D25 – DELIVERABLE C3.2.2

Project Acronym: OpenUp!
Grant Agreement No: 270890
Project Title: Opening up the Natural History Heritage for Europeana

D25 / C3.2.2 Domain specific vocabularies for EUROPEANA - final
Concept for inclusion of domain specific metadata vocabularies and contribution to improving access to scientific information via EDM

Revision: Version 1.1

Authors (in alphabetical order):
Benda Odo AIT Forschungsgesellschaft mbH
Höller Astrid AIT Forschungsgesellschaft mbH
Koch Gerda AIT Forschungsgesellschaft mbH
Koch Walter AIT Forschungsgesellschaft mbH

---

<table>
<thead>
<tr>
<th>P</th>
<th>Public</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Confidential, only for members of the consortium and the Commission Services</td>
<td></td>
</tr>
</tbody>
</table>
Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Author</th>
<th>Organisation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draft 0.1</td>
<td>2014-01-27</td>
<td>A. Höller</td>
<td>AIT</td>
<td>Concept and Draft</td>
</tr>
<tr>
<td>Version 0.2</td>
<td></td>
<td>O. Benda, A. Höller, W. Koch, G. Koch</td>
<td>AIT</td>
<td>Including comments</td>
</tr>
<tr>
<td>Version 1.0</td>
<td>2014-01-28</td>
<td>G. Koch</td>
<td>AIT</td>
<td>Finalisation of Version 1</td>
</tr>
<tr>
<td>Version 1.1</td>
<td>2014-01-29</td>
<td>Coordination Team (P. Böttinger, A. Michel)</td>
<td>BGBM</td>
<td>Minor Editing</td>
</tr>
</tbody>
</table>

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise.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

Distribution

<table>
<thead>
<tr>
<th>Recipient</th>
<th>Date</th>
<th>Version</th>
<th>Accepted YES/NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMG</td>
<td>29.1.2014</td>
<td>1.0</td>
<td>YES</td>
</tr>
<tr>
<td>Project Coordinator</td>
<td>29.1.2014</td>
<td>1.1</td>
<td>YES</td>
</tr>
</tbody>
</table>
Table of Contents

1 DESCRIPTION OF WORK ............................................................................................................. 4

2 COMMON NAMES VOCABULARY SERVICE ........................................................................... 5
   2.1 Extension of Pentaho Transformation and its Parameters .................................................. 5
      2.1.1 Extension 1: Voc URI? ................................................................................................. 7
      2.1.2 Extension 2: Add NS .................................................................................................. 7
      2.1.3 Extension 3: Voc Service ............................................................................................. 7
      2.1.4 Extension 4: use abcd + voc ....................................................................................... 9
   2.2 Transformation result ........................................................................................................... 9

3 BIODIVERSITY HERITAGE LIBRARY .................................................................................... 11
   3.1 Transformation result .......................................................................................................... 12

4 GEONAMES and DWC Type vocabulary ............................................................................... 14
   4.1 Transformation result .......................................................................................................... 14

5 LIST OF FIGURES ................................................................................................................ 17

6 LIST OF REFERENCES .......................................................................................................... 18
1 DESCRIPTION OF WORK

Based on the analysis of EDM and various domain specific vocabularies a concept for inclusion of metadata vocabularies and metadata enrichment was worked out. For this purpose existing tools for building and deploying semantic knowledge representations were evaluated.

The contextual classes of EDM support the modelling of semantic enrichment and allow to present information that is distinct from the actually provided cultural heritage object as giving additional details on eg. the common names of the depicted natural heritage object data, or the link to the place of gathering. Usually the values of the properties of these classes are taken from controlled vocabularies and thesauri in form of identifiers that link to further information to the vocabulary term (eg. the longitude/latitude of the place of finding, the various common names etc.) The enrichment processes in OpenUp! fulfil the tasks to provide the values of the properties of the EDM contextual classes edm:Place and skos:Concept, and OpenUp! enriches the data by linking to an object type vocabulary (edm:hasType) and to external bibliographic information from the Biodiversity Heritage Library (dc:relation).

Figure 1 Ingesting records into Europeana (technical components)
In this document the used services and results are presented.

2 COMMON NAMES VOCABULARY SERVICE

A style sheet is used in order to set the rules for displaying the mapped vocabulary information (see Figure 2).

This stylesheet defines that the vocabulary information is added to a record via the metadata field dc:subject. For each abcd:FullScientificNameString the vocabulary is searched for an appropriate common name. When one is found it is added to the metadata as a value of dc:subject.

Because Europeana does not display the skos:note that carries the references to the vocabulary sources so far in the portal OpenUp! added the to the common names web service the skos:editorialNote field that displays a link to the vocabulary information. The skos:editorialNote is added as final dc:subject field to the object description.

Figure 2 EDM Stylesheet

2.1 Extension of Pentaho Transformation and its Parameters

The transformation of the OpenUp metadata from the ABCD format into the ESE/EDM format is done with the Pentaho Kettle PDI tool. In order to generate the metadata suitable for Europeana's EDM format the ESE transformation routine is extended with the Ontology data. This is done by using the Ontology Data Gateway's REST service in the transformation program (see Figure 3).
Opening Up the Natural History Heritage for Europeana

Figure 3 Extended Pentaho Transformation with REST service (marked red)
2.1.1 Extension 1: Voc URI?
The first step called “Voc URI?” is a Filter Rows step. If there is a vocabulary URI the data is forwarded to step two “Add NS”. Otherwise it is directly sent to “Get Units from XML” (compare Figure 3). Figure 4 shows the configuration of this step with its condition.

![Figure 4 The Filter Rows step “Voc URI?”](image)

2.1.2 Extension 2: Add NS
When there is a vocabulary URI the data is sent to the second step “Add NS” (add namespace). This “Replace in String” step replaces `<biocase:response` with `<biocase:response xmlns:dc=http://purl.org/dc/elements/1.1/>` and adds the namespace (see Figure 6).

![Figure 5 Step 2: Adding a namespace](image)

2.1.3 Extension 3: Voc Service
The “Rest client” step named “Voc Service” accesses the Ontology Service (see Figure 6) and is defined with the variable URL $\{vocabulary_service_uri\}$ which is represented in the Transformation Parameters (see Figure 7, Parameter number 18 with URL `http://ait117:8080/Vocabulary/rest/~Mapping/NHMW_common_name/perform`).
Figure 6 Pentaho Ontology Service access

Figure 7 Pentaho Parameters
2.1.4 Extension 4: use abcd + voc

Finally the forth step “use abcd + voc” is a “Select Values” step which is executed in order to select other needed fields. As can be seen in Figure 8 the field abcdXMLwithVoc is renamed to abcdXML.

![Figure 8 Renaming “abcdXMLwithVoc” to “abcdXML”](image)

2.2 Transformation result

Figure 9 shows the butterfly *Papilio machaon Linnaeus, 1758*. The sample record for this specimen is shown in Figure 10. As can be seen the record includes many subjects with common names in different languages.

![Figure 9 Papilio machaon Linnaeus, 1758](image)
The stylesheet defines that the vocabulary information is added to a record via the metadata fields \textit{dc:subject}. For each \textit{abcd:FullScientificNameString} the vocabulary is searched for an appropriate common name. When one is found it is added to the metadata as a value of \textit{dc:subject}. In addition, the source of the common name with its reference link is added to the final \textit{dc:subject} metadata field (see Figure 10).

In Europeana the record with the subjects looks like shown in Figure 11.
3 BIODIVERSITY HERITAGE LIBRARY

The stylesheet also defines that for each scientific name the BHL API “bibliography by URL” is evoked and a link to the BHL bibliography is added in the dc:relation metadata field.

```xml
<xsl:variable name="bhllink"><xsl:for-each select="$unts/Unitt">
  <xsl:call-template name="BHLLink" />
</xsl:for-each></xsl:variable>
<xsl:for-each select="fn:distinct-values($bhllink/dc:relation)">
</xsl:for-each>
```

Figure 12 Mapping dc:relation
For this mapping a service of the Biodiversity Heritage Library is used¹.

3.1 Transformation result

Figure 14 shows the butterfly Luehdorfia japonica Leech, 1889. This record includes dc:relation with the BHL-service (see Figure 14).

![Figure 13 Luehdorfia japonica Leech, 1889](image)

When clicking on [http://www.biodiversitylibrary.org/name/Luehdorfia_japonica_Leech%2C_1889](http://www.biodiversitylibrary.org/name/Luehdorfia_japonica_Leech%2C_1889) (28. Jan. 2014.) the following information can be seen (see Figure 15).

Figure 15 Bibliographic information concerning Luehdorfia japonica Leech, 1889

When clicking on one of the result sets (Page # for example the first one) the bibliographic source is shown in detail (see Figure 16).

Figure 16 Bibliographic source in detail
In Europeana the relation looks like shown in Figure 17.

![Figure 17 The BHL-service in Europeana](image)

### 4 GEONAMES and DWC Type vocabulary

When a raw record contains geographic coordinates the [http://www.geonames.org/maps/google_%7BLatitudeDecimal%7D_%7BLongitudeDecimal%7D.html](http://www.geonames.org/maps/google_%7BLatitudeDecimal%7D_%7BLongitudeDecimal%7D.html) service is used. In the stylesheet the service is mapped to `edm:Place`. (see Figure 18).

```xml
<xstlvARIABLE name="edmPlace">  
  <xsl:for-each select="Sunits/Unit/Gathering/SiteCoordinateSets/SiteCoordinates/CoordinateLonLat">  
    <xsl:if test="LongitudeDecimal and not(LongitudeDecimal="8' and LatitudeDecimal="0")">  
      <edm:Place rdf:about="http://www.geonames.org/maps/google[LatitudeDecimal]_[LongitudeDecimal].html">  
        <xsl:if test="SpatialDatum="WGS84"">  
          <wgs84:lat>xsl:value-of select="LatitudeDecimal"/></wgs84:lat>  
          <wgs84:lon>xsl:value-of select="LongitudeDecimal"/></wgs84:lon>  
        </xsl:if>  
        <edm:Place>  
          <edm:Place>  
            <xsl:for-each>  
              <xstlvARIABLE>  
            </xsl:for-each>  
          </edm:Place>  
        </edm:Place>  
      </edm:Place>  
    </xsl:if>  
  </xsl:for-each>  
</xstlvARIABLE>
```

![Figure 18 Stylesheet containing geonames information](image)

#### 4.1 Transformation result

Figure 19 shows another specimen called *Ranunculus trichophyllus Chaix*. As can be seen it contains spatial information.
When looking at this record in Europeana (see Figure 20) the “Place Term” information is added by OpenUp! (see Figure 21).
In this sample record the link to the Darwin Core Type vocabulary\(^2\) can be seen as well.

5 LIST OF FIGURES

Figure 1 Ingesting records into Europeana (technical components) ........................................ 4
Figure 2 EDM Stylesheet........................................................................................................... 5
Figure 3 Extended Pentaho Transformation with REST service (marked red) ......................... 6
Figure 4 The Filter Rows step “Voc URI?” .............................................................................. 7
Figure 5 Step 2: Adding a namespace ....................................................................................... 7
Figure 6 Pentaho Ontology Service access .............................................................................. 8
Figure 7 Pentaho Parameters ................................................................................................... 8
Figure 8 Renaming “abcdXMLwithVoc” to “abcdXML” .......................................................... 9
Figure 9 Papilio machaon Linnaeus, 1758 ............................................................................. 9
Figure 10 Record with description of the object ....................................................................... 10
Figure 11 Europeana record with Subjects ........................................................................... 11
Figure 12 Mapping dc:relation .............................................................................................. 11
Figure 13 Luehdorfia japonica Leech, 1889 ......................................................................... 12
Figure 14 Record with dc:relation .......................................................................................... 12
Figure 15 Bibliographic information concerning Luehdorfia japonica Leech, 1889 ............. 13
Figure 16 Bibliographic source in detail.................................................................................. 13
Figure 17 The BHL-service in Europeana............................................................................. 14
Figure 18 Stylesheet containing geonames information .......................................................... 14
Figure 19 The specimen Ranunculus trichophyllus Chaix .................................................... 15
Figure 20 Europeana record with “Geographic coverage” and “Place Term” information .... 15
Figure 21 http://www.geonames.org/maps/google_47.7208333327_16.067222216.html .... 16
Figure 20 Europeana record with Link to Darwin Core Type vocabulary ................................. 16
6 LIST OF REFERENCES


