This project is funded under the
ICT Policy Support Programme part of the
Competitiveness and Innovation Framework Programme.

Project

Project Acronym:	AthenaPlus
Grant Agreement number:	325098
Project Title:	Access to cultural heritage networks for Europeana

Deliverable

Deliverable name:	**First Release Terminology Management Platform (TMP)**
Deliverable number:	**D 4.3**
Delivery date:	31-01-2014
Dissemination level:	Public
Status	Final
Authors (organisation)	Christophe Roche (UniSav), Luc Damas (UniSav)
Contributors (organisation)	Marie-Véronique Leroi (MCC), Florent André (UniSav)
Reviewers (organisation)	Erik Buelinckx (Kikirpa), Nikos Simou (NTUA), Sanja Halling (RA), Eva Coudyzer (KMKG)

Revision History

Revision	Date	Author	Organisation	Description
V0.1	2013-06-15	C. Roche L. Damas	UniSav UniSav	First Draft of requirements and specifications
V0.2	2013-09-01	F. André	UniSav	First release of the TMP
V0.3	2013-10-01	M.V. Leroi C. Roche	MCC UniSav	Specification updating
V0.4	2013-11-01	F. André L. Damas	UniSav MCC	Second release of the TMP
V0.5	2013-12-01	C. Roche M.V. Leroi	UniSav MCC	Updating specifications
V0.6	2014-01-01	L. Damas C. Roche M.V. Leroi F. André	UniSav UniSav MCC UniSav	Final integration and correction after peer reviewers comments
V1	2014-01-12	E. Coudyzer	KMKG	Revision and finalisation of the Deliverable
V1.1	2014-01-31	M.T.Natale	ICCU	Formal Check

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation

or both.

Project Coordinator: Istituto centrale per il catalogo unico delle biblioteche italiane
Address : Viale Castro Pretorio 105 – 00185 Roma
Phone number : +3906 06 49210 425
E-mail : info@athenaplus.eu
Project WEB site address : http://www.athenaplus.eu

Table of Contents

1 EXECUTIVE SUMMARY 2
2 INTRODUCTION 3
2.1 Background 3
2.2 Role of this Deliverable in the Project 3
3 REQUIREMENTS 4
3.1 Theoretical requirements 4
3.2 Implementation Requirements 19
3.3 Interface Requirements 21
3.4 Main Functionalities 22
3.5 Use Cases 24
4 SPECIFICATIONS 24
4.1 Logo 24
4.2 Starting with the TMP 24
4.3 Connection 25
4.4 Login 28
4.5 User management 29
4.6 Main page 31
4.7 Create or Import a Terminology 31
4.8 Skosification of the Terminology 36
4.9 Owner Management 36
4.10 Enrichment and Editing 38
4.11 Search and Navigate 44
4.12 Mapping 46
4.13 Publication 51
4.14 Export 51
4.15 Static Pages 51
5 NEW FEATURES IN ATHENAPLUS 53
6 CONCLUSION 54
7 APPENDIX 1: REFERENCES 56
8 APPENDIX 2: DEFINITION OF TERMS AND ABBREVIATIONS 60

1 EXECUTIVE SUMMARY

This document describes the activity done in the framework of Task 4.2 *Terminology management platform (TMP)* in work package 4 on terminologies and semantic enrichment. One of the objectives of WP4 is to provide a web service for online sharing of terminologies and semantic mappings between concepts of different terminologies in order to publish the terminologies as part of the Linked Data Web. It describes the theoretical basis and the technical requirements of the TMP.

2 INTRODUCTION

The deliverable contains functional and technical specifications for managing multilingual terminologies in the TMP. It includes a first production version release of the Terminology Management Platform, which is made available to the consortium as an open source web service.

The deliverable is two-fold. The first part describes the requirements of the TMP. The theoretical requirements are based on an ontology-oriented approach in order to clearly separate and link the conceptual and linguistic dimensions of terminology. The implementation requirements rely on a semantic web interchange format so terminologies can be shared across applications.

The second part presents TMP specifications through the description of use cases.

2.1 Background

The described functional and technical specifications rely on a prototype of the TMP developed in the Linked Heritage project (see the D3.2 *Functional and technical specifications of the terminology platform*[1]).

D4.3 takes into account returns on experience of TMP use, from a theoretical point of view as well as from an empirical and practical point of view[2].

2.2 Role of this Deliverable in the Project

The TMP is an important feature in the project scope of AthenaPlus. Using the TMP will:

- Allow a semantic mapping of terminologies with SKOS to achieve interoperability on the web, and thus greater visibility and valorisation of the partner collections
- Allow the reuse of the mapped terminologies in applications, linking with other relevant sources etc.
- Allow the creation of micro-thesauri for implementation in the MINT-tool, which will enhance multilingual search functionalities in the online catalogue Europeana

The TMP is a key element in the completion of Task 4.3 in WP4 on semantic enrichment of cultural metadata with Linked Open Data (see also D4.2). The mapped terminologies can also be of importance in WP5 *Creative applications for the re-use of cultural resources* and WP6 *Pilots for testing the creative use of cultural contents*.

The last years we have seen an increase in the development of tools for thesaurus management (OpenSKOS, Poolparty etc.). The TMP can be differentiated from these tools because it is an open source tool and easy accessible to non-professional users. The tool is also widely supported by 40 partners in the AthenaPlus project. The TMP registry will contain many multilingual terminologies from a variety of institutions and countries, which will increase the amount of possible resources when reusing the mapped terminologies.

[1] http://www.linkedheritage.com/index.php?en/142/documents-and-deliverables
[2] E.g. Seminar on multilingualism in Paris, April 18 2013; Workshop TMP, organized by PACKED, September 5 2013

3 Requirements

3.1 Theoretical requirements

3.1.1 Standards and Definitions

An important aim of the TMP is the collection of multilingual terminologies for content management systems. The TMP wants to unify functions of *terminologies* and *thesauri,* taking into account the latest versions of the ISO Standards describing terminologies and thesauri:

3.1.1.1 Terminology:

The standards on Terminology are under the responsibility of the Technical Committee 37 (TC 37) of the International Organization for Standardization (ISO). The scope of the ISO/TC37 is standardization of principles, methods and applications relating to terminology and other language and content resources in the contexts of multilingual communication and cultural diversity. It is a horizontal committee which provides guidelines for all other Technical Committees managing their own terminological problems.

The two main standards on terminology principles and methods are:

- *ISO 1087-1: Terminology work — Vocabulary — Part 1*: Theory and application. The main purpose of this international terminology standard is to provide a systemic description of the concepts in the field of terminology and to clarify the use of the terms in this field. All the ISO Standards rely on the definitions of the 1087-1.

- *ISO 704: Terminology work — Principles and methods*. This international standard establishes the basic principles and methods for preparing and compiling terminologies both inside and outside the framework of standardization, and describes the links between objects, concepts, and their terminological representations. It also establishes general principles governing the formation of terms and appellations and the formulation of definitions

3.1.1.2 Thesaurus:

- *ISO 25964-1* under the responsibility of the ISO Technical Committee 46 about Information and Documentation.

- *ISO 25964-1: Information and documentation — Thesauri and interoperability with other vocabularies — Part 1*: *Thesauri for information retrieval*. This part of ISO 25964 gives recommendations for the development and maintenance of thesauri intended for information retrieval applications. It applies to vocabularies used for retrieving information about all types of information resources, irrespective of the media used (text, sound, still or moving image, physical object, or multimedia) including knowledge bases and portals, bibliographic databases, text, museum or multimedia collections, and the items within them.

Because the TMP relies on the principles of *terminology*, the definitions used in this deliverable come from the ISO 1087-1 about terminology work:

Terminology	Definition
Designation / Designator	Representation of a concept by a sign which denotes it.
Term	verbal designation of a general concept in a specific subject field.
Concept	Unit of knowledge created by a unique combination of characteristics *Note*: Concepts are not necessarily bound to a particular

	language. They are, however, influenced by the social or cultural background which often leads to different categorizations.
Concept system	System of concepts: set of concepts structured according to the relations among them
Hierarchical relation	Relation between two concepts which may be either a generic relation or a partitive relation
Generic relation	Genus-species relation: relation between two concepts where the intension of one of the concepts includes that of the other concept and at least one additional delimiting characteristic
Partitive relation	Part-whole relation: relation between two concepts where one of the concepts constitutes the whole and the other concept a part of that whole
Associative relation	Pragmatic relation - relation between two concepts having a non-hierarchical thematic connection by virtue of experience

The ISO norms give guidelines on how to define and use *terminologies*, *thesauri* and *ontologies.* Another standard on which the TMP is built, is the W3C-standard SKOS. SKOS is an exchange format and not a language modelling format for thesauri, terminologies or ontologies, so the use of SKOS can result in a loss of information when "translating" e.g. a thesaurus in SKOS. However, it has the advantage of concentrating fully on the notion of the "concept", a principle which has been kept in the TMP.

3.1.1.3 SKOS[3]

The Simple Knowledge Organization System (SKOS) is a common data model standard for sharing and linking knowledge organization systems via the Semantic Web. It was developed by the world wide web consortium to support the use of knowledge organization systems, such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.
SKOS is an application of RDF (Resource Description Framework).

SKOS allows to create concepts and publish them on the web. These concepts can be linked with other sources on the web and can be automatically integrated in other concept schemes. A concept is identified in SKOS as a URI (Uniform Resource Identifier) and can contain following labels and properties, which are a reflection of the properties and relations in knowledge organization systems.

The table gives an overview of the most important SKOS types, class or properties, RDF tags and (ISO) definition and the specifications:

SKOS Type	Class/Property	Specifications	RDF tag	(ISO) Definition
Resource type	Class: Concept	• Fundamental element of SKOS [i.e. *units of thought* - ideas, meanings, or (categories of) objects and events] • Abstract entities independant from the terms they denote. • A skos:concept is defined by a *URI + rdf:type (rdf:resource)* to specifiy that the used resource is represented by a URI of the type skos:Concept. • Problematic: qualifiers in labels, e.g. technique (painting)	Rdf:about	

[3] Simple Knowledge Organization System, W3C, http://www.w3.org/2009/08/skos-reference/skos.html

Resource type	Class: ConceptScheme	• Description of the scheme itself • Expressed by a URI • Describes the meta-properties of the scheme, e.g. Dublin Core can be applied	Rdf:about	Title of the terminology, author, description etc.
Resource type	Property: inScheme	• A concept does not have to be a part of any scheme. It may also be a part of more than one scheme. To express that a concept is a part of a scheme, use the skos:inScheme property	Rdf:resource	
Resource type	Property: hasTopConcept	• Creates an efficient access to the broader/narrower conceptual hierarchies • Allows linking a specific concept scheme with the concepts it contains.	rdf:resource	
Link type	Property: prefLabel	• 1 preferred label per language • 2 identical preferred labels possible in different languages, but not advised when they belong to the same concept scheme • preferably one word (no compound words, adjectives etc.) • numbers must be expressed with skos:notations	rdfs:label	ISO preferred term
Link type	Property: altLabel	• more alternative labels possible per language • preferably one word (no compound words, adjectives etc.)	rdfs:label	ISO alternative term: synonyms, quasi-synonyms, acronyms, abbreviations, plurals, singulars etc.
Link type	Property: hiddenLabel	• more hidden labels possible per language • preferably one word (no compound words, adjectives etc.)	rdfs:label	Hidden label: concept not promoted, but included to increase correct search results, e.g. misspellings, concept id's of different terminologies. ISO USE and USE FOR
Link type	Property: broader term	• Links concepts semantically • subject is skos:Concept • Links concepts between one concept scheme (≠ concept schemes: use mapping properties) • concepts can have more than one broader term (polyhierarchy) • inverse of skos:narrower • non-transitive	rdf:resource	ISO Broader term
Link type	Property: narrower term	• Links concepts semantically • subject is skos:Concept • Links between two concepts in one concet scheme (≠ concept schemes: use mapping properties) • A concept can have more than one narrower term • Is inverse of skos:broader • Non-transitive	rdf:resource	ISO Narrower term

Link type	Property: related term	• Links concepts semantically • Links concepts which are associated • subject is skos:Concept • Links between two concepts in one concept scheme (≠ concept schemes: use mapping properties) • A concept can have more than one related term • Is symmetric • Non-transitive	rdf:resource	ISO Related term
Link type	Property: transitive	• Skos:broader and skos:narrower are used to make a direct hierarchical link between two concepts. • Skos:broaderTransitive and skos:narrowerTransitive are not used to make logical links between concepts • Is inverse • Skos:broaderTransitive and narrowerTransitive are transitive: they make links between two distant levels in the hierarchy	rdf:resource	Narrower transitive / Broader transitive
Link type	Property: scopeNote	• Documentary note about the concept • Gives information on: (conceptual) meaning and use • Linked to skos:Note • Multilingual		ISO Scope note
Link type	Property: definition	• documentary note about the concept • gives a full definition of the note • linked to skos:Note • multilingual		Concept definition
Link type	Property: example	• documentary note about the concept • gives an example of the concept use • Linked to skos:Note • meertalig		Concept example
Link type	Property: history note	• describes the evolution of the concept • linked to skos:Note • multilingual		Evolution of the concept in the terminology
Link type	Property: editorial note	• note for the terminology users • remarks on the management of the concepts or terminology • linked to skos:Note • multilingual		Editorial remarks
Link type	Property: change note	• note for the terminology users • documentary note about the concept • remarks on adaptations of the concept etc. • linked to skos:Note • multilingual		Keeping track of concept adaptations

The SKOS mapping properties are listed in the table below:

Mapping possibilities	**Concept expressed as:**	**SKOS definition**
skos:exactMatch	URI	• Equivalent or exact denotation between two concepts • Is transitive
skos:broadMatch	URI	• skos:broadMatch is a sub-property of skos:broader: every skos:broadMatch between concepts leads to skos:broader between these concepts

skos:narrowMatch	URI	• skos: narrowMatch is a sub-property of skos:narrowe: every skos:narrowMatch between concepts leads to skos:narrower between these concepts.
skos:closeMatch	URI	• Links two concepts that are sufficiently similar that they can be used interchangeably in some information retrieval applications. • Is non-transitive
skos:relatedMatch	URI	• Is used to state an associative mapping link between two concepts.

The ISO 25964-1, published in 2011, has been adapted to SKOS users. They have developed a set of linkages between the elements of the ISO 25964 data model and the ones from (among others) SKOS.

The TMP will use the SKOS format for representing and linking the terminologies (see also infra).

3.1.2 An Ontology-Oriented Approach

3.1.2.1 A double dimension: conceptual and linguistic

It is important to bear in mind that a *terminology* is not a *thesaurus*, because the main goal of a terminology is not to index documents. Neither is a *thesaurus* a *terminology*, because the main goal of a terminology is to define terms in relation to the domain conceptualisation.

In the same way, a *term* must not be confused with a *concept name*. Terms belong to a linguistic system, where concept names or *identifiers* belong to a conceptual system.

Both the latest versions of ISO standards on terminology[4] and thesauri emphasize that concepts and terms must be separated and that a priority should be given to concepts, because they are supposed to be *linguistically independent*.

Terminologies and thesauri - in a concept-oriented approach – have two distinct functions:

- A terminology contains the terms of the domain in a given language;
- A thesaurus specifies their meanings, i.e. the concepts they designate.

Furthermore, there is a difference between *a term* and *a concept* : a concept is by definition *extra linguistic*. Also, linguistic relationships between terms, like *synonymy* and *hyponymy*, must not be confused with the relationships between concepts like “is a kind of” and “ is part of”.

In fact, a terminology is a combination of several non-matching networks:

- a *concept network* which represents the shared and common conceptualization of a domain. This conceptualisation is linguistic independent[5];
- as many *linguistic networks* as there are languages. Each of them is linked to the conceptual network because each term denotes a concept.

The example below shows two vocabularies sharing the same conceptual model of the same *building* classification, one in English and one in French:

[4] Since the ISO/TC37 meeting in Madrid in 2012, the ISO 1087-1 and ISO 704 are under the process of revision in order to take into account new approaches coming from knowledge engineering and in particular ontology. C. Roche of University of Savoie is the Project Leader of the IOS 1087-1 Standard about Terminology of Terminology.
[5] The conceptualisation of a domain is assumed to be linguistic independent, but not necessarily social or cultural independent.

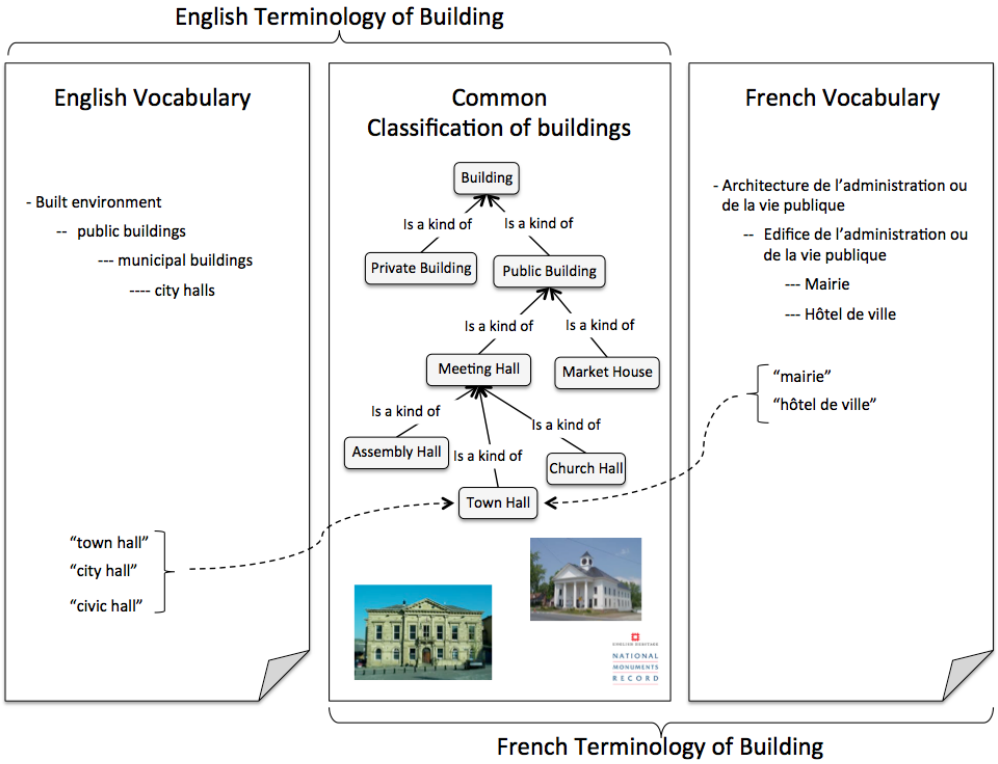


Figure 1: Structure of a terminology

3.1.2.2 An Ontology

From the semantic point of view, the quality of a terminology or a thesaurus directly depends on the quality of its *conceptual model*. This is the reason why the production version of the TMP should represent a conceptual system of a terminology as an *ontology*.

An ontology is, according to the famous Gruber's[6] definition, "a specification of a conceptualisation", *i.e.* a shared description of concepts of a domain including their relationships expressed in a formal and computer readable language.

In other words, an ontology is a system of concepts:

- linked by relationships: a kind of, a part of, an associative relation etc.
- defined and described by characteristics either essential or descriptive.

The formal language used for concept definition, e.g. description logic, allows to guarantee "good" properties in the logical sense. Definitions are objective, coherent, precise, consensual, reusable, sharable and readable.

The two following figures are examples of ontologies. The first one is a domain ontology where concepts are defined by specific differences, e.g. a <Watch> is a 'portable' <Timepiece>. A 'portable' is a specific difference.

[6] Tom Gruber, 1993

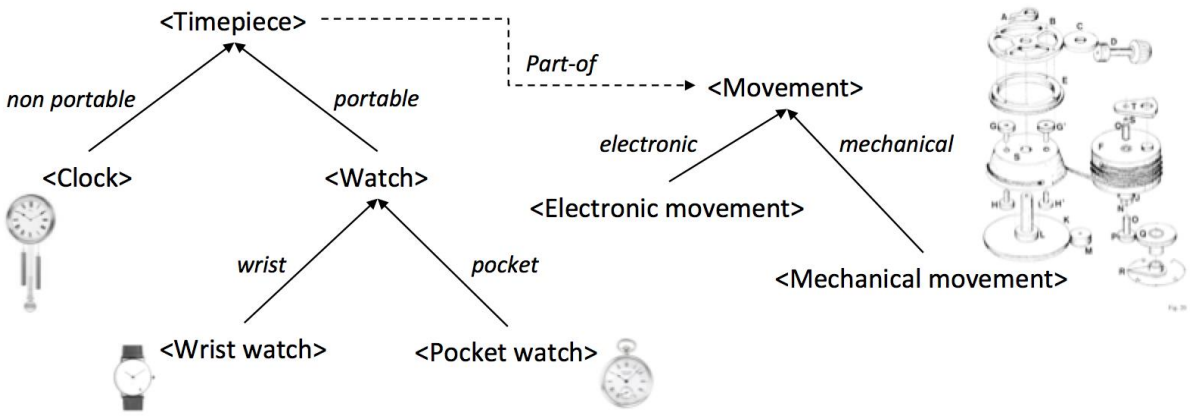


Figure 2: The TimePiece Ontology defined by a specific difference

The second one is a top ontology whose aim is to describe everything. The representation language is a frame-based language where every concept (also called classes) is defined by a set of attributes and a structure following an inheriting sub-class relationship.

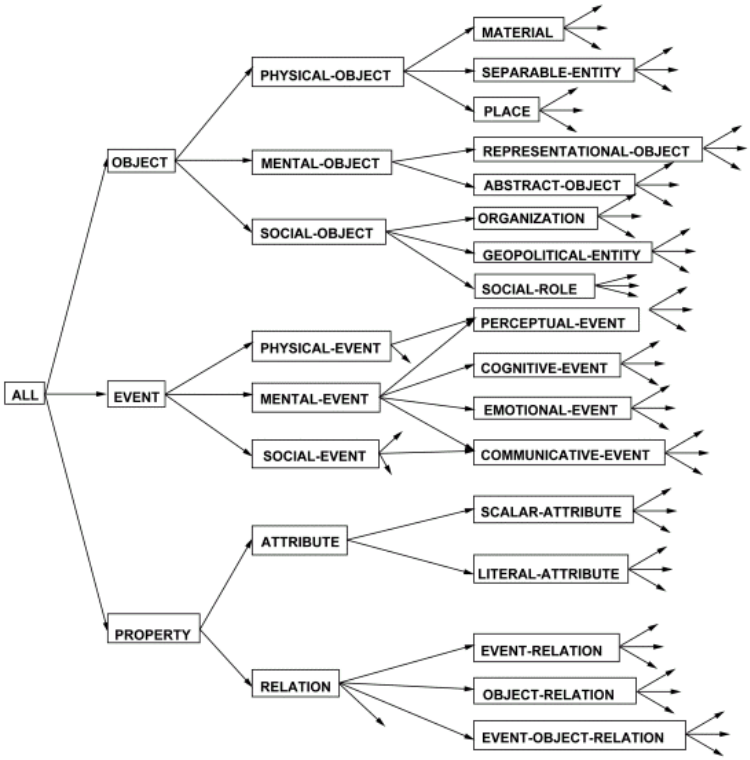


Figure 3: The Mikrokosmos ontology

Combining ontology and terminology leads to the *ontoterminological* approach. This approach is based on:

- a clear separation between the conceptual dimension – which is supposed to be common in whatever the language - and the different linguistic dimensions – one per language;
- an ontology-oriented approach for the conceptual model.

The consistency of the terminology, and therefore its real usefulness and sustainability, is guaranteed by the ontology.

The extra-linguistic representation of the conceptual system allows to define and link multilingual terminologies as illustrated by the following figure.

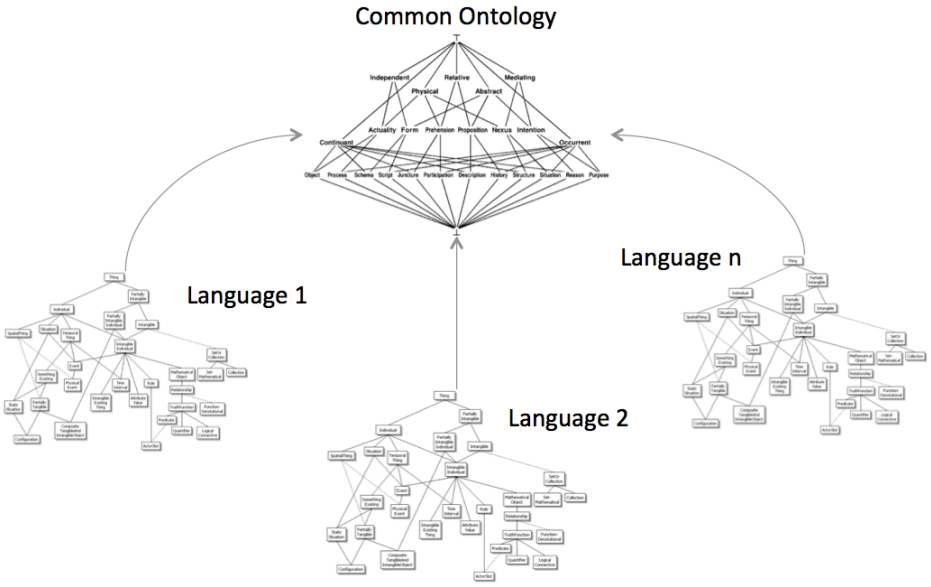


Figure 4: The Ontoterminological Approach

3.1.2.3 Methodological guidelines

If the main goal of the TMP is to define multilingual terminologies, the TMP must also provide guidelines for building terminologies, i.e. provide good principles for building the conceptual system.

Concept modelling means building a conceptual system where concepts are linked with three kinds of relationships:

- the “kind of” (or “is a”) relationship between concepts whose meaning is to express the fact that a concept (e.g. <Town Hall>) is more specific than other one (e.g. <Meeting Hall>);
- the “part of” relationship whose meaning is to express that an object is made up of different parts;
- associative relationships, i.e. a relation between two concepts having a non-hierarchical thematic connection by virtue of experience.

These types of relationships must not be confused, especially the two first ones, since they do not express the same kind of knowledge. As a matter of fact, the “kind of” relationship links concepts of the same nature (one is more general, one is more specific), where the “part of” relation links a part and a whole which are not necessary of same nature, for example a vase and its handle.

Furthermore the “kind of” relationship is a hierarchical relation which defines a *strict order** relationship between concepts. This is not the case in the “part of” relationship.

Let us recall that a strict order relationship is a binary relationship which is by definition:

- *irreflexive*, which means that an object cannot be put in relation with itself - a concept cannot be more (or less) specific than itself;
- *asymmetric*, which means that if a concept C1 is more specific than a concept C2 then C2 cannot be more specific than C1;

- *transitive* which means that if C1 is more specific than C2 and C2 is more specific than C3 then C1 is more specific than C3.

The logical properties of these relationships define useful guidelines for the terminology building process and must be verified at each step. Therefore the TMP will propose only subconcepts which will not create any cycle
Also, the TMP emphasizes on definition by the *specific difference* where a concept is defined from a super concept with adding a specific difference e.g. <Market House> is a <Public House> for 'commercial' purposes.

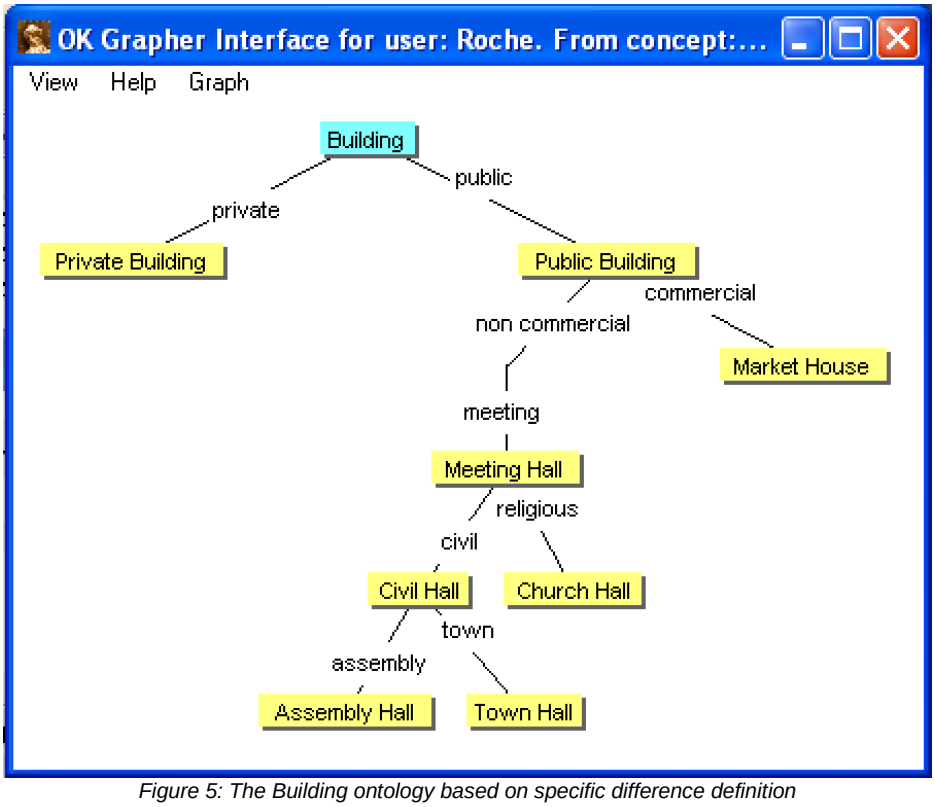


Figure 5: The Building ontology based on specific difference definition

3.1.3 Indexing and Information Retrieval

In separating the conceptual dimension from the linguistic ones, ontoterminology defines a new approach for indexing and retrieval information. Terms are used to describe the content which will be classified by concepts denoted by the terms. In a similar way, a document will be automatically analysed and automatically classified by concepts denoted by the terms contained in the documents.

The figure below illustrates this approach:

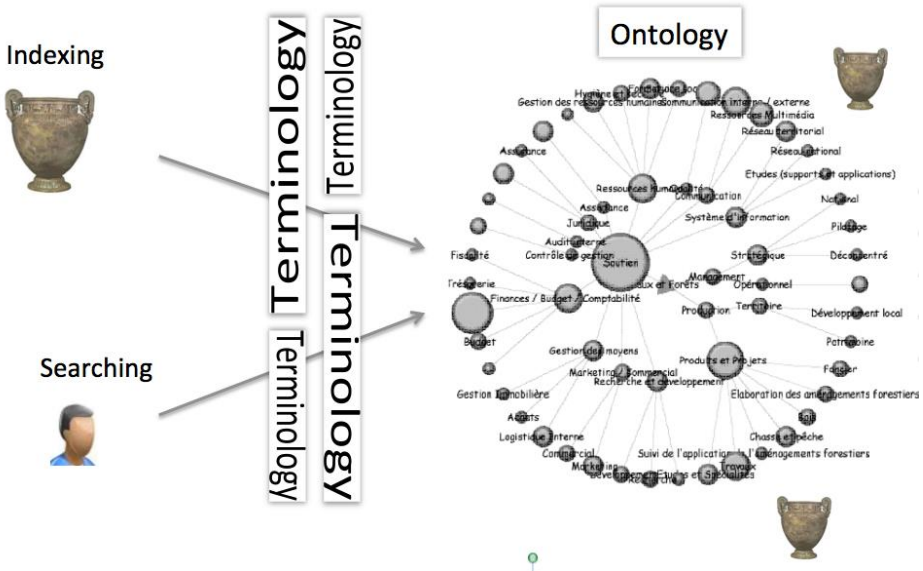


Figure 6: Semantic and multilingual information retrieval principles

From a thesaurus point of view, concept names can be considered as *descriptors* when terms can be considered as *non-descriptors*.

It is important to notice that content can be indexed by a concept in a given language and can get it back through another language. Furthermore, the search engine will be able to exploit the logical properties of the conceptual relationships in order to improve the search results.

The ontoterminology-oriented approach for content management systems was validated by the ASTECH FP6 project: “Advanced Sustainable Technologies for Heating and Cooling Applications” (2006-2009). One of the goals of the project was to carry out a multilingual information retrieval system in renewable energy based on ontology and multilingual terminologies.

The figure belows shows an example where the query is given in French through the ontology (concept <Transfert de chaleur>) and the first results are documents written in English indexed through English terms denoting the same extra-linguistic concept of <Transfert de chaleur>.

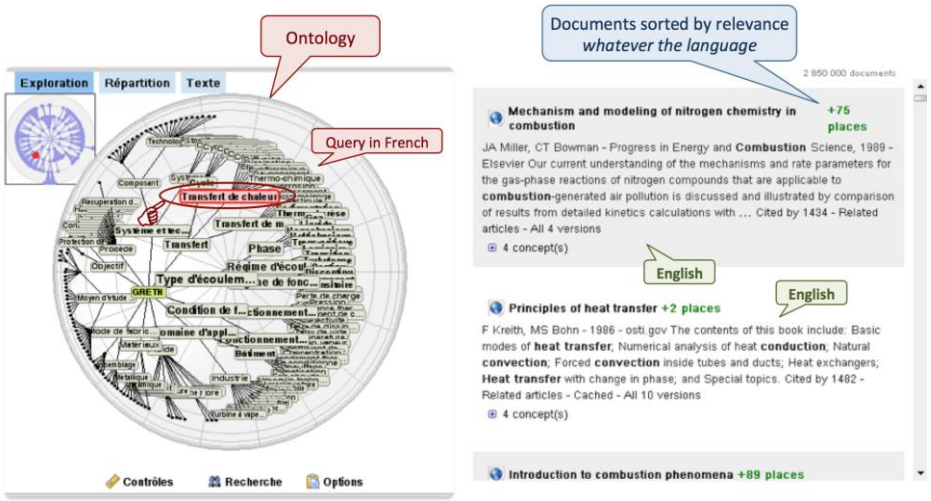


Figure 7: An application of ontoterminology-oriented information retrieval for content management system: the FP6 ASTECH project

3.1.3.1 Mapping

One of the outcomes of the Linked Heritage project was to adopt a mapping terminology approach for interrelating information from diverse sources rather than to use a single and universal terminology which would be difficult to create and maintain. Mapping terminology in creating a network of terminologies allows their interoperability.

In an ontology-oriented approach, mapping two terminologies consists in determining the relationships between concepts.

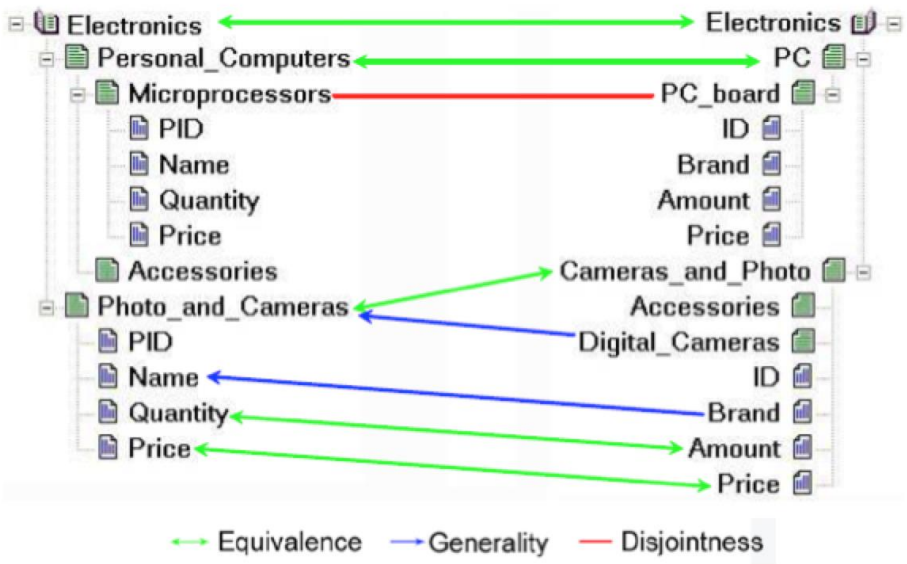


Figure 8: "Tutorial onSchema and Ontology Matching"
Pavel Shvaiko Jérôme Euzenat ESWC'05 – 29.05.2005

Two concepts can be either equivalent, more specific (versus general), overlapping (sets of instances) or different. This information is embedded into a mapping element (c_1, c_2, R, n) where:

- c_1 and c_2 are concepts to be mapped;
- R is one of the relations ≡ (equivalent), > (more general) , < (less general), ∩ (overlapping), ⊥ (disjoint);
- n is a confidence measure.

The result of the matching process of 2 terminologies is the alignment of the terminologies defined as a set of mapping elements.

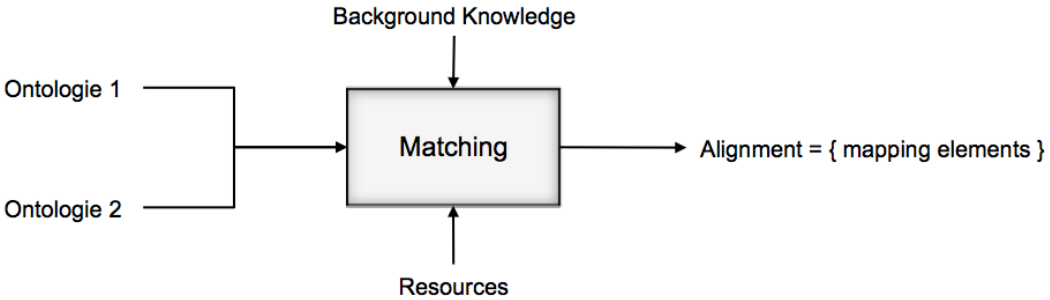


Figure 9: The matching process

Terminology mapping must deal with three kinds of problems:

- *terminological (lexical) mismatch* occurs when the concept names are different including synonymy (“car”, “automobile”), homonymy (same names with different meanings), terminological variation (same names with slight differences, ellipsis…)
- *language mismatch* (language heterogeneity) occurs when the terminologies are expresses in different languages, e.g. frame language, logic, XML, RDF, SKOS etc.
- *conceptual mismatch* occurs when there are different conceptualizations of the same domain, for examples “similar” concepts with different sets of instances (scope mismatch), “similar” concepts describing objects with different levels of details (granularity mismatch).

The TMP handles terminologies written in SKOS, which has the advantage that the language mismatch does not exist anymore.

Furthermore, terminology mapping can combine two types of methods, linguistic and semantic:

3.1.3.2 The Linguistics Models

The linguistic methods compare the concept names using string-based techniques based on the principle that the more the strings are similar, the more they denote the same concept.

After lemmatization of the concept names, a distance between the two strings is calculated using different techniques: prefix (e.g. “int”, “integer”), suffix (e.g. “phone”, “telephone”), same letters or n-gram (common sequences of n characters), string metric (Levenshtein, Jaccard, etc.)

Linguistic resources like dictionaries, thesauri or systems like WordNet can be mobilized to calculate the “distance” or the “similarity” between terms. For example “digital camera” is a hyponym of “camera”.

3.1.3.3 The Semantic Methods

The semantic methods differentiate two kinds of approaches:

- The *extensional approach* relies on the sets of instances of the concepts postulating that if two concepts have the same extension they are identical[7]. The relations **≡ , > , <,** ∩ and ⊥ are defined according to the set operators.

- The *intensional approach* gathers methods based on the internal structure of concepts, i.e. on their attributes. These methods compare the attributes names (using linguistic methods) as well as their “semantics” i.e. the data type of the value of attributes (e.g. *date* and *working date*). The intensional approach also relies on the external structure of concepts, i.e. on relationships between concepts involving the graph structure (depth of concepts), the connected nodes (two nodes are all the more similar since their connected nodes are similar), etc.

[7] This mathematical definition of identity between sets (identity of extensions) is not really applicable in cultural heritage applications since a same set of objects can be “viewed”, and then conceptualized, in different ways.

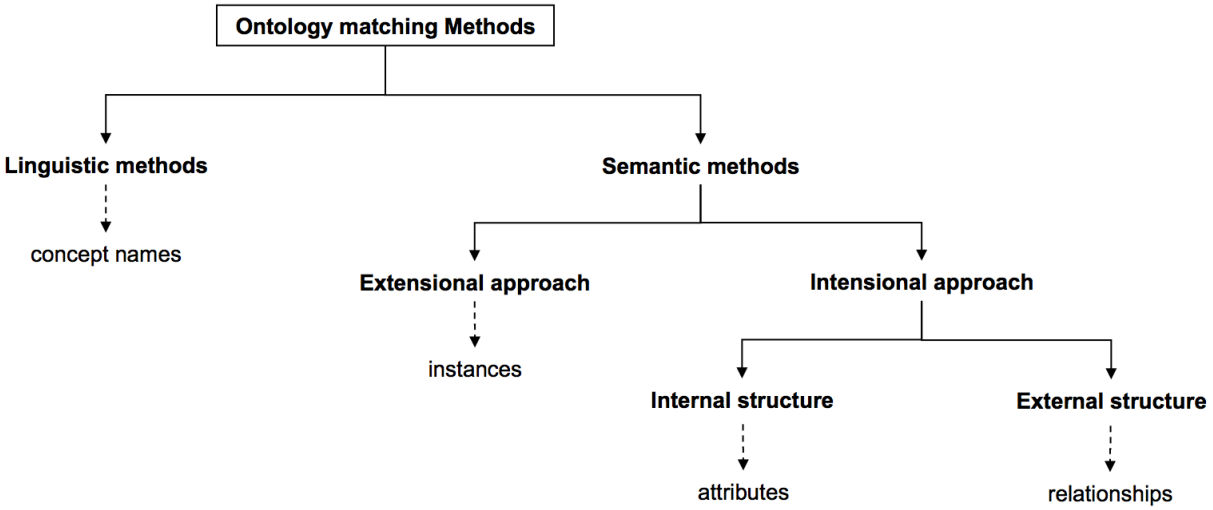


Figure 10: The Ontology matching methods

3.1.4 Interactive Navigation and Browsing

There are different paradigms of interactive navigation based on an ontological structure. The following ontology-oriented navigation paradigms have been tested:

- The *Eye-Tree View paradigm* is based on an hyperbolic geometry which allows to focus on the selected concept and its neighbouring concepts putting away the most distant concepts on the borders. The selected concept can move on the Eye surface.
- On the opposite, the *Radial-Tree View paradigm* displays the selected concept in the centre of the view when the sub-concepts move around it. A magnifying glass slide allows to focus on some sub-concepts.
- The *Building-Tree View paradigm* is a means to highlight the borders between sub trees.
- At last several classical Tree View Paradigms where *nodes can be folded and unfolded*.

A classical Tree View with a *à la* Windows display appeared as one of the simplest means of navigation. Furthermore the Tree View will be also used for the graphical mapping of terminologies.

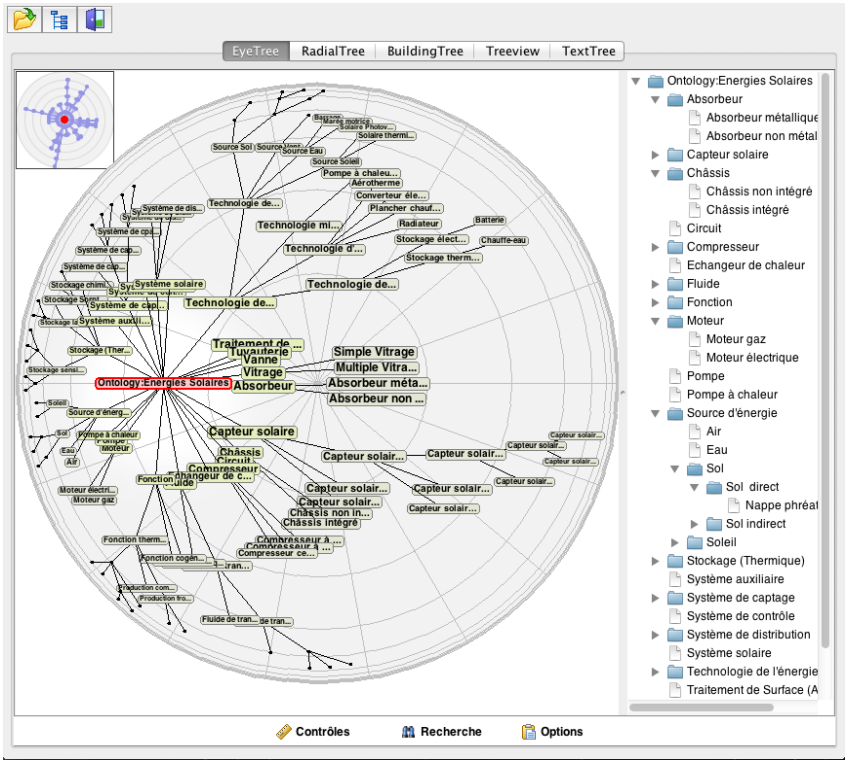


Figure 11: The Eye-Tree View paradigm

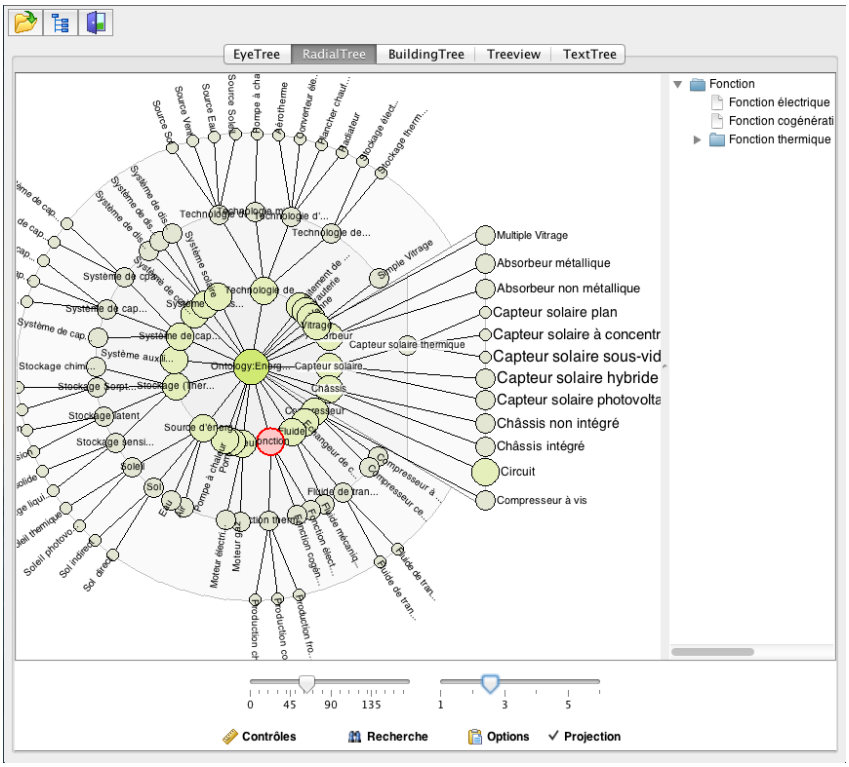


Figure 12: The Radial-Tree View paradigm

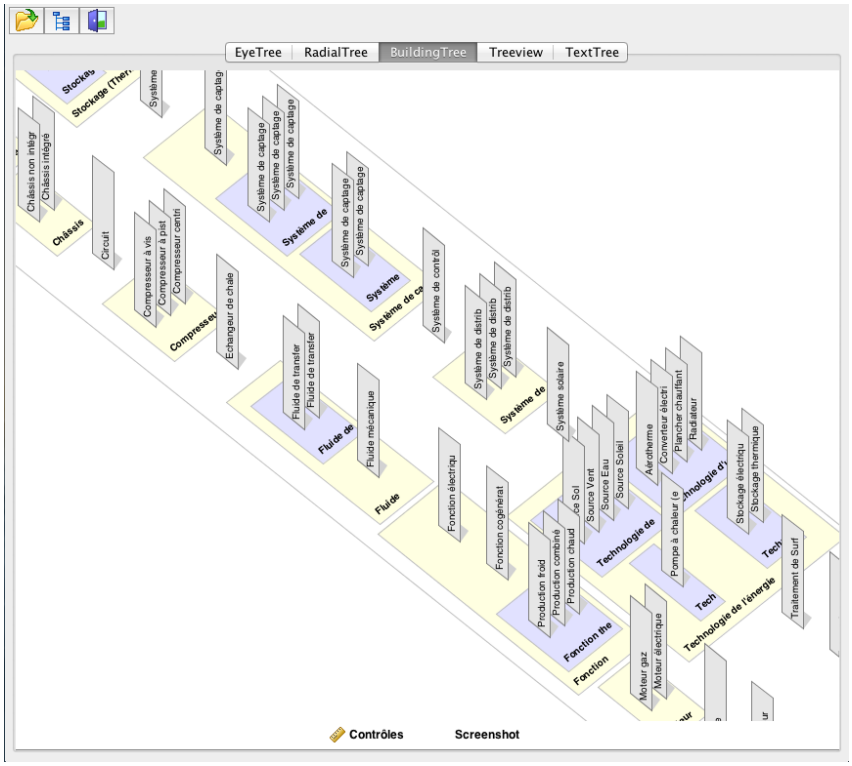


Figure 13: The Building View paradigm

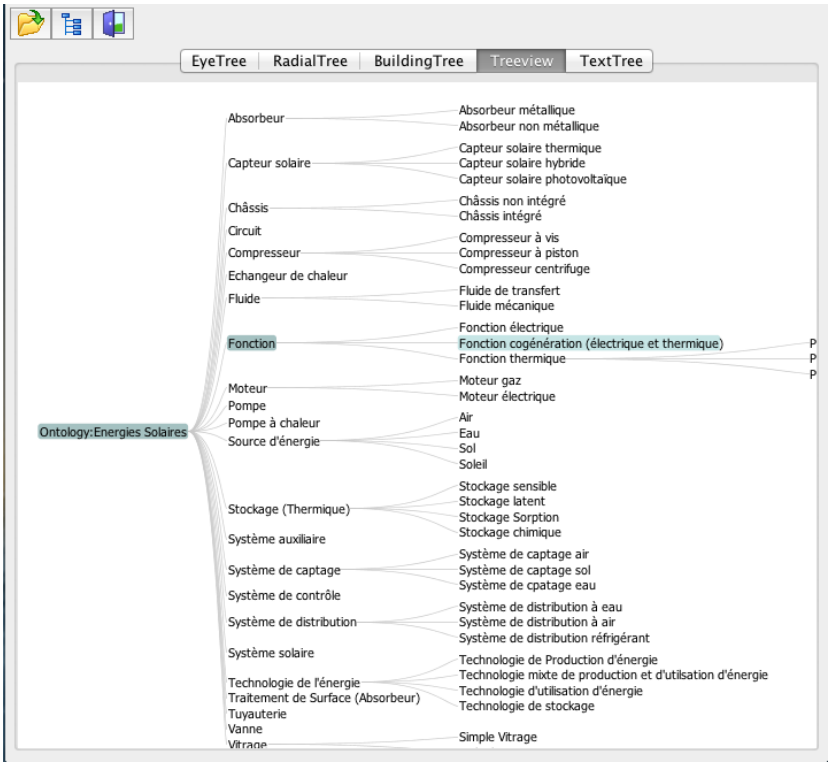


Figure 14: A Tree View paradigm

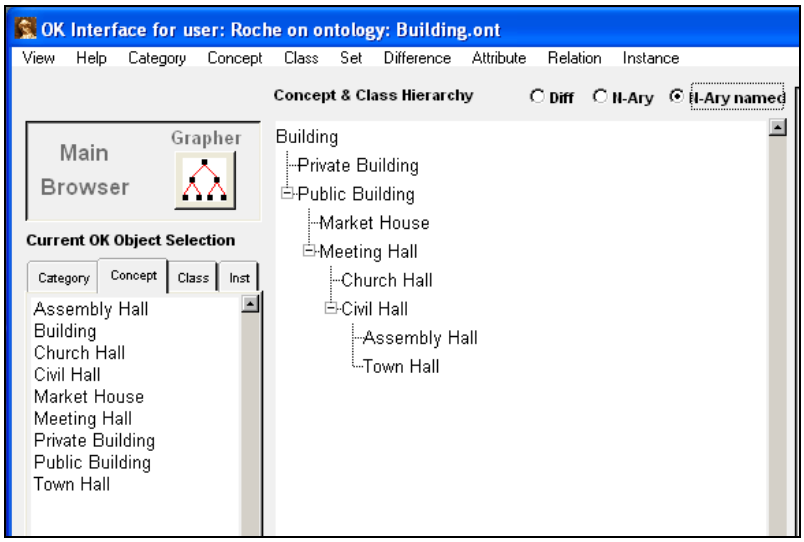


Figure 15: A Tree View paradigm

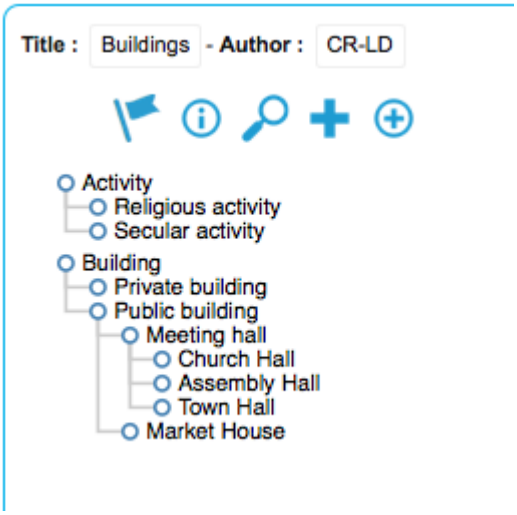


Figure 16: A Tree View paradigm in the TMP

3.2 Implementation Requirements

3.2.1 Technical Architecture

Some main guidelines of what a TMP architecture could be were specified during the Linked Heritage project (see the D 3.2 Deliverable “Functional and technical specifications of the terminology platform”).

The new TMP architecture takes into account the returns on experience of the first version of the TMP. In particular three new systems have been specified and integrated: a *user management system*, a *dedicated registry* and a *Tree view editor*.

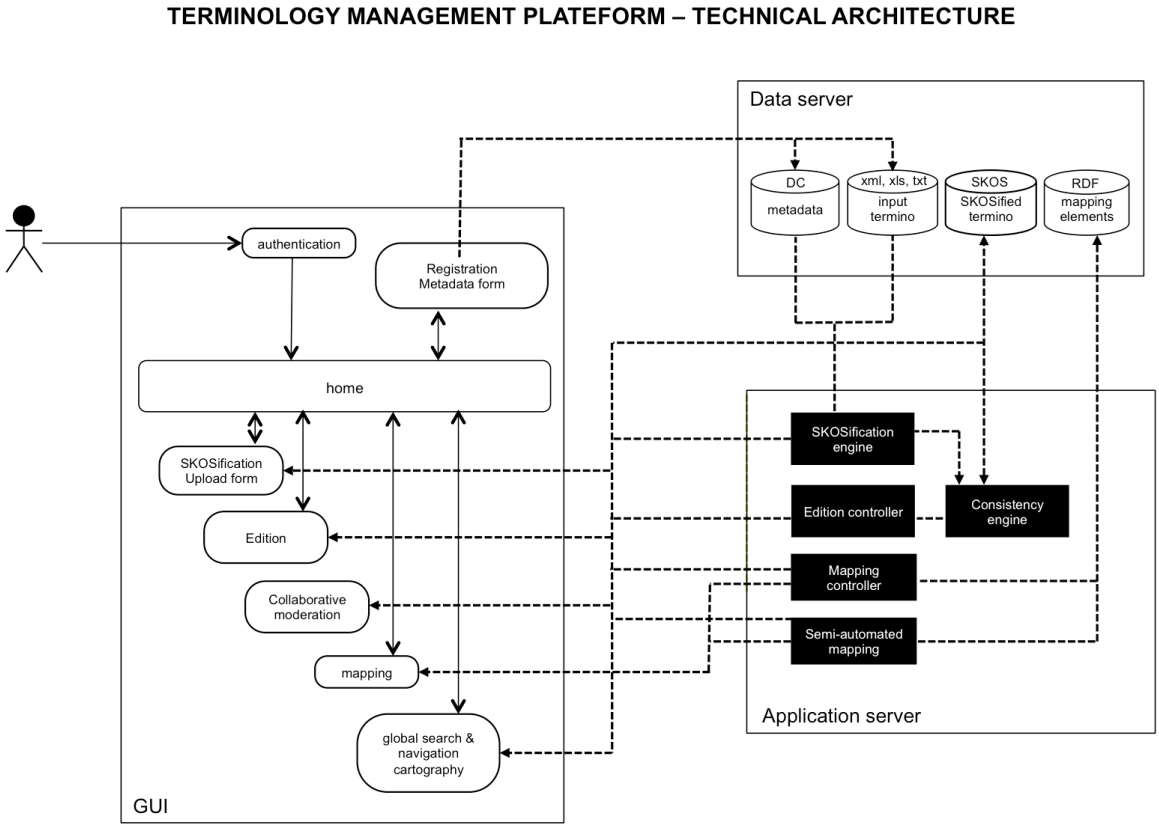


Figure 17: The architecture of the TMP prototype

The three main elements of the TMP are distributed by three separated components.

- On the left, the GUI manages the main application logic, with graphical user interface for skosification, searching and mapping.
- On the top right, the terminology registry is in charge of storage and versioning of all data (terminologies content, mapping information, …).
- On the bottom right, the application server manage data and all behaviour. This part also contains engines that automate some tasks.

The communication between servers will be managed with RESTful web services.

3.2.2 Internal Format

One of the outcomes of the Linked Heritage Project was to choose SKOS language for the internal representation of TMP terminologies (see the D 3.2 Deliverable “Functional and technical specifications of the terminology platform”).

As mentioned before, SKOS (for *Simple Knowledge Organization System*) is a common data model for sharing and linking knowledge organization systems via the Web and appears as one of the most interesting format for thesaurus representation. SKOS is more and more required and used for sharing thesaurus like in the online catalogue Europeana.

Although SKOS is not a language dedicated to terminology - let us recall that a terminology is not a thesaurus - it was decided to keep SKOS as pivot language for the first version of the TMP. It means that terminologies are uploaded[8], edited, stored and mapped under this formalism based on the principles of this language.

[8] Importing terminologies in TMP is a means to skossify them.

Nevertheless, it is important to note that if terminologies can be imported and exported in W3C standards (SKOS and RDF), the TMP principles do not rely on them and they are independent of the different versions of these standards.

At last, it is important to bear in mind that SKOS is first of all an interchangeable format, developed for thesaurus sharing. It is not a modelling language for terminology since some properties cannot be guaranteed due to the SKOS relationships.

For example, the skos:broader relationships (and its inverse relationship skos:narrower) do not distinguish between the "kind of" relationship and the "part of " relationship, the two fundamental relationships used in concept modelling building.

Let us take the following example in SKOS describing Berlin as the capital of Germany:

```
ex:country rdf:type skos:Concept;
        skos:prefLabel "Country"@en.
ex:germany rdf:type skos:Concept;
        skos:prefLabel "Germany"@en;
        skos:broader ex:country.
ex:berlin rdf:type skos:Concept;
        skos:prefLabel "Berlin"@en;
        skos:broader ex:germany
```

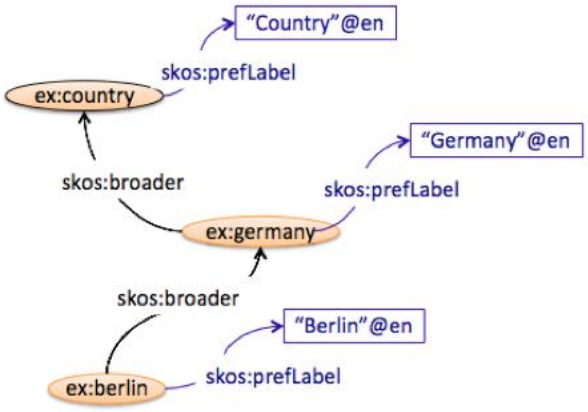


Figure 18: Example of a SKOS graph

Berlin is not "a kind of" country. It is the reason why the skos:broader and the skos:narrower are not defined as transitive in SKOS.

In the same way, the skos:broader and the skos:narrower are not defined as irreflexive so they can be imported into SKOS ontologies written in OWL – the reflexive rdfs:subClassOf statement of OWL will be rewritten as a askos:broader relationship (the "subClassOf" relies on a set-inclusion meaning which is reflexive).

Since SKOS is the pivot language of the TMP, it means that the TMP must verify that the use of the SKOS semantic relationships is such that these relations are irreflexive, asymmetric, and acyclic. This way a consistent terminology can be built.

3.3 Interface Requirements

3.3.1 Structure

The TMP terminology-editing interface follows the architecture of the TMP model. A clear separation is done between the *conceptual dimension* and the *linguistic dimension*. The figure below illustrates this structure. The main TMP-editing interface displays:

- the linguistic dimension of the terminology in a frame dedicated to the SKOS linguistic relationships, i.e. the labels (skos:prefLabel, skos:altLabel, skos:hiddenLabel) and the

documentary notes (skos:note, skos:scopeNote, skos:definition, skos:example, skos:historyNote, skos:editorialNote, skos:changeNote);

- the conceptual dimension of the terminology in a frame dedicated to the SKOS semantic relationships, i.e. the hierarchical relationships (broader and narrower) and the associate relationship (related).

Insofar as the TMP is concept-oriented, the main access point to the terminology for editing, creating, etc. is the *hierarchical structure of concepts* (left frame of the interface) displayed through a Tree view interface.

Let us notice that the language used for displaying the concept names *can be different from the language used for editing the attached labels*. Such a feature allows defining labels in a different language from the language used for the ontology building. The latter is called the *preferred* language – a new metadata has been introduced in order to take into account the preferred language.

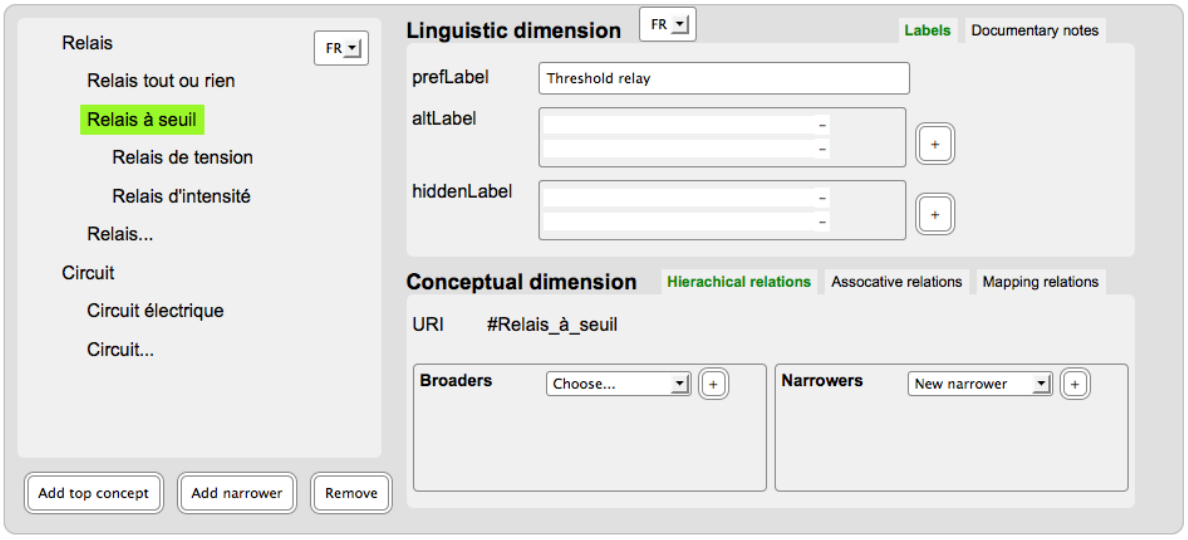


Figure 19: TMP-Editing interface

The logical properties in the TMP are enhanced as useful guidelines for terminology building. They are integrated into the TMP interface. For example, the irreflexivity and acyclic properties of the SKOS semantic relationships allows proposing to the user only the concepts that verify these properties.

3.3.2 Representation

The internal representation of the interface is written in RDF and takes into account the logical properties of the conceptual model.

3.4 Main Functionalities

The functionality map below shows the flow of tasks from the user point of view:

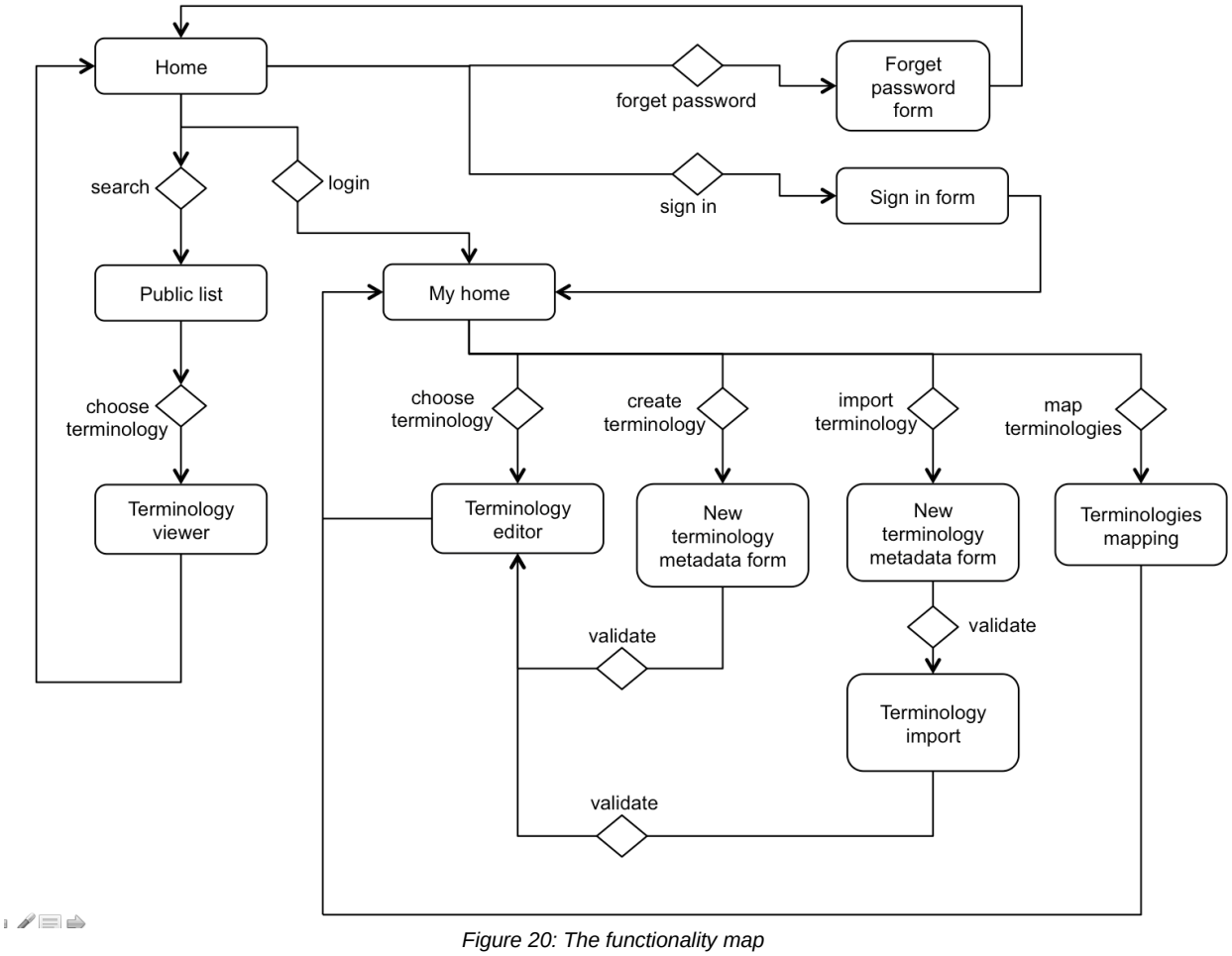


Figure 20: The functionality map

The main functionalities for a terminology management system are :

3.4.1 Registration

The user must *log in* or *sign in* to access the application. Different kinds of users will be introduced according to their rights on terminologies. These rights are defined in terms of *C(reate), R(ead), U(pdate) or D(elete).*

3.4.2 Terminology Definition

The user can create a terminology in the TMP using two features:

- Creation of a terminology from scratch;
- Creation of a terminology by importing an existing terminology. It means that the propriety terminology is imported into the TMP format (SKOS). Such a functionality is a way to skossify terminology defined in a different format, such as a CSV-file.

The definition of the terminology is done according to the theoretical requirements previously defined (see supra).

3.4.3 Search, navigation

The terminology can be accessed by a textual search and by navigating a tree view of concepts.

3.4.4 Mapping

Two terminologies can be linked by association of their concepts (match).

3.4.5 Collaboration moderation

Suggestions can be done on elements of the terminology in an online collaborative environment.

3.5 Use Cases

The functionalities will be described through use cases.

4 Specifications

The specifications of the TMP are described through use cases.

4.1 Logo

The logo of the TMP shares the same owl symbol common to the Athena Plus and Linked Heritage projects.

Figure 21: The TMP Logo (AthenaPlus)

4.2 Starting with the TMP

The TMP is a web application whose URL is http://www.culture-terminology.org/

The TMP Home page allows users to:

- Search into public terminologies;
- Login or Sign in for new users.

It also includes a brief description of the TMP as well as some information about the project and partners.

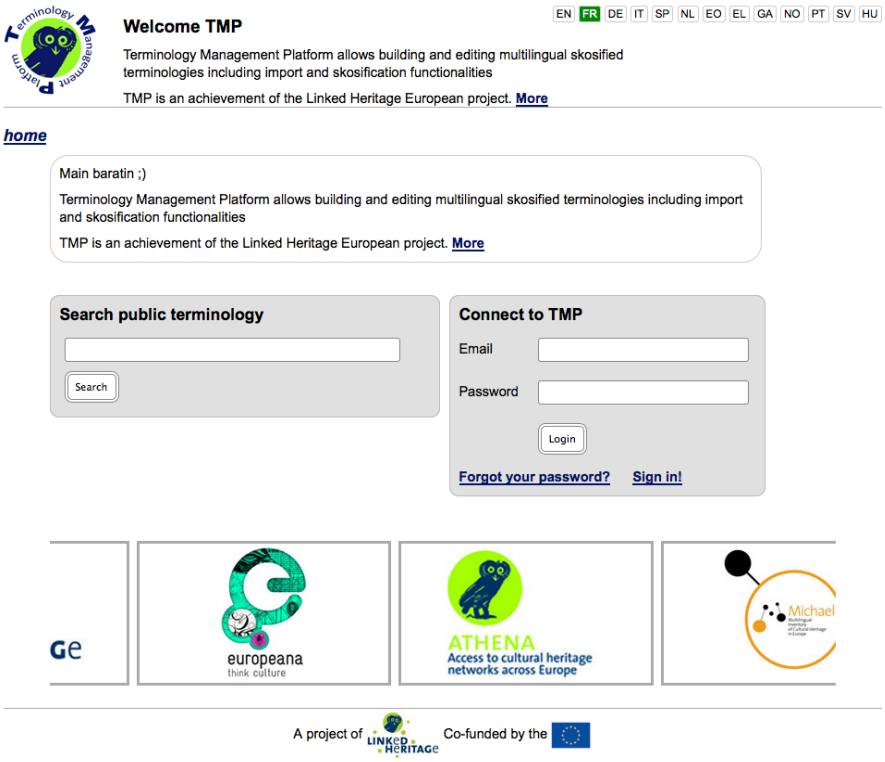


Figure 22: The TMP Home page

4.3 Connection

There are public terminologies that do not need authentication to be visualised. However, it is recommended to be logged in to make the most of the tool. A username and password can be obtained in the sign up page.

4.3.1 Roles

Four kinds of users (roles) will be implemented within the TMP. They all have different rights:

- **General user** : The general user can browse and navigate through the terminology registry. The general user can search into the registry and view terminologies that are registered as public ones. The general user is not logged in.

- **Contributor** : The contributor is allowed to view terminologies and make suggestions to modify or comment them. The contributor is a user who has an account and is logged in. The contributor may belong to another institution and is then able to make suggestions to a public terminology.

- **Editor** : The editor works under the authority of the owner of the terminology. The editor has the same rights than a contributor but can also have management rights on his terminologies. The editor is then able to Create, Read, Update and Delete concepts and terms.

- **Owner** : The owner has a role of administrator of terminologies. The owner is responsible for all the terminologies of his institution. He can Create, Read, Update and Delete concepts and terms and also accept or refuse the suggestions made by the other users. The owner is also the one responsible for managing rights for the users and editing the metadata of the terminologies. He is the one able to publish the terminology and set it as private or public.

	Actions								
Role	View Public terminology	Make suggestions	CRUD Concept	CRUD outcoming mapping	Accept suggestions	CRUD Terminology metadata	Manage Rights	Publish Terminology	Set public/ private
Owner	X	X	X	X	X	X	X	X	X
Editor	X	X	X	X	X				
Contributor	X	X							
General User	X								

Figure 23: The four types of users

Every contributor, editor or owner belongs to an **institution** and must be logged in in order to access to the TMP functionalities.

Institution: The institution will be created by the owner of the terminologies. There is one owner per institution. The first user from a specific institution to connect to the TMP is the one able to give details about the institution. He will be then the owner of the terminologies and the one to grant access and rights to other users from the same institution.

Administrator: the administrator of the institution is the one who can accept or reject the requests from users to be affiliated to the organization. There is one administrator per institution who validates the users but do not provide rights to them. Each user who is creating a terminology is an owner. The owner can provide editor rights to other users affiliated to the same institution.

The roles are not exclusive. The owner from a museum A can be a contributor for a library B.

4.3.2 Sign up

There are two possibilities when creating a new account, depending on whether the institution exists or not. Be aware that a third possibility exists: an account for a sole “personal” use of the TMP.

1st Use Case : New User and Undefined Institution

The user is signing up for the first time to the TMP and her/his institution is not registered yet. She/he has to create the account for the first time and will identify the institution by providing all the details about it.

First step : Creation of a New User

A User is defined by a *name*, a *first name*, an *email address* and a *password*. The password will be confirmed.

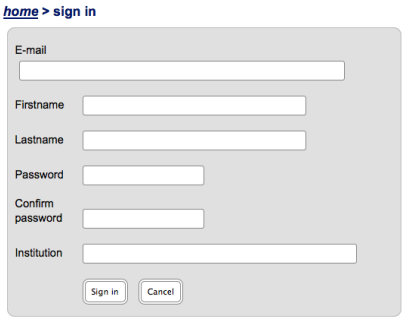


Figure 24: Creation of a New User

When flying over, the link "create new" displays a tooltip: “Your institution is not in the list. Add it now.” You can add the name of the institution in the list, following the second step.

Second step: Creation of a New Institution

An institution is defined by a *name*, an optional *acronym*, a *legal status*, an *address*, a URL and a description of the institution (giving more details about collections, target groups etc.). You can define affiliated organisations and primary contact users by defining them in a drop down list.

A modal window (lightbox) opens (blocking user interface).



Figure 25: Creation of a New Institution

Once validated, the Institution and the User are created and it returns to the previous age "Creation of a New User".

The selected item in the list is the new institution. The link "create new" disappears and is replaced by "edit". If another institution is selected, the link "edit" disappears. This link "edit" re-opens the modal editing window of the institution.

Finally, the user clicks on "sign-in". If the information is correct, he can log in.

Note: If a user is not affiliated to a specific institution, the user will have to create an organization with his own name and set as “Private” the type of organization. This allows any owner to manage suggestions made to their terminology.

2nd Use Case: New User and Existing Institution

The user has not yet an account but her/his institution is already registered. Only information about the new user is then required: *email address*, *first name*, *last name*, *password* and *institution*.

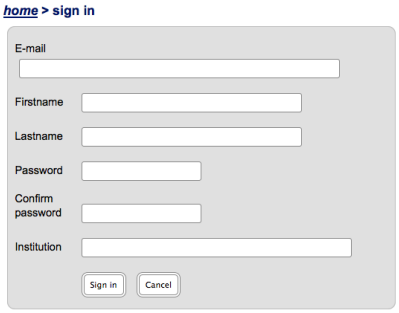


Figure 26: Creation of a New User

The user selects her/his institution (drop down list with all the institutions already registered) and sign in. If the information is correct the user is logged in.

3rd Use Case : New User without any Institution

The user is not affiliated to an institution but wants to use the TMP.

The affiliation to an institution is mandatory. When a user wants to contribute to terminologies from other institutions and/or create his own terminology as personal and not institutional, he must select a "Personal use" item in the dropdown list with all the institutions.

4.3.3 Information Validation

The information must be checked and verified for all use cases. The control is done on the fly (onblur, onchange...), for example:

- Control mechanism to check the user is not already registered
- Verification of password equality
- Verification of empty fields (all required)

4.4 Login

The user must be logged in in order to access all the functionalities of the TMP. From the Home page, the User fills in the Login Form with her/his email address and password.

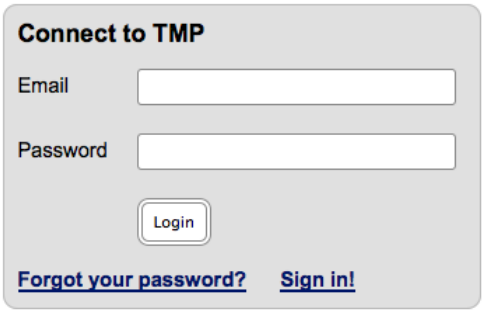


Figure 27: The Login Form of the TMP Home page

1st Use Case: The User is already registered in the TMP

The User can access the TMP Main page.

2nd Use Case: Registration failed

When the registration fails, the system will redirect the user to the Login Form with failed email and password pre-filled. The user had to re-enter the password and/or email address. When the user has forgotten the password, he has to recover it by clicking the *forgot your password?* tab.

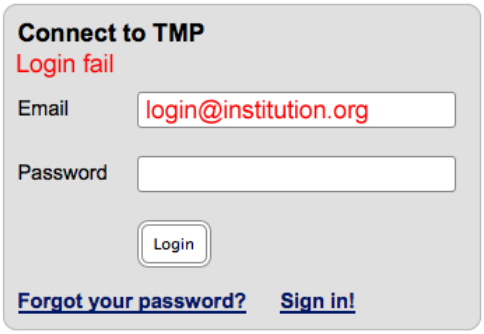


Figure 28: Failed Login

3rd Use Case: Forgotten password

By clicking the *forgot password* link, the user is redirected to a password recovery form:

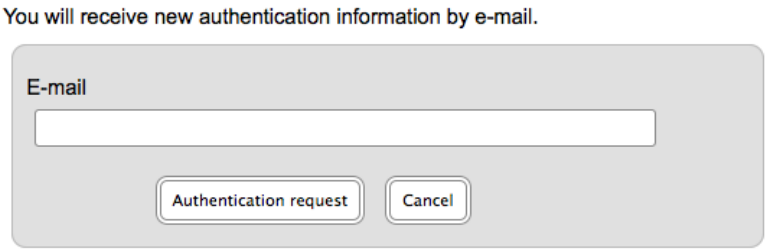


Figure 29: Recovery password form

When validated, the server sends an e-mail with a new generated password. The web site returns back to the Home page where he/she can log in.

4.5 User management

Once logged in, a User can use the features as defined by her/his role:

- Administrator: can accept or reject the requests from users to be affiliated to the organization. Let us recall that there is only *one administrator per institution* who validates the users, but he/she does not provide rights to them. Each user creating a terminology is an owner of this terminology.

- Owner: is responsible for all the terminologies of her/his institution and attributes rights for editors. She/he is the one able to publish terminology and set it as private or public.

- Non-owner: can only request for a change of rights.

4.5.1 Owner

The interface gives access to the *editing profile*, *managing users* and *logout*.

Use Case: Editing Profile

All data can be updated, except for *institution* and *email address*. She/He can use two independent forms:

- General Information (first name, last name)
- Password (with confirming password)

Use Case: User Management

The owner has access to all the users of the institution (array of users).

For each of them the owner can access their:

- email, last name and first name;
- rights (General user, Contributor, Editor, Owner) which can be updated;
- request for a change of rights (accept or reject)

He can also delete a user.

Use Case: Logout

The user is redirected to the Home page

4.5.2 Non-Owner

The interface gives access to the *editing profile*, *change of rights* and *logout*.

Use Case: Edit Profile

Like in the case of the owner, all data can be updated, except for the institution and email address. He/She can use two independent forms:

- General Information (first name, last name)
- Password (with confirming password)

Use Case: Right Request

The Right Request page displays the *current rights* of the user and the *available rights*.
The request form with the requested rights is sent to the administrator of the institution.

Use case: logout

The user is redirected to the Home page.

4.6 Main page

When the user is logged in, the TMP displays the Main page which offers four possible actions:

- Search for terminologies

This action allows searching and browsing terminologies, either set public or personal. Search queries can be done entering *titles*, *authors* or *any other metadata* for quick access and relevance.

- Create a new terminology

This action allows creating a new terminology in the TMP or importing a terminology from an existing repository (format in Excel, text, XML etc.).

- Map two terminologies

This action allows mapping concepts from different terminologies. Both terminologies are displayed and the selected concepts can be linked in pairs.

- Manage an existing terminology

Depending on her/his role for a given terminology, the user can:

- Edit the terminology: navigate or modify concepts, terms, notes etc.
- Export the terminology in different formats: RDF, SKOS etc.
- Delete the terminology

The figure below is an example of a possible TMP Main page.

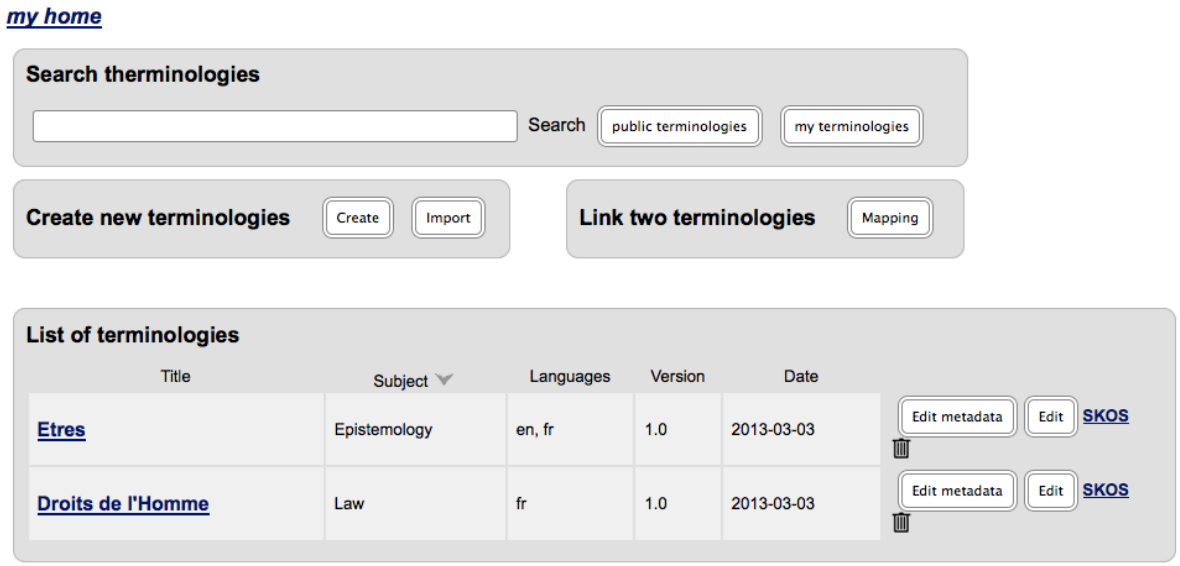


Figure 30: The TMP Main page

4.7 Create or Import a Terminology

The creation of a new terminology can be done either from scratch or from an existing terminology which must be downloaded.

In both cases, it is necessary to define some metadata describing the terminology.

Use case 1 : Create a terminology from scratch

The user has to fill in metadata on the terminology. These can be required (mandatory) or optional.

Required fields:

- Title (free text): short and readable. A title is the name of the terminology by which the terminology is formally known;
- Subject (choice in a list and free text): main domain describing the terminology;
- Languages (multiple choices): language(s) of the terminology. The languages are represented by codes based on the ISO-6391 norm;
- Edition (free text): the edition date or version of the terminology;
- Creator (free text): name of a person or organisation responsible of the terminology;
- Publisher (free text): if different from the creator, name of an organisation, a service or a person responsible for making the resource available;
- Rights (list choice): information about rights held in and over the resource defined from Creative Commons[9] or other copyright licences.

Optional fields:

- Alternative title (free text): official or more descriptive title including abbreviations or translations of the title;
- Description (free text): more information than the title. It can include an abstract, table of contents, etc.
- Creation date (date field): date of creation of the terminology;
- Date of issue (date field): date for the terminology to be set as public;
- Contributor (free text): the person contributing to the terminology;
- Identifier (free text): unambiguous reference to the terminology within a given context;
- Source (free text): sources used to create the terminology;
- Amount of concepts (choice in list: 10 or less, between 11 and 100, between 101 and 500, between 501 and 1000, etc.): the size of the terminology defined as an approximate number of concepts.

Once validated, the form continues to the Terminology Editor web page (see infra).

[9] Creative Commons (CC) is a non-profit organization that enables the sharing and use of creativity and knowledge through free legal tools. The CC copyright licenses provide a simple, standardized way to give the public permission to share and use creative work on conditions of authors' choice. CC licenses let authors easily change their copyright terms from the default of "all rights reserved" to "some rights reserved" .
[9]http://creativecommons.org/

Use case 2 : Import a terminology

This specification describes the process of importing a terminology starting with the metadata form, as specified above. The only extra metadata is the format of the data (CSV-file or SKOS).

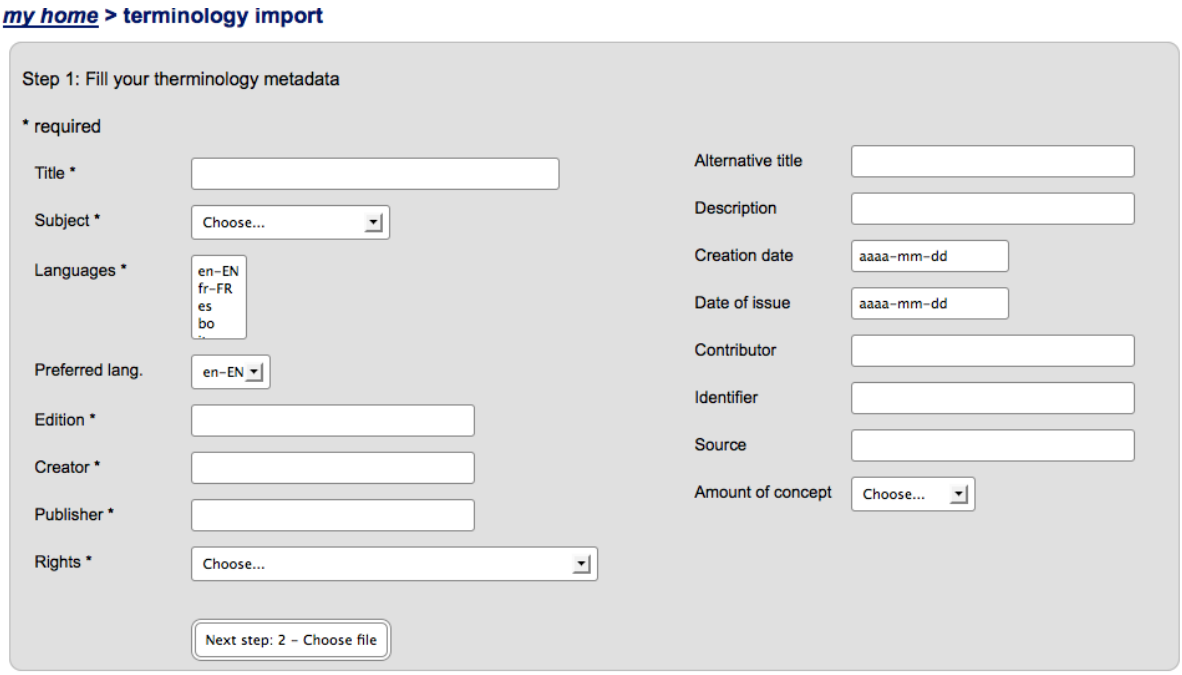


Figure 31: The Metadata Form

Once the form is validated, the user can import her/his repository.

First step: select the file

The user has to select the file with the terminology. Make sure that the file is saved as a CSV file, e.g. when using an Excel. For more information, the user can click on the *More about CSV rules* link. The user has to choose the type of the imported file.



$

Figure 32: File selection

Second step:

When the file has been successfully imported, the user must now map each field of his terminology with the SKOS labels and properties.
The data are visualized as a table and each column will a provide a list of mapping choices.

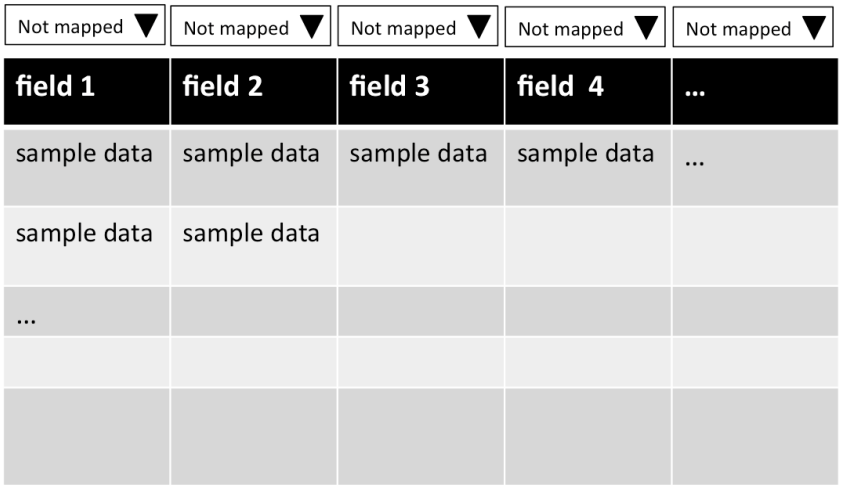
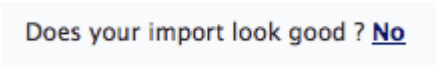


Figure 33: Configuration of import

If (some) columns do not show mapping choices for each column, this means that the CSV file has not been recognized. The main cause is the character used as column separator in the CSV (MS Excel, OpenOffice and other software generates different kind of CSV). To change this, the user can click on the link below (before the table):



and then choose the appropriate separator:

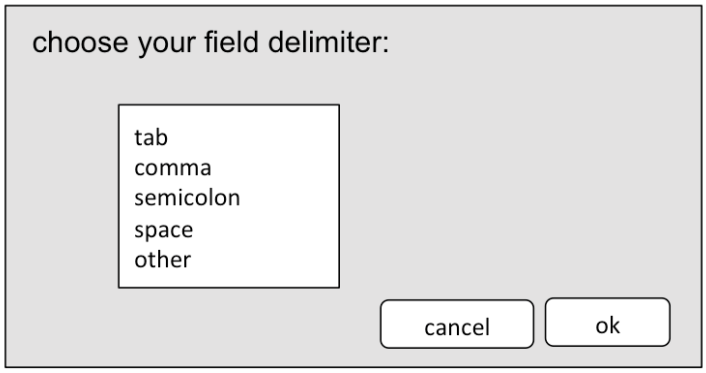
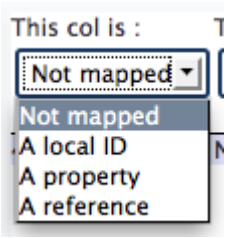


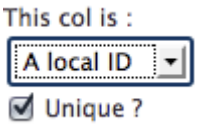
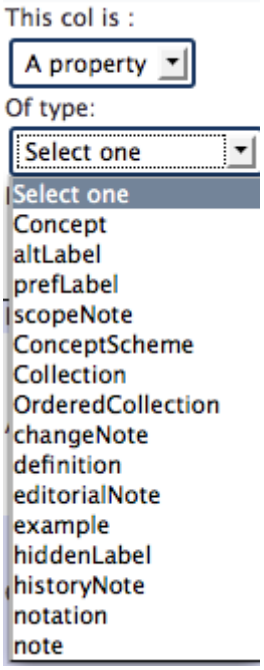
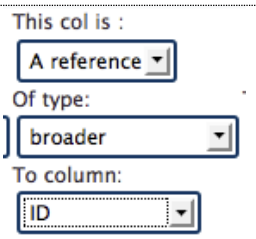
Figure 34: CSV configuration

When the file has been recognized successfully, the user must define each column of the file. He can choose between:

- A local ID: the unique identifier of each terminology element
- A property: a SKOS property (e.g. preferred label)
- A reference: a SKOS reference (e.g. broader or narrower)



The possibilities are specified in detail:

Local ID	Mainly, a number, single representative of an entry	
A SKOS property	A SKOS field containing text.	
A reference	A local mapping: defines how concepts are linked together. This reference is another column already managed. For example, the broader reference is a concept identifier, already listed (ID).	

At the end of the second step, each column must be mapped with a SKOS field. The terminology is then ready to be imported.

Third step:

When the last form is validated, data are processed by the TMP and a new terminology is created. The user is redirected to the Editor page.



Figure 35: Import

4.8 Skosification of the Terminology

Importing a terminology into the TMP consists in translating the terminology into the internal format of the TMP. Since this format is SKOS, importing is a means to skossify an existing terminology defined in a different format, for example in CSV stemmed from an Excel file.

The import process will use the structure of data to automatically define the correct concepts with their predefined relationships. For example, from the terminology below the TMP will automatically generate the concepts <Building>, <Private Building>, <Public Building>, <Meeting Building>, etc. with the correct relationships between them: <Private Building> and <Public Building> have <Building> as broader concept, etc.

	A	B	C	D
1	Building			
2		Private Building		
3		Public Building		
4			Meeting Hall	
5				Assembly Hall
6				Town Hall
7				Church Hall
8			Market House	
9				

Figure 36: Terminology in Excel format

4.9 Owner Management

Every terminology owned by a user can be fully managed using a specific page accessible in the terminologies list.

- Edit metadata
- Attribute rights
- Manage versions
- Publish
- Delete

Edit metadata:

This interface allows to edit metadata of a terminology which were already created.

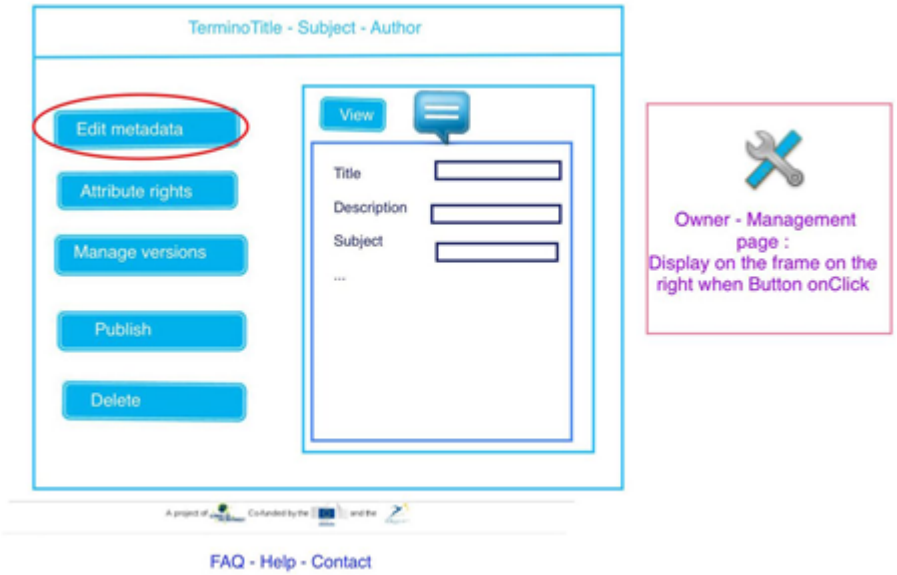


Figure 37: The Metadata Form

Attribute rights:

This interface allows the owner to modify rights claimed on his/her terminologies.

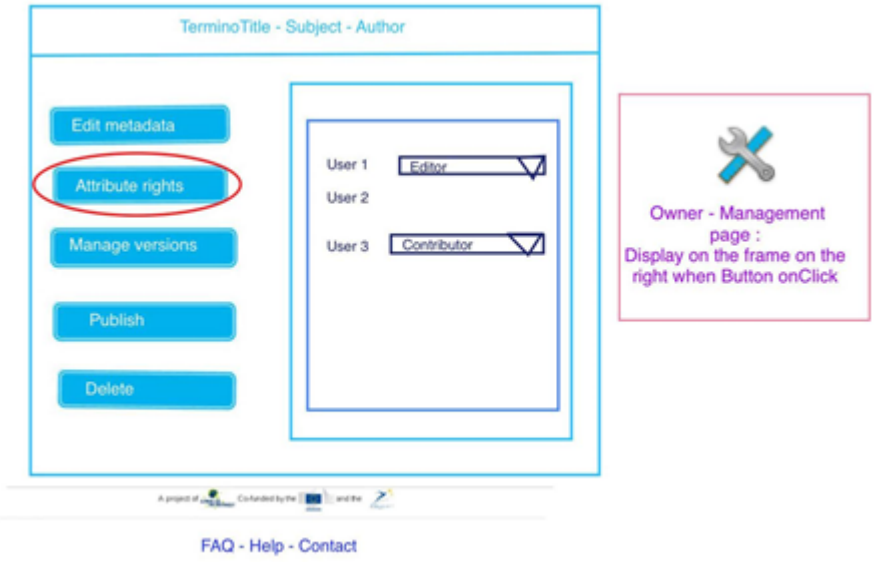


Figure 38: Right management

Manage versions

This interface list all versions of a terminology. The available actions on versions are:

- Change number: For convenience only, a version number can be modified.
- Create version: fix the current terminology with a version number and duplicate it as a new “current” terminology.
- Set current: choosing an old version and set as current, so it can be edited. This create a new branch.
- Delete: an obsolete version can be deleted

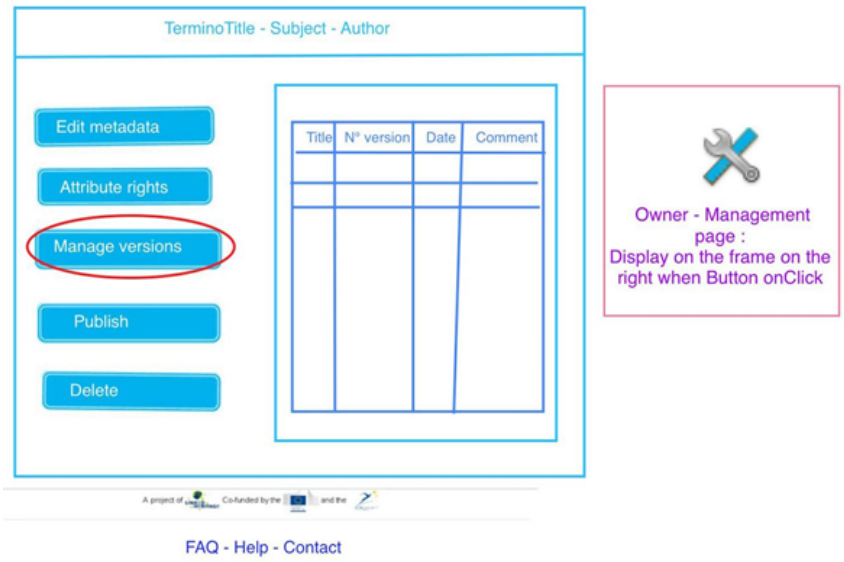


Figure 39: Versioning management

Publish

By clicking on the “publish” button, the current version of the terminology becomes public. This means that any user can freely browse and search the terminologies.

Delete

Delete all versions of a terminology. Cannot be undone. Require confirmation.

4.10 Enrichment and Editing

The main goal of TMP is to manage the terminology. The terminology editor is thus an important part of the TMP. It provides a lot of functionalities such as:

- Adding concepts
- Naming (labelling) concepts in each language
- Alternative naming concepts (means “other term”)
- Organising concepts into a hierarchy (broader/narrower)
- Linking related concepts
- Defining, documenting, adding notes about concepts

The interface will be divided in two parts.

4.10.1 The Tree View Part

On the left, the tree view shows the concept hierarchy. Managing terminology is done through this part: selecting and searching for a concept, creating new ones, organizing them as well as deleting concepts. This is done with easy symbols created for each action.

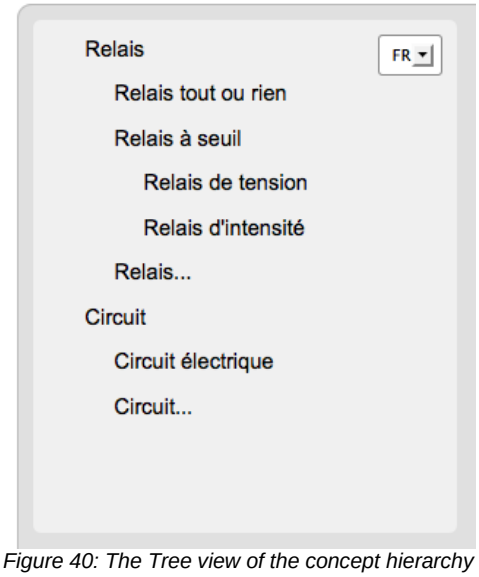


Figure 40: The Tree view of the concept hierarchy

4.10.2 The Linguistic and Conceptual Parts

On the right, a set of fields show all information about the concepts. This pane is also divided in two:

- The top part of the pane is dedicated to the linguistic dimension. Here you can define the preferred label, hidden labels and alternative labels.
- The bottom part of the pane is dedicated to the conceptual dimension, which shows concepts are linked together.

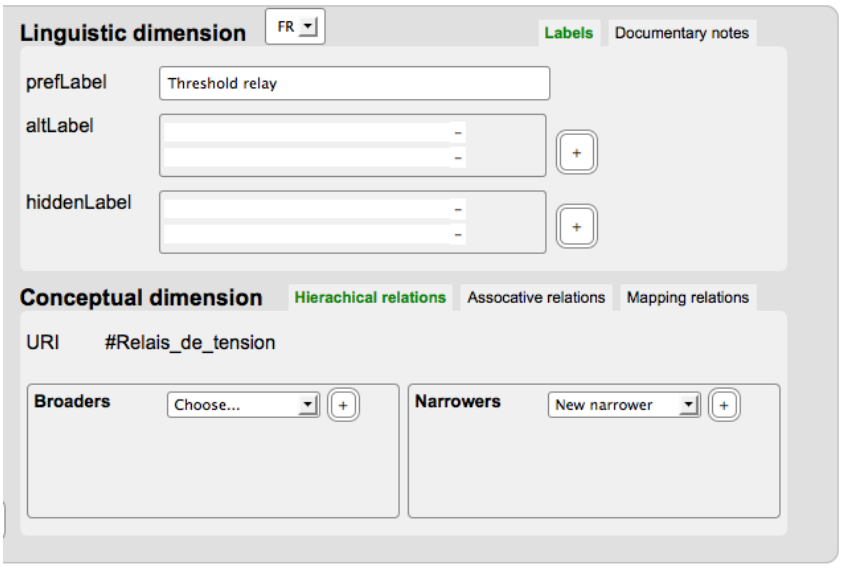


Figure 41: The linguistic and conceptual cards associated to the selected concept

4.10.3 Language Selection

Because the conceptual network (common to every language) and the linguistics networks (one per language) are separated, there are two locations where the user can change the language:

On the left, the blue flat allows to change the language of the tree view, i.e. the language of the preferred labels of the concepts to be displayed.

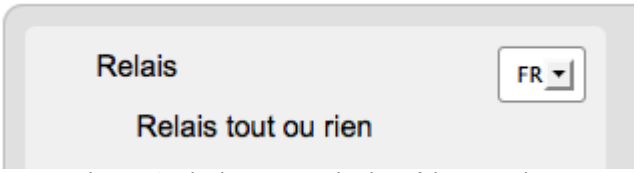


Figure 42: The language selection of the Tree view

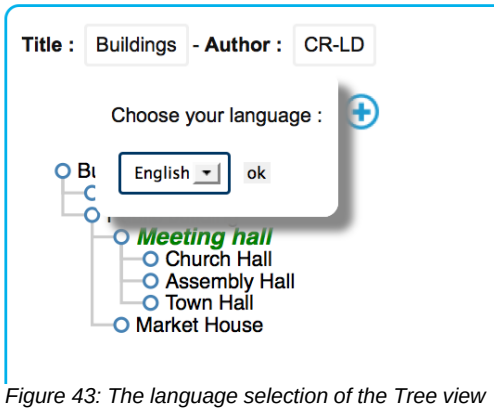


Figure 43: The language selection of the Tree view

On the right, the language field concerns the language in which information about concept are displayed.

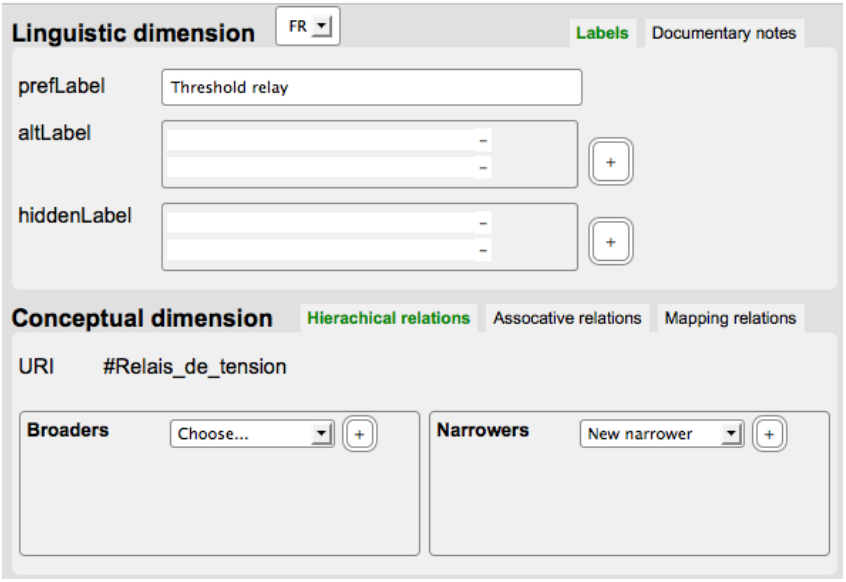


Figure 44: The language selection of the Linguistic and Conceptual cards

These two language selectors are independent. The user can navigate the tree view in his own language, for convenience and good locating in the tree, and he can modify information in another language. It is especially useful for translation.

4.10.4 Creating a New Concept

Creating a new concept is done in the Tree view Part using one of the “plus” button. This action creates a new and unnamed concept, which then can further be defined.

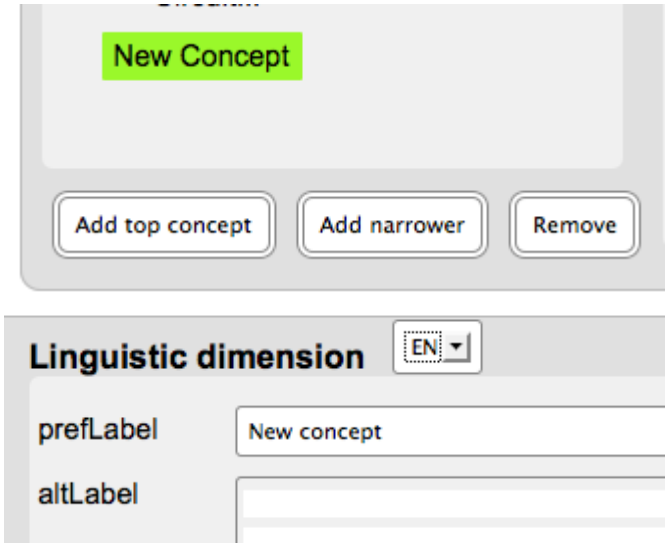


Figure 45: Creation of a new and unnamed concept

4.10.5 Naming A Concept

The unnamed concept must first be given a name, by using the linguistic part editor. The language selection must be the same for both the Tree View Part and the Linguistic Part.

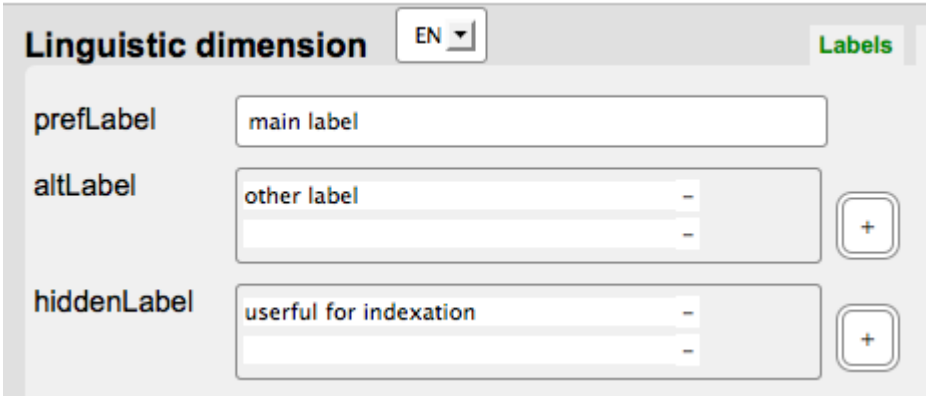


Figure 46: Naming a new concept

4.10.6 Linking concepts: Defining a hierarchy

Defining the place in the hierarchy consists in linking concepts through the broader (super concept) or narrower (sub concept) relationships. This can be done either using the Tree View Part (*drag and drop*) or using the Conceptual Part Editor (has broader, has narrower). The Tree View Part and the Conceptual Part Editor are two different means to access and modify the ontology. These two parts display the same information: if the ontology is modified by one view, the other view is automatically updated.

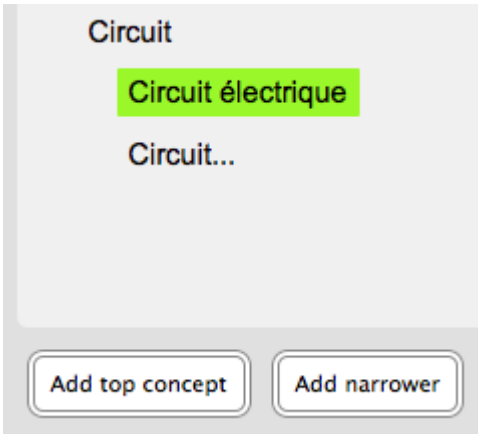


Figure 47: Defining an hierarchy by clicking on the "Add narrower button"

Logical properties of the hierarchy of concepts:

The hierarchical relationship is a *strict order* relation, i.e. a binary relation which is:

- irreflexive: this means that a concept cannot be directly linked with itself. As a matter of fact, a concept cannot be more generic (broader) neither specific (narrower) than itself;
- asymmetric. this means that if a concept A is more general (*versus* specific) than a concept B, the concept B cannot be more general (*versus* specific) than the concept A.

Therefore there is no cycle in a hierarchy structure.

The TMP will propose only the possible concepts which respect the properties for defining new broader or narrower concepts.

4.10.7 Linking concepts: Defining a multiple hierarchy

It is possible to define a multiple hierarchy in TMP because one concept can have several broader concepts (polyhierarchy). The same sub concept can be displayed as many time as necessary.

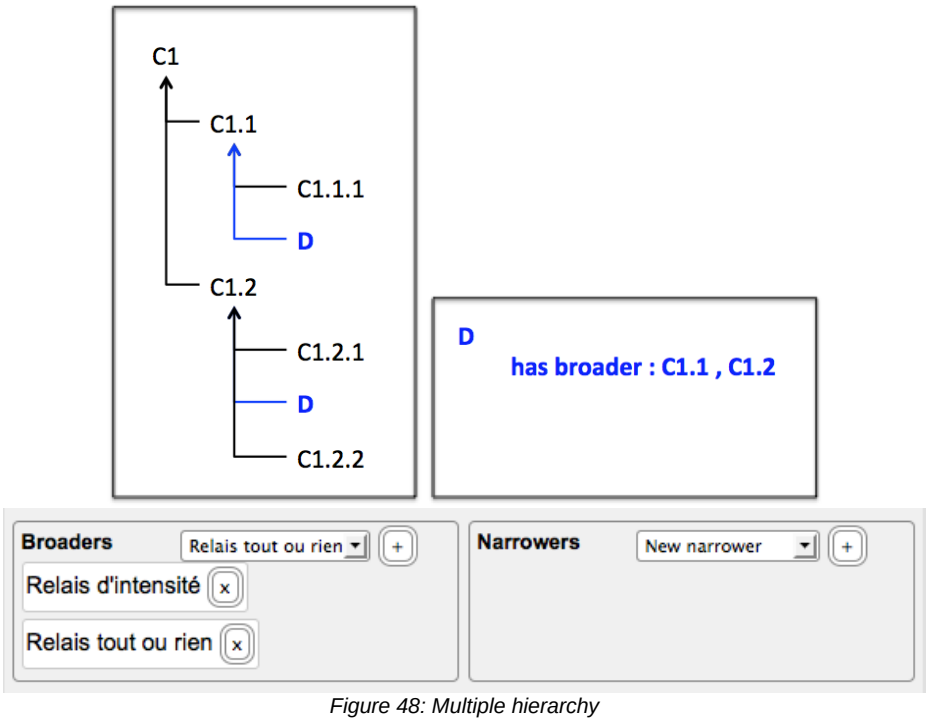


Figure 48: Multiple hierarchy

4.10.8 Linking concepts: Defining associative relationships

The related relation is, in SKOS, an associative symmetric relation between two concepts that indicate they are linked, but that one is not in any way more general than the other.

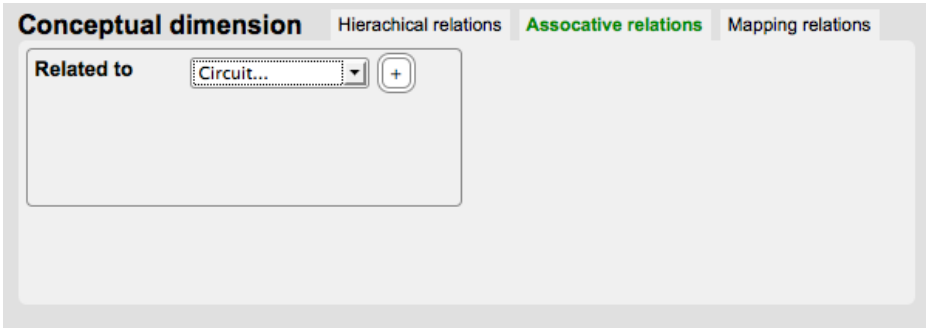


Figure 49: Associative relationships

4.10.9 Linguistics editing

Linguistics concern all information about the concept like labels and notes. This information is displayed on the top-right of the tool, with two tabs:

1 - Labels:

- preferred label: the main expression for a concept, is unique
- alternative label: other expressions for a concept (synonyms, acronyms, used for…)
- hidden label: expressions used for indexing, but which the user doesn't want to promote (e.g. regular misspellings)

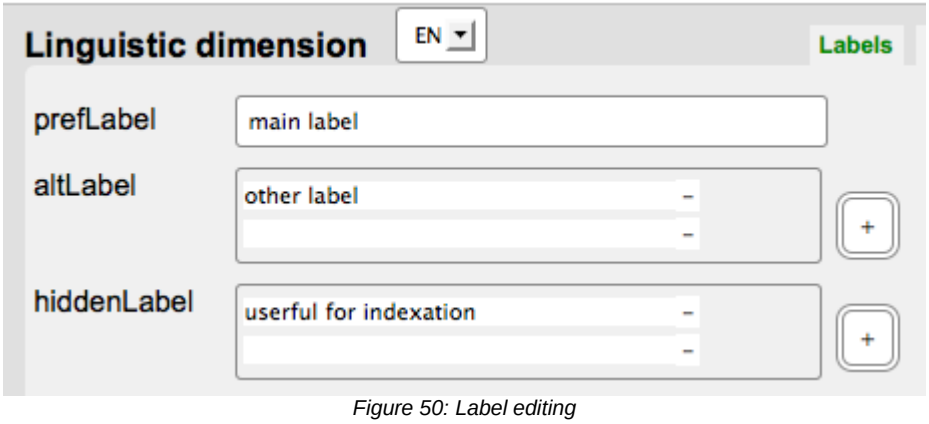


Figure 50: Label editing

2 - Notes:

- Scope notes supply some, possibly partial, information about the intended meaning of a concept, especially as an indication of how the use of a concept is limited in indexing practice.
- Definitions supply a complete explanation of the intended meaning of a concept.
- Examples supply a user case of a concept.
- History notes describe significant changes to the meaning or the form of a concept.
- Editorial notes supply information that is an aid to administrative housekeeping.
- Change notes document fine-grained changes to a concept, for the purposes of administration and maintenance.



Figure 51: Note editing

4.10.10 Linguistics editing: other language

Linguistic information is available in every language by selecting the language in the list (top).

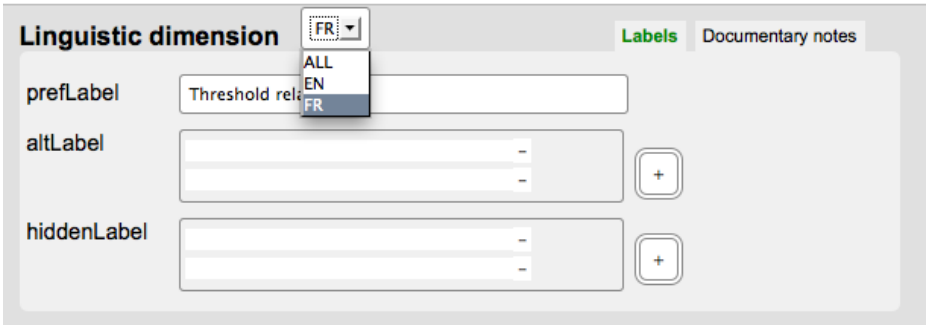


Figure 52: Language selection

4.11 Search and Navigate

4.11.1 Simple Search

The general user (not logged in) or a registered user can browse the public terminologies.

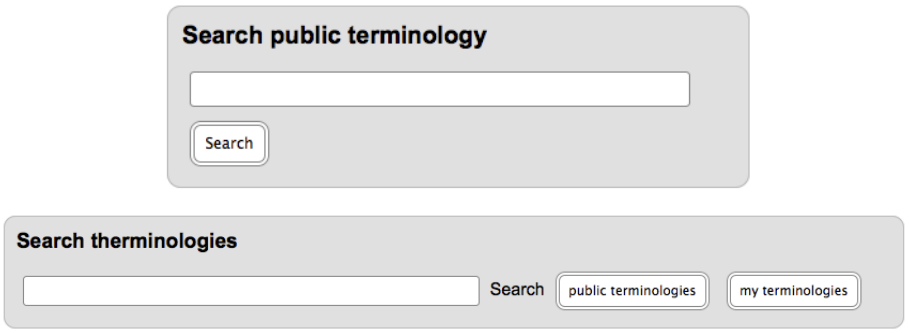


Figure 53: Search forms

Search fields can be title, subject and author.

The results of the search query are displayed in a similar table than the initial terminologies list.

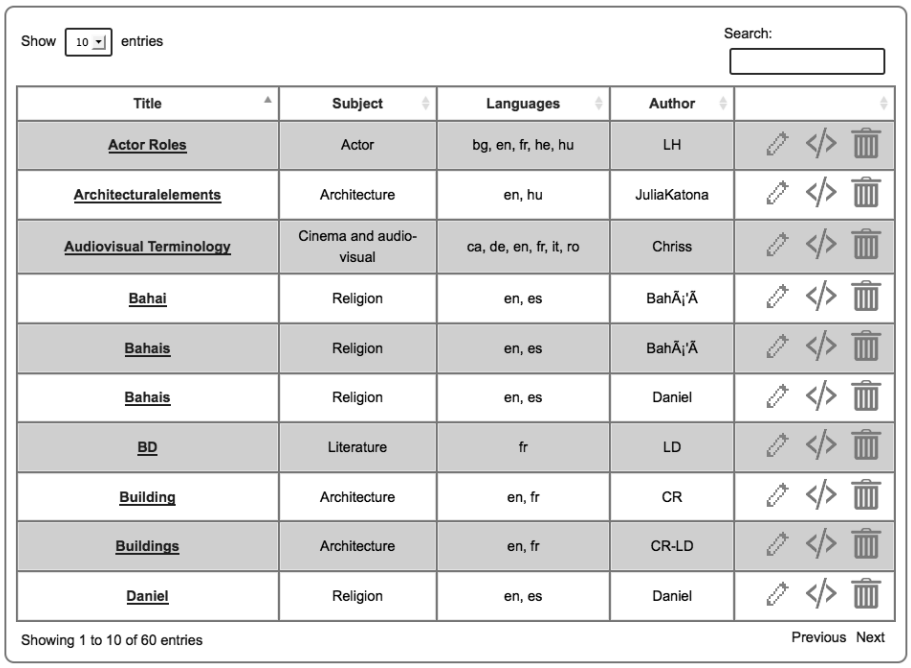


Figure 54: Results

The search field at the top of the result allows to dynamically filter with keywords.

4.11.2 Advanced Search

For more specific search queries, an appropriate form is available:



Figure 55: Advanced search

Each line of the form, added with the "+" button, is a constraint in the search (conjunction). For each constraint, the user can choose the SKOS field on which the search is performed, and the inclusion (contains) or exclusion (does not contain) of the specified expression. The string matching is case insensitive.

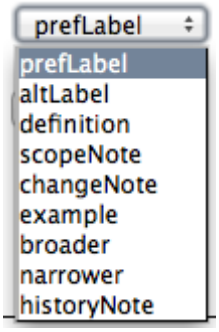


Figure 55: SKOS vocabulary

The results are displayed in the same way as in the classical search query.

4.12 Mapping

4.12.1 Principles

The TMP terminology mapping relies first on a linguistic approach applied to the concept names. This is based on the principle that the more concept names are similar, the more concepts are equivalent.

The similarity between concepts corresponds to a distance between the concept labels. The distance is all the more smaller as the labels are similar.

There are different measures of similarity between names (string of characters). One of the most popular string metric is the *Levenshtein distance* defined as the minimum number of necessary single-character edits to change one word into the other (insert, delete, substitute). For example, the distance between “examinetion” and “examination” is one since only one substitution is required (“e” for “a”). For longer strings (e.g. compound words), there are other measures like the Jaccard distance.

The TMP allows a semi-automatic mapping. It means that for each source concept (concept of the first terminology) the TMP proposes an ordered list of possible target concepts (concepts of the second terminology). The target concepts are ordered according to the TMP string metric.

Distance:

The TMP distance is based on a Levenshtein distance applied to all labels, either preferred labels or alternative labels, for all common languages. The minimum of all the calculated distances will be kept as the final distance. If there is no common language between the two terminologies, English will be chosen as pivot language (intermediary language). The labels of both terminologies will be translated into the pivot language.

Logical properties:

The TMP terminology matching relies also on a semantic approach[10] since it respects the logical properties of the hierarchical relationship of the two ontologies to be matched. It means that, for example, if a concept A exactly matches (*skos:exactMatch)* with a concept B, the narrower concepts of A can be aligned only with the narrower concepts of B.

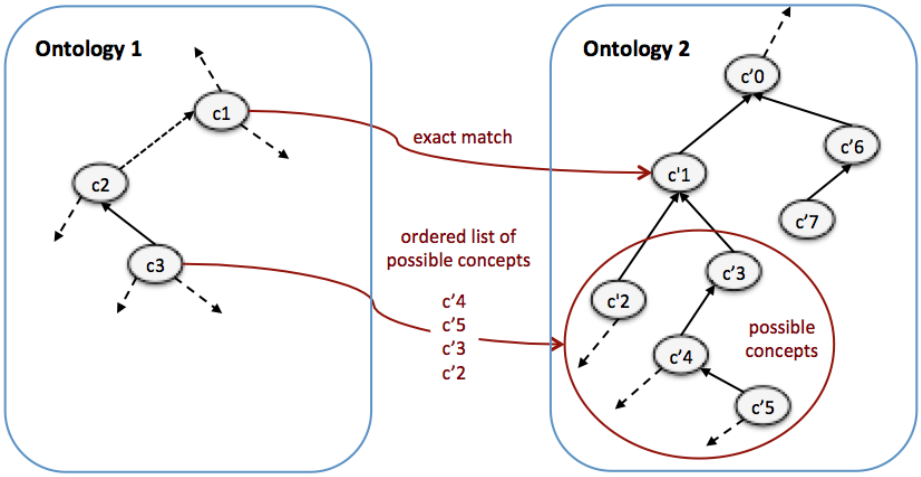


Figure 56: Terminology mapping

[10] An intensional approach based on the hierarchical structure of the concept system.

Use Case

The editor/owner of an institution wants to map his terminology with another one.

The TMP proposes on the left part only the terminologies belonging to the institution and proposes on the right side other terminologies from the registry and external terminologies.

A filter field is displayed to help concept search.

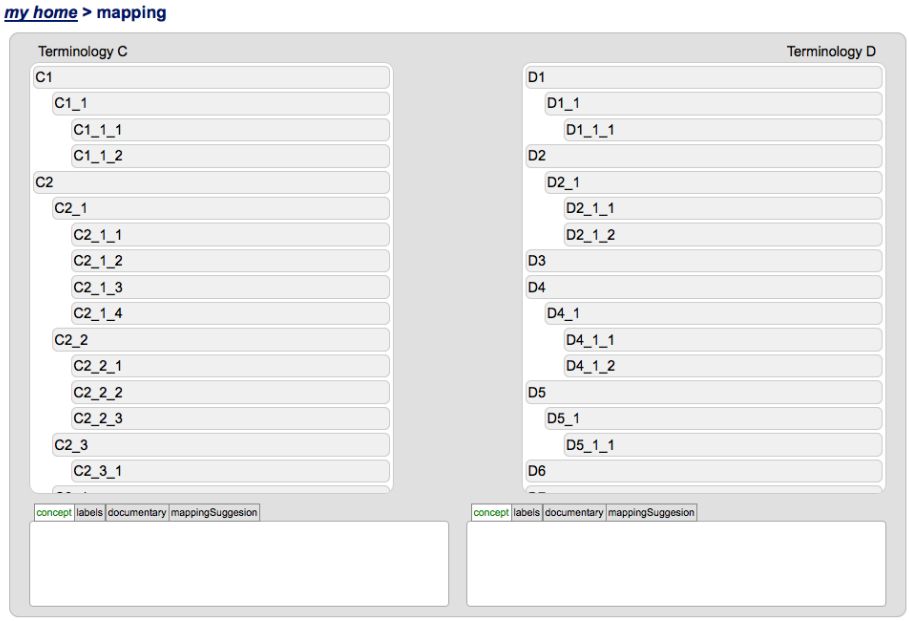


Figure 57: The Mapping view

Choosing a mapping relation:

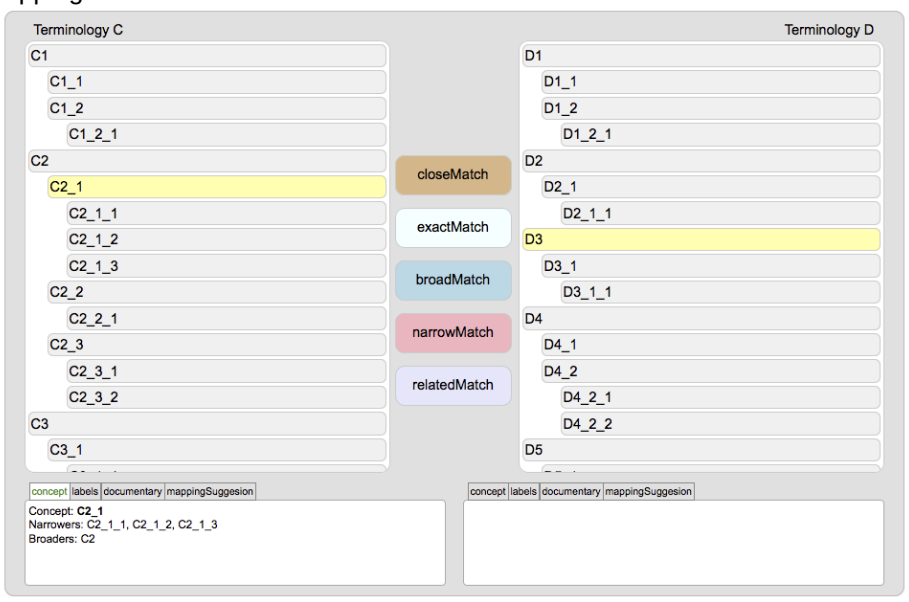


Figure 58: The SKOS mapping relationships

The mapping relationships are those of SKOS:

- *Close match*: This property is used to link two concepts that are sufficiently similar that they can be used interchangeably in **some** information retrieval applications. In order to avoid the possibility of "compound errors" when combining mappings across more than two concept schemes, skos:closeMatch is **not** declared to be a transitive property.

- *Exact match*: This property is used to link two concepts, indicating a high degree of confidence that the concepts can be used interchangeably across a wide range of information retrieval applications.

- *Broad match*: This property is used to state a hierarchical mapping link between two concepts: a super concept.

- *Narrow match*: This property is used to state a hierarchical mapping link between two concepts: a sub concept

- *Related match*: This property is used to state an associative mapping link between two concepts.

4.12.2 Forbidden mapping

The TMP is a collaborative space where everyone can interact by proposing moderation.

In all cases, the owner will get a notification of the mapping suggestion made. Then, he can validate or reject the suggestion from the list of suggestions.
The owner of the terminology can access to a single page from his homepage where all the mapping and enrichment suggestions are listed. Notifications are also displayed on all pages of the TMP.

The data displayed in a notification are:

- Terminology name (link)
- User who did the suggestion, with its institution and email
- Type of suggestion (editing or mapping)
- Editing case:
    - Concept (link)
    - Involved field
    - Value suggested
- Mapping case:
    - Terminology linked (link)
    - Concept mapped (link)
    - Type of mapping (see mapping relations in the mapping section)

4.12.3 Terminology Editing

A contributor viewing the terminology from another institution can make a suggestion for the enrichment of the terminology.

A specific form for the suggestion is available, where the contributor can add or modify details on a concept. Suggestion for modification of the hierarchy can be done via the form (no drag and drop feature)

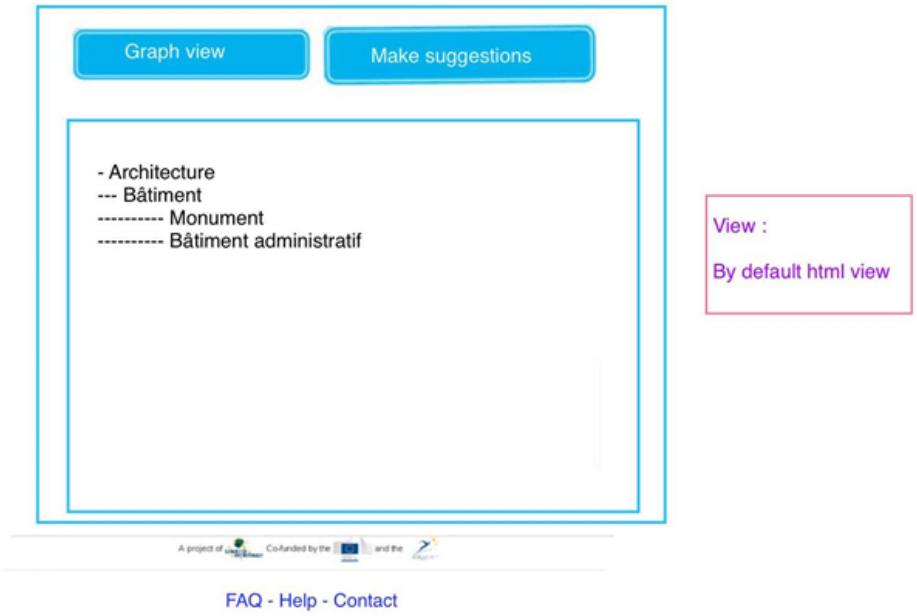


Figure 59: Collaborative view



Figure 60: Collaborative view

4.12.4 Mapping

Every user can open two non-propriety terminologies for mapping and making concept links. These connections as considered to be *suggestions of mapping*.

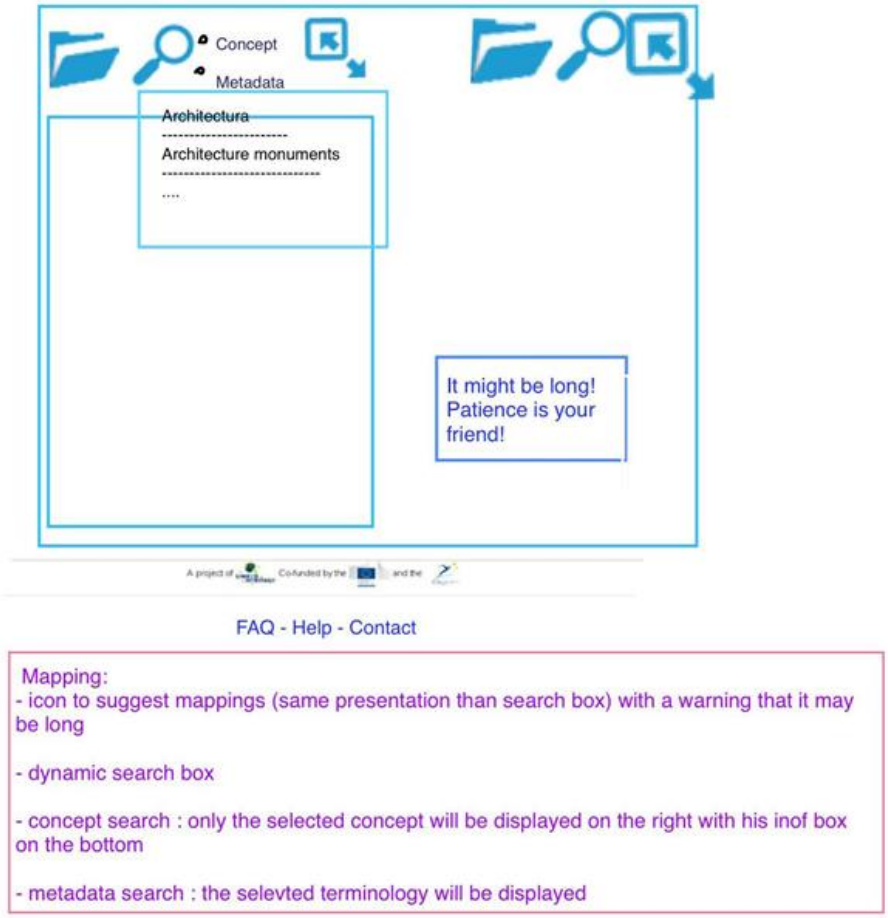


Figure 61: Collaborative view

4.12.5 Forum / chat

The owner and other users/contributors can communicate via the forum.

The forum is accessible as a link on the top of the interface.

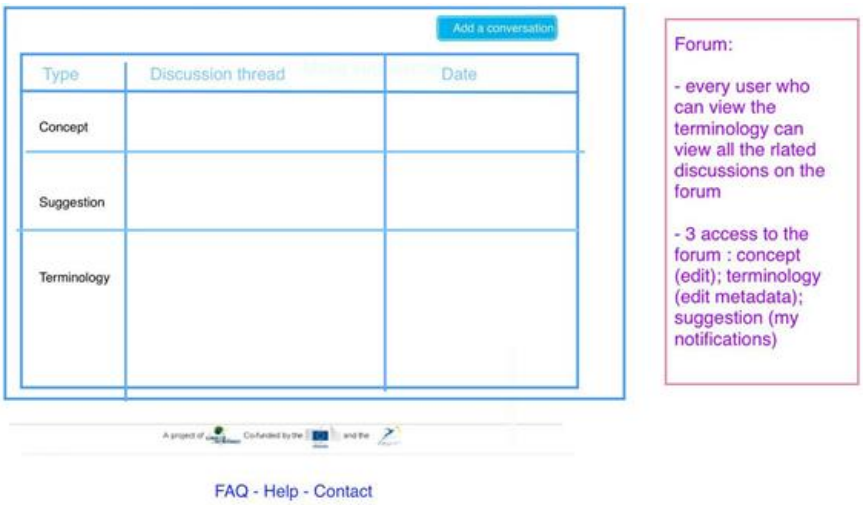


Figure 62: The Forum view

4.13 Publication

The owner is the one who can publish the terminology. The terminology is considered private until it is published.

A publish button is available to the owner from the list of terminologies:

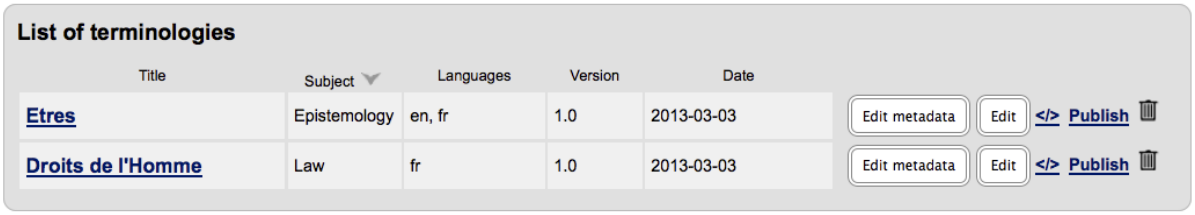


Figure 63: Publication

All published terminologies are available at the root of the application, for all users, including not connected ones.

4.14 Export

At every moment, a terminology can be exported in SKOS/RDF format, by clicking on the tag icon .



The generated file can be opened in classical XML editor (text editors, eclipse, oXyen, …) and can be reused in specific applications.

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#" >
  <rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/7568f0e4-abba-48f0-bb8f-3e664b6531f3">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings"/>
    <skos:prefLabel xml:lang="fr">Change me</skos:prefLabel>
    <skos:prefLabel xml:lang="en">Meeting hall</skos:prefLabel>
    <skos:narrower rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/6a136987-b71b-4fb5-92ec-3dede0a72b9b"/>
    <skos:narrower rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/f6365f34-eaa6-4b42-8d25-ef9b36a79297"/>
    <skos:narrower rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/6f972193-fd9b-4140-b8ed-6ffcf4dfd5a5"/>
    <skos:broader rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/4ddae43a-3c84-4858-ae05-25dc6a24cf2d"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/6f972193-fd9b-4140-b8ed-6ffcf4dfd5a5">
    <skos:broader rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/7568f0e4-abba-48f0-bb8f-3e664b6531f3"/>
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings"/>
    <skos:prefLabel xml:lang="fr">Change me</skos:prefLabel>
    <skos:prefLabel xml:lang="en">Church Hall</skos:prefLabel>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/b48fb080-d8ca-4dee-8437-310c4e807fa3">
    <skos:broader rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/9bcd7dd0-76a1-470c-bf42-a479d605cd81"/>
    <skos:related rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings/6f972193-fd9b-4140-b8ed-6ffcf4dfd5a5"/>
    <skos:prefLabel xml:lang="en">Religious activity</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Change me</skos:prefLabel>
    <skos:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/CR-LD/Buildings"/>
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  </rdf:Description>

  ...

</rdf:RDF>
```

Figure 64: Export

4.15 Static Pages

A set of links is available at the bottom of the application.

4.15.1 Credits

The TMP is developed within the framework of AthenaPlus, Linked Heritage, Europeana, co-funded by European comission and Competitiveness and Innovation Framework Programme (CIP):

- www.linkedheritage.eu
- www.europeana.eu
- http://ec.europa.eu/index_en.htm
- http://ec.europa.eu/cip/



4.15.2 About the TMP

This section gives information on the history and development of the TMP with links and logos, e.g.:

The TMP is developed by Université de Savoie (France)



4.15.3 FAQ

A classical list of questions with answers.

4.15.4 Help !

A link to the user guide available as a PDF file and to the online helpdesk (see D4.4).

5 New features in AthenaPlus

A prototype version of the TMP was developed in the Linked Heritage project[11]. The development of the prototype allowed us to identify some important principles for thesauri and terminologies, as well as the main functionalities of a terminology management system. Functionalities such as skosification of a thesaurus were tested, the interest in SKOS as a pivot language was discussed etc.. The Linked Heritage prototype relied on a Terminology Registry developed by IST[12] and on XTree, a thesaurus-editing software developed by DigiCult.

The TMP was extended in AthenaPlus with new features both from a theoretical point of view as of a practical point of view.

From a theoretical point of view, the TMP unifies principles coming from different ISO standards in Terminology (ISO 1087-1 and ISO 704) and in Thesaurus (ISO 25964-1 and ISO 25964-2) focusing on the notion of "concept". The principles of ontology[13] have become the backbone of the TMP allowing a more simple approach for multilingual terminologies. The logical properties of the relationships between concepts had to be defined, since they are not specified in the SKOS pivot languages. And at last a semi-mapping functionality has been specified based on linguistic methods (Levenshtein distance applied to labels) and semantic methods (relationships between concepts).

From a practical and implementation point of view, the TMP included in AthenaPlus:

- A new Terminology Registry based on a RDF Data Base including versioning in order to keep track of different versions of terminologies;
- A User Management System with specific actions reserved to user roles (including a new kind of users: administrator), login service with password and username, metadata scheme for new users and organizations;
- An Editing tool based on a distinction between the linguistic and the conceptual dimensions taking into account the logical properties of the hierarchical relationships between concepts (for example the notion of "strict order" relationship);
- New functionalities including:
    - Creating/importing terminologies: metadata scheme for creating/import, extension of the skossification table;
    - Exporting: new feature (exporting in SKOS/RDF);
    - Semi-automatic mapping features (under development);
- New interfaces: adaptation of icons, etc.

In AthenaPlus the University of Savoie will continue to work on the TMP theoretically as well as practically. Existing functionalities will be improved and finalized, and feedback from users will be considered in the developing phase. A second release of the TMP is planned for January 2015 with deliverable D4.5. It will explain improvements of technical and functional specifications in this last version of the TMP.

[11] D3.2: *Functional and Technical specification of the Terminology Management Platform*, Linked Heritage 2012 and D3.3: *Terminology management and terminology registry (Demonstrator)*, Linked Heritage 2012.
[12] Instituto Superior Técnico, Lissabon, Portugal, LH partner
[13] Defined as a specification of a conceptualisation

6 Conclusion

This deliverable, the third one of the WP 4, describes the specifications of the first release of a production version of the TMP (Terminology Management Platform) a system for the management of multilingual terminologies.

It describes the functional and technical specifications of TMP, taking into account the returns on experience of a prototype version carried out during the Linked Heritage Project.

The conclusion of the deliverable can be explained theoretically and technically (specifications):

The theoretical results are:

the conceptual approach of the TMP

Even if a terminology is not a thesaurus (the main goal of terminology is not to index documents) and a thesaurus is not a terminology (the main goal of thesaurus is not to define terms in relation to the domain conceptualisation), both of them emphasize the role of concept (in their latest versions of ISO Standards). The properties of these systems directly depend on the properties of their conceptual system. It is the reason why the TMP is ontology-oriented with a clear separation between the conceptual dimension (the ontology) and the different linguistic dimensions (one per language).

the internal representation language for terminologies

The SKOS interchangeable language is the pivot language of the TMP. It means that terminologies are represented, uploaded, edited, stored, exported and mapped under SKOS. This choice has been justified because SKOS is more and more used for sharing thesauri like in Europeana.

terminology mapping

The TMP provides a semi-automatic mapping based on linguistic and semantic methods. The linguistic method is based the principle that the more the concept names are similar, the more the concepts are equivalent. The TMP distance between concept names is a based on a Levenshtein distance applied to all labels, either preferred labels or alternative labels, for every common language of the two terminologies. The semantic methods respect the ontology structure in order to preserve the hierarchy (if a concept A matches with a concept B, the narrower concepts of A can be only aligned with the narrower concepts of B.

The technical results or specifications are:

the architecture of the TMP

The TMP architecture is split into three components. The Graphical User Interface (GUI) is the launcher of the TMP functionalities including the authentication service. The Registry is in charge of storage and versioning of data (terminologies content, mapping information…). The last component, the Application server, gathers the TMP functionalities: editing, mapping, skosification, import/export etc.

interface specification

The TMP interface follows the TMP principles. A clear separation is done between the conceptual dimension and the linguistic dimension. The Tree View applied to the conceptual structure is the only means to access to the elements of the terminology.

the specification of the TMP functionalities

The functionalities are specified and described through use cases: Authentication, User management, Edition of terminology, Import/Export, Skosification, Mapping, Collaborative moderation.

6.1 Suggestions and Next steps

Improvements of the TMP functionalities

The second and last release of the TMP is scheduled in month 23. This time will be dedicated to the improvements of the TMP functionalities. To this end, a close collaboration will be set up between the technical partner and the pilot group of Content Providers including physical meetings, Skype meetings and communication by distribution mailing lists.

Some work will also be done on other ISO Standards in relation to the TMP principles like the ISO 860 (Terminology work - Harmonization of concepts and terms) and the ISO 22274 (Systems to manage terminology, knowledge and content — Concept- related aspects for developing and internationalizing classification systems).

Training and Diffusion

Some work remains to be done so the TMP principles can be explained, the ontology-oriented approach, the problems raised by the unification of terminology and thesaurus, the role of ISO and W3C standards, their differences and similarities, and the role of interchange and pivot format in relation with the TMP principles.

This will be achieved through training material, articles and workshops.

7 APPENDIX 1: REFERENCES

H. Al-Feel, R. Schafermeier, A. Paschke (2013)
"An Inter-lingual Reference Approach For Multi-Lingual Ontology Matching"
IJCSI, 2013.

F. Baader, D. Calvanese, D. McGuinness, D. Nardi and P. Patel-Schneider (2003)
"The Description Logic Handbook"
Cambridge University Press, 2003.

Z. Bellahsene, A. Bonifati, E. Rahm (eds.) (2011)
"Schema Matching and Mapping"
Springer, 2011.

Berners-Lee T., Hendler J. and Lassila O. (2001)
"The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities"
Scientific American Magazine, May 17.

Brachman R.J., Levesque H.J. (1985)
"Readings in Knowledge Representation"
Morgan Kaufmann Publishers, Inc. 1985.

Budin, G. (2001)
"A critical evaluation of the state-of-the-art of Terminology Theory"
ITTF Journal, 12. Vienna. TermNet (2001)

Cabré T. (2003)
"Theories in terminology"
Terminology 9:2. pp. 163-199 (2003)

Daille B., Kageura K., Nakagawa H. and Chien L.F. (2004)
"Recent Trends in Computational Terminology"
Special issue of Terminology 10:1 (Benjamins publishing company).

S. Dalbin, N. Yakovleff, H. Zysman (2013)
"Livre Blanc. ISO 25964-1 – Thésaurus pour la recherche documentaire".
AFNOR, Edition Janvier 2013.

Dean M. and Schreiber G. (2004)
"OWL Web Ontology Language Reference"
W3C Recommendation, http://www.w3.org/TR/owl-ref/.

Euzenat, Jérôme, Shvaiko, Pavel (2013)
"Ontology Matching"
Springer-Verlag, 2013

J. Euzenat, C. Meilicke, H. Stuckenschmidt, P. Shvaiko, C. Trojahn (2011)
"Ontology Alignment Evaluation Initiative: six years of experience"
Journal on Data Semantics, 2011.

Felber, H. (1984)
"Terminology Manual"
Unesco (United Nations Educational Scientific and Cultural Organization) – Infoterm (International Information Centre for Terminology) (1984)

Genesereth M.R. and Fikes R.E. (1992)
"Knowledge Interchange Format Version 3.0, Reference Manual"

Report Logic 92-1, Computer Science Department, Stanford University, June 1992

Gomez-Perez, A., Corcho. O., Fernandez-Lopez, M. (2004)
"Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web".
Springer (2004)

M. Granitzer, V. Sabol, K. Weng Onn, D. Lukose, K. Tochtermann (2010)
"Ontology Alignment - A Survey with Focus on Visually Supported Semi-Automatic Techniques"
Future Internet, 2010.

Gruber, T. (1993)
"A Translation Approach to Portable Ontology Specifications".
Knowledge Systems Laboratory September 1992. Technical Report KSL 92-71. Revised April 1993. Appeared in Knowledge Acquisition, 5(2):199-220 (1993)

Guarino N., Carrara M. and Giaretta P. (1994)
"An Ontology of Meta-Level Categories of Knowledge Representation and Reasoning"
Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR94) (Morgan Kaufmann).

ISO
International Organization for Standardization is the world's largest developer of voluntary International Standards.
http://www.iso.org

ISO 704 (2009), Terminology work -- Principles and methods.

ISO 1087-1 (2000), Terminology work -- Vocabulary -- Part 1: Theory and application.

ISO 25964-1 (2011) Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval

Y. Kalfoglou, M. Schorlemmer (2003)
"Ontology Mapping: The State of the Art"
The Knowledge Engineering Review Journal, 2003.

Kiryakov, Popov, Terziev, Manov and Ognyanoff (2005)
"Semantic Annotation, Indexing, and Retrieval"
Elsevier's Journal of Web Sematics, 2 (1).

Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2008)
"Terminological Principles Used for Ontologies."
Managing ontologies and lexical resources. TKE 2008. Copenhagen: ISV. (2008)

Linked Heritage (LH)
The main goals of the Linked Heritage project were: 1) to contribute large quantities of new content to Europeana, from both the public and private sectors; 2) to demonstrate enhancement of quality of content, in terms of metadata richness, re-use potential and uniqueness; 3) to demonstrate enable improved search, retrieval and use of Europeana content.
http://www.linkedheritage.org/

Philipp Mayr, Vivien Petras (2008)
Cross-concordances: terminology mapping and its effectiveness for information retrieval, http://arxiv.org/abs/0806.3765

Ontology matching
This web site provides a repository of information devoted to different aspects of ontology matching. Ontology matching is a promising solution to the semantic heterogeneity problem. It finds correspondences between semantically related entities of the ontologies.

http://www.ontologymatching.org/

OWL (2009)
"OWL 2 Web Ontology Language. Document Overview".
W3C Recommendation 27 October 2009. http://www.w3.org/TR/2009/REC-owl2-overview-20091027/.

Pavel Shvaiko Jérôme Euzenat (2005)
"Tutorial onSchema and Ontology Matching"
ESWC'05 – 29.05.2005

PubMed Tutorial.
PubMed uses an Automatic Term Mapping feature to search for unqualified terms http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_040.html

E. Rahm, P. Bernstein (2001)
"A Survey of Approaches to Automatic Schema Matching"
The VLDB Journal, 2001.

RDF Schema (2004)
"Resource Description Framework"
W3C Recommendation, http://www.w3.org/TR/rdf-schema/ February 2004

Roche C. (2001)
"The *specific-difference* principle: a methodology for building consensual and coherent ontologies »
IC-AI 2001, Las Vegas USA, June 25-28 2001

Roche C. (2005)
"Terminologie et ontologie"
Revue Langages, 157, mars 2005, pp. 48-62 (Éditions Larousse).

Roche, C. (2007)
"Le terme et le concept : fondements d'une ontoterminologie".
TOTh 2007. Terminologie & Ontologie : Theories et applications. pp. 1-22, Annecy. France. 1er juin (2007)

Roche C., Calberg-Challot M., Damas L., Rouard P. "Ontoterminology: A new paradigm for terminology", KEOD 2009 (International Conference on Knowledge Engineering and Ontology Development), 5-8 October, Madeira (Portugal)

Roche Christophe (2012)
"Should Terminology Principles be re-examined"
TKE 2012 (Terminology and Knowledge Engineering), Madrid, 20-21 June 2012

Sager, J. (1990)
"A Practical Course in Terminology Processing"
John Benjamins Publishing Company (1990)

Sowa J. (2000)
"Knowledge Representation"
Brooks/Cole

Staab, S., Studer, R. (2004)
"Handbook on Ontologies"
Steffen Staab (Editor), Rudi Studer (Editor). Springer (2004)

Temmerman R. (2000)
"Towards New Ways of Terminology Description. The sociocognitive approach"
Amsterdam/Philadelphia : John Benjamins.

Ushold, M., Gruninger, M. (1996)

“Ontologies: Principles, Methods and Applications”
Knowledge Engineering Review, Vol. 11, n° 2, June 1996. Also available from AIAI as AIAI-TR-191 (1996)

W3C.
The World Wide Web Consortium (W3C) is an international community where Member organizations, a full-time staff, and the public work together to develop Web standards. http://www.w3.org/

WordNet
WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations.
http://wordnet.princeton.edu/

Wright, S.E., Budin, G. (1997)
“Handbook of Terminology Management”
volume 1 and 2. John Benjamins Publishing Company (1997)

Zeng, M. L., & Chan, L. M. (2004)
“Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems”
Journal of the American Society for Information Science and Technology, 55(3), 377-395.

8 APPENDIX 2: DEFINITION OF TERMS AND ABBREVIATIONS

- **ASYMMETRIC:** see “strict order relation”
- **CONCEPT**: unit of knowledge created by a unique combination of characteristics (ISO 1087-1)
- **CSV**: Comma Separated Value
- **IRREFLEXIVE**: see “strict order relation”
- **ISO 1087-1**: Terminology work — Vocabulary — Part 1: Theory and application
- **ISO 704**: Terminology work — Principles and methods
- **ISO 25964-1**: Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval
- **KIND OF**: generic relationship between two concepts
- **KMKG**: Koninklijke Musea voor Kunst en Geschiedenis (Belgium)
- **LEMMATIZATION**: process of determining the canonical form (dictionary form) of a word
- **LEVENSHTEIN**: string metric between two words defined as the minimum number of necessary single-character edits to change one word into the other
- **LH**: Linked Heritage
- **MCC**: Ministry of Culture and Communication (France)
- **ONTOLOGY**: specification of a conceptualisation - shared description of concepts of a domain with their relationships expressed in a formal and computer readable language
- **ONTOTERMINOLOGY**: terminology whose conceptual system is a formal ontology
- **PART-OF**: partitive relationships - relation between two concepts where one of the concepts constitutes the whole and the other concept a part of that whole
- **RDF**: Resource Description Framework
- **SKOS**: Simple Knowledge Organization System
- **SKOSIFICATION**: process of translation a terminology into SKOS
- **STRICT ORDER RELATION**: binary relationship between two concepts wich is irreflexive (a concept cannot be in relation with itself), asymmetric (if a concept C1 is more specific than a concept C2 then C2 cannot be more specific than C1) and transitive if C1 is more specific than C2 and C2 is more specific than C3 then C1 is more specific than C3)
- **TERM**: verbal designation of a general concept in a specific subject field (ISO 1087-1)
- **TMF**: ISO 16642:2003: Computer applications in terminology - Terminological markup framework
- **TMP**: Terminology Management Platform
- **TRANSITIVE**: see “strict order relation”
- **UNISAV**: University of Savoie (France)
- **WordNet**: lexical database for the English language.

- **W3C**: World Wide Web Consortium
- **XML**: Extensible Markup Language