Hierarchical objects Task force

Case Study from the HOPE Project

1.Name

Dr. Donald Weber
Social History Researcher / Head ICT, Amsab-Institute of Social History, Ghent.
http://www.amsab.be/index.php?lang=english
Amsab-ISH is work package leader of *WP2 – Agreed Standards and Best Practices.*

Alessia Bardi
PhD student in Information Engineering, University of Pisa
Graduate Fellow at the Istituto di Scienza e Tecnologie dell'Informazione (ISTI – CNR)
alessia.bardi@isti.cnr.it
ISTI-CNR is work package leader of *WP4 –HOPE Aggregator Service*.

2. Background information

HOPE (Heritage of the People's Europe) is funded byFP7 EU eContentplus, Best Practice Networks Project: Grant Agreement N. 250549
The HOPE web site is available at http://www.peoplesheritage.eu/.

HOPE (Heritage of the People's Europe, FP7 EU eContentplus, grant agreement: 250549, http://www.peoplesheritage.eu) is a "Best Practice Network" for archives, libraries, museums and institutions operating in the fields of social and union history. The goal of the project is providing a unified access to materials about the European social and labour history from the 18th to 21st centuries, proposing guidelines and tools for the management, aggregation, harmonisation, curation and provision of digital cultural heritage content. HOPE unites about 900,000 digitised objects and makes them available through web-based platforms like Europeana and the Labour History portal.

3. Definition of hierarchies for the domain

- Collection: A set of items with one or more common factors, such as material type, author, publisher, provenance, and/or subject. In HOPE, Collections are provided by CPs in the form of metadata records and, if available, Digital Objects. Collections are used as the basis for submission and management of records and objects; and serve as a key access point for end users. As such, collections must meet certain technical requirements.
  Source: HOPE (http://igwiki.peoplesheritage.eu/index.php/Glossary#Collection)

- Level of Description:Level of Granularity of a Descriptive Unit that is part of a hierarchical description. The designation of the level is generally specific to the collection domain. (E.g. for archival collections, this might include fonds, series, files, and items, while for library collections, series, titles, and issues.) HOPE does not limit

the number and type of Levels of Description and can also support idiosyncratic descriptive levels.
Source: HOPE (http://igwiki.peoplesheritage.eu/index.php/Glossary#Level_of_Description)

- Archival Finding Aid: Descriptive metadata on the records composing an archival collection. The Archival Finding Aid is generally hierarchic, describing the collection from general to specific, starting with the whole then proceeding to the components (fonds, series, folders, and items). Such metadata are usually created and captured in an archival management system.
  Source: HOPE (http://igwiki.peoplesheritage.eu/index.php/Glossary#Archival_Finding_Aid)

4. Use Case Scenario

HOPE Content Partners have archival collections of very different sizes and contents. Generally these collections are structured into so-called ‘fonds’. A fonds is a set of records that have been generated by a single person, family or corporate body – called the ‘creator’ – in the course of its activities and functions.

A fonds is typically structured in a hierarchical way: the records are grouped into series on multiple levels.

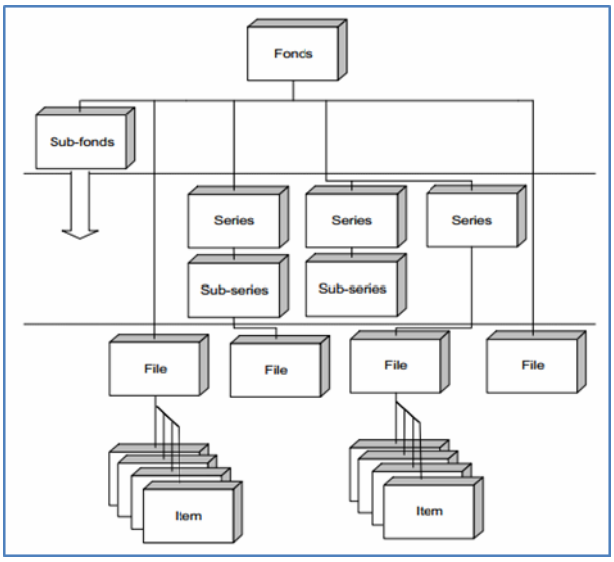

Figure 1 ISAD (G) Model of the levels of arrangement of a fonds

Descriptions of the content of a fonds (metadata) will reflect the hierarchical structure of a fonds. There will be descriptions on every level. Metadata will start at the top of the structure, then descend to the lower levels, thus describing the fonds from the general to the specific. Each described unit must identify its own hierarchical level, and identify its next higher unit of description (‘parent’). Metadata on any level must only contain information relevant to that specific level. Information relevant to multiple levels must only be given at the highest appropriate level, and should not be repeated on a lower level.

Figure 2 Hierarchical representation of archival metadata (Amsab-Institute of Social History)

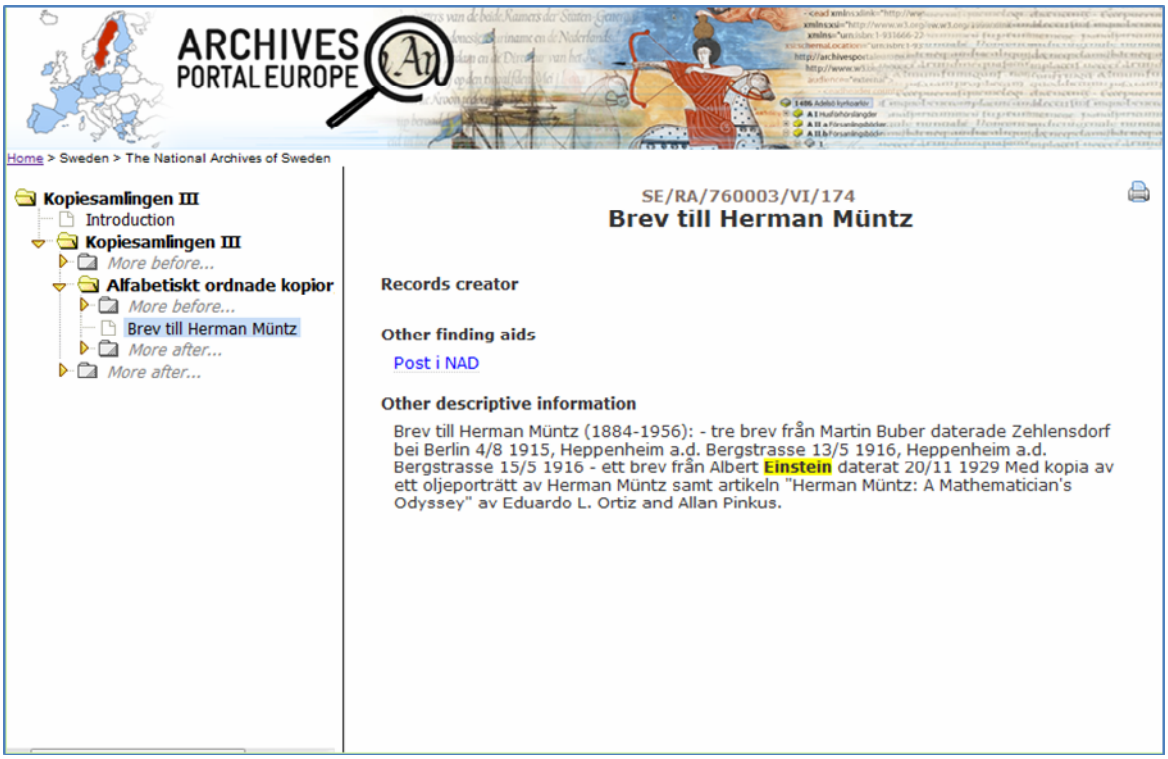

Figure 3 Hierarchical representation of archival metadata (Archives Portal Europe)

Typical for such hierarchical metadata is that the lowest level descriptions contain not enough information to allow for a stand-alone representation, i.e. the record's meaning and content are understandable and relevant, only when it is presented as part of a hierarchical 'tree'.

On the part of the data model, in order to allow for hierarchical descriptions, it must minimally allow for two elements to be included into each record: (1) an element describing the level of description, and (2) a link to the parent record. Optionally, a next-in-sequence element may be added, to support intra-level sorting of the metadata (although usually a sorting order will already be determined by the reference code element).

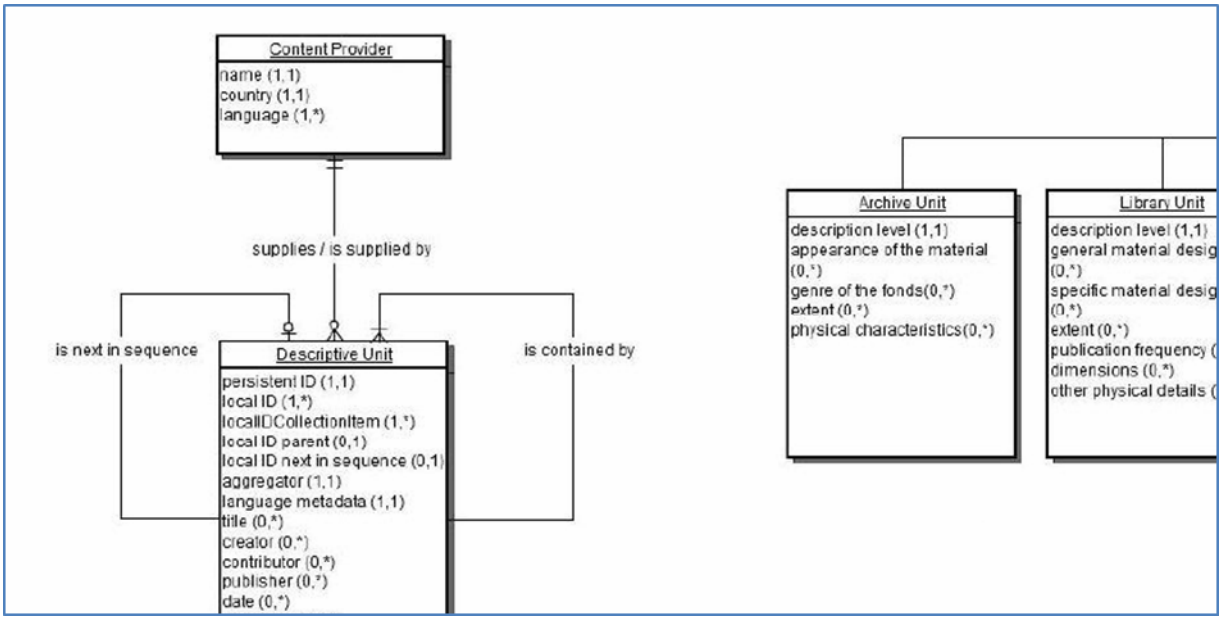

Figure 4 Hierarchical elements in the HOPE Data Model

In the HOPE Data Model, each Descriptive Unit contains an element hope:isContainedBy, allowing to identify a parent next up in the hierarchy, and an element hope:descriptionLevel, identifying the position of the unit's level in the hierarchy. There is also an optional element hope:isNextInSequence.

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

5. Problems and limitations

- Digital content

Usually digital content will be linked only to the lower levels of description. This means there will be records that are not linked to digital content, usually the higher-level descriptions. This could conflict with Europeana's policy of focusing on direct access to digital content.

- Inherited values

An alternative solution to hierarchical tree representations could be the automatic inheriting of values of higher level records into their lower level (grand) children. However, this might lead to 'heavy' descriptions. The massive scale copying of values into lower level records will also create data redundancy and requires technical solutions to ensure data consistency among different metadata records.

- GUI

Looking at the data model alone will not solve the issue of hierarchical metadata. Providing the end user with sufficient relevant information will depend as much on the user interface as on the underlying data model. Great care should be given to the representation of hierarchies in the GUI.

6. Proposed solutions for Europeana

Figure 6 sketches the mapping logic among HOPE and EDM entities. In summary:

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

[1]Europeana Data Model MappingGuidelines, http://pro.europeana.eu/documents/900548/ea68f42d-32f6-4900-91e9-ef18006d652e, pages 14-16.

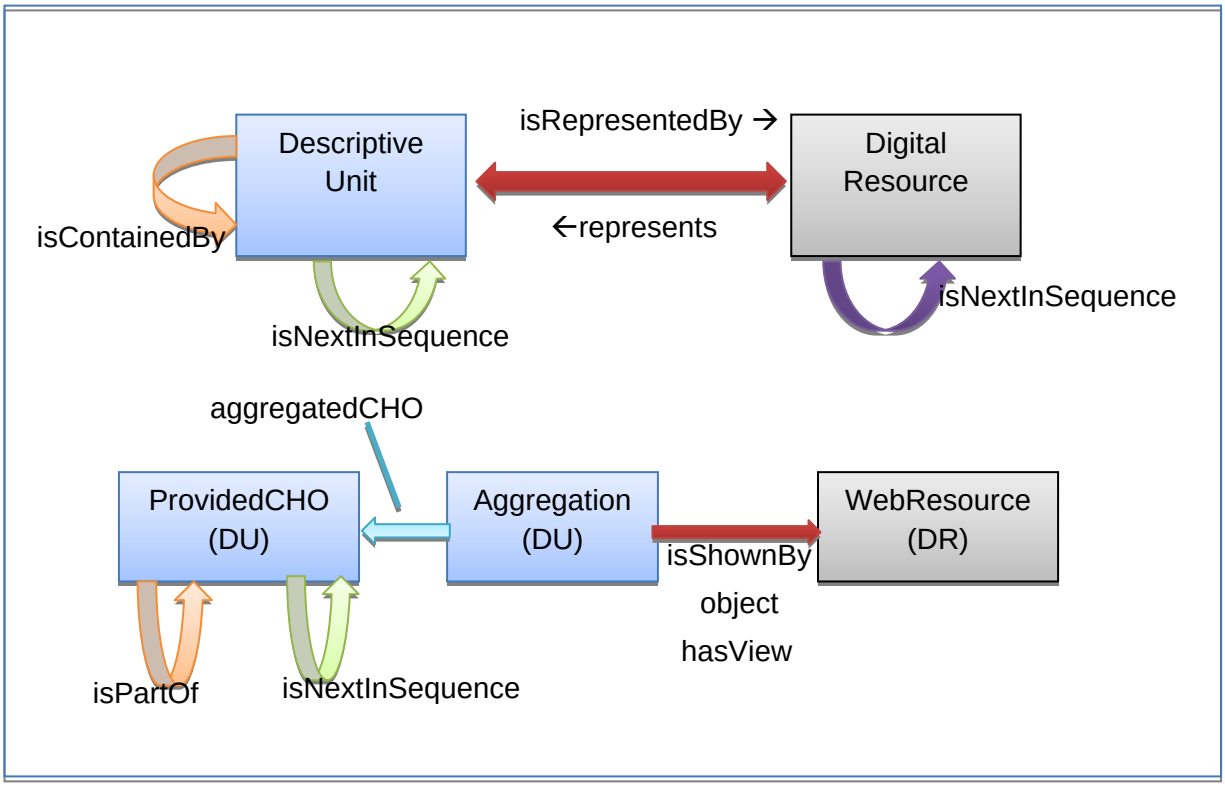

Figure 6 Logical Mapping from the HOPE Data Model to EDM

The following XML is an example of an EDM file provided by HOPE to the Europeana Ingestion Team. XML namespace declarations are omitted for sake of simplicity.

```
<rdf:RDF>
<ore:Aggregationrdf:about="10796/A400000627_1067">
<edm:aggregatedCHOrdf:resource="http://hdl.handle.net/10796/A400000627_1067"/>
<edm:dataProvider>Amsab-Institute of Social History</edm:dataProvider>
<edm:provider>Hope - Heritage of the People's Europe</edm:provider>
<edm:rightsrdf:resource="http://www.europeana.eu/rights/rr-f/"/>
<edm:isShownBy
rdf:resource="http://hdl.handle.net/10796/1982_38?locatt=view:derivative2"/>
<edm:isShownAtrdf:resource="http://hdl.handle.net/10796/A400000627_1067#page"/>
<edm:object rdf:resource="http://hdl.handle.net/10796/1982_38?locatt=view:derivative2"/>
</ore:Aggregation>

<edm:ProvidedCHOrdf:about="http://hdl.handle.net/10796/A400000627_1067">
<!—the dcterms:isPartOf refers to the container edm:ProvidedCHO -->
<dcterms:isPartOfrdf:resource="http://hdl.handle.net/10796/A400000627_22"/>
<dcterms:issued>1982</dcterms:issued>
<dc:identifier>http://hdl.handle.net/10796/A400000627_1067</dc:identifier>
<dc:identifier>PV-MTSF 258</dc:identifier>
<dc:title>Combat (1982)38</dc:title>
<dc:language>fra</dc:language>
```

```
<dc:relation>www.peoplesheritage.eu</dc:relation>
<dc:type>issue</dc:type>
<edm:type>TEXT</edm:type>
</edm:ProvidedCHO>
<edm:WebResource
rdf:about="http://hdl.handle.net/10796/1982_38?locatt=view:derivative2">
<edm:rightsrdf:resource="http://www.europeana.eu/rights/rr-f/"/>
</edm:WebResource>
</rdf:RDF>
```

7. References

- *Glossary – The HOPE Wiki*: http://igwiki.peoplesheritage.eu/index.php/Glossary
- *ISAD(G) - General International Standard Archival Description*: http://www.icacds.org.uk/eng/standards.htm
- *The Common HOPE Metadata Structure, including the Harmonisation Specifications*: http://www.peoplesheritage.eu/content/news.htm
- *The APENET-EAD Schema*: http://www.apex-project.eu/index.php/outcomes