Thesaurus Maintenance WG

Developing a model for sustainable interoperable thesauri maintenance
History

Rome, 17-19 Sep. 2014: Proposal at the General VCC meeting

Berlin, 18/12/2014: Launching of the WG (Kick-off) held in DAINST

Early March: Approval of WG

17/3/2015: Skype meeting

Ljubljana
Work packages

1. Intellectual principles of thesaurus maintenance methodology
2. Social collaboration (Workflows, Decision procedures, Volunteers for running services)
3. Tools
4. Empirical material for testing
Backbone thesaurus: Methodological outlines

Martin Doerr, Helen Katsiadakis, Maria Daskalaki, Helen Goulis
**Designing the backbone thesaurus**

Our main task: to design a backbone thesaurus for the Humanities that is valid irrespective of the specific branch of its application.

The main issue in this process: each field has developed its own system of classification so that it is impossible to have access to all the information related to a user’s query or to map the information from different fields so that the same word will be used to denote the same meaning.

For that reason we pay a lot of attention to build the proper *methodological principles*!
Planning the method

Our goals in designing the method:

- Interdisciplinarity
- Intersubjectivity
- Provision of a consistent, comprehensive, reasonable and logically correct classification system

Misleading criterion: the affinity of the meaning e.g.: Bride (person), Wedding dress (material object)

In order to achieve these goals we appeal to categorical semantics!
Categorical semantics

Categorical semantics provide us with the theoretical background for the detection and the determination of the top-level concepts in a principled way in order to build a **common, coherent and consistent** thesaurus within the framework of DARIAH.

Categorical semantics are based on primary concepts (facets) which could be seen as the preconditions of the way we perceive our reality (Person, Matter, Activity, etc.)

**In this context we search for the intentional properties of the terms.**
Intentional properties

- Characteristics expressing the **nature/substance** of a concept and providing an **unambiguous recognition** of an item as belonging to a category.

- Intentional properties are essential: **necessary** and sufficient conditions for belonging to a category, cannot be replaced without loss of meaning.
  e.g. bachelor is defined as 'unmarried man'. Not being married is an essential property of a bachelor, because one cannot be a bachelor unless he is an unmarried man (necessary condition) and any unmarried man is a bachelor (sufficient condition).

- Recognition must be based on **accessible** information

- When sufficient intentional properties are implicit or not commonly accessible, the term is defined through confining or referring to commonly known phenomena:
  e.g. human being, necessary: DNA, genetical. Accessible: confining morphological characteristics.
Intentional properties

Why is the detection of the intentional properties a crucial task?

- It leads us to the fundamental concepts i.e. to the top-level concepts (facets and hierarchies) that are used in categorical semantics in order to classify terms from every field of application. They are concepts, in which all the other concepts and terms can be analysed. Such concepts are, for example, person, place, activity, conceptual object, time-span, etc.

- It keeps the interrelationship between the top-level concepts and the specific terms of each field alive. The upper-level concepts are the result of the application of a bottom-up method, by virtue of which we can define the intentional properties of the terms. The bottom-up method prevents us from imposing a priori top-level concepts on the experts and makes the links-bonds which relate a term to a concept obvious.
Bottom-up method

- Clearly defined relationships between the top-level concepts and the terms emerging from each specific field guarantees the preservation of the specific meaning of each term.
- The bottom-up method helps us to select and define the top-level concepts which are suitable enough to cover the requirements within each classification system.
- The interactive relationship between the terms and the top-level concepts helps us to find out any shortcomings with respect to the hierarchies and the facets we propose (it could be the case that our proposed hierarchies need to be complemented with new ones in order to cover the requirements of a classification).
The inviolable rules

In this procedure of the bottom-up method the inviolable rules are:

- That we apply the **IsA relationship**, by means of which we can avoid the categorical errors resulting from the subsuming of terms under hierarchies and facets which present different properties than those of the indexed terms.
- That we **never define by negation** (e.g. “the object of perception is all the things that a consciousness is not!”).
“Good” top-level concepts

A top-level concept could be used successfully not only when...

it can be deduced from the intentional properties of a term,

BUT also when...

new properties can be deduced from the already known and accessible: these properties are called potential properties.

potential properties are consequences of the nature of a thing. They may be confined to a category or not. They may appear at some instances at some time.

e.g.: potential properties of the bachelor: no children, is male or female (not a child), live alone etc.
An “open world” classification

If we follow the methodological guidelines mentioned above we end up with a classification system that does not divide the world in closed spheres of meanings according to specific characteristics, but brings to light hidden connections between the terms and establishes concept relationships!
The golden rules of hierarchy building

- Levels of hierarchies are **never absolute**. Even Facets may have **generalizations**.
- Levels of hierarchy are **never complete**.
- Generalizations are **never unique**.
- Sets of sibling concepts **are never complete**. Anything that does not fit goes into the **next level** category, until a better specialization is found.
- Don’t complete levels by “other objects” or “elephants and none-elephants”.
- **Particulars** (gazetteers, person lists) are **NOT terminologies** (but other KOS).

Based on this, we avoid most **arbitrariness** and context dependency. **Collaborative** development of an upper level becomes **feasible**.
Benefits of the faceted classification:

- Reveals the complexity of a term and reduces it to its fundamental components.
- It is not an artificial classification of the terms or a “top to bottom” classification, but is generated from the analysis/decomposition of one term in its elementary characteristics.
- A term can be classified in multiple hierarchies (e.g. doll toys/visual works).
- Is independent of the context, within which each a term appears although the context is crucial for the classification of a term in facets.
- Is based only in a restricted number of fundamental concepts.
- Can be expanded without disrupting or disorganizing existing facets and hierarchies and enables thus compatibility between different classification systems from different domains without imposing terms on the experts.
- It does not presuppose knowledge regarding the exact context of the terms.
Facets and terms: A selection
Facet: Activities

The “Activities” facet comprises types of intentional actions that result in the preservation, creation, production, modification or destruction of an entity (living beings, conceptual/material objects, groups, social, intellectual, physical etc. phenomena).
Activities

Hierarchy

Disciplines: This hierarchy comprises types of branches of professional or potentially professional occupations socially and/or legally acceptable under the criteria of sector self-subsistence, practice efficiency, adoption of common methods and transferability of knowledge and expertise. Each sector includes types of unified activities that express some sort of professional or potentially professional specialization.

Narrower terms:
- a) Construction of material objects and installations
- b) Conception and comprehension of phenomena.
- c) Provision of knowledge and expertise.
- d) Production of works and/or phenomena of aesthetic value

Events: This hierarchy comprises types of intentional activities carried out by at least one actor causing or changing phenomena or states of affairs on the social, political, financial, cultural and intellectual level.

Narrower terms:
- a) Social events
- b) Confrontations, conflicts
- c) Political, social and economic occurrences
- d) Group management
**Activities Hierarchy**

**Intentional destruction:** this hierarchy comprises intentional activities causing the end of existence of an entity or of a valid state of affairs.

**Functions:** This hierarchy comprises types of activities that are structural parts of a relatively stable complex system of permanent and self-contained procedures that repeat themselves within this system and thus contribute to its preservation. Although functions are part of a wider system, each function is completely distinct from the rest. As structural parts of a complex system, functions are types of actions that play a certain role within a system and aim at a specific goal, which they must accomplish. In this respect it is not possible that the purpose which a certain function has to achieve is different from that for which the function is performed. In other words, the purpose of a function is one of its identity criteria. Consequently, the notion of the function univocally relates the actions performed and the target achieved by these actions in such a way that, if some other target is achieved due to external factors, we speak of a different function or activity.
Facet: Activities  
Hierarchy: Disciplines  
Narrower term: Production of works and/or phenomena of aesthetic value  
This term comprises types of branches of professional or potentially professional occupations regarding actions which, while they may involve the construction of material objects or of installations, are primarily designed to attach aesthetic value to them or to foster aesthetic values through the creation of corresponding phenomena.

Performances  
Scope note: In the performing arts, a performance is the artistic enactment of one or more works of art before an audience by a performer or a group of performers. Performances can take place at designated performance spaces (such as a theatre), or in a non-conventional space (on the street, in warehouses or in open-air spaces).

Discipline: Theatre Studies, Contributor: Helen Goulis (Academy of Athens)
Facet: Activities
Hierarchy: Events
Narrower term: Confrontations, Conflicts
This term comprises complex activities (a combination of activities) that presuppose at least two actors or groups of actors, who understand their interests and demands as competitive and thus aim at their satisfaction through their involvement in situations of controversies (coups d’e tat, legal actions, wars, revolutions, strikes etc.)

Wars
Scope note: Large-scale, most often armed, conflicts between two or more parties, nations, or states.

Civil Wars
Scope note: Wars between organized groups within the same nation state or republic, or, less commonly, between two countries created from a formerly united nation state. The aim of one side may be to take control of the country or a region, to achieve independence for a region, or to change government policies.

Wars of national liberation
Scope note: Conflicts fought by nationalities to gain independence. The term is used in conjunction with wars against foreign powers to establish separate sovereign states for the rebelling nationality. From a different point of view, these wars are called insurgencies, rebellions, or wars of independence.

Discipline: History, Contributor: Helen Katsiadakis (Academy of Athens)
Facet: Activities
Hierarchy: Events
Narrower term: Political, social and economic occurrences
This term comprises complex activities (a combination of activities) that presuppose at least two actors or groups of actors, aiming at the emergence or change of political, social and economic conditions.

Migration (human)
Scope note: Human migration is the movement of people from one place to another with the intention of settling temporarily or permanently in the new location. Typically migration refers to movement over long distances and from one region, country or continent to another. It does not refer to movements with no intention to settle in the new place, such as nomadic movements, traveling, pilgrimage etc. Migration applies to movements of individuals, family units or large groups. It covers both voluntary and forced movement owing to political, economic or other causes.

Discipline: History, Contributor: Helen Katsiadakis (Academy of Athens)
Facet: Material Objects

The Material Objects facet comprises things with physical substance that constitute complete units and have a relatively stable form with identifiable boundaries. Such units can be natural or man-made (with regard to origin), simple or complex (with regard to composition) or consist of parts. In this latter case it is possible that the parts are either distinct and independent from the unit of which they are part (e.g. a cave on a mountain) or that they have to be defined with reference to the sum of the parts (e.g. chess-chessmen).
Mobile objects
This hierarchy comprises a) objects originally manufactured as artifacts e.g. figurines, statuettes, quivers etc. b) objects that have been transported as a whole to a museum e.g. altars, columns, fountains, temples etc. c) objects that have been detached from monuments e.g. mosaic inlays, column capitals etc., d) objects that may exist either as movable or as immovable monuments e.g. tombs transported to museums.
**Facet:** Material Objects  
**Hierarchy:** Mobile Objects

**Stelae**
Scope note: Concrete pieces of stone erected usually upright as monuments, bearing inscriptions

**Discipline:** Classical Archaeology, **Contributor:** Gerasimos Chrysovitsanos (Academy of Athens)
Facet: Conceptual Objects

This facet comprises objects whose essence remains the same regardless of the carrier. They are products of the human activity supported by the use of technical or electronic devices (digital photos, geometric measurements etc.) or without it (concepts, thoughts). The fact that they are materially produced does not determine their identity. Conceptual objects have the ability to exist on more than one particular carrier at the same time (paper, electronic signals, photos, human memories etc.), without the latter changing or altering their identity. On the contrary, any alteration of the conceptual object itself (removal of a part, revision etc.) changes the definition of its identity.

Conceptual objects exist as long as they can be found on at least one carrier (human memory included). Their existence ends when the last carrier and the last memory are lost.
Symbolic Objects
This term comprises identifiable symbols and/or any aggregation of symbols, that have an objectively recognisable structure and that are documented as single units (sets or arrays of signs). Symbolic objects may serve to designate something, or to communicate some propositional content, but they don’t depend from what they designate or communicate.
They can exist on one or more carriers simultaneously without this feature adding to or removing from the identity of the symbols. Being objectively recognizable and documented as single units, symbolic objects are independent from the material carrier and the symbolized content as well. Consequently Symbolic Objects may or may not have a specific meaning.

Narrower term: Information Object
Propositional Objects
This term comprises mental contents produced by human activity, representing data of physical, logical or psychological realities or even fiction. These objects, though expressed by the use of symbols, do not depend on them, matter of fact that allows their transcription to another set of symbols. However a propositional object sometimes can not be differentiated distinctively from the used symbols (e.g. a poem).

Narrower term: Information Object
This term comprises all products of human mental activity having an objectively recognizable structure and being documented as single semantic units. Information Objects give us information on the content of a Conceptual Object. The content and the symbols used are linked in such a way that any attempt of transcription to another set of symbols normally leads to alteration of the information. Being a semantic unit that connects the information and the symbol, every Information Object assumes in advance the Symbolic one, but not vice versa. That’s why the Informational Object is a subdivision not only of the Propositional Object but of the Symbolic as well. It allows us to draw conclusions about the socio-historic and cultural framework of the Object, its creator and the way it was created. For e.g. the various forms of primary reference data (archives etc.) used in scientific research.
Facet: Conceptual Objects
Hierarchy: Propositional Objects
Narrower term: Information Objects

Decrees
Scope note: Texts drawn up and balloted by the institutional bodies of the Greek city-states (boulē, the assembly [ekklēsia]) through varying procedures in the different city-states

Honorific decrees
Scope note: Decrees attributing honour (materially or morally) to an individual for the services he provided to the city-state (polis)

Related Terms: Epigraphy, Inscriptions, Stelae

Discipline: Classical Archaeology, Contributor: Gerasimos Chrysovitsanos (Academy of Athens)
Identity
Scope note: The distinct features, whose combined conceptualisation and expression substantiate persons and/or groups

Social Identity
Scope note: An individual’s self-concept based on specific biological, social and cultural parameters (ethnicity, religion, language, social class, gender etc.), as derived from his or her interaction with the broader social environment.

Gender Identity
Scope note: Gender identity is an aspect of the social identity referring to stereotypes about the characteristics and the roles of its bearers, according to the dominant social and cultural norms regarding their biological sex.
TheBaH SUBMISSION TOOL

USER MANUAL

Report
Version 1.0
Status: DRAFT

Institute of Computer Science
Foundation for Research and Technology – Hellas
Academy of Athens

Martin Doerr, Maria Daskalaki, Helen Katsiadakis, Helen Goulis

DARIAH-GR
Digital Research Infrastructure for the Arts and Humanities

δνάς
Greek Research Infrastructure Network for the Humanities
# Table of Contents

1. Introduction ................................................................................................................. 4  
2. The Contributor Role ................................................................................................. 5  
3. Login ............................................................................................................................ 5  
4. Description of the User Interface ................................................................................. 6  
5. Functionalities ............................................................................................................... 7  
   5.1 Submission Functionalities ..................................................................................... 7  
      5.1.1 Create Submission for Classes ...................................................................... 7  
      5.1.2 Create Submission for Relations ................................................................. 13  
      5.1.3 List Submissions ............................................................................................. 16  
      5.1.4 View a submission ........................................................................................ 17  
      5.1.5 Add a new note to the Submission ............................................................... 18  
      5.1.6 Search Submission ........................................................................................ 19  
   5.2 Class / Relation Functionalities ............................................................................. 21  
      5.2.1 Search a Class ............................................................................................... 21  
      5.2.2 Search Relation ............................................................................................. 23  
   5.3 Profile Functionalities ............................................................................................. 25  
      5.4 Change Profile information ................................................................................ 25  
   5.5 Ontology Viewer ...................................................................................................... 26  
   5.6 Logout ..................................................................................................................... 26  
6. List of Submission Statuses ....................................................................................... 27  
   6.1 Submitted .............................................................................................................. 27  
   6.2 Under discussion, wait for reply ........................................................................... 27  
   6.3 Under discussion, replied ..................................................................................... 27  
   6.4 Implementation ...................................................................................................... 27  
   6.5 Wait for release .................................................................................................... 27  
   6.6 Released ............................................................................................................... 27  
   6.7 Postponed ............................................................................................................. 27  
   6.8 Rejected ................................................................................................................ 27
Introduction

It is now widely acknowledged that ontologies can make a significant contribution to the design and implementation of information systems in all the different science fields. The use of ontologies in the Humanities is mainly focused on the representation and (re-) organisation of humanities terminologies. Information systems need to be able to communicate complex and detailed concepts unambiguously.

Due to the complexity of the Ontologies, there is a need for feedback from the end Users. Furthermore, the Humanities Ontologies are difficult to define owing to the varying needs of researches across discipline and cultural boundaries. So, there is a growing need to track variations in the Ontology, to view the history of the changes as well as retrieve the dialogues that lead to the changes.

The TheBaH Submission system is designed to answer these needs. It is a communication system for the end users to review and submit their own opinion on existing Humanities Master Ontologies. Moreover it stores the history of master ontology and requests on changes. In this way it assists ontology experts in their work of building and maintaining Ontologies, by reviewing change submissions, along with retrieving the version history of the Ontology.

What is more, it notifies all interested parties when a new version of the Ontology is released.
Description of the User Interface

After logging into the System, the user is directed to the welcome page of the Submission Tool. The page is divided into 4 sections. The first section is the upper menu. This is the main menu, where the user can choose between the “Submission” functionalities, “Classes/ Relations” functionalities, “Profile” functionalities, view the “Ontology Viewer” and logout of the System. The second section is the left menu. Depending on what the user has chosen in the main menu, the left menu offers the user different options. These options are explained in detail in the Functions paragraph.

The third section is the results area. It is located in the center of the page and it shows the results of the actions in the left menu.

The fourth section is the footer section. In the footer the user can see information about the user name that is logged in and information about the system copyright restrictions.
TheBaH Master Ontology Submission Tool
Functionalities

Submission Functionalities
Create Submission for Classes

In order to create a new Submission for Classes, the user has to choose “Submissions” from the top menu and “Create new Class submission” from the left menu.

In the center of the page there appears the form of a new Class Submission as shown in the next slide:
Create Class submission

The user has to insert the action of the change request that he/she wants to add. The available submissions actions are change a Class, delete a Class, merge Classes, split classes, create new Class and rename a Class.
Create new class submission

TheBaH Master Ontology Submission Tool

User: Contributor
Create new class submission

When the user chooses to create a new Class submission he/she has to provide the following information for the submission:

**Parent Node**: This field provides the information on where the new Class should be inserted into the tree. It is an auto-fill field, which means that the user is provided with the available classes list once he types the first letters of the chosen class.

**Suggested Name**: The new class name.

**Explanation**: In this field the user gives an explanation why he/she believes that the new Class is necessary to be added into the TheBaH Master Ontology.

**Version**: Automatically inserted from the system, depending on which TheBaH Master Ontology version is inserted into the system.

**Date**: Automatically inserted from the system. It is the date and the time that the submission is added.

**Submitter**: automatically inserted from the system, it is the user who is logged into the system.
Create Submission for Relations

In order to create a new Submission for Relations, the user has to choose “Submissions” from the top menu and “Create new Relation submission” from the left menu.

In the center of the page there appears the form of a new Relation Submission, as shown in the figure below:

The user has to insert the action of the change request that he/she wants to add. The available submissions actions are create new Relation and delete a Relation.
Submission for adding a new Relation

TheBaH Master Ontology Submission Tool

[Image of the submission interface with fields for Action, Domain Name, Relation Name, Range Name, Is the Domain Name a new Class?, Is the Range Name a new Class?, Add an explanation, Version, Date, Submitter.]

User: Contributor
New Relation Submission

When the user chooses to create a new Relation submission he/she has to provide the following information for the submission:

**Domain Name:** This field provides the information of the domain name of the relation. It is an auto-fill field, which means that the user is provided with the available classes list once he types the first letters of the chosen class.

**Relation Name:** The new Relation name.

**Range Name:** This field provides the information of the range name of the relation. It is an auto-fill field, which means that the user is provided with the available classes list once he types the first letters of the chosen class.

**Is the Domain name a new class?** The user can either select to insert as the domain of the relation a new class or an existing class. Depending of what he/she has selected in the field “Domain Name”, he/she should answer YES or NO accordingly.

**Is the Range name a new class?** The user can either select to insert as the range of the relation a new class or an existing class. Depending of what he/she has selected in the field “Range Name”, he/she should answer YES or NO accordingly.

**Explanation:** In this field the user gives an explanation why he/she believes that the Relation should be added in the TheBaH Master Ontology.

**Version:** Automatically inserted from the system, depending on which TheBaH Master Ontology version is inserted into the system.

**Date:** Automatically inserted from the system. It is the date and the time that the submission is added.

**Submitter:** automatically inserted from the system, It is the user who is logged into the system.
List of Submission Statuses

Submitted: This is the first status of a change request. Once the End User sends a submission it takes the status “Submitted”. This status shows that the submission has not been checked from the ontology expert. 

Wait for reply: After the submission the ontology expert checks the submission, he might need some explanations or even more information about the submission. If that is the case, then he sends a submission back to the submitter or the domain expert, and the submission is taking the status “wait for reply”.

Replied: The submitter or the domain expert reviews the received submissions and replies giving explanations or more information about the submission. The submission gets the status “replied”.

Implementation: When a submission has the status “Implementation”, means that the change request is in the phase of implementation at the TheBaH Master ontology.

Wait for release: After the Implementation phase and when the submission is ready to be released it takes the status “Wait for release”. During this phase, changes can still occur to the submission until it comes to it final state.

Released: The submission is released in the new version and all the interested parties have been informed about the final status of the request for change.

Postponed: The request for change will be reviewed later in time.

Rejected: The request for change is considered as not implementable or implementable in the future, and all the parties are informed.
Sequence Diagram of the submission statuses
Counterexamples

1. Example: TaDiRAHA: http://tdirah.dariah.eu/vocab/

**Broader term:** Research Activities

**Narrower terms:** capture (conversion, data recognition, discovering, gathering, imaging, recording, transcription) creation, enrichment, analysis, interpretation, storage, dissemination, meta-activities.

Scope Note: Research Activities “are usually applied to one or several research objects. An article about modelling of manuscript properties would therefore be tagged with the tags “Modelling” and “Manuscript”. A plain text editor would be tagged with the tags "Writing" and "Code" and "Text””.

---

**DARIAH-G**

Digital Research Infrastructure for the Arts and Humanities
Counterexamples

But:
The definition of activities is based on another other concept, that of “research objects”.
- The definition makes no reference whatsoever about the intensional properties of the term.
- The hierarchical division of the broader term does not demonstrate a necessary connection between the narrower (capture, data recognition, discovering, gathering imaging recording clean up) and broader terms.
- Narrower terms are not actually narrower since they are not specializations of the term “research activities” but they also appear in other contexts.
**Counterexamples**

2. Example ([www.openfolklore.org/et/tree.htm](http://www.openfolklore.org/et/tree.htm)). Context-oriented classification system:

**Top of the hierarchy:** Dance

**Narrower term:** Dancer.

But:
Dance is an activity while dancer is a Person. The only relationship between them is external, on the basis of the context and not of the essential characteristics of the concepts.
Counterexamples

3. Example: (www.openfolklore.org/et/tree.htm).

Top of the hierarchy: Health
Scope note: “Knowledge, beliefs, and practices that concern conditions of the body and the mind, including diagnosis and prognosis, treatment, and care”.

Narrower term: Death

But:
death is a negation of health!
Health and death are not “knowledge and beliefs” but situations/state of affairs.
Counterexamples

4. TaDiRAHA: http://tadirah.dariah.eu/vocab/

Broader term: Digital Humanities Research Objects
Narrower terms: Persons (but persons are not objects!!)
Research (but research is an activity and not an object!!)
Interaction (but interaction is also an activity!!)
Methods (is a process and not an object!!)

5. TaDiRAHA: http://tadirah.dariah.eu/vocab/
Broader Term: Technique > related Goal-Method
Narrower terms: Brainstorming (in an activity and not a method!)
The Art And Architecture Thesaurus (AAT Example)

Note: The Functions hierarchy contains descriptors for activities that are conducted in order to accomplish specific purposes, as well as methodologies associated with specific areas of endeavor. It includes descriptors for activities relating to the manipulation of data, the collecting of objects, human communication, economics, business, law, and government, as well as other professional activities. Relation to Other Hierarchies: Descriptors describing branches of learning, (e.g., history) and areas of specialization, (e.g., law), are found in the Disciplines hierarchy. Descriptors for operations and processes performed on or with objects and materials (e.g., polishing), are found in the Processes and Techniques hierarchy. Descriptors denoting occasions and happenings of a social, cultural, religious, or personal nature, (e.g., exhibitions), are found in the Events hierarchy.

But:
this Note enumerates some cases (particulars) we are dealing with functions but does not say what are the functions, which are their essential properties!!!