Thesaurus maintenance WG workshop, Heraklion, Crete report (16-17/6/2015)

During the workshop a detailed presentation of the facets (definitions) and hierarchies/narrower terms developed so far was given.

The overall goals have been:

- To define principles which allow us to have a basic agreement about basic concepts and their generalisations/ specialisations at least at the upper level, across the research infrastructure,
- to match the principles against the current definitions,
- to develop a common understanding of how the BBT should work, be implemented and be maintained,
- to agree on the verification process of BBT content and correct use

Some participants expressed concerns about the overlapping with the CIDOC-CRM Ontology concepts. The distinction between ontology and controlled vocabularies was made clear: In the backbone thesaurus we deal with types of entities for classification according to disciplinary aspects, and not describe facts about instances for information integration of cultural-historical contexts. This creates a difference in scope which is not further analyzed currently. The BBT is developed from the analysis of terminological systems of partners and their intended application. Later, a mapping will be provided about which BBT facets are compatible to classify instances (as "P2_has type") of which CRM classes.

In the "Thesaurus Maintenance" workshop we aim at classifying terms emerging in the humanities under broader concepts. In our classification we follow the IsA relationship according to which the narrower terms inherit the properties of the broader terms. Each term is a subclass of a

facet inheriting all its characteristics. Particulars are not terminology, so we don't deal with them.

In many cases the example of Art & Architecture Thesaurus (AAT) was brought up in comparison with the Backbone Thesaurus (BBT). The AAT has to do mainly with material objects and does not cover for instance types of concepts. The BBT follows a similar, basically compatible logic in the way the AAT organizes facets and hierarchies. The BBT is an attempt to improve this upper level of the AAT and provide a wider scope. It is being developed in complete awareness of the AAT. Later, a mapping of AAT terms/facets and hierarchies to the BBT should be provided.

During the detailed discussion of examples and cases that followed, improvements in the definitions were made, narrower terms were refined (in activities, functions, materials, natural processes, immobile objects, monuments, complexes, propositional objects, methods etc.) while areas that are not covered by the facets/hierarchies developed so far were detected (biology, states or changes of the inner psyche in psychology). If a distinction is not clear in the domain then narrower terms should be removed from the BBT.

A term (word) may appear under two or more facets but not with the same scope note, i.e. there must be a conceptual distinction. We have cases of strong polysemy (e.g. museum can be: a building, its people) which constitute distinct concepts. It is good practice to disambiguate the term with a parenthesis: "museum(building)", "museum(organisation)", "museum(holdings)". Cases of true polyhierarchy have to stay within a facet. For instance, "carmine" is a "dye", a "pigment", a "substance of animal-origin", a "red colorant".

The forming of hierarchies is not permanently fixed. It may become useful to split or merge hierarchies, or to introduce facets on top of existing facets and thus downgrading existing facets to hierarchies, or to introduce a hierarchy between a term and a facet or hierarchy. Some of these operations are monotonic – they will not affect the term base

under the BBT, some will affect it. Respective update operations are common practice for instance by the LCSH team. A complete description of these maintenance operations and their resolution will be provided.

Discussion about tools:

We discuss two kinds of tools:

- submission tool to manage user requests to improve the BBT by an expert team.

- Thesaurus management tool to implement and access the BBT content itself in SKOS form
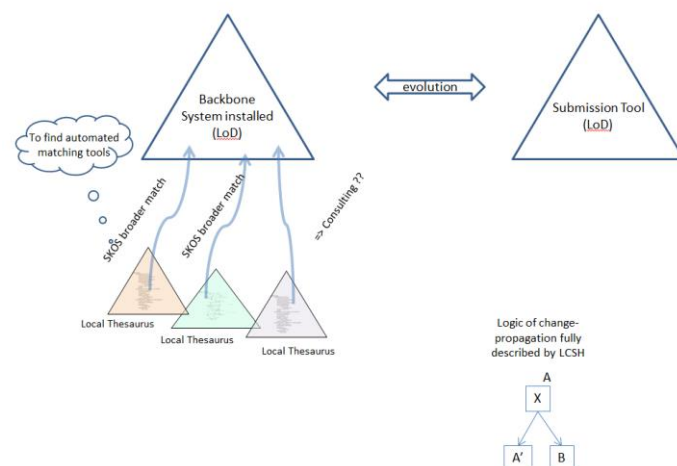

Figure 1:

Requirements for the Submission tool:

The existing tool from ICS-FORTH was discussed, and how the current system of roles can be adapted to the maintenance requirements of the BBT. Provisional maintenance requirements of the BBT are the following

a) Change scope notes

b) Disambiguation remarks

c) New hierarchy

d) merging facet up (Facet => hierarchy), "bad change": split facet, delete hierarchy

e) foresee mappings to high level resources (e.g. AAT), and support roles such as coordinator, submitter, experts

f) no "skos related" (XM notes: we do not have relations terms submission) in the backbone, linking in the backbone by URI and skos broad match (interpret as transitive)

g) registry of linking vocabularies and responsible people require "links to backbone" flags

The submission tool from FORTH currently is configured to interact with Protégé-OWL as ontology implementation. It needs to be adapted to a SKOS-based thesaurus management tool.

Requirements for the thesaurus management tool:

1. Multiple languages

2. A way to present the versus thesaurus as linked dataset on the web (URI)

3. view backbone as part of a local hierarchy, view a local vocabulary as part of a larger

4. SKOS vocabulary on the edge of running service

5. Embeditor that allows us to bring new

6. Flexibility for editing. We want more than 5.000 terms

7. Print out ability (export)

8. Viewing mechanism (visualisation tool, local terminology development support)

9. Distribution of local thesauri as part of the Backbone

Todos:

Practical overview of hierarchies for having a clear status of terms (Graphic representation of hierarchies)- ICS

- List of all update operations in classification of terms in the thesaurus management tool
  - Provide a method of verification and improvement of definitions of facets hierarchies

- Development of a hierarchy for processes and states of the psyche
  - Provide psychology terms - AA
- Mapping of BBT to compatible concepts in the CIDOC CRM. Exploration how differences of scope and function of CRM and BBT are reflected in the way BBT concepts are defined.
- Comprehensive document with examples of terms to be presented (AA, 1st week of August)
- Method description (scholarly guidlines making terms, facets etc)
- Cookbook version (short version 2 pages with explanation 10 rules)
- **Get partners** match terms to the BBT

Developing tools
1. Excel form of BBT to ÖAW (ICS, 2weeks)
2. RDF form to AA/FORTH (ÖAW, 2weeks)
3. RDF form uploaded on the web as LOD (ÖAW, 2 weeks)
4. Link a local term or more into BBT as LOD

Getting feedback from partners regarding terms
Problem reports
· Scope note ambiguous
· Missing hierarchy
· Ambiguity of classification

Document on Submission / Maintenance