

Assisting Backbone Thesaurus use and maintenance

Methodology and Infrastructure

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

Version 2.0
Status: working document

March 2017
(last update: 30/03/2017)

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



Table of Contents

Table of Contents	2
Introduction	3
1 Objects, Actors and Proposed Workflow.....	4
1.1 Handled objects.....	4
1.2 Involved parties	4
1.3 Overall Workflow	5
2 Proposed tools.....	7
2.1 BBT - Thesaurus Access tool.....	7
2.2 BBT - Thesaurus Maintenance tool.....	8
2.3 BBT - Submission and Connection Management tool.....	8
2.4 Federated thesauri viewer.....	9

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 “*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 concepts 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 independent local thesauri maintainers 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 BBT curators of this common scheme of abstract concepts (hereafter BackBone Thesaurus, or BBT), to support and maintain the BBT, as a central thesaurus which would provide the general concepts under which *local thesauri maintainers* can link/map their thesauri.
- enable potential users (public, scientific community, etc.) to browse, navigate, visualize and use this very rich thesaurus that would incorporate the wealth of the different thesauri.

¹ “*A model for sustainable interoperable thesauri maintenance*”, produced by Thesaurus Maintenance Working Group, VCC3, DARIAH EU, Version 1.1, January 2016, (http://83.212.168.219/DariahCrete/sites/default/files/dariah_bbt_v_1.1.pdf)

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:

- 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 concepts 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 concepts and their generalizations/ specializations as represented in the “backbone thesaurus”. Currently the Deutsches Archäologisches Institut, (DAI, <http://www.dainst.org/el>) uses the BBT (version 1), and discusses questions and proposals of improvement of the BBT with the BBT curators.

BBT-curators is the group of thesaurus experts responsible for changes in the BBT model. They take requests for changes, regarding concepts, from different users of the BBT and decide upon their validity. Among the *BBT-curators* there is a curator that coordinates the group (*BBT-coordinating-curator*); 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. *BBT-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 *BBT-curators* are the members of the Thesaurus Maintenance Working Group, VCC3, DARIAH EU and the BBT management tool used is Synthesis (https://www.ics.forth.gr/isl/index_main.php?l=e&c=271).

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 concepts of the BBT, in order to be referenced by the

local thesauri. Currently the *BBT access providers* are ACDH-OEAW members (<http://www.oeaw.ac.at/acdh>) and the BBT - thesaurus access tool used is OpenSKOS (<http://openskos.org/>).

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

We propose the following workflow:

- **Making requests for BBT changes.** The BBT is expected to get updated or extended with the addition of new concepts. *Local thesaurus maintainers*, and *BBT-curators* alike, may suggest changes in the BBT (modification², addition, or deletion of concepts of the BBT). We propose the use of a tool (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.
- **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 *BBT-curators*. The *BBT-curators* will use the BBT Submission and Connection Management tool in the decision making process: while processing a submitted change the *BBT-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 BBT Submission and Connection Management tool will keep track of the different versions of the BBT and the history of the submissions (related past discussions). Notice that in this process *BBT-curators* 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 BBT Submission and Connection Management tool and take part in specific change-related discussions.
- **Making a BBT change.** After a change is approved and agreed upon, the *BBT-curators* will have to introduce the change into the database using the BBT Management tool. Since the BBT Submission and Connection Management tool 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 *BBT-curators* will use BBT Management tool 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 *BBT-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 BBT Definition Document (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 BBT Access Service). Exposing the new BBT version to the public requires that consistent identification (LOD identification) is maintained for all concepts of the BBT,

² Notice that modification of a concept, may mean the change of its scope note, the change of its label, change of its relations to other concepts, etc. A complete list of all the possible changes of concepts should be defined later in this report.

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 linked to BBT, therefore we propose that *local thesauri maintainers* should be notified as described below.

- **Linking/connecting local thesauri concepts to BBT concepts.** *Local thesauri maintainers* create their own local thesauri, using their own workflow and software. We encourage *local thesaurus maintainers* to use concepts from BBT as top-concepts in their thesauri. This will enable the alignment of their vocabularies and terminologies (thesauri) under one shared thesaurus, the BBT.

The first step in linking/connecting local thesauri with the BBT in general means deciding which of the upper level concepts of the local thesauri should be classified under the general concepts of the BBT. This should be performed only by the *local thesaurus maintainers*. We propose, that *local thesauri maintainers* should include in their local thesauri general BBT concepts, by using local concepts (declared as “same as”/“exact equivalence” or “narrower of” to the BBT concepts, by their LOD identifiers as these are provided by the [BBT Access Service](#)). This would constitute a one-direction link from the local thesaurus to the BBT.

Additionally we propose a second link to be created, originating from the BBT concept to its connected concept in the local thesauri. This is possible by using the [BBT Submission and Connection Management tool](#), which enables *local thesaurus maintainers* to create this link/connection (LOD identification of the local thesaurus concept which is declared as “same as” or “narrower of” the BBT concept) and stores contact information of the *local thesauri maintainers* in order to keep them updated for changes on the specific BBT concept (e.g. contact e-mail, organization info, etc.).

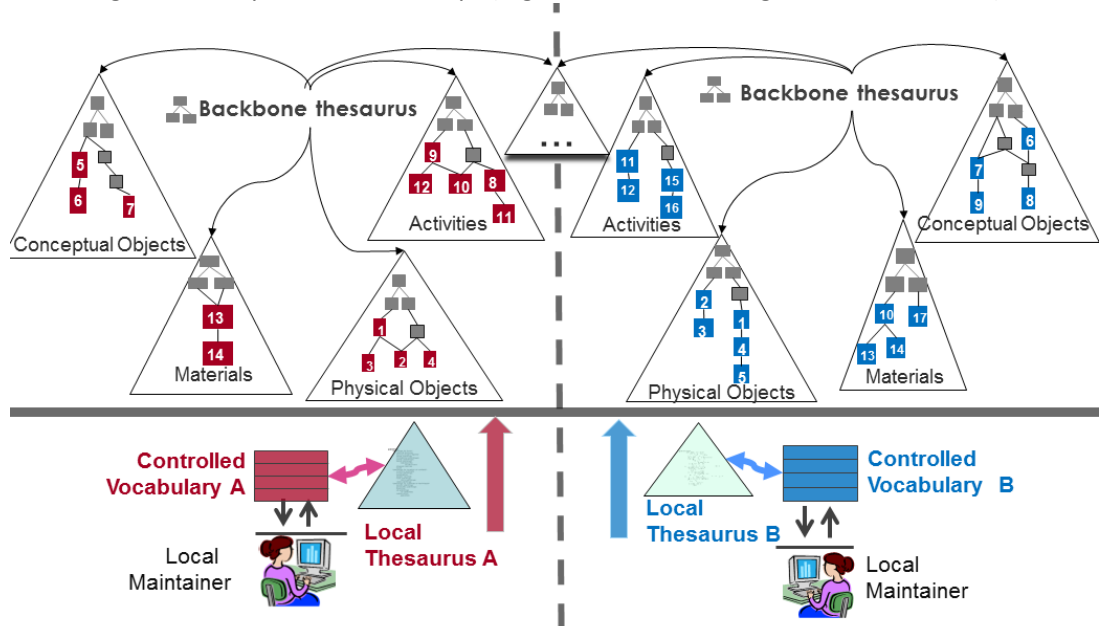


Figure 1: Linking to the BBT

- **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 concept is modified (e.g. its scope note is updated, thus its meaning is altered), all local thesauri developed that are linked/connected to the specific BBT concept as a top-concepts in their thesauri,

should be notified about the change in order to verify if the specific change affects their local thesauri.

- **Unlinking local thesauri from the BBT.** *Local thesauri maintainers* may also decide to un-link their thesauri from BBT. For that, they should remove the local “clone” concepts (declared as “same as” to the BBT concepts (removal of the one-direction link from the local thesaurus to the BBT). Additionally removing the second link is possible by using the BBT Submission and Connection Management tool, which would remove the link/connection originating from the BBT concept to its connected concept in the local thesauri (and maybe would also remove the related *local thesauri maintainers* contact information).
- **General requirements**
 - All thesauri (local and BBT) should use consistent LOD identifiers for referencing concepts and their relations. These identifiers should not change across thesauri versions.
 - All tools should be able to export and import data (thesauri, or parts of thesauri) in SKOS format, under a scheme that is currently under discussion. This scheme should be described in detailed and later be part of this report.

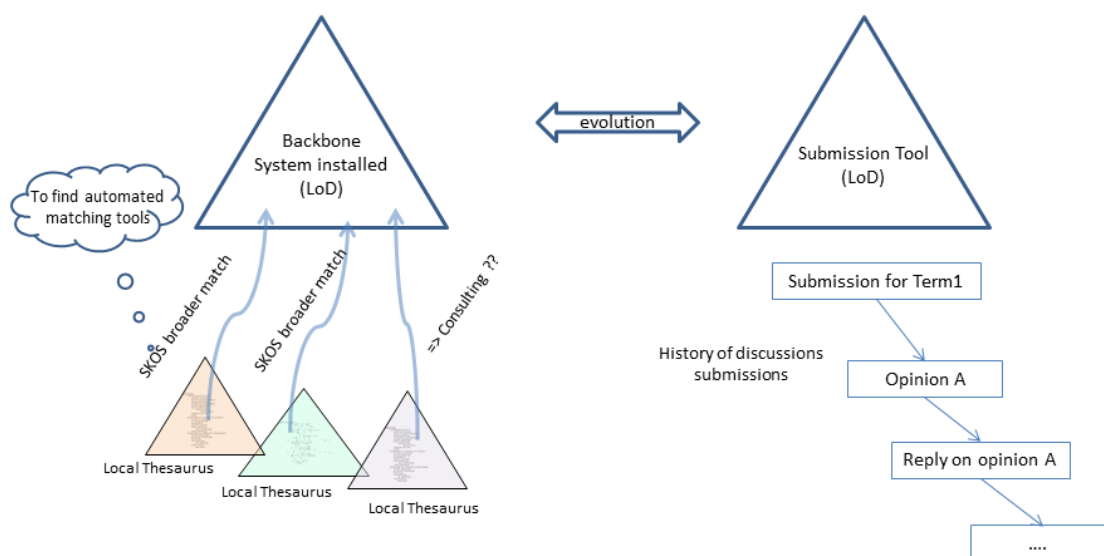


Figure 2: BBT Management Tool and BBT Submission and Connection Management tool

2 Proposed tools

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

2.1 BBT - Thesaurus Access tool

This is the tool responsible for hosting and providing access to the current version the BBT. It includes a service to expose the current official BBT version to the public (BBT Access Service). It should provide consistent identification (LOD identification) for all concepts of the BBT, in order to be referenced by the local thesauri.

Currently the proposed BBT - thesaurus access tool used is OpenSKOS (<http://openskos.org>), developed by ACDH-OEAW (www.oew.ac.at/acdh). It runs as a service for the research community in the context of research infrastructures CLARIN-ERIC and DARIAH-EU. It comprises of: the thesaurus browser (may be used as the proposed BBT Access Service), which

provides an overview of the concepts (and concept schemes) (*see also* <https://clarin.oeaw.ac.at/vocabs/browser/>) and the editor for manual updates (*see also* <https://clarin.oeaw.ac.at/vocabs/editor/login/>).

Once decided, that an official version of the BBT model should 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 – Thesaurus Maintenance tool

This is the tool responsible for thesaurus management (BBT management tool) to maintain the BBT thesaurus database. It should communicate with the BBT Access Service to ensure consistent identification (LOD identification) for all concepts of the BBT, in order to be referenced by the local thesauri.

The proposed BBT - thesaurus maintenance tool, is based on THEMAS system (<http://www.ics.forth.gr/isl/THEMAS>), an open source Web based system for creating, managing and administering multi-faceted multilingual thesauri according to the principles of ISO 25964-1 and ISO 25964-2 standards, developed by FORTH-ICS (www.ics.forth.gr).

This tool will be used by the *BBT-curators* who maintain the BBT model. Once decided, that an official version of the BBT model should 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 should be sent to the BBT – Thesaurus Access and BBT Submission and Connection Management tools to be loaded in order to be publicly accessible and to be able to receive submissions and link/connections.

2.3 BBT – Submission and Connection Management tool

The BBT Submission and Connection Management tool is a communication system, developed by FORTH-ICS (www.ics.forth.gr), that supports discussions regarding the changes proposed for the BBT (changes related to concepts 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* when they want to suggest changes for the BBT (contributors); it provides a form by which they can request modifications/additions/deletions regarding the concepts of the thesaurus. The tool is also used by the *BBT-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 versions of the thesaurus and the history of all the submissions in order to facilitate *BBT-curator's* job. The *BBT-curators* may also forward a submission to users that are experts in specific domains (*thesaurus domain experts*), for further consultation. Finally the tool is used by *thesaurus domain experts* that take part on specific change-related discussions.

As mentioned earlier publishing a new version of the BBT may also affect the local thesauri that are linked to BBT, therefore we propose that *local thesauri maintainers* should be notified. The BBT Submission and Connection Management tool additionally provides functionality that mainly consist of (a) a service for creating (and removing) links originating from BBT concepts to local thesauri concepts (LOD identified), (b) the storage of contact info, (c) along with a notification mechanism.

Finally, the BBT Submission and Connection Management tool should be able to provide BBT version history.

The proposed BBT - Submission and Connection Management tool is based on BBT – Submission tool (version1.0)³, a submission system, developed by FORTH-ICS (www.ics.forth.gr).

2.4 Federated thesauri viewer

Since the different thesauri of the proposed thesauri federation (BBT and local thesauri) are located (as their common concepts are only linked) 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 cached) providing a single interface.

(This document works as a design document, so we welcome any suggestions).

³ BBT – Submission tool (version1.0) is described in more detail in document http://83.212.168.219/DariahCrete/sites/default/files/bbt_submissiontool_v1.0_draft.pdf.