical carrier used for the manifestation.
1030:                 Digital manifestations may be bound to a physical carrier (such as DVD) or
1031:                 exist without a defined carrier (such as online streaming media, podcasts,
1032:                 etc.). </xs:documentation>
1033:             </xs:annotation>
1034:         </xs:element>
1035:         <xs:element maxOccurs="unbounded" minOccurs="0" name="originalDuration"
1036:             type="defaultType">
1037:             <xs:annotation>
1038:                 <xs:documentation> The total physical, logical or temporal extent of the
1039:                 manifestation of a cinematographic work, or of the item from a
1040:                 manifestation. Physical extent is the total length of the medium carrying
1041:                 the manifestation or item. Logical extent is the number of discrete units,
1042:                 e.g. image frames, bytes (octets), etc., of which the manifestation or item
1043:                 is composed. Temporal extent is the running time under normal or specific
1044:                 playback conditions (e.g. a specific frame rate). </xs:documentation>
1045:             </xs:annotation>
1046:         </xs:element>
1047:         <xs:element maxOccurs="unbounded" minOccurs="0" name="originalLength"
1048:             type="defaultType">
1049:             <xs:annotation>
1050:                 <xs:documentation>The running time of the first known manifestation of a
1051:                 cinematographic work, measured in minutes and second. </xs:documentation>
1052:             </xs:annotation>
1053:         </xs:element>
1054:         <xs:element maxOccurs="unbounded" minOccurs="0" name="aspectRatio" type="defaultType">
1055:             <xs:annotation>
1056:                 <xs:documentation> The ratio between width and height of the image. For
1057:                 anamorphic formats, this element should contain the aspect ratio of the
1058:                 projected image rather than that on the carrier. Some aspect ratios are
1059:                 known not only by numeric values, but also by names (e.g. "full frame",
1060:                 "Cinemascope"). </xs:documentation>
1061:             </xs:annotation>
1062:         </xs:element>
1063:         <xs:element maxOccurs="unbounded" minOccurs="0" name="colourSystem" type="defaultType">
1064:             <xs:annotation>

```

```

1065:     <xs:documentation> A name for the colour system or process. </xs:documentation>
1066:   </xs:annotation>
1067: </xs:element>
1068: <xs:element maxOccurs="unbounded" minOccurs="0" name="gauge" type="defaultType">
1069:   <xs:annotation>
1070:     <xs:documentation> The width of the film stock or other carrier (such as
1071:       magnetic tape) used for the manifestation. Should include value and unit
1072:       (e.g. "35 mm", "1/2 in.") </xs:documentation>
1073:   </xs:annotation>
1074: </xs:element>
1075: <xs:element maxOccurs="unbounded" minOccurs="0" name="soundSystem" type="defaultType">
1076:   <xs:annotation>
1077:     <xs:documentation> The name of the system by which sound is recorded either on
1078:       the carrier, on a separate medium, or as part of the digital encoding (e.g.
1079:       Western Electric (Westrex) Movietone, RCA Photophone, Tobis, Dolby Digital,
1080:       etc.). </xs:documentation>
1081:   </xs:annotation>
1082: </xs:element>
1083: </xs:sequence>
1084: </xs:complexType>
1085: <xs:complexType name="dcProfileType">
1086:   <xs:sequence>
1087:     <xs:element maxOccurs="1" minOccurs="1" name="descriptionLevel">
1088:       <xs:annotation>
1089:         <xs:documentation>The level of arrangement of the descriptive unit.
1090:       </xs:documentation>
1091:       </xs:annotation>
1092:       <xs:complexType mixed="true">
1093:         <xs:attribute name="normalised" use="required">
1094:           <xs:simpleType>
1095:             <xs:restriction base="xs:string">
1096:               <xs:enumeration value="collection"/>
1097:               <xs:enumeration value="mid-level"/>
1098:               <xs:enumeration value="item"/>
1099:             </xs:restriction>
1100:           </xs:simpleType>
1101:         </xs:attribute>
1102:         <xs:attribute name="label" use="required"/>
1103:         <xs:attribute name="encoding" use="required"/>
1104:       </xs:complexType>
1105:     </xs:element>
1106:     <xs:element maxOccurs="unbounded" minOccurs="0" name="extent" type="defaultType">
1107:       <xs:annotation>
1108:         <xs:documentation>The size or duration of the resource. </xs:documentation>
1109:       </xs:annotation>
1110:     </xs:element>
1111:     <xs:element maxOccurs="unbounded" minOccurs="0" name="medium" type="defaultType">
1112:       <xs:annotation>
1113:         <xs:documentation>The material or physical carrier of the
1114:         resource.</xs:documentation>
1115:       </xs:annotation>
1116:     </xs:element>
1117:   </xs:sequence>
1118: </xs:complexType>
1119: </xs:schema>

```