Assisting Backbone Thesaurus

use and maintenance

Methodology and Infrastructure

Produced by the
Thesaurus Maintenance Working Group,
VCC3, DARIAH EU

Version 2.3
Status: working document

April 2018
(last update: 19/04/2018)

Contributors: Christos Georgis, Martin Doerr, Konstantina Konsolaki, Maria
Daskalaki, Chryssoula Bekiari, Lida Charami, Ilias Tzortzakakis, Evagelia
Daskalaki, Helen Katsiadaki, Helen Goulis, Hella Hollander, Matej Durco,
Vanessa Hannesschläger, Wolfgang Schmidle, and others

Table of Contents

Introduction

The aim of this work is to design and develop a maintenance methodology, along with a toolset to assist that methodology, following the proposal of how existing thesauri and ontologies will become interoperable and can be maintained in a sustainable and scalable way. This work follows the work proposed in the report "*Definition of a model for sustainable interoperable thesauri maintenance".[1]* This model proposal has been undertaken by the Thesaurus Maintenance WG, which was established in 2014 in the framework of DARIAH EU: *The Digital Research Infrastructure for the Arts and Humanities - a research infrastructure*. This Research Infrastructure aims at enhancing and supporting digitally-enabled research and teaching across the arts and humanities.

The idea proposed in the above report is to design and establish a coherent overarching thesaurus for the humanities, a "backbone" or "metathesaurus", under which all the vocabularies and terminologies in use in the domain can be aligned. The proposed approach is bottom-up; top-level terms are developed by adequate abstraction from existing local terminological systems.

We need to support all the stakeholders in this endeavor, by proposing a maintenance methodology, along with an assisting toolset that would:

- enable <u>independent local thesauri maintainers</u> to create and maintain their thesauri, and at the same time incorporate them, while still maintaining their independence, into a shared common thesaurus, that will be available to the public,
- enable the <u>curators</u> of this common scheme of abstract terms (hereafter BackBone Thesaurus, or BBT), to support and maintain the BBT, as a central thesaurus which would provide the general terms under which *local thesauri maintainers* can link/connect their thesauri,
- enable <u>potential users (public, scientific community, etc.)</u> to browse, navigate, visualize and use this very rich thesaurus that would incorporate the wealth of the different thesauri.

1   Objects, Actors and Proposed Workflow

1.1   Handled objects

Local thesauri existing (or new) thesauri and ontologies, independently developed and maintained. These thesauri that would like to become part of a coherent overarching thesaurus for the humanities, a "backbone" or "metathesaurus", under which all the vocabularies and terminologies in use in the domain can be aligned.

BackBone Thesaurus (BBT) is a coherent overarching thesaurus for the humanities, the "backbone" or "metathesaurus", under which all the vocabularies and terminologies (local thesauri) in use in the domain can be aligned. The BBT model is maintained in a thesaurus database with the use of **BBT Management tool.** The official description of the BBT model is automatically exported from the BBT management tool database, in two forms:

---

[1] "*Definition of a model for sustainable interoperable thesauri maintenance", p*roduced by Thesaurus Maintenance Working Group, VCC3, DARIAH EU, Version 1.2, September 2016, (http://backbonethesaurus.eu/sites/default/files/DARIAH_BBT%20v%201.2%20draft%20v4_0.pdf)

- the **BBT Definition Document**, that describes the BBT model in textual form. In the appendix are described all changes between the current and the previous versions of BBT.
- the **BBT LOD representation** of the BBT model (SKOS RDF description), which is available through a service: **BBT Access Service.** A thesaurus system (software) hosts and provides public access to the official BBT LOD version. This system maintains consistent identification (LOD identification) for all terms of the BBT, in order to be referenced by the local thesauri, or to be accessed by the public.

DARIAH Thesaurus Federation (BBT and Local thesauri) is the federated thesaurus for the humanities, which comprises the BBT and all the local thesauri that are aligned with the BBT.

1.2   Involved parties

Local thesauri maintainers may already have built one (or more) thesaurus(ri) or wish to create a new one. We do not intent to interfere with the existing thesaurus creation workflows or practices, but we, nevertheless, need to have a basic agreement regarding the basic terms and their generalizations/ specializations as represented in the "backbone thesaurus".
Currently the Deutsches Archäologisches Institut, (DAI, www.dainst.org) uses the BBT (version 1) and discusses questions and proposals of improvement of the BBT with the curators.

Curators is the group of thesaurus experts responsible for changes in the BBT model. They take requests for changes, regarding terms, from different users of the BBT and decide upon their validity. Among the *curators* there might be a curator that coordinates the group; for instance he is responsible to select a submission regarding a change and initiate the discussion on this change, and also to end the discussion (e.g. concluding that a common agreement is reached or by asking a voting to take place, etc.). Once a decision on a change is made they are responsible to introduce the change to the BBT model using the BBT Management tool. *Curators* are also responsible to decide on the publication of a new version of BBT model, by making available the BBT Definition Document (the official description of the BBT model) and the BBT LOD model (a SKOS RDF document).
Currently the c*urators* are the members of the Thesaurus Maintenance Working Group, VCC3, DARIAH EU.

BBT access providers are responsible to host and provide access to the current version the BBT. They load the exported BBT LOD model (RDF description) to the BBT Access Service thus exposing the current official BBT version to the public and maintaining consistent identification (LOD identification) for all terms of the BBT, in order to be referenced by the local thesauri.
Currently the *BBT access providers* are ACDH-OEAW members (www.oeaw.ac.at/acdh).

Potential thesaurus users (public, scientific community, etc.) are the users of all the vocabularies and terminologies that are (or could be) aligned under the BBT. These users should be provided with tools that browse, navigate, visualize and use this very rich thesaurus infrastructure that would incorporate the wealth of all the different thesauri in the DARIAH Thesaurus Federation.

1.3   Overall Workflow

The evolution and maintenance of the BBT involves suggestions for additions and changes in the thesaurus, discussions, decision making and finally the implementation and publishing of new versions of the BBT. In this section, we propose a maintenance workflow:
- **Making requests for BBT changes.** The BBT is expected to get updated or extended with the addition of new terms. *Local thesaurus maintainers,* and *curators* alike, may

suggest changes in the BBT (modification[2], addition, or deletion of terms of the BBT). We propose the use **BBT Submission and Connection Management tool** that would facilitate submission of such requests and would also enable the discussion on such requests, hereafter also called submissions. Furthermore, this tool enables to connect local thesauri terms to BBT terms, which will be discussed later.

- **Deciding upon requests for BBT changes.** Since the BBT is a common thesaurus scheme, any change (modification, addition, or deletion) has to be commonly decided by the *curators*. The *curators* will use the <u>BBT Submission and Connection Management tool</u> in the decision making process: while processing a submitted change the *curators* might need to review past discussions regarding the proposed change, in order to accept, reject or postpone it. They might also need to go back in the BBT version history and consult the differences between the different BBT versions. The <u>BBT Submission and Connection Management tool</u> will keep track of the different versions of the BBT and the history of the submissions (related past discussions). Notice that in this process c*urators* may also forward a submission to third parties (external to WG) that are considered to be experts in specific domains (thesaurus-domain experts), for further consultation. These experts will also use <u>BBT Submission and Connection Management tool</u> and take part in specific change-related discussions.

- **Making a BBT change**. After a change is approved and agreed upon, the c*urators* will have to introduce the change into the database using the <u>BBT Management tool</u>. Since the <u>BBT Submission and Connection Management tool</u> keeps track of all the involved parties in the discussion, it will also notify them, about the progress of a submission, as well as the release of the new versions of the BBT (see below).

- **Publishing a BBT new version**. The *curators* will use <u>BBT Management tool</u> to update the current BBT version in the thesaurus database. A new version of BBT may include several minor or few major changes of the BBT. The *curators* are responsible to decide upon the publication of a new version of BBT. Once decided, an official version of the BBT is released: both the <u>BBT Definition Document</u> (the official textual description of the BBT model) and the BBT LOD model (a SKOS RDF document) are exported and made public (the new BBT LOD model is loaded and made accessible by the <u>BBT Access Service</u>). Exposing the new BBT version to the public requires that consistent identification (LOD identification) is maintained for all terms of the BBT, in order to be referenced by the local thesauri, without loss of their referential integrity. Publishing a new version of the BBT may also affect the local thesauri that are connected to BBT; therefore, *local thesauri maintainers* are notified as described below.

- **Connecting local thesauri terms to BBT terms.** *Local thesauri maintainers* create their own local thesauri, using their own workflow and software. We encourage *local thesaurus maintainers* to use terms from BBT as top-terms in their thesauri. This will enable the alignment of their vocabularies and terminologies (thesauri) under one shared thesaurus, the BBT.
  The first step in connecting local thesauri with the BBT in general means deciding which of the upper level terms of the local thesauri should classified[3] under the general terms of the BBT. We propose, that *local thesauri maintainers* should include in their local thesauri general BBT terms, by using local terms (declared as "same

---

[2] Notice that modification of a term, may mean the change of its scope note, the change of its label, change of its relations to other terms, etc.

[3] This linking/connection of the local thesaurus with the BBT should be performed by the *local thesaurus maintainers* using their own thesaurus maintenance workflow and software.

as"/"exact equivalence" or "narrower of" to the BBT terms, by their LOD identifiers as these are provided by the BBT Access Service). This would constitute a one-direction link/connection from the local thesaurus to the BBT.

Additionally we use the BBT Submission and Connection Management tool to maintain a second connection originating from the BBT term to its connected term in the local thesauri. The tool enables *local thesaurus maintainers* to create this connection (LOD identification of the local thesaurus term which is declared as "same as" or "narrower of" the BBT term) and stores contact information of the *local thesauri maintainers* in order to keep them updated for changes on the specific BBT term (e.g. contact e-mail, organization info, etc.).
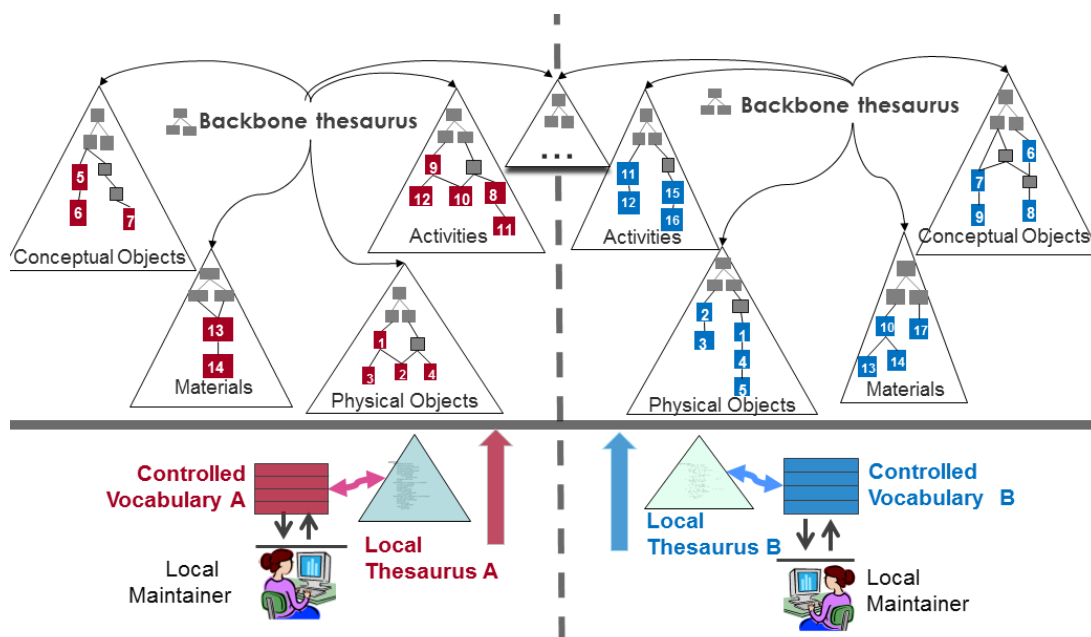

Figure 1: Connections to the BBT

- **Notifying *local thesauri maintainers* about new BBT version changes that may affect them.** The BBT Submission and Connection Management tool also includes a service that notifies the *local thesauri maintainers* about changes in the new version of BBT that may affect them. For instance, if a BBT term is modified (e.g. its scope note is updated, thus its meaning is altered), all local thesauri developed that are connected to the specific BBT term as a top-terms in their thesaurus, are notified about the change in order to verify if the specific change affects their local thesauri.
- **Disconnecting local thesauri from the BBT.** *Local thesauri maintainers* may also decide to disconnect their thesauri from BBT. For that, they should remove the connections from the local term (declared as "same as" to the BBT terms (removal of the one-direction connection from the local thesaurus to the BBT). Additionally removing the second connection is possible by using the BBT Submission and Connection Management tool, which removes the connection originating from the BBT term to its connected term in the local thesauri.
- **General requirements**
  o All thesauri (local and BBT) should use consistent LOD identifiers for referencing terms and their relations. These identifiers should not change across thesauri versions.

o All tools should be able to export and import data (thesauri, or parts of thesauri) in SKOS format, under a consistent scheme.
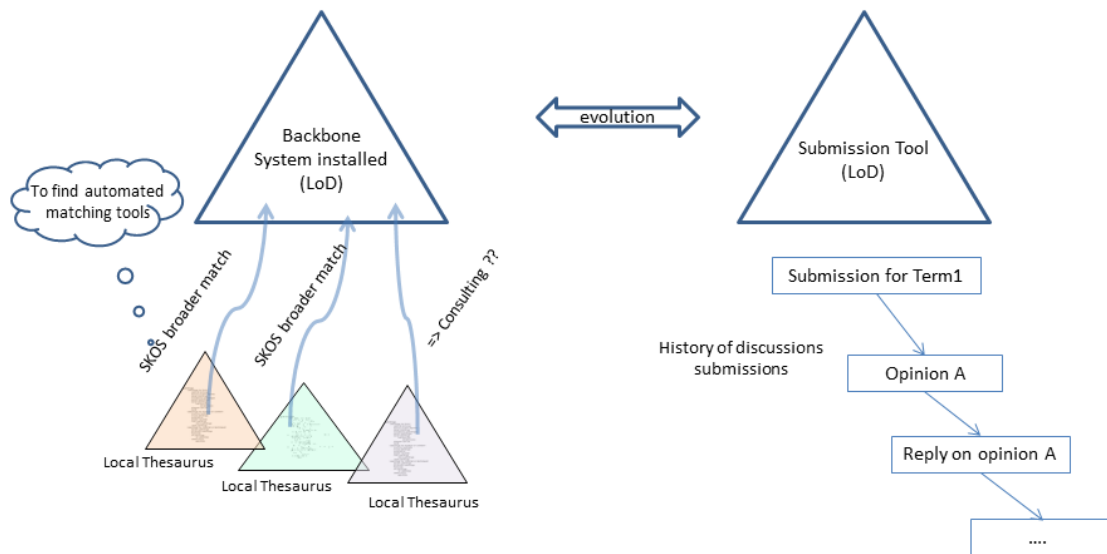

Figure 2: BBT Management Tool *and* BBT Submission and Connection Management tool

2   Proposed tools

The following sections give an overview of the infrastructure-components / tools and we describe their basic features.

2.1   BBT Access Service

The BBT Access Service is a service responsible for hosting and providing access to the current version the BBT. It provides consistent identification (LOD identification) for all terms of the BBT, in order to be referenced by the local thesauri.

The proposed BBT Access Service, is based on Scosmos (http://skosmos.org) an open source tool, customized and maintained by ACDH-OEAW (www.oeaw.ac.at/acdh). The BBT Access Service, runs at https://vocabs.dariah.eu/backbone_thesaurus, providing web-presence and LOD access to the current version the BBT.

Once decided, that an official version of the BBT model is to be released, the BBT LOD model (SKOS RDF document) is exported from the BBT management tool and made public (the new BBT LOD model is loaded and made accessible by the BBT Access Service).

2.2   BBT Management tool

The BBT Management tool is the system responsible for BBT management. It communicates with the BBT Access Service to ensure consistent identification (LOD identification) for all terms of the BBT, in order to be referenced by the local thesauri.

The proposed BBT Management tool, is implemented as an additional functionality of BBT Submission and Connection Management tool (see below). It enables the management and administration of BBT according to the principles of ISO 25964-1 and ISO 25964-2 standards, developed by FORTH-ICS (www.ics.forth.gr).

This tool is used by the *curators* who maintain the BBT model. Once decided, that an official version of the BBT model is to be released: both the BBT Definition Document (the official

textual description of the BBT model) and the BBT LOD model (a SKOS RDF document) are exported from the BBT management tool and are made public. The new version of BBT LOD model is sent to the BBT Access Service to be loaded in order to be publicly accessible and to able to receive submissions and connections.

2.3 BBT Submission and Connection Management tool

The BBT Submission and Connection Management tool (a.k.a. BBTalk) is a communication system[4] (www.backbonethesaurus.eu/BBTalk), developed by FORTH-ICS (www.ics.forth.gr), that supports discussions regarding the changes proposed for the BBT (changes related to terms and their relations), hereafter called submissions. It keeps track of the different versions of the BBT and the history of the submissions (related past discussions). It also notifies all the interested parties, about the progress of a submission, and the release of the new versions of the BBT.

The BBT Submission and Connection Management tool is used by *local thesauri maintainers* to suggest changes for the BBT (*contributors*); they can request modifications/additions/deletions regarding the terms of the thesaurus. The tool is also used by the *curators* to browse and review submissions, and decide whether they agree to the suggested changes or disagree and ignore/reject/postpone them. The system also provides access to the previous BBT versions of the thesaurus and the history of all the submissions in order to support *curator* decision. The *curators* may also invite in the discussion on a submission users that are experts in specific domains (*reviewers*), for further consultation*. Reviewers* use the tool to take part on the above discussions once they are invited.

Finally, the BBT Submission and Connection Management tool provides functionality for creating (and removing) connections originating from BBT terms to local thesauri terms (LOD identified), for storing contact info and for notifying all interested parties.

2.4 Federated thesauri viewer

Since the different thesauri of the proposed thesauri federation (BBT and local thesauri) are located in different namespaces/URIs (note that common terms of BBT and local thesauri are only inter-connected) and accessed by different systems (as each thesauri may provide its specific viewer) we should provide a federated thesauri viewer that would enable to browse, navigate, visualize and use the different thesauri of the proposed thesauri federation. This viewer should be able to work with the different thesauri (them being either available online or cashed) providing a single interface.

We propose that the Federated thesauri viewer (this single/unified interface) to be implemented based on Scosmos, the open source platform, customized and maintained by ACDH-OEAW, which is also used by BBT Access Service. The local thesauri or parts of local thesauri (exported in SKOS files[5]), along with the BBT LOD model and the (LOD identified) connections between BBT terms and local thesauri terms, would be loaded to the viewer platform and thus would be made accessible to the public. The viewer would enable browsing, navigation, visualization and maybe even querying the terms of the proposed thesauri federation (BBT and local thesauri).

---

[4] A detailed description of the proposed system can be found in "*BBTalk - BBT Submission and Connection Management tool – version (2.1) - System Description*", January 2018: http://backbonethesaurus.eu/sites/default/files/BBT-SubmissionConnectionMgntTool_v2.1_draft.pdf

[5] As discussed in section "1.3 Overall Workflow", local thesauri, or parts of local thesauri should be exported in SKOS format, under a consistent scheme and consistent LOD identifiers.