

Definition of VLO facets

1. Summary Extraction Process

The metadata importer basically uses the following procedures to identify and extract relevant entries from CMDI files into the respective facet.

1. Identification of relevant CMDI elements

Every facet is defined by a set of data categories (mostly based on ISOcat) and XPath expressions. At the import process all occurring CMDI profiles are checked for elements that fit at least one of these criteria (i.e. appropriate concept link or XPath expression) for every facet.

2. Extraction

The content of all matching CMDI elements is extracted.

3. Post Processing

For some facets specific post processing procedures are used to adapt the extracted values (like assigning language names to ISO639-3 codes, mapping collection names to project names etc.). Right now post processors are used for the facets *Country*, *Language*, *ResourceType*, *Nationalproject*, *ComponentProfile*.

4. Storage

After post processing (if any) the values are stored in the database.

For a more detailed description see: [Semantic metadata mapping in practice: the Virtual Language Observatory](#) (LREC 2012)

2. Facet List

The following summary lists the actual mapping used by the VLO. Due to legacy reasons some of these facets contain mappings that do not fit the intended meaning (like DC-2591 „unknown“ for facet *Subject*). These are already under consideration for a removal in the future.

1. Id
[/CMD/Header/MdSelfLink/text\(\)](#)
[DC-2573 \("persistent identifier"\)](#)
2. Collection
[/CMD/Header/MdCollectionDisplayName/text\(\)](#)
3. ProjectName
[DC-2536 \("project name"\)](#)
[DC-2537 \("project title"\)](#)
4. Name
[DC-5428 \("title statement"\)](#)
[DC-5127 \("Original"\)](#)
[DC-4160 \("web service name"\)](#)
[DC-4114 \("book title"\)](#)
[DC-2544 \("resource name"\)](#)
[DC-2545 \("resource title"\)](#)
[DC-4160 \("web service name"\)](#)
[http://purl.org/dc/terms/title](#)
[http://purl.org/dc/elements/1.1/title](#)
[/CMD/Components/LrtInventoryResource/LrtCommon/ResourceName/text\(\)](#)
5. Year
[DC-2509 \("completion year"\)](#)
[DC-2510 \("creation date"\)](#)
[DC-2538 \("publication date"\)](#)

6. Continent
[DC-2531 \("location continent"\)](#)
7. Country
[DC-2532 \("location country"\)](#)
/CMD/Components/OLAC-DcmiTerms/spatial[@dcterms-type="ISO3166"]/text()
/CMD/Components/OLAC-DcmiTerms/coverage[@dcterms-type="ISO3166"]/text()
8. Language
[DC-2482 \("language ID"\)](#)
[DC-2484 \("language name"\)](#)
/CMD/Components/OLAC-DcmiTerms/language/@olac-language
/CMD/Components/OLAC-DcmiTerms/subject/@olac-language
/CMD/Components/LrtInventoryResource/LrtCommon/Languages/ISO639/
iso-639-3-code/text()
9. Organisation
[DC-2459 \("organization"\)](#)
[DC-2979 \("Organisation"\)](#)
<http://purl.org/dc/terms/publisher>
<http://purl.org/dc/elements/1.1/publisher>
10. Genre
[DC-2470 \("genre"\)](#)
[DC-3899 \("sub genre"\)](#)
/CMD/Components/OLAC-DcmiTerms/type/@olac-linguistic-type
11. Subject
[DC-2591 \("unknown", will be removed\)](#)
<http://purl.org/dc/terms/subject>
<http://purl.org/dc/elements/1.1/subject>
/CMD/Components/OLAC-DcmiTerms/subject/@olac-linguistic-field
/CMD/Components/OLAC-DcmiTerms/subject[@dcterms-type="LCSH"]/text()
12. Description
[DC-2520 \("description"\)](#)
<http://purl.org/dc/terms/description>
<http://purl.org/dc/elements/1.1/description>
13. ResourceType
[DC-2571 \("mime type"\)](#)
<http://purl.org/dc/terms/type>
<http://purl.org/dc/elements/1.1/type>
/CMD/Components/LrtInventoryResource/LrtCommon/ResourceType/text()
/CMD/Components/OLAC-DcmiTerms/type/text()
14. NationalProject
/CMD/Header/MdCollectionDisplayName/text()
15. ComponentProfile
/CMD/Header/MdProfile/text()
16. Tag
[DC-5436 \("metadata tag"\)](#)
[DC-5436 \("metadata tag", alternative link to ISOcat datcat, will be removed\)](#)