

## Project

|                                |  |
|--------------------------------|--|
| <b>Project Acronym:</b>        | AthenaPlus   |
| <b>Grant Agreement number:</b> | 325098   |
| <b>Project Title:</b>          | Access to cultural heritage networks for Europeana |

## Deliverable

|                                    |  |
|------------------------------------|--|
| <b>Deliverable name:</b>           | <b>Specialist Training material : TMP user manuals and semantic mapping procedures</b> |
| <b>Deliverable number:</b>         | <b>D4.4</b>  |
| <b>Delivery date:</b>              | December 2013  |
| <b>Dissemination level:</b>        | Public   |
| <b>Status</b>                      | Final  |
| <b>Authors (organisation)</b>      | Marie-Véronique Leroi (Ministère de la Culture et de la Communication)                 |
| <b>Contributors (organisation)</b> | Contributor(s) (affiliation)   |
| <b>Reviewers (organisation)</b>    | Maria Teresa Natale (ICCU), Barbara Dierickx (PACKED)                                  |

## Revision History

| Revision | Date    | Author                | Organisation | Description           |
|----------|---------|-----------------------|--------------|-----------------------|
| V0.1     | 2013-10 | Marie-Véronique Leroi | MCC          | 1 <sup>st</sup> draft |
| V0.2     | 2014-01 | Marie-Véronique Leroi | MCC          | 1st version           |
| V0.3     | 2014-01 | Barbara Dierickx      | PACKED vzw   | Peer review           |
| V0.4     | 2014-01 | Maria Teresa Natale   | ICCU         | Peer review           |
| V0.5     | 2014-02 | Marie-Véronique Leroi | MCC          | Final version         |
| V0.6     | 2014-02 | Andrea Tempera        | 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This deliverable contains original unpublished work except where clearly indicated otherwise.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 .....   | 4  |
| 3        | TMP USER MANUAL AND TRAINING MATERIAL .....   | 5  |
| 3.1      | Presentation .....  | 5  |
| 3.2      | TMP user manual .....   | 5  |
| 3.2.1    | <i>Introduction</i> .....   | 5  |
| 3.2.2    | <i>Terminology design and editorial choices</i> .....   | 6  |
| 3.2.3    | <i>Rights and Licenses</i> .....  | 6  |
| 3.2.4    | <i>SKOS</i> .....   | 7  |
| 3.2.5    | <i>SKOS and ISO-25964</i> .....   | 14 |
| 3.2.6    | <i>Users and roles</i> .....  | 15 |
| 3.2.7    | <i>Getting started</i> .....  | 16 |
| 3.2.8    | <i>Connection to the TMP</i> .....  | 18 |
| 3.2.9    | <i>Search terminologies</i> .....   | 21 |
| 3.2.10   | <i>Create new terminologies</i> .....   | 22 |
| 3.2.11   | <i>SKOSification</i> .....  | 25 |
| 3.2.12   | <i>Edit a terminology</i> .....   | 29 |
| 3.2.13   | <i>Exporting a terminology</i> .....  | 31 |
| 3.2.14   | <i>Linking multiple terminologies (mapping)</i> .....   | 32 |
| 3.2.15   | <i>Semantic Web and Linked Data</i> .....   | 33 |
| 3.2.15.1 | <i>Linked Data, Open data and Linked Open Data</i> .....  | 35 |
| 3.2.15.2 | <i>The Semantic Web Stack</i> .....   | 36 |
| 3.3      | Online material .....   | 45 |
| 3.3.1    | <i>Help online</i> .....  | 45 |
| 3.3.2    | <i>FAQ (Frequently Asked Question)</i> .....  | 46 |
| 3.3.3    | <i>Video tutorials</i> .....  | 48 |
| 4        | DISSEMINATION .....   | 50 |
| 5        | TRAINING SESSIONS AND WORKSHOPS .....   | 52 |
| 6        | CONCLUSIONS .....   | 55 |
| 7        | APPENDIX 1: REFERENCES .....  | 56 |
| 8        | APPENDIX 2: DEFINITION OF TERMS AND ABBREVIATIONS .....   | 57 |
| 9        | ANNEXES .....   | 58 |
| 9.1      | Evaluation survey for the TMP .....   | 58 |
| 9.2      | Presentation made for the training session in Berlin, the 8 <sup>th</sup> of October 2013 ..... | 62 |

## **1 EXECUTIVE SUMMARY**

This deliverable presents the material prepared by WP4 of the AthenaPlus Project for cultural institutions managing multilingual terminologies. Cultural institutions from the GLAM sector need to deepen their knowledge and understanding of the Semantic Web principles, the existing standard and supporting technologies. The documentation and the training schedule proposed in this deliverable intends to fill this need.

This deliverable provides a detailed introduction to the Semantic Web with a presentation of each of the technologies that are currently available for it. A focus is given to the SKOS format which is the interchange format expected for the cultural institutions' terminologies.

This theoretical introduction is followed by a detailed user manual on the Terminology Management Platform. Each functionality is described in a detailed way so institutions can use the tool as intuitively as possible.

The last parts of this deliverable present the dissemination material available and the aspects that need to be further elaborated to foster the use of the TMP in the future.

## 2 INTRODUCTION

This deliverables provides technical documentation and training material on Semantic Web principles and technologies, and on the TMP (Terminology Management Platform). The idea of developing a tool such as the TMP started within the Athena project where it appeared that moving towards the Semantic Web would require some efforts from the cultural institutions. The TMP was intended as a toolset which could limit these efforts and facilitate the evolution of institutions' documentary system.

As mentioned in the Description of Work, this deliverable will provide *“Specialist training material: TMP user manuals and semantic mapping procedures. This deliverable, coordinated by MCC, will present documentation and extensive training material for terminology providers, including documentation on how to manage one's terminology and prepare it for the semantic web, manuals to use the TMP for the semantic mappings between multilingual terminologies and a detailed planning for the organisation of the pilot session and training workshops.”*

### 2.1 Background

The Semantic Web, the evolution following the Web of documents, has changed the way to manage the knowledge and information offering improved services for multilingual information retrieval. Private companies have been the first to implement enhanced web services using Semantic Web technologies. Moreover the use of Semantic Web and Linked data technologies has been defined as crucial for the development of Europeana to provide standard and structured but especially to enable multilingual access to Europeana collections.

The Athena project (2008-2011)<sup>1</sup> was one of the first European projects to address the issue of the use of Semantic Web technologies for cultural data and especially for providing structured and semantic museum data to Europeana.

The work package 4 (WP4) of the Athena project made a state of the art of the terminologies in use in European museums. The study<sup>2</sup> showed that most of the institutions were using in-house terminologies built from scratch or adapted from reference terminologies. The cost of the licences for reference terminologies and the need for specific terms were the reasons why specific terminologies were in use rather than standard reference terminologies. The study also showed that most of the terminologies were monolingual. The Athena WP4 made a benchmark of tools and standards that could be used by cultural institutions to move towards the Semantic Web and it appeared that no software solutions were available to tackle all the functional needs that a museum would have to convert its terminology to a format compliant with the Semantic Web and benefit from the results of this conversion. Guidelines and recommendations on how to conceive and manage a terminology as part of the Semantic Web have been established within the activity of the Athena WP4. These guidelines and recommendations were then updated and brought together in a booklet *“Your terminology as part of the Semantic Web: recommendations for design and management”* published within the Linked Heritage project<sup>3</sup>.

The Linked Heritage project (2011-2013) followed the Athena project with a specific work package dedicated to terminology and multilingualism. The WP3 of Linked Heritage built on the work done within the Athena project. The WP then took into account the state of the art and the benchmark on terminologies and tools to specify and develop the Terminology Management Platform, TMP, a complete software solution available online and providing the main functionalities needed by cultural institutions for using Semantic Web technologies in relation to their terminology (and broader collection) management.

The TMP is available at the following URL : <http://www.culture-terminology.org>

---

<sup>1</sup> Athena Project : <http://www.athenaeurope.org>

<sup>2</sup> Athena D4.1 Identification of existing terminology resources in museums : <http://www.athenaeurope.org/getFile.php?id=398>

<sup>3</sup> Linked Heritage : <http://www.linkedheritage.org>.

English edition of the booklet: <http://www.linkedheritage.eu/getFile.php?id=244>

French edition of the booklet: <http://www.linkedheritage.eu/getFile.php?id=355>

## **2.2 Role of this Deliverable in the Project**

This deliverable is about the documentation of the TMP and the material that will be used for the training sessions and workshops dedicated to the TMP and more widely to giving background information on the Semantic Web.

This deliverable is subsequent to the D4.3 which defined the technical and functional specifications required for developing the first release of the TMP.

This deliverable is aiming at professional experts in terminology management who are not familiar with Semantic Web technologies. The TMP user manual provides only material for understanding the purpose of the Semantic Web and its technologies and a detailed description of each functionality.

This deliverable is connected partially to the work package dedicated to dissemination (WP7).

### 3 TMP user manual and Training material

This chapter introduces the core of the deliverable with the different material that have been drawn up for documenting the TMP and for training.

#### 3.1 Presentation

This chapter introduces the TMP user manual. During the training sessions and workshops organised within the Linked Heritage project it appeared that most of the cultural institutions have heard about Semantic Web but do not know exactly what it is and what could be its benefits. This is the reason why each presentation and documentation for the TMP begins with a general introduction to the Semantic Web, its main concepts and its technologies. Therefore a chapter on SKOS and a general introduction on the Semantic Web have been included in the TMP user manual.

The online documentation and video tutorials have also been included in this chapter in order to get an overview of all the material that is available.

#### 3.2 TMP user manual

This chapter presents the manual and user guides for the Terminology Management Platform (TMP). It starts with a general introduction and also some key elements for terminology management such as editorial and conceptual choices for the terminology design and definition of rights and licenses for each terminology before getting into detail for each functionality.

##### 3.2.1 Introduction

The Terminology Management Platform (TMP) has been developed within the Linked Heritage and the AthenaPlus project in order to provide to cultural institutions a set of tools which would allow them to easily publish their terminologies as a part of the Semantic Web in the form of Linked Data. « Terminology » in the context of the TMP refers to any Knowledge Organisation System or controlled vocabulary. A focus is given to thesauri as most of the cultural institutions are using thesauri. Classifications, simple lists of terms or glossaries can also be managed within the TMP. The TMP can help in structuring and documenting terminologies that might need to be enriched.

The final goal of this step is to enhance the multilinguality and the semantic understanding of the metadata available on Europeana. Europeana as a matter of fact strongly recommends the use of the SKOS as a Semantic Web compliant format<sup>4</sup>.

The tool is mainly dedicated to the institutions from the GLAM sector as structuring their data as Semantic Web data is very time-consuming and requires some specific skills that might not be present in every institution. The TMP is open to all contributors - professionals as well as the general interested public, as it proposes some collaborative features. The idea is that cultural institutions contributing metadata to Europeana manage their terminologies using the TMP and these terminologies can be enriched with the end user suggestions. The user role organisation provided for the TMP will allow each owner of a terminology to moderate and manage safely these suggestions.

TMP is a Web service that is freely accessible to anyone who is willing to publish one's terminology or anyone searching for some terminologies that might be of interest for research or cataloguing purposes.

SKOSify, Edit, Link, Publish and Collaborate : these are the main tasks the TMP can help you with.

---

<sup>4</sup> Publishing vocabularies on the Semantic Web using SKOS <http://www.few.vu.nl/~aisaac/talks/Isaac-Talk-SemAst09.pdf>

### 3.2.2 Terminology design and editorial choices

The purpose of the Semantic Web is not to present you more time consuming and technical work. The idea is really to help you structuring your data to make them meaningful so you can exploit them at best.

By terminology we mean any kind of Knowledge Organisation System (KOS). A terminology can be a thesaurus, a flat list of terms, a classification or any other controlled vocabularies.

TMP offers you the toolset to transform your terminology as structured and meaningful data thanks to the SKOS format. This means that even if an institution does not have a thesaurus, or any other structured vocabulary, it can still use the TMP to publish its terminology in SKOS with some preliminary conceptual work.

Before presenting the tool in detail, it is important to keep in mind that the TMP does not intend to force the institutions to give up their terminologies and their collections management system and move entirely to SKOS. The TMP is there to provide you with a set of tools if you intend to provide your collection metadata as Semantic Web material. Moreover the approach of the TMP is very new and respectful of the professional needs and current processes in the collection management. Indeed the approach of the TMP is rather bottom-up than top-down. Instead of building a big reference terminology that would be then imposed to the institutions, we propose here to build a network of terminologies which is then a network of concepts. The purpose of the TMP is to help cultural institutions getting into the Semantic Web without investing over-the-top financial and human resources to do so. Therefore the TMP does not impose any modelling and design for building and/or adapting terminologies except using as much as possible the standards and recommendations that were already mentioned in this documentation.

The objective of a tool such as the TMP is also to foster to collaboration between the institutions, contributing to the network of concepts.

The TMP and this user manual do not intend to provide guidelines or material for the preliminary conceptual work on your terminology. You can refer to the ISO norm 25964 - The international standard for thesauri and interoperability with other vocabularies<sup>5</sup> and the booklet on terminology « Your terminology as part of the Semantic Web : recommendations and guidelines » available online<sup>6</sup>, to get a guidance on how to design a terminology.

### 3.2.3 Rights and Licenses

The rights on the terminologies are a very important aspect of the TMP. TMP is aligned with the vision of the Semantic Web and therefore is willing to provide Linked Open Data. The organisational reality is however not yet completely aligned with this vision.

A special attention must be paid to the licenses that will be attributed to the terminologies. The institution is the one to decide which will be the license applied to its terminology. Europeana has adopted a specific CC0 license, which is fully in line with the Open Data vision but the license applied to the metadata can be different from the one proposed for the terminologies. This is also the reason why the TMP offers the possibility to publish a terminology. Once a terminology is published it will be freely available on the Web under a CC-BY-SA license.






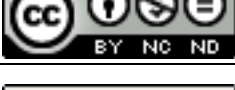
**The institution is responsible for the terminologies it uploads and publish via the TMP. If the institution is using a restrictive license but still intends to publish its terminology through the TMP, it will be published under the CC-BY-SA license.**

---

<sup>5</sup> <http://www.niso.org/schemas/iso25964/>

<sup>6</sup> « Your terminology as a part of the semantic web recommendations for design and management » : <http://www.linkedheritage.eu/getFile.php?id=244>

The table below provides details on each Creative Commons licenses that are the most convenient ones to share data on the Web with a minimum of restrictions. The descriptions proposed here are based on the one from the Creative Commons<sup>7</sup> :

| License logo   | License name  | Description  |
|--|---|--|
|   | Attribution<br>CC BY                                | <ul style="list-style-type: none"> <li>- Distribution, remix, tweak, and build upon your work, even commercially ;</li> <li>- Mention of the author as credit for the original creation.</li> </ul>  |
|   | Attribution-ShareAlike<br>CC BY-SA                  | <ul style="list-style-type: none"> <li>- Remix, tweak, and build upon your work even commercially ;</li> <li>- Mention of the author as credit for the original creation.</li> <li>- The new work will itself be distributed with the same license and terms</li> </ul>        |
|   | Attribution-NonCommercial<br>CC BY-NC               | <ul style="list-style-type: none"> <li>- Redistribution commercial or non commercial</li> <li>- mention of the author as credit for the original creation.</li> <li>- No modification at all on the original work</li> </ul>   |
|   | Attribution-NonCommercial-ShareAlike<br>CC BY-NC-SA | <ul style="list-style-type: none"> <li>- Remix, tweak, and build upon your work non-commercially,</li> <li>- Mention of the author as credit for the original creation.</li> <li>- The new work will itself be distributed with the same license and terms</li> </ul>          |
|   | Attribution-NonCommercial-NoDerivs<br>CC BY-NC-ND   | <ul style="list-style-type: none"> <li>- Mention of the author as credit for the original creation.</li> <li>- No modification at all on the original work</li> <li>- Non commercial distribution of the work</li> </ul>   |
|  | CC0-“No Rights Reserved”                            | <ul style="list-style-type: none"> <li>- Others can completely and freely build upon your work without any copyright restriction.</li> <li>- The original work is considered as public domain one.</li> <li>- No more copyright or credit for the original creation</li> </ul> |

**Table 1 : Creative Commons licenses description**

The CC-BY-NC-ND is the most restrictive Creative Commons license as it is not possible to build on the original work and commercial redistribution is forbidden.

The most widely used license is the CC-BY-SA which allows reusing and redistributing the work with copyright mention and distribution under the same license terms. This is basically the license used by the Wikimedia foundation which is managing Wikipedia.

The CC0 license is also described in the table. This license is the least restrictive one as modification, commercial use and the like are allowed without licensing under the same terms and without mentioning the author of the work.

It is up to the institutions to decide which is the most convenient license for their terminology resource, knowing that each country can have its proper rights terms, though we recommend the use of the CC-BY-SA license for guaranteeing the intellectual work done for terminologies and sharing this work under the same licence, leaving it free for modification.

### 3.2.4 SKOS

SKOS stands for Simple Knowledge Organisation System and is a recommendation of the W3C since August 2009. The aim of SKOS is to provide a simple datamodel for sharing and linking unstructured data on the Web to make this data meaningful.

<sup>7</sup> <http://creativecommons.org/licenses/?lang=en>



What is called a Knowledge Organisation System (KOS) could be assimilated to what we call a terminology resource, e.g. a thesaurus, a taxonomy, or a classification.

As presented in the SKOS primer, the reference document from the W3C, "*using SKOS, concepts can be identified using URIs, labeled with lexical strings, in one or more natural languages, assigned notations (lexical codes), documented with various types of notes, linked to other concepts and organised into informal hierarchies and associated networks, aggregated into concepts schemes grouped into labeled and/or ordered collections and mapped to concepts in other schemes*".<sup>8</sup> This paragraph summarises well the potential of SKOS.

Where OWL and RDF work well together to describe the world or a specific domain, a datamodel like SKOS is more adapted for the cultural field where institutions have to deal with objects and collections. Institutions from the GLAM sector (Galleries, Libraries, Archives and Museums) use thesauri, classifications or taxonomies to describe in a controlled way their objects and collections. Providing metadata on these objects and collections as Semantic Web based data on the Web would be a first major step but providing also these Knowledge Organisation System in a compliant format would be even crucial to enhance the semantic interoperability between the metadata and the controlled vocabularies used to describe cultural collections and objects.

A simple and lightweight datamodel as SKOS provides a low-cost and intuitive means to migrate from classical terminology to a Semantic Web compliant one. SKOS offers the same potential as OWL but is simpler and more appropriate to terminologies used within cultural institutions such as GLAMS. SKOS can be used on its own or combined with other standards. There are some use cases of use of SKOS combined with OWL.

SKOS is used as an interchange format where the norm on thesaurus interoperability ISO 25964 gives guidelines and recommendations to produce a thesaurus that could respect the standards and technologies of the Semantic Web.

SKOS, just like OWL, relies on RDF, XML and Unicode. And like RDF, SKOS is composed of classes and properties and the interdependency between those classes and properties and their characteristics is what brings the integrity and consistency of the datamodel.

## ➤ Introduction to SKOS

The fundamental element of terminology designed in SKOS is the concept. As it is the case for most of the Semantic Web standards, there is a strong distinction between the concept and the term(s) that is(are) used to express this concept. As SKOS is highly related to RDF, SKOS data are also expressed as RDF triples. SKOS Concepts can be either subject, object and related thanks to a SKOS property which would operate as a predicate.

In conformity with the principles of the Semantic Web and Linked Data, each SKOS concept can be identified by a URI. As we saw in the previous sections, the role of the URI is to provide a unique and unambiguous identifier to a resource, in this case to a concept. As mentioned above, there are some works and study investigating guidelines and recommendations for PIDs (Persistent Identifiers). The booklet produced within the Athena project for example provides a good reference of the existing systems : « Persistent identifiers: Recommendations for institutions » : <http://www.athenaeurope.org/getFile.php?id=779> . This document has been updated with a deliverable produced within the Linked Heritage project : D2.2 State of the art report on persistent identifier standards and management tools (September 2013)<sup>9</sup>

The three main components of the SKOS datamodel are the following ones :

- Classes
- Properties
- Relations

<sup>8</sup> <http://www.w3.org/TR/skos-primer/>

<sup>9</sup> D2.2 State of the art report on persistent identifier standards and management tools (September 2013) : <http://www.linkedheritage.org/getFile.php?id=556>

As for any Semantic Web standard, the datamodel is specified using a prefix, for example owl :Class or rdf :Class, in this case the prefix is "skos:". The case is important to distinguish classes, properties and relations ; skos :Class with an upper case and skos : property with a lower case.

### ➤ **Concept, Concept Schemes and Collections**

**Concept** : SKOS is a concept-oriented data model, therefore the concept is the central element of the terminology. The ISO-25964 norm about thesauri for information retrieval defines a concept as *"Unit of thought; concepts can often be expressed in a variety of different ways. They exist in the mind as abstract entities independent of terms used to express them."*<sup>10</sup>

A skos concept is introduced as a class **skos :Concept**. SKOS concepts can be aggregated into two other classes **skos :Collection** and **skos :ConceptScheme**.

Concepts can be organised related to each other by semantic relations which can be associative or hierarchical.

**Concept schemes** : A concept scheme is introduced by the class **skos:ConceptScheme** and allows grouping several concepts. As presented in the SKOS reference, « *an individual concept scheme roughly corresponds to the notion of an individual thesaurus, classification scheme or any other knowledge organization system* ». <sup>11</sup> Declaring different concept schemes allows keeping track of the provenance of the concepts presented in the concept scheme. The SKOS data model does not restrict the possibility to have the same concept in more than one concept scheme. This flexibility can be useful especially in the context of cultural institutions who can use the same concepts but organise their concepts differently, according to their professional point of view.

The properties skos :inScheme or **skos :hasTopConcept** are the two main properties used in the context of concept schemes.

**Concept collections** : Concept collections are introduced by the class skos :Collection. SKOS concepts can be grouped in a labeled or ordered collection. The SKOS collection is different from the concept scheme since collections allow grouping concepts according to a theme or a meaningful order. The grouping in SKOS collections does not impact the hierarchy or relations of the concepts. For the migration of a thesaurus for example, the whole could be considered as a concept scheme where several thematic groups of concepts could be designed as collections.

In the case of concepts that are grouped in a specific order, the class **skos :OrderedCollection** will be used.

### ➤ **Labels and notation**

As we already said, the SKOS datamodel is concept-oriented and concepts are expressed by terms. These terms can be expressed by lexical labels. A lexical label is a string of Unicode characters in any natural language, no matter which character set it uses.

Lexical labels are the elements helping to understand the meaning of concepts ; they are very useful for search and retrieval. The use of labels is not mandatory according to the SKOS datamodel but it is highly recommended for maintenance purposes and also for enabling multilingualism and general understanding of the concept.

The SKOS data model defines three types of lexical labels:

- **Preferred label** :

A preferred label, introduced in the SKOS data model as the **skos:prefLabel** property, corresponds to the notion of descriptor from the standards for the elaboration of thesauri (ISO 25964). There can be as many preferred labels as needed but there should be only one preferred label per language. The SKOS will not be considered as valid if there are two preferred labels in the same language present.

- **Alternative label**

---

<sup>10</sup> <http://www.w3.org/TR/skos-primer/>

<sup>11</sup> <http://www.w3.org/TR/skos-primer/>

The **skos:altLabel** property is used to define alternative labels. The distinction between an alternative label and a preferred label mainly consist of the disitinction between the descriptor and the non-descriptor in a thesaurus. Alternative labels are generally used to give synonyms or acronyms expressing the concept. The SKOS model does not forbid the exclusive use of alternative labels, instead of one preferred label and many alternative labels.

#### • Hidden label

The **skos:hiddenLabel** property introduces the hidden labels that are generally used for mentioning the misspelled forms of concepts. These hidden labels can be useful to improve the search and retrieval by using a misspelled expression to retrieve the relevant concept. Hidden labels are not meant to be visible.

Alternative and hidden labels correspond roughly to the USE and UF (Used For) indicators defined in the ISO standards for thesauri.

The SKOS data model has an intrinsic consistency checking system that forbids the use of the same string of character for a preferred, alternative or hidden label in the same language.

#### Notation

A notation is a symbol or a code that is not expressed in a natural language. Another property is available for expressing notations which are different from labels. This is the reason why notations differ from labels. The notation introduced by the **skos:notation** property can be used in the case of classifications or any resource where a code stands as a local identifier to refer to the concept. The notation in some Knowledge Organisation System can have the role of identifier and then be unambiguous.

#### ➤ Documentation

Documentation of concepts is really the key for fully understanding a concept. As the use of labels is recommended but not mandatory, documentation notes can be very useful to understand or remove any ambiguity about a concept.

The SKOS data model provides different types of documentation notes that can be used without any restriction.

Here are the different types of notes :

- Note (**skos:note**)
- Change note (**skos:changeNote**)
- Definition (**skos:definition**)
- Editorial note (**skos:editorialNote**)
- Example (**skos:example**)
- History note (**skos:historyNote**)
- Scope note (**skos:scopeNote**)

The **skos:note** property introduces the most generic kind of note which can provide a very general information on the concept. All the other notes are specialised but can be used complementary to the general property **skos:note**.

The **skos:changeNote** and **skos:editorialNote** are mainly useful for the maintenance and for keeping track of the modifications and versions on the concept.

The **skos:definition**, **skos:example** and **skos:historyNote** are useful for providing information on the concept for a better understanding of its meaning. As for labels, documentation properties can be provided in different languages by using language tags with the **xml:lang** attribute.

#### ➤ Semantic relations

There are two types of semantic relations : hierarchical and associative. Each concept can be related to another one based on these semantic relations. This is where the power lies of the SKOS data model and the conformity with the principles of the Semantic Web and Linked Data. Data needs to be connected to be meaningful. The semantic relations are the mechanisms for producing meaningful content.

#### Hierarchical relations

The **skos:broader** and **skos:narrower** properties are the ones used for expressing hierarchical relations between two concepts. **skos:broader** property is used to assert that a concept has a more general meaning. **skos:narrower** is the inverse property, used to assert that a concept has a more specific meaning. One concept can have more than one broader concept or more than one narrower concept. As these properties are the inverse of each other, if a broader relation is asserted, the inverse narrower property will be inferred without specifying it.

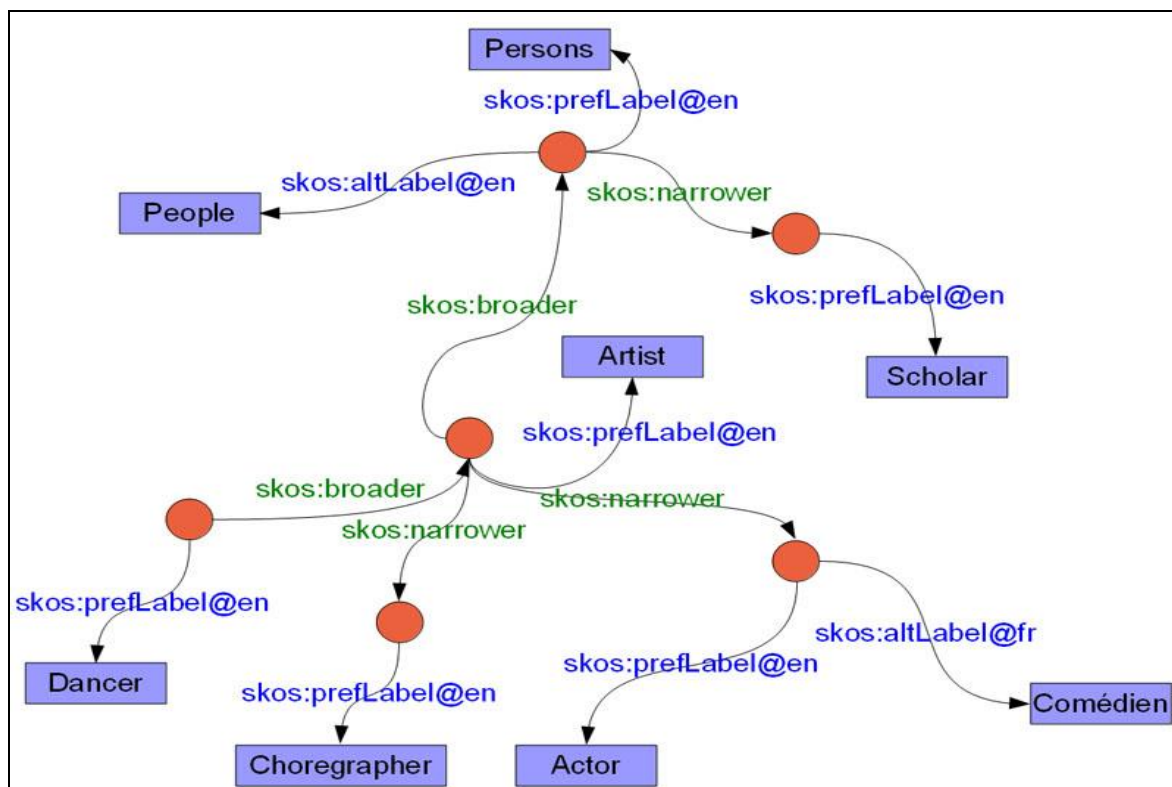


Figure 1 : SKOS graph illustrating the hierarchical relations

The SKOS graph above illustrates the **skos:broader** and **skos:narrower** properties. We can see on this graph that « Dancer » and « Choreographer » are at the same level of the hierarchy as narrower of the concept « Artist » but they are represented via different properties.

A very important aspect for the hierarchical relations between concepts and the logical consistency of the SKOS graph as a whole relies on the transitivity of these hierarchical relations. The **skos:broader** and **skos:narrower** properties are not transitive and therefore the hierarchical link between two concepts is direct and immediate. The transitivity in hierarchical relations is indicated by two other SKOS properties : **skos:broaderTransitive** and **skos:narrowerTransitive**.

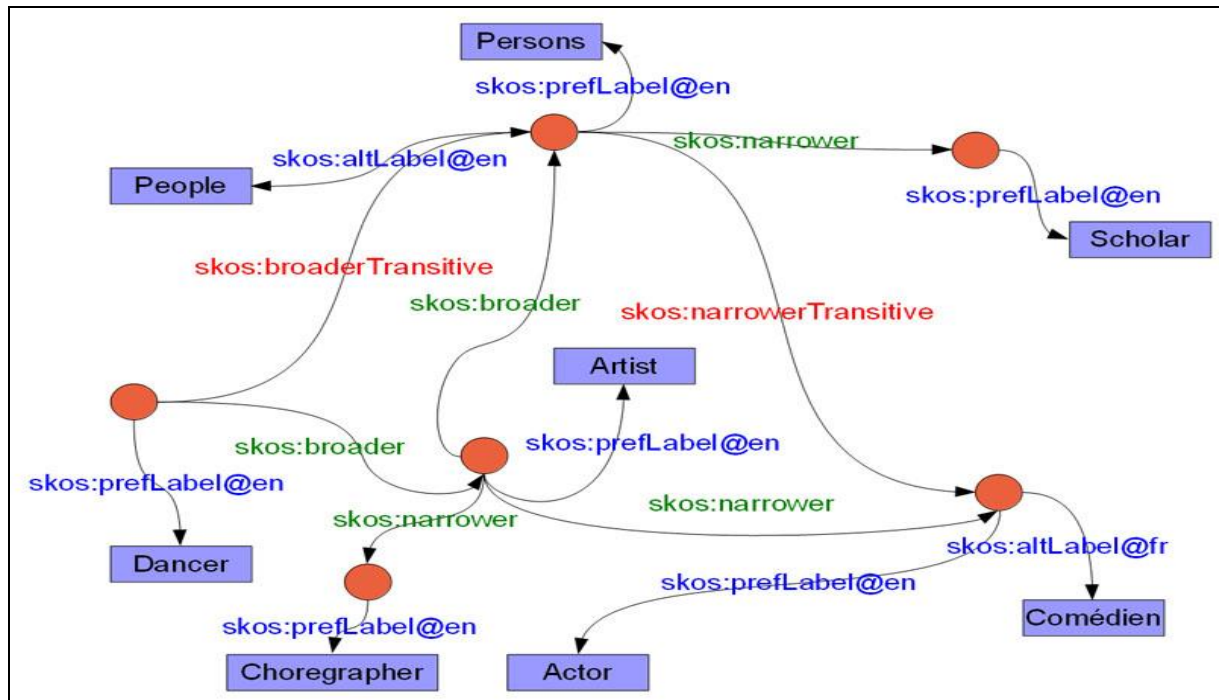


Figure 2 : SKOS graph illustrating the transitive hierarchical properties

The SKOS graph above illustrates the use of the transitive properties **skos:broaderTransitive** and **skos:narrowerTransitive**, which allows asserting that « Dancer » is a narrower concept of « Persons ». This information could not be inferred from the relation between « Persons »-« Artist » and « Artist »-« Dancer ».

The `skos:broaderTransitive` and `skos:narrowerTransitive` work, like the **skos:broader** and **skos:narrower**, as inverse property of each other.

### Associative relations

Associative relations in SKOS are introduced by the `skos:related` property. A concept can be considered as related to another one when an equivalent or hierarchical relation between these concepts can not be made.

The `skos:related` property is symmetric. Therefore as in the example presented in the graph below, stating that « Musician » is related to « Composer » is equivalent to stating that « Composer » is related to « Musician ».

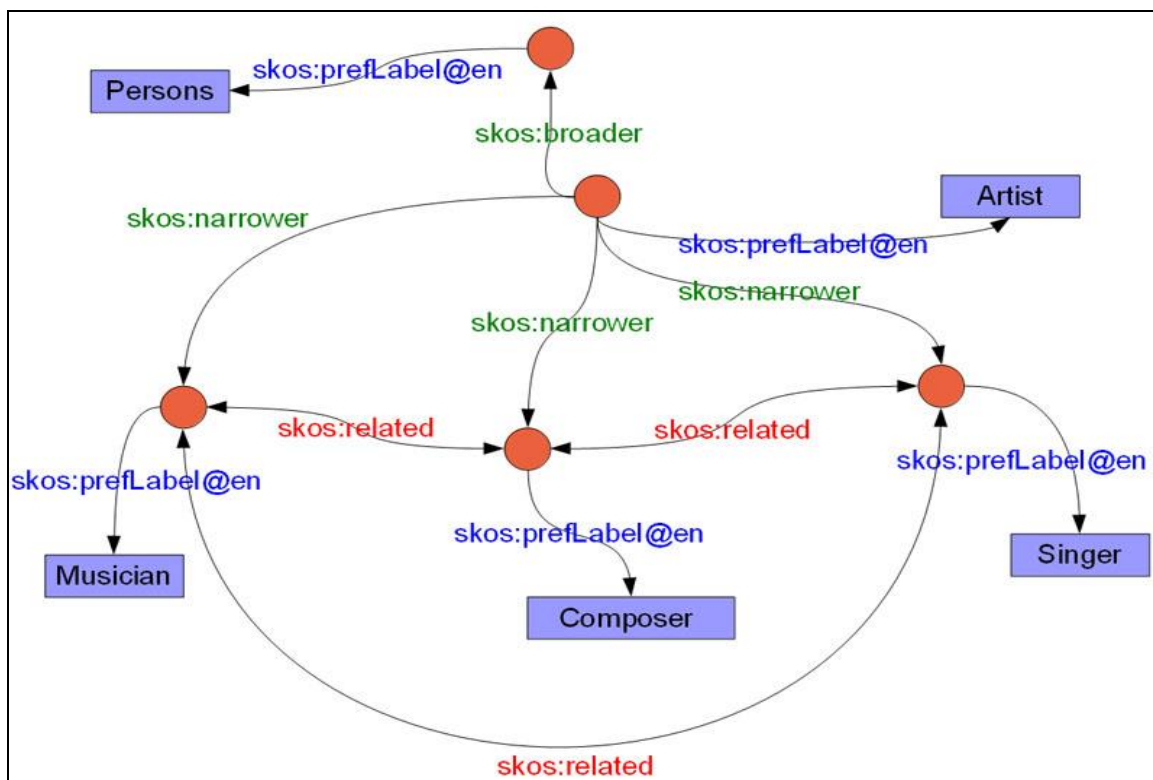


Figure 3: SKOS graph illustrating associative relations

The associative relations are a direct and immediate type of link as `skos:related` is not a transitive property. As represented in the example graph, if we state that « Musician » is related to « Composer » and that « Composer » is related to « Singer », the associative relation between « Musician » and « Singer » can not be inferred. To establish this associative relation we have to state with another assertion that « Musician » is related to « Singer » via the `skos:related` property.

Associative relations and hierarchical ones are two different kind of semantic relations. Mixing these relations would not be logical and is not consistent neither with the SKOS datamodel nor with the guidelines defined in the ISO 25964.

As an example, it would not be possible to state that « Musician » is related to the concept « Artist » if there is already a hierarchical relation between these two concepts.

### ➤ Mapping

Mapping enables making links between concepts from different terminologies or concept schemes. The mapping properties provided by the SKOS datamodel are the key to respect the Linked Data principles. The mapping process can be defined as making alignments or matching concepts from different sources that are equivalent or similar. The connected concepts can help to improve search and retrieval of relevant results for a given query.

The mapping properties provided by SKOS are the following ones :

- **skos:closeMatch**
- **skos:exactMatch**
- **skos:broadMatch**
- **skos:narrowMatch**
- **skos:relatedMatch**

The two essential mapping properties are the **skos:exactMatch** and **skos:closeMatch** as they are the ones which allow asserting a perfect or partial equivalence between concepts from different concept schemes. The `skos:exactMatch` property is transitive and symmetric. This means that if a concept « A » is



mapped as an exact match to the concept « B », all the child-concepts of B could be mapped as narrow match of A.

Mapping properties work as semantic relations, except that the concepts that are linked come from different concept schemes. As for semantic relations, the mapping properties can be either associative or hierarchical. The hierarchical mapping relations are introduced by the **skos:broadMatch** and **skos:narrowMatch** properties and these two properties are the inverse of each other. The associative mapping relation is introduced by the **skos:relatedMatch** property.

Mapping properties are used rather than semantic relations in order to make mapping links between concepts from different concept schemes. Therefore in the case of a same concept scheme, semantic relations have to be used instead of mapping properties.

As for semantic relations, mixing hierarchical and associative mapping for the same concepts can create conflicts and affect the logic and the consistency of the terminology.

### 3.2.5 SKOS and ISO-25964

ISO 25964 is the very recent norm on thesauri elaborated by the International Organisation for Standardisation (ISO) in order to upgrade two previous norms on monolingual and multilingual thesaurus ISO 2788<sup>12</sup> and ISO 5964<sup>13</sup>. *ISO 25964 Information and documentation – Thesauri and interoperability with other vocabularies* takes into account the evolution of the technological environment and the emergence of the Semantic Web. A major change in this norm lies also in the distinction between a concept and the terms used to express this concept. This distinction as presented in the sections above, is perfectly compliant with the general principles of the Semantic Web.

This norm has been published in 2011-12 while SKOS was acknowledged as a W3C recommendation in 2009. SKOS has provided the technology that could help cultural institutions to make the first step towards the Semantic Web environment but one of the criticisms is exactly this too simplified model that does not fit the complexity of the tools used by museums, libraries and the like. SKOS, as presented in the W3C recommendation, was already presented as a core that could be completed with some specific extensions. This is the case with the SKOS-XL extension that allows including morphological and/or syntactic elements to the description of labels expressing concepts.

The working group of ISO 25964 and Antoine Isaac co-editor of the SKOS recommendation have elaborated correspondence<sup>14</sup> between the thesaurus model and the SKOS one in order to align them and enrich the core of SKOS and the SKOS-XL with some new properties using the prefix « iso-thes ». The purpose of this work is to provide all the missing elements from the ISO norm in the SKOS core data model as properties introduced with the prefix « iso-thes ».

This work is not yet implemented as a RDF graph and still has to be acknowledged as a standard but it proposes some very interesting features for keeping track of different versions for concepts and also providing details on the composition of terms expressing concepts.

---

<sup>12</sup> [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=7776](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=7776)

<sup>13</sup> [http://www.iso.org/iso/fr/catalogue\\_detail.htm?csnumber=12159](http://www.iso.org/iso/fr/catalogue_detail.htm?csnumber=12159)

<sup>14</sup> Correspondence between ISO 25964 and SKOS/SKOS-XL Models

[http://www.niso.org/apps/group\\_public/download.php/9627/Correspondence%20ISO25964-SKOSXL-MADS-2012-10-21.pdf](http://www.niso.org/apps/group_public/download.php/9627/Correspondence%20ISO25964-SKOSXL-MADS-2012-10-21.pdf)

### 3.2.6 Users and roles

The user management of the TMP has been designed and specified in order to take into account the organisational reality in institutions. Very few institutions have full-time employees for terminology management; generally this task is attributed to the curator or the person responsible for collections. The TMP does not intend to be a tool especially dedicated to experts of one domain but to anyone who has a professional or scientific knowledge of the domain they are going to describe through their terminologies.

Therefore five user roles have been designed within the TMP<sup>15</sup>. The organisational aspect is also taken into account as the affiliation to an institution is managed in the TMP.

The table below summarizes the actions allowed for each user role.

| Actions   | Roles         |       |        |             |              |
|---|---------------|-------|--------|-------------|--------------|
|   | Administrator | Owner | Editor | Contributor | General user |
| Validate affiliations                               | X             |       |        |             |              |
| Attribute user rights                               |               | X     |        |             |              |
| Manage rights                                       |               | X     |        |             |              |
| View public terminologies                           |               | X     | X      | X           | X            |
| Search public terminologies                         |               | X     | X      | X           | X            |
| Make suggestions to the terminology                 |               | X     | X      | X           |              |
| Make suggestions to the metadata of the terminology |               | X     | X      | X           |              |
| Accept suggestions                                  |               | X     | X      |             |              |
| CRUD concept  |               | X     | X      |             |              |
| CRUD terminology metadata                           |               | X     |        |             |              |
| CRUD outcoming mapping                              |               | X     | X      |             |              |
| Set as public or private terminology                |               | X     |        |             |              |
| Publish terminologies                               |               | X     |        |             |              |
| Export terminologies                                |               | X     | X      | X           |              |

**Table 2 : TMP - User roles and actions**

**Institution:** the institution is the public or private organisation or company which the terminology belongs to. The first user who creates an account and declares the institution as an entity is considered to be the owner of the terminology.

**Administrator:** the administrator is the user that is defined by the owner as the primary contact of the institution. This role is more an administrative role than a managerial one. As a primary contact, the administrator is the one who can accept or reject the requests from users for affiliation to the institution. The role of the administrator is only to validate the users who want to belong to his institution. But the administrator is not in charge of providing rights to these users. Once affiliated to the institution, the owner will be the one to give rights to the users. There can only be one administrator per institution.

**Owner:** the owner has a role of administrator of terminologies. He is the user who can manage the terminologies from the creation to the publication or deletion. He is also the one who can give rights as editor or contributor to the other users affiliated to the same institution. He can create, read, update and delete concepts and terms, accept or reject the suggestions made by other users. He is the one who can modify the metadata of the terminologies and publish these terminologies. Each user affiliated to an

<sup>15</sup> These user roles and organisations have been defined according to the actions that would be needed and the privileges required for these actions by the Linked Heritage WP3 and the AthenaPlus WP4.



institution who is creating a terminology becomes owner of the terminology. There can be more than one owner per institution.

**Editor:** the editor works under the authority of the owner of the terminology. The editor has the same rights as a contributor but can also have some management rights on his terminologies. The editor is then able to create, read, update and delete concepts and terms.

**Contributor:** the contributor is allowed to view public thesauri and make suggestions to modify them. The contributor is a user who has an account on the TMP and can be logged in. The contributor may belong to another institution and is then able to make suggestions to a public terminology. Note that this role only allows suggestions for modification, and does not grant direct modification rights.

**General User:** The general user is the visitor who may only browse and navigate through the terminology registry. The general user does not have an account (for he does not need one) and is not logged in. He can search in the registry and see terminologies that are registered as public ones. Editing, mapping and export functionalities are not available for the general user.

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

These user roles are not exclusive; one single individual can take on each of these roles, depending on the context and the terminologies being used. The owner from a museum A can be a contributor for a library B. This is exactly what makes the TMP so interesting; it really allows for cross-domain terminology interaction.

In order to have access to the main features of the TMP, the user will have to be logged in to the TMP. The general user is not able to make suggestions because details on the person making the suggestion are necessary to assess if the suggestion is relevant. This is the reason why any user willing to contribute directly on terminologies and metadata will have to create an account even if they don't belong to an organisation or they are contributing in their personal, non-professional name. This procedure allows for moderation – which is highly needed in this kind of system.

### 3.2.7 Getting started

The following figure presents the homepage of the TMP.

The logo of the TMP has been designed following the designs of the Athena, Linked Heritage and AthenaPlus project, as the idea of developing such tool emerged from these projects. The banner and the dots are inspired by the MICHAEL project<sup>16</sup> and the Michael Culture Association<sup>17</sup>, as an important part of their activity is dedicated to multilingualism.

A breadcrumb is available to help the user find his way when navigating and working in the tool.

A short headline presents the main purpose of the TMP. Further details are presented in static pages.

TMP intends to provide a terminology registry freely available and accessible, therefore any user can search and navigate through the terminology registry. A search box is available on the homepage of the TMP from the block “Search public terminology”.

---

<sup>16</sup> MICHAEL : <http://www.michael-culture.org/>

<sup>17</sup> Michael Culture Association ( <http://www.michael-culture.eu/> ) is not-for-profit organisation which has been created to sustain the activity of the European project Michael (Multilingual Inventory of the Cultural Heritage in Europe). The aim of Michael is to provide a multilingual access to European digital collections. Michael Culture Association gathers a strong network of more than 100 public and private organizations from all over Europe. Michael Culture leads different working groups (<http://www.michael-culture.eu/ressources-working-groups-427>) which are related to Research. One of this working group is dedicated to multilingualism (development on standards, linguistic diversity, language technology, language learning)

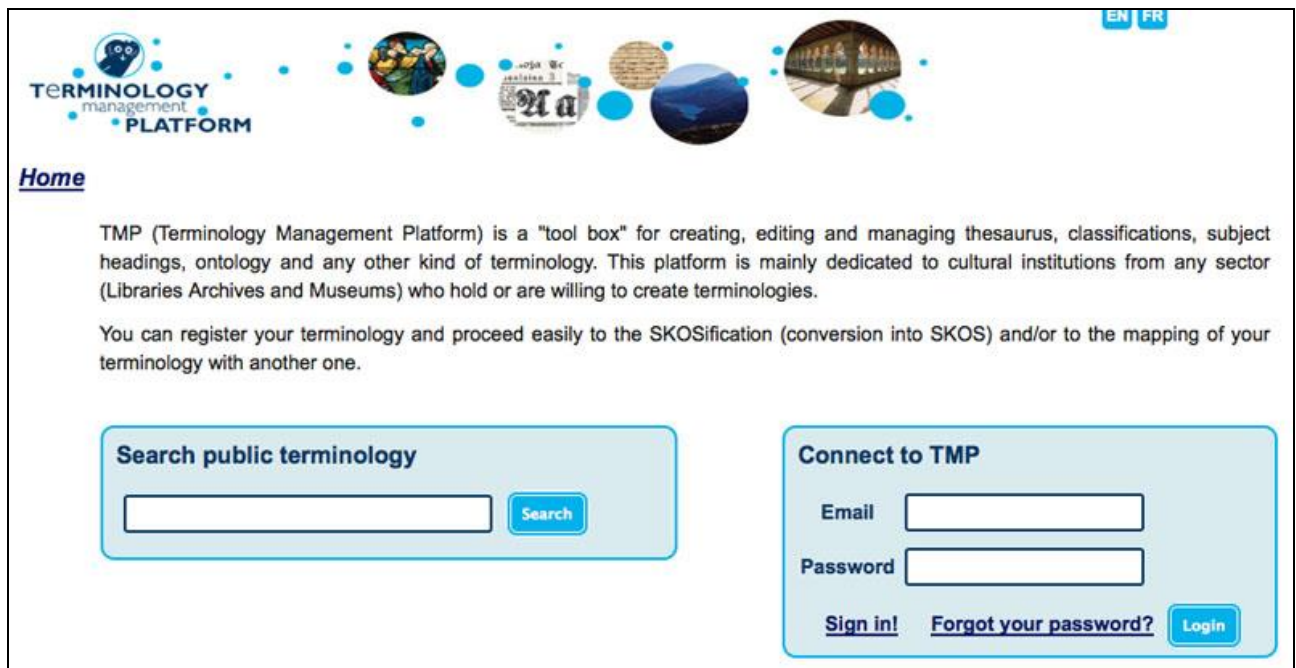


Figure 4 : Homepage of the TMP

If you search for the concept “Architecture” for example, you will get a list of results :

[My home](#) > [search public terminologies](#)

Show  entries Search:

| Title ▲                               | Subject ⇅    | Languages ⇅            | Author ⇅    |                          |
|---------------------------------------|--------------|------------------------|-------------|--------------------------|
| <a href="#">Architecturalelements</a> | Architecture | en, hu                 | JuliaKatona | <a href="#">View</a> </> |
| <a href="#">Building</a>              | Architecture | en, fr                 | CR          | <a href="#">View</a> </> |
| <a href="#">Buildings</a>             | Architecture | en, fr                 | CR-LD       | <a href="#">View</a> </> |
| <a href="#">Edifices</a>              | Architecture | fr                     | LD          | <a href="#">View</a> </> |
| <a href="#">PICO Thesaurus</a>        | Archaeology  | en, it, en, it, en, it | mibac       | <a href="#">View</a> </> |
| <a href="#">SMM</a>                   | Architecture | nl                     | SMM         | <a href="#">View</a> </> |
| <a href="#">testimportPacked</a>      | Architecture | de                     | 0.1         | <a href="#">View</a> </> |

Figure 5 : Results of public search


The search will be made through the title and the metadata of the terminology. A general user will then be able to view and/or export the terminology.

As mentioned above, the user has to create an account and be connected to be able to fully benefit from the functionalities of the TMP.

### 3.2.8 Connection to the TMP

When the user is trying to connect for the first time to the TMP, he will have to click on the « Sign-in ! » link available from the homepage on the « Connect to TMP » block.

The user will then have to fill in a form asking for authentication information.



The image shows a 'Sign in' dialog box with the following fields and values:

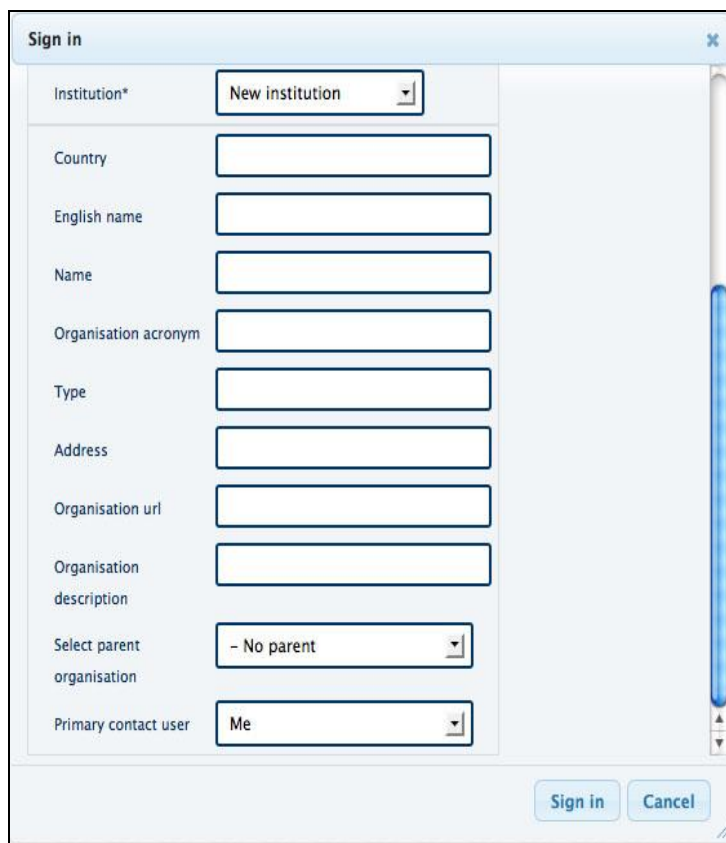
| Field             | Value                               |
|-------------------|-------------------------------------|
| Email*            | rie-veronique.leroi@culture.gouv.fr |
| Firstname*        | Marie                               |
| Lastname*         | Leroi                               |
| Password*         | ••••••                              |
| Confirm password* | ••••••                              |
| Institution*      | New institution                     |

Buttons: Sign in, Cancel

Figure 6 : Sign in form

Mandatory fields such as the email address, first name, last name, password are asked and marked with an asterisk (\*). Confirmation of the password is necessary to guarantee that bots or spamming tools will not be able to get into the applications.

The last information asked is the institution. Existing institutions, the ones that are already registered at the time of your registration, are proposed as drop-down list. If your institution is not on the list the last element proposed is « new institution ».



**Figure 7 : Extended form for signing in**

When the user clicks on the « New institution » option, the sign-in form is extended in order to fill in the details about the institution. The informations asked are similar to the one asked in the MINT tool<sup>18</sup> in order to have a consistent institution information. A distinction is made between the name in the language of the institution and the name in English in order to have a unique and understandable appellation.

As presented in the screenshot above, the user will be asked to indicate the country, the name of the organisation in the primary language and in English and the acronym. The type of institution is also asked ; this could be a museum, a library, an archive service, a ministry or a private company. The access to the TMP is not restricted - therefore the organisation can be of any type, even if the main users are from the cultural sector. If the organisation is depending on another organisation this can be specified with the « select parent organisation » field which proposes a drop-down list with all the registered organisations. This could be of interest for example for a national museum who is legally depending on a ministry.

The primary user mentioned in the form will have the role of administrator presented in the previous section. In the case of a first connection and first registration of institution, the primary contact would be by default « me ». If the person who had the role of administrator is not willing to continue or is moving to another organisation, the administrator will be able to set another user affiliated to the organisation as primary contact.

After this registration process, the user will be able to log in to the TMP using his email and the password defined during the registration. In the case the user does not remember his password a link “Forgot your password?” is available to send again the password to the user by email.

There are different scenarios possible:

- I am a user and have already registered

<sup>18</sup> <http://mint-projects.image.ntua.gr/athenaplus> The MINT platform provides content holders with the ability to perform the required mapping of their own metadata schemas to LIDO as well as its publication to Europeana. It addresses the ingestion of metadata from multiple sources, the mapping of the imported records to a target metadata schema and the transformation and storage of the metadata in a repository.

AthenaPlus D4.4 Specialist Training material : TMP user manuals and semantic mapping procedures procedures

Please fill in the authentication form with your email-address and password. When successfully logged in, you are redirected to the **Main page** of the TMP.

When login has failed, the system redirects to the login form with failed email and/or password pre-filled.

- I am a new user and my organization has already registered

Click on the tab “New User” and fill in the empty cases.

|                  |
|------------------|
| E-mail           |
| First name       |
| Last name        |
| Password         |
| Confirm password |
| Organization     |

**Note!** The password should be different from the username and should contain at least 6 characters.

The organization must be selected from a drop down list. If the organization is not figured in the list, a tooltip will appear: “Your organisation is not in the list”. You will have to register your organisation first before being able to connect to the TMP.

**Note!** The first user will automatically be attributed the role of *administrator*. The administrator can change his/her role and designate another user as administrator.

- I am a new user and my organisation has not yet registered

When flying over the link “create new”, a modal window (lightbox) opens. Enter following details about your organisation

| Field                      | Explanation  |
|----------------------------|--|
| Country                    | Enter the country where your organisation is located                         |
| English name               | Enter the English name of your organisation (translation)                    |
| Acronym                    | Enter the acronym of your organisation (e.g. MCC)                            |
| Type                       | Enter the type of organisation, e.g. library, museum.                        |
| Address                    | Enter the legal address of your organisation                                 |
| Organisation URL           | Enter the persistent URL of your organisation                                |
| Organisation description   | Describe in short the role, collection and target group of your organisation |
| Select parent organisation | If applicable, select parent organisation from a drop-down list              |
| Primary contact user       | Select the name of the primary user from the drop                            |

|  |           |
|--|-----------|
|  | down list |
|--|-----------|

If the creation is successful, please click “Sign-in”. You can edit the information about the organisation by clicking the “Edit” tab.

### 3.2.9 Search terminologies

Once logged in, the search box allows the user to make a full-text search on publicly available terminologies: metadata, concepts and documentation on concepts are searched and displayed in a list.

**My home > search public terminologies**

Show  entries Search:

| Title ▲                               | Subject ⇅    | Languages ⇅            | Author ⇅    |  |
|---------------------------------------|--------------|------------------------|-------------|--|
| <a href="#">Architecturalelements</a> | Architecture | en, hu                 | JuliaKatona |  |
| <a href="#">Building</a>              | Architecture | en, fr                 | CR          |  |
| <a href="#">Buildings</a>             | Architecture | en, fr                 | CR-LD       |  |
| <a href="#">Edifices</a>              | Architecture | fr                     | LD          |  |
| <a href="#">PICO Thesaurus</a>        | Archaeology  | en, it, en, it, en, it | mibac       |  |
| <a href="#">SMM</a>                   | Architecture | nl                     | SMM         |  |
| <a href="#">testimportPacked</a>      | Architecture | de                     | 0.1         |  |

**Figure 8 : TMP list of search results**

For each terminology the title, the subject(s), language(s) and the author is visible. It is possible to sort each of these fields and then view the terminology. The number of terminologies to be displayed on the same page can also be set via the drop-down menu « Show XX entries ».

EN FR

**My home**

**Search therminologies**

Search

**Figure 9 :TMP : Search box for owner, editor and contributor**

When the user is logged in to the TMP, he can search through his own terminologies or through all the public terminologies. The results are then displayed with the same organisation and sorting facilities.

### 3.2.10 Create new terminologies

The owner can create a terminology through the « Create new terminologies block ».



Figure 10 : TMP : Create new terminologies

Before starting the creation of a terminology using the TMP, the conceptual work on the terminology has to be finalised. The work involving all experts, curators and cataloguers who will be using the terminology should be completed. They are also the group who will keep the terminology evolving after a first release has been published.

Cultural institutions can refer to the booklet « Your terminology as part of the Semantic Web : recommendations and guidelines » available online<sup>19</sup> and to the norm ISO 25964 on thesauri for information retrieval for the conceptual design of their terminologies.

The owner has to decide if the terminology will be built from scratch or if an existing terminology will be imported. The TMP allows the owner to either create from scratch its terminology or importing an existing one that will then by the import process be converted into SKOS.

- **Creating a terminology manually :**

The first step to create a new terminology is to provide some metadata on this terminology. In order to respect standards for most of the existing registry, the metadata fields have been defined using the Dublin Core description scheme. The most relevant fields are marked as mandatory with an asterisk (\*), the other fields are optional but the more detailed the description of a terminology is, the more « linkable » and reusable it will be.

Therefore, we recommend you to document these metadata with attention and detail.

The figure below represents a screenshot of the metadata form when creating a new terminology.

---

<sup>19</sup> « Your terminology as a part of the semantic web recommendations for design and management » : <http://www.linkedheritage.eu/getFile.php?id=244>





[My home](#) > Terminology creation

Create a terminology from scratch: fill in general information (metadata) on your terminology (\* mandatory fields)

|             |   |                   |   |
|-------------|---|-------------------|---|
| Title *     | Designation                                     | Alternative Title | architecturales et des espaces aménagés |
| Subject *   | Decorative arts ✕                               | Description       | véritable outil documentaire numérique. |
| Languages * | French ✕  | Creation Date     | 2013-04-23                              |
| Edition *   | 1   | Date of issue     |   |
| Creator *   | MCC   | Contributor       |   |
| Publisher * | ère de la culture et de la communication        | Identifier        | e.fr/thesaurus/resource/ark:/67717/T96  |
| Rights *    | Attribution ShareAlike (CC BY-SA) ▾             | Source            | MCC                                     |
| Format *    | Select a value<br>Select a value<br>CSV<br>SKOS | Amount of Concept | 101 and 500 ▾                           |

[Save Metadata](#)

A project of Co-funded by the and the

**Figure 11 : TMP: Metadata on the terminology**

The table below lists the metadata fields and their description. The descriptions below are based on those provided by the Dublin Core metadata initiative.<sup>20</sup>

| Field             | DC property                   | Description  | Mandatory/optional |
|-------------------|-------------------------------|--|--------------------|
| Title             | DC:title                      | The name given to the resource. Typically, a title will be a name by which the resource is formally known.   | Mandatory          |
| Alternative title | DC:Terms<br>Alternative Title | An alternative name for the resource; this can be any form of the title that is used as a substitute or an alternative to the formal title of the resource including abbreviations or translations of the title. Refinement of dc:title; | Optional           |
| Subject           | DC :Subject                   | The topic of the resource. Used for: describing the subject of the terminology (description domain / application domain); the subject list proposed for this field is based on the Michael subject terminology <sup>21</sup>             | Mandatory          |
| Description       | DC:Description                | An account of the content of the resource. Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.                      | Optional           |
| Languages         | DC:Language                   | The language of the resource ; the languages' codes are based on the ISO-63922 norm.   | Mandatory          |
| Creation date     | DC:Terms<br>Date created      | Date of creation of the resource.  | Optional           |
| Date of issue     | DC:Terms:                     | Date of formal issuance (i.e., publication) of the   | Optional           |

<sup>20</sup> <http://dublincore.org/documents/2005/11/07/usageguide/elements.shtml>

<sup>21</sup> Michael Terminology : [www.michael-culture.org](http://www.michael-culture.org)

<sup>22</sup> [http://www.iso.org/iso/home/standards/language\\_codes.htm?="](http://www.iso.org/iso/home/standards/language_codes.htm?=)



|                    |                 |   |           |
|--------------------|-----------------|---|-----------|
|                    | Date issued     | resource;   |           |
| Edition            | Edition         | Semantic versioning of the resource ( <a href="http://semver.org/">http://semver.org/</a> ). Used for: defining the version of this particular terminology. If no versioning was kept before, use v.1;  | Mandatory |
| Contributor        | DC:Contributor  | An entity responsible for making contributions to the resource. Used for: the name of contributors to the terminology. This could be a person, an organisation or a service;  | Optional  |
| Creator            | DC:Creator      | An entity primarily responsible for making the resource; this may be a person, an organisation or a service. Used for: the name of the creator(s) (person of organisation) of the terminology   | Mandatory |
| Identifier         | DC:Identifier   | An unambiguous reference to the resource within a given context. Used for : reference to the resource (paper version, URL, ...)   | Optional  |
| Publisher          | DC:Publisher    | An entity responsible for making the resource available. Examples of a publisher include a person, an organisation and a service. Used for: the name of the publisher of the terminology, if different than the creator   | Mandatory |
| Source             | DC:Source       | A related resource from which the described resource is derived in whole or in part. Used for: reference to several sources that are related or where used in the creation of the terminology;  | Optional  |
| Rights             | DC:Rights       | Information about rights held in and over the resource. Used for: information about the current intellectual Property Rights, access rights or license arrangements for the terminology (digitised or born digital);  | Mandatory |
| Amount of concepts | DC:Terms Extent | The size of the resource. Used for: describing the size of the terminology; a list with the following options is proposed :<br><ul style="list-style-type: none"> <li>o 10 or less,</li> <li>o Between 11 and 100,</li> <li>o Between 101 and 500,</li> <li>o Between 501 and 1000,</li> <li>o Between 1001 and 5000,</li> <li>o Between 5001 and 10000,</li> <li>o 10001 or more.</li> </ul> | Optional  |
| Format             | /               | The input format of the terminology : either SKOS or CSV in the case of the import of an existing terminology   | Mandatory |

**Table 3 : TMP – Metadata fields description**

The owner of a terminology will be the one responsible for providing or updating these metadata.

As we stressed it in the introduction to the Semantic Web part, the notion of ‘identifier’ is crucial to have. It is a unique and unambiguous source for terminology resources. Within the TMP, a unique identifier is applied to terminologies but also to concepts. Therefore the information you provide in the field ‘identifier’ is very important as it can have an impact on the building of the URI for the terminology. Basically, the URI generated within the TMP for a terminology when it is created or imported is based on the following scheme:

Domain-Name/thesaurus/(Producer ID)/(Terminology ID)

For example : <http://www.culture-terminology.org/thesaurus/76320/T09430987654>

If an existing terminology is imported and already carries a persistent identifier, this identifier will be reused instead of creating another URI for the same resource.

Once the metadata have been filled in, the terminology metadata are registered and the terminology is created. After saving the metadata form, TMP will directly lead the user to the Editing interface. (See figure 27)

- **Import an existing terminology**

For importing an existing terminology, the user will have to click on the “Import” button. TMP is able to import SKOS files or CSV files.

SKOS files can have either a .xml or .rdf extension.

As for the creation of a terminology, the user will have to fill in a form with metadata fields. See the details in the paragraph above. The metadata form will be the same than the one for creating a terminology apart from one field. The “format “ field (see Figure 19) can have two possible values : SKOS or CSV.

If the terminology is already available in the SKOS format then there will be a simple import, the terminology can be directly uploaded after going through a syntax and validity checking. In conformity with the SKOS data model, each concept must have one and only one preferred label per language. Each property must have a language attribute. The SKOS validation tool also checks whether the hierarchical and associative relations are mixed. In this case, TMP will point it out as an error and the user will not be able to proceed with the import. If the import has been done successfully without any SKOS validation error, TMP will directly lead the user to the Editing interface (see Figure 27).

### **3.2.11 SKOSification**

By SKOSification we mean the conversion of input terminologies (thesaurus or any other kind of terminology resource) into SKOS format. When an institution has an existing terminology, the first step is to check if the collection management system used by the institution has a SKOS export functionality. In the case the terminology is stored in a proprietary format, an export in the CSV standard format will be needed to be imported in the TMP.

CSV stands for Comma-Separated Values. As presented on the CSV Wikipedia page « CSV is a delimited data format that has fields/columns separated by the comma character and records/rows terminated by newlines »<sup>23</sup>.

Terminologies exported in the CSV format will have to be processed according to a specific scheme in order to be imported and mapped correctly to the SKOS data model.

First the spreadsheet file must be exported in the CSV format and not the XLS format (which is a non standard Microsoft proprietary format). If the export of the terminology is only available in the ODS or XLS format, the user will have to open it respectively in the programmes OpenOffice or Excel and save it as a CSV file. Attention must be paid to the encoding system that will be used during the CSV export. In order to be compliant with Semantic Web technologies it is important that the content is made available in the Unicode encoding<sup>24</sup>. One of the parameters should then be UTF-8 for this export, another parameter should be the separator. The separator can be a comma, a tabulation, or a semicolon. The text is generally encapsulated with single or double quotes.

---

<sup>23</sup> [http://en.wikipedia.org/wiki/Comma-separated\\_values](http://en.wikipedia.org/wiki/Comma-separated_values)

<sup>24</sup> See on page 39 of this deliverable a short introduction of Unicode.

|    | C                                   | D                                   | E       | F   |
|----|-------------------------------------|-------------------------------------|---------|---|
| 1  | prefLabel@nl                        | prefLabel@eng                       | broader | SN@nl   |
| 2  | Objectnaam                          | Object name                         |         |   |
| 3  | Architectuur                        | Architecture                        | 41875   | Topterm voor architecturale elementen en onderdelen v     |
| 4  | Bouwwerk                            | Single built work                   | 41877   | Algemene benaming voor alle soorten van (in aanbouw       |
| 5  | Deel van een gebouw                 | Building division                   | 41879   | Wordt gebruikt voor het benoemen van afzonderlijke ka     |
| 6  | Funerair bouwwerk                   | Funerary structure                  | 41879   | Wordt gebruikt voor gebouwen die betrekking hebben op     |
| 7  | Monument                            | Monument                            | 41879   | Bouwwerken die getuigen van een (vroegere) cultuur, k     |
| 8  | Architecturaal element              | Architectural element               | 41877   | Structurele of decoratieve vormen die oorspronkelijk of l |
| 9  | Antefix                             | Antefix                             | 41878   | Dakornamenten, die het open eind van een rij dekregel     |
| 10 | Boog en boogonderdeel               | Arch and arch component             | 41878   | Gidstern voor bogen en onderdelen van bogen               |
| 11 | Korbeel                             | Bracket (structural element)        | 41878   | Wordt gebruikt voor architecturale elementen die diene    |
| 12 | Haard en haardonderdelen            | Hearth and hearth component         | 41878   | Gidstern voor haarden en onderdelen van haarden           |
| 13 | Zuil en zuilonderdeel               | Column and column component         | 41878   | Gidstern voor zuilen en onderdelen van zuilen             |
| 14 | Onderdeel van een pyramide          | Pyramid component                   | 41878   | Wordt gebruikt voor onderdelen van pyramiden              |
| 15 | Kraagsteen                          | Corbel                              | 41878   | Wordt gebruikt voor een uit de muur vooruitspringende     |
| 16 | Oppervlakte-element en onderdeel va | Surface element and surface element | 41878   | Gidstern voor architecturale elementen die een deel va    |
| 17 | Entablement en entablementonderdeel | Entablature and entablature compone | 41878   | Wordt gebruikt voor de omlijsting van een raam of deur    |

Figure 12 : Example of a terminology in an Excel spreadsheet file

The terminology has to be reorganised in order to have each SKOS property as a header of a column and one line per concept, as shown in the table below:

| SKOS property   | SKOS property   | SKOS property   | SKOS relation             | SKOS relation  |
|-----------------|-----------------|-----------------|---------------------------|----------------|
| skos:prefLab:EN | skos:prefLab:FR | skos:altLabEN   | skos:Broader              | Skos:Related   |
| Boat            | Bateau          | Craft           | Watercraft                | Boatbuilding   |
| Spinning wheel  | Rouet           | Spinning wheels | Textile working equipment | Spinning jenny |

In order to be parsed correctly by the SKOSification module of the TMP, this Excel file will have to be exported as a CSV file. This can be done quite easily just clicking on Menu "File"> Save as. Then the Excel file will be exported as a CSV file.

If you have a specific layout for your spreadsheet file (indented or all linear presentation), you will have to reorganise your terminology in order to have each SKOS properties as a header of a column and one line per concept.

Once the file is exported in CSV, you can see how it looks like in the following figure :

| "ID"  | "prefLabel@fra"          | "prefLabel@nl"  | "prefLabel@eng"   | "broader" | "SN@nl"   | "SN@eng" | "SN@fra" |
|-------|--------------------------|---|---|-----------|---|----------|----------|
| 41875 | "Nom de l'objet"         | "Objectnaam"  | "Object name"   |           |   |          |          |
| 41877 | "Architecture"           | "Architectuur"  | "Architecture"  | 41875     | "Topterm voor architecturale elementen en onderdelen v                            |          |          |
| 41879 | "Construction"           | "Bouwwerk"  | "Single built work"   | 41877     | "Algemene benaming voor alle soorten van (i                                       |          |          |
| 45278 | "Composant de bâtiment"  | "Deel van een gebouw"   | "Building division"   | 41879     | "Wordt gebruikt voor het be   |          |          |
| 46319 | "Construction funéraire" | "Funerair bouwwerk"   | "Funerary structure"  | 41879     | "Wordt gebruikt voor  |          |          |
| 42134 | "Monument"               | "Monument"  | "Monument"  | 41879     | "Bouwwerken die getuigen van een (vroegere) cultuur                               |          |          |
| 41878 | "Élément d'architecture" | "Architecturaal element"  | "Architectural element"   | 41877     | "Structurele  |          |          |
| 45650 | "Antéfixe"               | "Antefix"   | "Antefix"   | 41878     | "Dakornamenten, die het open eind van een rij dekte                               |          |          |
| 42134 | "Plaque de couverture"   | "Plaques closing the outer end of the final cover tile in each row of overlapping | "Plaques closing the outer end of the final cover tile in each row of overlapping |           | "Plaques closing the outer end of the final cover tile in each row of overlapping |          |          |

Figure 13 : CSV export of the terminology

Some tools such as Google Refine<sup>25</sup> can be used to help to reorganise the CSV file as presented above before importing it in the TMP. Google Refine helps cleaning and transforming data from one format to another.

The import process can start with uploading the terminology from local directories. A control is done at this point to check whether a CSV file has been uploaded in conformity with what was declared during your initial registration (submitted metadata information).

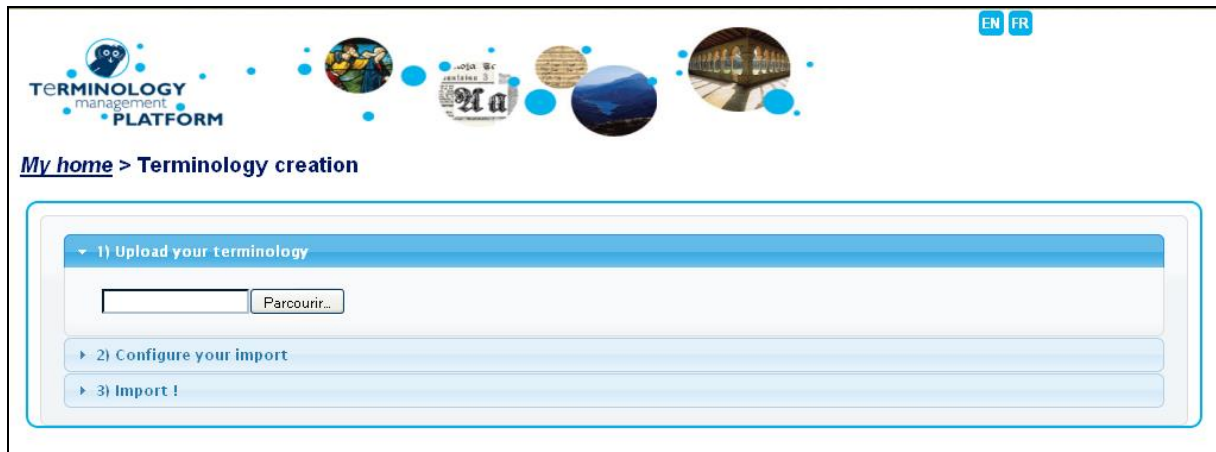


Figure 14 : TMP Import terminology screen

Each column should show an element from the terminology. If the content from the CSV file is not well displayed, the choice of the separator must be revised according to the separator you have chosen for exporting your Excel file as a CSV file. These parameters can be changed answering the question right under the “Configure your import” tab: “Does your import look good”, if not, then click on “no” to change the parameters.



Figure 15 : TMP CSV parameter dialog box

The dialog box above will show up. Once the parameters have been correctly changed, the imported CSV file should display each concept as a line and each SKOS property as a column. The SKOSification process relies on the correct mapping from the CSV input file to the SKOS data model.

<sup>25</sup> <http://code.google.com/p/google-refine/>

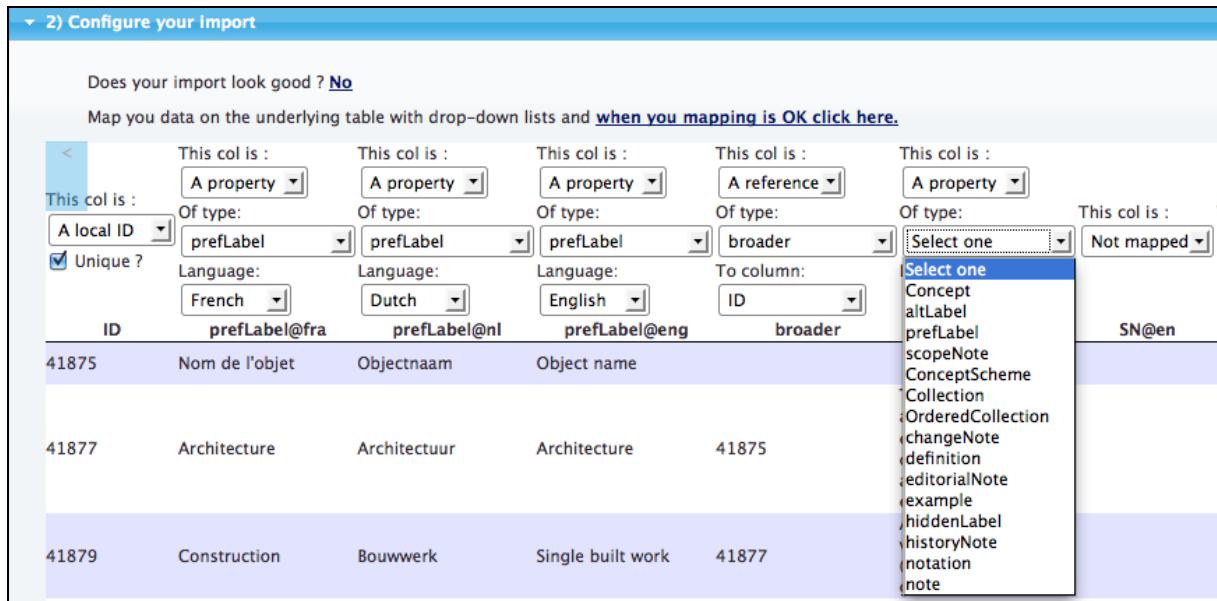


Figure 16 : TMP: SKOSification process

Now that the import is well formed you can proceed with the mapping of your input CSV file with the SKOS properties.

Each column can be a local ID, a property or a reference. In the case of a property, the common skos:properties will be suggested, for example: a concept, labels (preferred, alternative or hidden), notes (scope note, editorial or history note,...). For each property, the language attribute is mandatory. By default, the languages that have been declared during the registration of the terminology in the metadata will be proposed to the user. In the case of a reference, all the properties for SKOS relationships will be suggested, for example: broader, narrower, related; mapping properties such as exactMatch or closeMatch will be also suggested. This mapping completely depends on the structure of the input CSV file and then on the structure of your terminology.

When the mapping is done, the import process can start and the SKOSification will be done simultaneously.

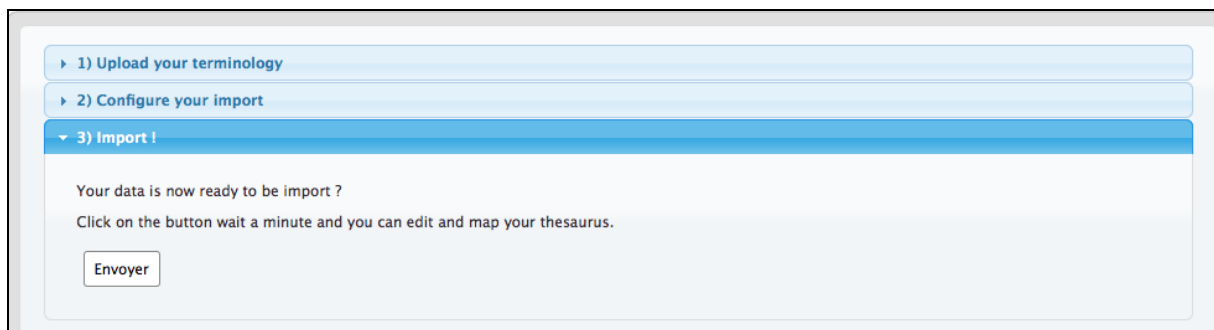


Figure 17 : TMP : Import



### 3.2.12 Edit a terminology

Once the terminology is registered (and imported and SKOSified), it can be edited via the editing module. Only the editor and the owner are able to edit a terminology. The owner is the only user who can create or delete a terminology.

| Title                            | Subject         | Languages | Author       |  |
|----------------------------------|-----------------|-----------|--------------|--|
| <a href="#">Michael Siera</a>    | Performing arts | ar, en    | Siera        |  |
| <a href="#">nm-thesauri</a>      | Decorative arts | cs        | NM           |  |
| <a href="#">Ornamental Art</a>   | Decorative arts | hu        | Julia Katona |  |
| <a href="#">Ornamental Art 2</a> | Decorative arts | en        | Julia Katona |  |

Figure 18 : TMP: access to the editing module

The user can access the Edit icon when browsing his terminologies. The Graphical User Interface has been designed in order to be very intuitive and easy to understand for the users.

TERMINOLOGY management PLATFORM

EN FR  
Antoine

[My home](#) > Edit the terminologie

Title : EUPHOTO - Author : EUPT

English

Linguistic dimension

Labels Documentary notes

preferred label fundamental rights

alternative label  +

hidden label  +

Conceptual dimension



Hierarchical relations Associative relations Mapping relations

has broader politics +

has narrower human rights +  
freedom of the press  
freedom of religion  
civil rights  
censorship







Figure 19 : TMP: Screenshot of the edition of a terminology

The figure above presents the user interface for editing the terminology. The screen is divided into two major blocks. The block on the left displays a tree view of the terminology you have chosen to edit. Insofar as the TMP is concept-oriented, the main access point to the terminology for editing, creating, etc. is the hierarchical structure of concepts. Therefore the left block displays a tree of concepts. The concept will turn orange when the mouse is moved over the concept. Two actions are then suggested, with specific icons:

|   |   |
|---|---|
|  | Move the concept up or down in the tree of concepts |
|  | Delete the concept                                  |

The concepts which have children concepts have their circle tinted in blue in the tree of concepts. The hierarchy of concepts can be modified very easily with a simple drag and drop using the appropriate icon.

On the very left part of the screen some features allow to set the display or modify the terminology.

|   |  |
|---|--|
|    | Change the language of the terminology (the languages that were declared in the metadata form will be suggested to the user) |
|    | Display the metadata that were declared for the terminology as a tool-tip  |
|   | Search for a concept in the terminology  |
|  | Add a new concept at the top level of the tree of concepts   |
|  | Add a child concept to the selected concept  |
|  | Edit metadata, save and return to the general editing screen   |

When a concept is selected, the selected concept is then displayed in green and the related information is displayed in the right block with two specific dimensions. Indeed a clear separation is done between the conceptual and the linguistic dimensions. As it is reminded in the D4.3 dedicated to this first release of the TMP, a concept is an abstract idea which is language independent and defined by the semantic relations that link it to other concepts.

- The linguistic dimension of the terminology is displayed in a dedicated frame for 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 is displayed in another dedicated frame for the SKOS semantic relationships e.g. the hierarchical relationships (broader and narrower) and the associative relationship (related). The mapping relations resulting from the mapping procedure will be displayed on the dedicated tab.

The flag icon on the left block will allow the user to change the language of the tree of concepts (according to the languages that were declared in the metadata) but the language can also be changed in the right block. This distinction between the language of the tree of concepts and the one for editing/enriching the selected concept is foreseen in order to enable the multilingualism of the terminology. Such a feature allows defining labels in a different language from the language used for ontology building.

Regarding the conceptual dimension, the TMP logical properties define 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 user only the concepts that verify these properties.

Modifying the labels and the documentation on the right part of the screen only requires a double click on each field to become editable.

One might need to edit a terminology to modify its structure (e.g. change the hierarchy of concepts), add or modify labels that are expressing a concept, or add a translation or documentation to a concept.

It is important to note that each modification is automatically saved, there is no need to save the work done. The undo icon allows canceling the last actions.

### 3.2.13 Exporting a terminology

Once the edition of the terminology is finished, the user can get back to the list of his terminologies thanks to the breadcrumb trail. The breadcrumb is available from any page of the TMP and allow the user to get back to his own user homepage ("My home"). The owner, editor and contributor can then be able to export the terminology.

[My home](#) > search public terminologies

Figure 20 : TMP Breadcrumb



is the icon representing the export functionality. The terminology will be exported in SKOS as a whole; the mapping elements will also be included in the export.

```
-<rdf:RDF>
- <rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/e8a73120-2f27-4b89-8067-3a0f32ce9188">
  <j.0:broader rdf:resource="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/76c932d7-b62f-48d6-810d-5de19019909d"/>
  <j.0:prefLabel xml:lang="hu">liliom</j.0:prefLabel>
  <j.0:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art"/>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
</rdf:Description>
- <rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/0277e97b-e37d-42af-8a4f-410625b1da79">
  <j.0:broader rdf:resource="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/595a69be-5276-4cfd-9059-31057df81613"/>
  <j.0:prefLabel xml:lang="hu">láng</j.0:prefLabel>
  <j.0:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art"/>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
</rdf:Description>
- <rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/0b441f39-0074-4611-a406-99550556919b">
  <j.0:narrower rdf:resource="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/216fe432-1919-432d-9eee-877eb7641970"/>
  <j.0:narrower rdf:resource="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/621a9621-bebf-4a71-a845-ab6dd16c313b"/>
  <j.0:narrower rdf:resource="http://www.culture-terminology.org/thesaurus/Julia%20Katona/Ornamental%20Art/e6bedcaa-52f7-49d5-83e3-f7978f3bc7d8"/>
```

Figure 21 : TMP : SKOS export

A SKOS export can be very useful to any application which could use a SKOSified terminology to improve a search and retrieval system. For a professional from a cultural institution it could be interesting to export his SKOSified terminology, enriched with all the mapping links done within the TMP to be reimported in his own collection management system.



### 3.2.14 Linking multiple terminologies (mapping)

The SKOS data model provides several mapping properties for making alignment between concepts from different concept schemes. Within TMP, we can consider that each terminology from an institution is a SKOS:ConceptScheme. Therefore the SKOS mapping properties will be used for linking two concepts from different terminologies. Although the mapping happens in a one-to-one way, eventually several one-to-one mappings can be performed and this way, several terminologies can be linked using the TMP.

TMP provides a graphical tool for the mapping process. The graphical user interface has been defined in order to be the most intuitive possible since the concepts can be mapped together easily with a simple drag and drop and the type of mapping can also be specified.








My home > Map terminologies

The screenshot displays the TMP mapping tool interface. At the top, two columns represent the terminologies to be mapped. The left column is titled "Title : PICO%20Thesaurus - Author : m bac" and the right column is titled "Title : Euphoto - Author :". Below the titles are icons for document, flag, information, and search. The left column contains a hierarchical tree of concepts, with "Italian Government" highlighted in blue. The right column contains a similar tree, with "national government" highlighted in green. A line connects the two highlighted terms. In the center, a dialog box titled "Select the type of the link" is open, showing a dropdown menu with "Select a value" and a "clickMatch" button with a trash icon. At the bottom, two panels show the mapping details for each term. The left panel shows "preferred label: Italian Government" and the right panel shows "preferred label: national government". Both panels also have fields for "alternative label" and "hidden label".

Figure 22 : TMP: mapping

The figure above illustrates the mapping tool of the TMP. The terminology of the institution will be displayed on the left and the terminology to be mapped to will be displayed on the right. The title and the author of the terminologies to be mapped are displayed on the top of the columns. The icons right under the title of the terminology can perform the following actions:

|   |  |
|---|--|
|  | Upload a terminology for mapping   |
|  | Change the language of the terminology (the languages that were declared in the metadata form will be suggested to the user) |
|  | See the metadata of the terminology  |
|  | Search for a concept in the terminology  |
|  | Add a new concept  |

When a concept is selected, a block displaying the details on the selected concept appears on the bottom of the column. The information on the concept that is displayed in that block is separated in different tabs for the labels, the documentary notes, hierarchical relations, associative relations and mapping relations. We will see in the edit part that each concept is displayed in this way. This information on the selected concept can help the user during the mapping process as it can reduce the ambiguity or guide the user to define the type of link that will be chosen.

The procedure to make the link is quite easy since the user must select the concept in his terminology (left column) and the concept to be mapped with in the other terminology (right column) and then a pop up box will suggest the type of link that must be made between the two selected concepts.



**Figure 23 : TMP: dialog box for the mapping link selection**

Once the mapping is done the link between the two concepts appears in grey.

The search icon available for each of the terminologies opens for the mapping. This allows searching for concepts that can be mapped together. The semi-automated mapping will also be accessible from the search icon which will calculate the proximity of two concepts in the same language.

### 3.2.15 Semantic Web and Linked Data

We have started this user manual with an introduction to SKOS. But it is important to have the whole framework of the Semantic Web to understand what it is its general purpose and how this context is evolving thanks to technologies and organisational initiatives.

Before introducing the Semantic Web it could be worth reminding what is the Web. Tim Berners-Lee<sup>26</sup> is well known as the creator of the Web since he had the vision and developed a first web client and server

<sup>26</sup> <http://www.w3.org/People/Berners-Lee/>

in 1990. The Web stands actually for World Wide Web, abbreviated in WWW. The Web is a system of interlinked hypertext documents accessed via the Internet. The Web relies on three principles:

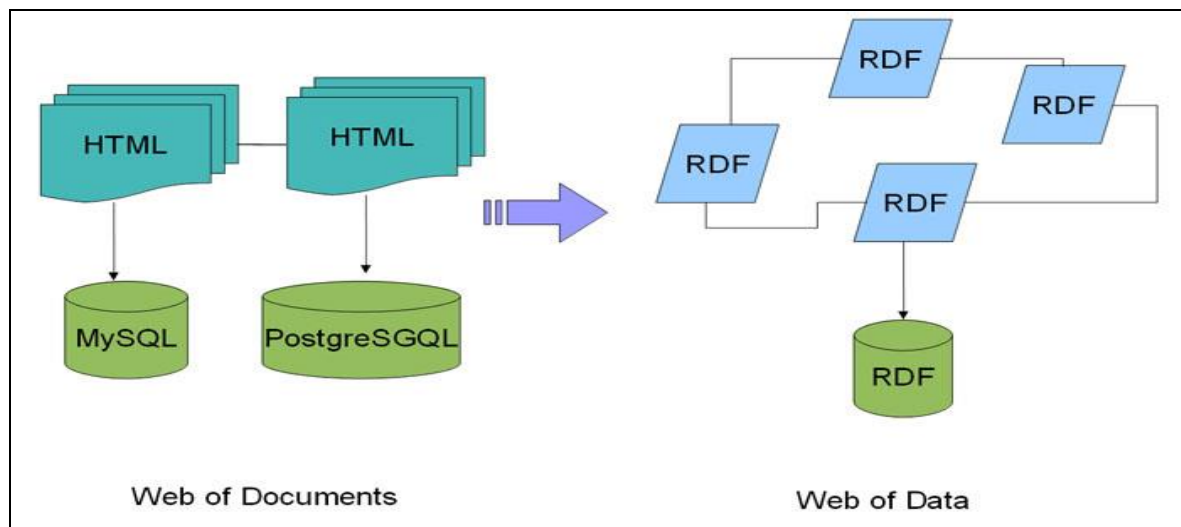
- Each document/Website is accessed from a unique point identified with a URL (Uniform Resource Locator)
- The publishing language used is HTML (Hypertext Markup Language)
- HTTP (HyperText Transfer Protocol) is the protocol used to query the Internet.

These specifications defined by Berners-Lee have contributed to spread the Web. Thanks to a Web browser, human users are able to read documents through the hypertext web. As nowadays the hardware solutions are bigger, faster and more performant, Tim Berners-Lee had imagined the Semantic Web for improving this Web and its data process. The Semantic Web is not a new version of the Web but an evolution. Tim Berners-Lee expressed in 2001 the idea of the Semantic Web as follows:

*"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation." It is a source to retrieve information from the web (using the web spiders from RDF files) and access the data through Semantic Web Agents or Semantic Web Services.*<sup>27</sup>

The principle is to maximize the possibilities of information retrieval through the Web thanks to technologies and structuration of information. The Web become semantic by using these technologies since machines can interpret and understand the meaning of data. The way to get to this understanding is to put available on the Web semantically enriched and structured data.

With the Web in its simple form, we were in the Web of Documents where only documents have an identifier (URLs) and only human beings were expected to read these documents. With the Semantic Web, we are no longer in the Web of Documents but in the Web of Data where human beings can read the data displayed in their Web browser and machines can also process these data for providing higher quality services to the users.



**Figure 24: Web of Documents versus Web of Data**

Moving from the Web of Documents to the Web of Data allows to structure data semantically. Where one document was identified by its URL, many data in one document are now identified by unique identifiers, URI (Uniform Resource Identifier).

HTML describes documents and the links between them while other coding systems such as RDF (Resource Description Framework), XML (eXtensible Markup Language) or OWL (Web Ontology Language) are specifically designed for data description and organisation. RDF, XML or OWL can be used to describe arbitrary things or concepts.

<sup>27</sup> Source: "The Semantic Web" by Sir Tim Berners-Lee, James Hendler, and Ora Lassila, Scientific American. 2001.

With things and concepts semantically structured in documents, computers are able to infer the meaning from content on the basis of metadata and assertions that have been already made. Semantic Web is usually called Web 3.0 or Web of Linked Data. Web 3.0 might suggest that it is subsequent to the Web 2.0, eg. the social web with its collaborative aspects but in fact Semantic Web and Social Web are evolving and growing in parallel.

### 3.2.15.1 *Linked Data, Open data and Linked Open Data*

These three expressions are often used to refer to Linked Data. Linked Data is really the core of the Semantic Web: this is the linked data that allows machines to infer meaning from the data. Data that are not connected or mapped with other related data would be meaningless and not usable for machines. Linked data is not mandatorily open. "Open data" is used to describe data that are freely available through the Web. The notion of Open Data only refers to data with open licenses, but it does not imply that these data are also linked. In the case of a combination of data that are open and linked, the expression Linked Open Data would be appropriate.

Linked Data has been defined as follows by Tim Berners-Lee:

*" The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data."*<sup>28</sup>

It is important to make clear that Semantic Web and Linked Data rely on specific technologies but there are also and most of all a set of best practices for publishing and connecting structured data on the Web. At this point in time, only humans are able to make successful links between data in order to get relevant results. Tim Berners-Lee describes Linked Data as "Semantic Web done right"<sup>29</sup>. Technologies such as URI, RDF and HTTP mentioned above are the key components to implement these best practices.

Tim Berners-Lee has identified four principles that should be respected in order to provide Linked Data:

- ★ Use URIs as names for things
- ★★ Use HTTP URIs so that people can look up those names
- ★★★ When someone looks up a URI, provide useful information, using the standards (RDF\*, SPARQL)
- ★★★★ Include links to other URIs so that they can discover more things

These four principles clearly state that each data should be identified by a URI and that the use of standards such as RDF can help make inference from the links that are made between two data thanks to their unique identifier (URI).

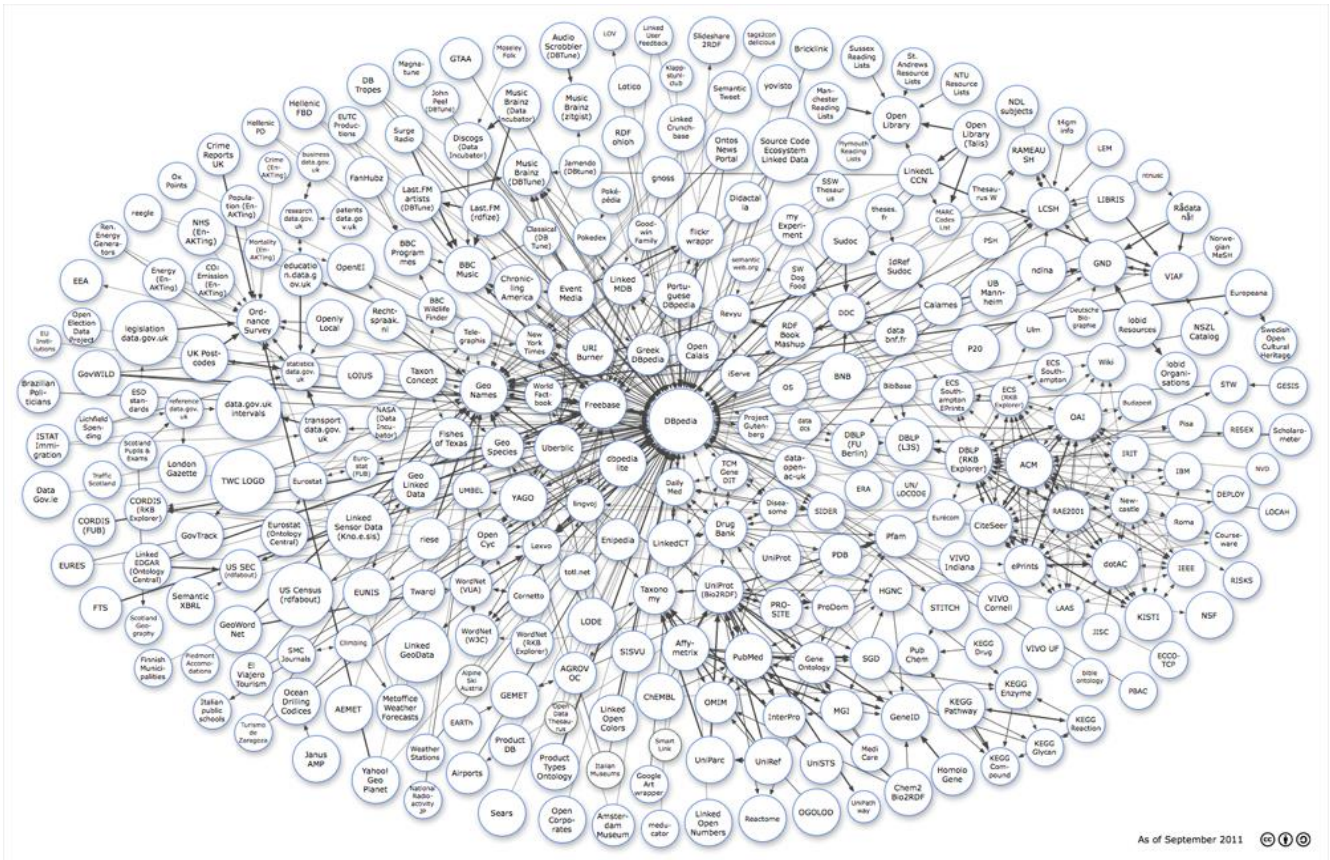
The figure below is a well-known representation of this Linked Open Data cloud.

---

<sup>28</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

<sup>29</sup> <http://linkeddata.org/faq>





**Figure 25 : Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch.**  
<http://lod-cloud.net/>

### 3.2.15.2 The Semantic Web Stack

The Semantic Web Stack<sup>30</sup>, also known as the Semantic Web Cake has been created by Tim Berners-Lee and illustrates the architecture of the Semantic Web. It should be read from the bottom to the top as it describes the standards and technologies that are achieved and that still need to be developed to reach the full vision of the Semantic Web.

<sup>30</sup> <http://www.w3.org/2006/Talks/1023-sb-W3CTechSemWeb/#%2819%29>

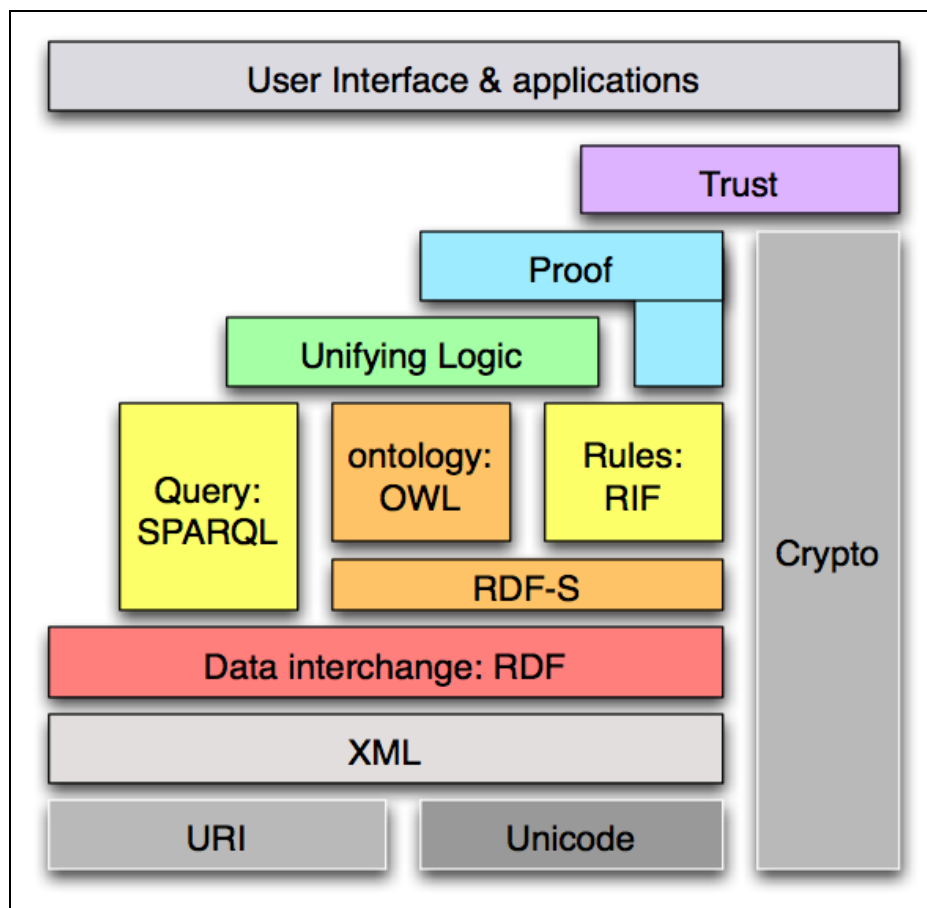


Figure 26 : The Semantic Web Stack

## ➤ URI

URI stands for Uniform Resource Identifier. This is the foundation of the Semantic Web as without URI, things and concepts cannot be named and located in a unique way. URIs are strings of characters used to identify a name of a web resource. URLs (Uniform Resource Locator) and URNs (Uniform Resource Name) are specific kinds of URIs.

The role of URIs is to provide a unique and unambiguous name to a resource and to locate it on the Web.

URLs are used to locate documents, images, or any kind of file on the Web. URIs' scope is wider since each resource can be identified and a resource can either be a kind of file, a thing or a concept.

This principle is important especially in the cultural field as every cultural object or collection is usually attributed with an identifier in a collection management system. Moving from this local identifier understandable only by the collection management system that generated it, towards a URI that can be processed by any system is the first key step to get into the Semantic Web.

Applying a URI to a thing or a concept implies that this URI would be unique and persistent. Choosing a URI implies specific technologies to be able to provide a unique identifier. This identifier might be explicit or opaque, the costs for these identifiers should be free of charge. Specific technologies have been developed in order to provide persistent identifiers, each of them have their particularities that cultural institutions should consider. ARK (Archival Resource Key)<sup>31</sup>, PURL (Persistent URL)<sup>32</sup> or DOI (Digital Object Identifier)<sup>33</sup> are well known technologies and organisations providing persistent identifiers. Each of these systems offer specific features therefore cultural institutions do have to think which system would be more appropriate for their digital objects and collections and its own internal policy. The Athena booklet

<sup>31</sup> ARK : <https://confluence.ucop.edu/display/Curation/ARK>

<sup>32</sup> PURL : <http://purl.oclc.org/docs/index.html>

<sup>33</sup> DOI : <http://www.doi.org>

“ Persistent identifiers: recommendations”<sup>34</sup> and the Linked Heritage deliverable D2.2 “State of the art report on persistent identifier standards and management tools”<sup>35</sup> presents the different systems available and give some best practice recommendations on how to use a PID service.

The Linked Heritage D2.2 states that A PID service should fill in the following requirements :

- “Evaluate and be assured of the technical reliability of a PID system (including their own) before adopting it;
- Evaluate and be assured of the authority and credibility of a PIDs service’s provider before adopting that system;
- Make sure that the service it uses is flexible enough to represent the granularity their collections;
- Make sure that the service uses intellectually open standards for the implementation of PIDs.”

The issue for persistent identifiers is not only technical since the maintenance, the costs for maintenance and the sustainability of the organisation providing PIDs are crucial for the uniqueness and the persistency of the identifiers.

### ➤ **Unicode**

Unicode<sup>36</sup> is the universal character encoding, maintained by the Unicode Consortium. Before Unicode was developed many encoding systems were in use. An encoding system attributes a number for each character to store it. These various encoding systems were different according to the programs where they were used, the platform or the language they were covering. Unicode has brought a universal and standard encoding system since each character has a unique number. Unicode covers all the characters for all the writing systems of the world, modern and ancient. It also includes technical symbols, punctuations, and many other characters used in writing text. Unicode is then able to support a huge number of languages especially for languages with non-latin characters.

The most famous Unicode encoding systems are UTF-8 and UTF-16. UTF-8 and UTF-16 stand for Unicode Transformation Format 8 or 16 bit. Each number given by Unicode is converted to fit in 8 or 16 bits according to the encoding system chosen.

Unicode is another important building block of the Semantic Web since it facilitates the multilingualism but also the basic interoperability between different systems (programs, platforms, etc...)

### ➤ **XML**

XML stands for eXtensible Markup Language. XML<sup>37</sup> is a markup language for documents containing structured information. A markup language is a mechanism to identify structures in a document. The XML specification defines a standard way to add markup to documents. HTML is also a markup language but all the tags are predefined so a Web browser can interpret these tags and their content in an appropriate way to display documents. XML provides a strong support to Unicode for the languages and scripts all over the world. XML is used for structuring documents but also for representing arbitrary data structures that are used by Web services and/or programming interfaces. This is the case of SKOS or RDF that may be represented in an XML file.

The syntax of an XML document can be quite free as the creator of the document defines himself the tags that will be used for describing and structuring this document. Thus it has to respect at least one principle, which is to be ‘well formed’. Each opening tag has to be closed, each space or special character has to be considered carefully in order to avoid any syntax error. A DTD (Document Type Definition)<sup>38</sup> or an XML schema<sup>39</sup> can be applied to the XML document so it can state which will be the tags used in the document and which will be the rules for these tags to obtain a valid document.

---

<sup>34</sup> Persistent identifiers: recommendations, 2011 : <http://www.athenaeurope.org/getFile.php?id=779>

<sup>35</sup> Linked Heritage D2.2 :State of the art report on persistent identifier standards and management tools

<sup>36</sup> Unicode : <http://www.unicode.org>

<sup>37</sup> XML is a standard acknowledged by the W3C and its specification has been designed in the recommendation of the 26th of november 2008. See <http://www.w3.org/TR/REC-xml>

<sup>38</sup> DTD : <http://www.w3schools.com/dtd>

<sup>39</sup> XML Schema : <http://www.w3schools.com/schema>

```
<books>
- <book>
  <title xml:lang="en">Romeo and Juliet</title>
  <title xml:lang="fr">Roméo et Juliette</title>
  <author>William Shakespeare</author>
</book>
- <book>
  <title xml:lang="fr">Crimes et châtements</title>
  <author>Victor Hugo</author>
</book>
</books>
```

Figure 27 : Example of XML

XML is the bottom of the Semantic Web stack since RDF relies on an XML-based serialisation and also because it supports fully Unicode.

## ➤ RDF

RDF stands for Resource Description Framework and it is a standard model for data interchange on the Web. RDF is used as a conceptual description or a modelling of information that is implemented in Web resources, using a variety of syntax formats. As defined by the W3C, RDF is “a general-purpose language for representing information in the Web”<sup>40</sup>. First RDF was intended to represent metadata about web resources with the very common details such as title, author, date, etc... but it has been extended to describe anything that can be identified on the Web even if this thing cannot be retrieved directly from the Web, for example arbitrary things such as concepts or a person.

RDF is not intended for human reading but it is clearly meant to be used and processed by applications. As RDF provides a common framework for modelling information, any application should be able to process this information even if those applications were not the one which generated this information.

What we call information is represented as statement in RDF in the form of triple. All the resources are represented as simple statements with simple properties and property values. Triples are the expressions of statements about resources which are presented as subject-predicate-object expressions.

Considering this, the sentence “William Shakespeare wrote the drama ‘Othello’ ” will be turned into a simple statement as “‘Othello’ has a creator whose value is William Shakespeare. As each subject, predicate and object have to refer to an URI, this statement can also be represented as follow:  
<http://en.wikipedia.org/wiki/Othello> has a creator whose value is <http://en.wikipedia.org/wiki/Shakespeare>

The combination of these different statements creates an RDF graph.

RDF-S stands for RDF-Schema and is directly connected to RDF as it sets the RDF vocabulary. This vocabulary defines classes and properties that may be used in the RDF. RDF-schema<sup>41</sup> works as XML-schema works for XML.

<sup>40</sup> RDF : <http://www.w3.org/RDF>

<sup>41</sup> RDF-Schema : <http://www.w3.org/TR/rdf-schema>



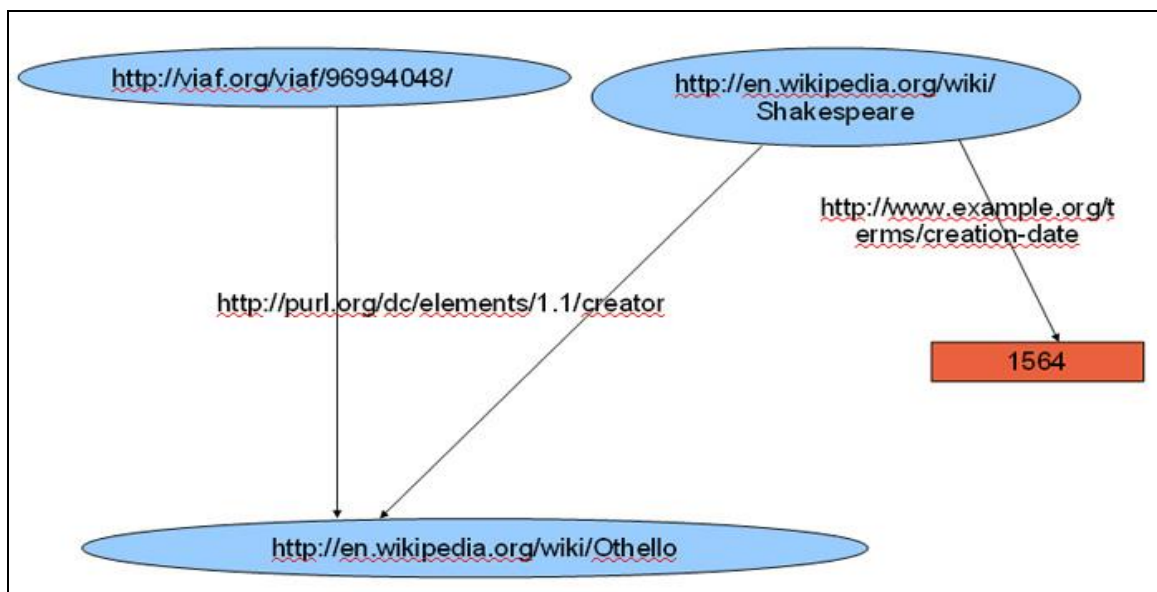


Figure 28 : Example of an RDF graph

RDF is described as “data:interchange” in the Semantic Web Stack since it provides the technical interoperability needed to enhance the Semantic Web. It is also one of the building blocks as it relies on URI and is represented in a XML-based file. RDF defines a specific XML markup language, referred to as RDF/XML, for use in representing RDF information, and for exchanging it between applications.

OWL or SKOS build on RDF and provide language for defining structured, Web-based ontologies which allow richer integration and interoperability of data among descriptive communities.

### ➤ OWL

OWL stands for Web Ontology Language and is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is designed specifically for the Semantic Web as it is developed as a vocabulary extension of RDF. OWL is a recommendation of the W3C since February 2004, it has been revised and published in a second version as OWL2 in December 2012 as mentioned in the OWL2 quick reference document produced by the W3C<sup>42</sup>. The aim of OWL is to build an ontology for the Web. According to the W3C recommendation<sup>43</sup>, an ontology defines “*the terms used to describe and represent an area of knowledge. Ontologies are used by people, databases, and applications that need to share domain information (a domain is just a specific subject area or area of knowledge, like medicine, tool manufacturing, real estate, automobile repair, financial management, etc.). Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them*”.

An OWL ontology is an RDF graph, therefore it respects the same syntax as RDF and the same structure in triples. OWL2 ontologies provide classes, properties, individuals, and data values and are stored as Semantic Web documents. OWL2 ontologies can be used along with information written in RDF, and OWL2 ontologies themselves are primarily exchanged as RDF documents.

The specificity of OWL relies on the classes, instances of classes and relations between these instances. The W3C recommendation<sup>44</sup> describes classes as follow:

*“Classes provide an abstraction mechanism for grouping resources with similar characteristics. Like RDF classes, every OWL class is associated with a set of individuals, called the class extension. The individuals in the class extension are called the instances of the class. A class has an intensional meaning (the underlying concept) which is related but not equal to its class extension. Thus, two classes may have the same class extension, but still be different classes.”*

<sup>42</sup> OWL : <http://www.w3.org/TR/2012/REC-owl2-quick-reference-20121211/>

<sup>43</sup> OWL use cases and requirements : <http://www.w3.org/TR/webont-req/>

<sup>44</sup> <http://www.w3.org/TR/2004/REC-owl-ref-20040210>

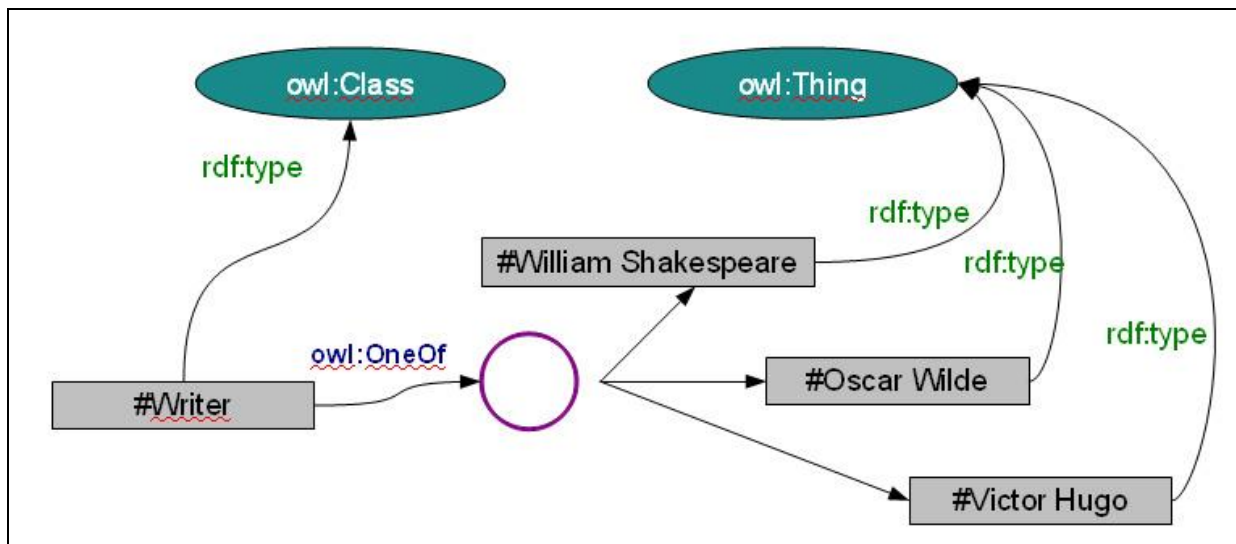


Figure 29 : Example of an OWL graph

In the example above, we can see that Writer is an extension of the class owl:Class, while William Shakespeare or Oscar Wilde are extensions of the owl:Thing class.

Like each part of the Semantic Web Stack, OWL is tightly connected to RDF as it provides a specific ontology vocabulary for RDF.

## ➤ RIF

RIF<sup>45</sup> stands for Rule Interchange Format and is a recommendation from the W3C since December 2012. As it is very recent, RIF is not as well-known as technologies such as OWL or RDF.

Each language has its own rules for declaration and production. RIF is intended with the same normalisation purpose than the Semantic Web as a whole.

RIF is a standard using a verbose XML syntax that has been developed in order to facilitate ruleset integration and synthesis. RIF is based on various versions that are called dialects and that are interoperable. RIF is also compatible with RDF and OWL which allows semantic interoperability in a normalised way.

For example, let us consider the following statement:

- If a writer writes a drama then he is a dramatist;
- "Othello" is a drama
- William Shakespeare wrote "Othello"

The fact that Shakespeare is a dramatist can logically be inferred from this statement. The translation of this statement into a rule will provide a generalisation where the instances "Othello" and Shakespeare will become logical variables and then logical assertions can be inferred from this rule.

```

Document(
  Base(<http://example.com/people#>)
  Prefix(cpt <http://example.com/concepts#>)
  Prefix(drmst <http://example.com/dramatist#>)
  Prefix(drm <http://example.com/drama#>)

```

```

Group
(
  Forall ?Writer ?Item (
    cpt:writeDrama(?writer ?Item)
  )
)

```

<sup>45</sup> RIF : <http://www.w3.org/TR/2013/NOTE-rif-primer-20130205/>

```
cpt:writeDrama(<William Shakespeare> drm:Othello)
)
```

With this general rule, William Shakespeare and “Othello” could be replaced by Oscar Wilde and “Salomé”, then the same inference could be made.

### ➤ SPARQL

SPARQL<sup>46</sup> stands for SPARQL Protocol and RDF Query Language. SPARQL is a query language for RDF, which means that it can browse and retrieve information that is stored in databases supporting RDF. SPARQL is a W3C recommendation since January 2008. SPARQL syntax reminds of other databases query language such as SQL. SPARQL allows formulating simple queries through simple RDF graphs but also very complex queries through very complex RDF graphs. As RDF, a SPARQL query result is not intended to human reading. Results have to be processed thanks to stylesheets to be displayed on a Web browser.

As any query language SPARQL is composed of a verb, clauses and variables. SPARQL offers 4 types of query:

- SELECT
- CONSTRUCT
- ASK
- DESCRIBE

SELECT is the most commonly used query form as it allows returning all results matching the query pattern. The clause WHERE is used for all the types of query and provides the graph pattern that should match the data graph.

SPARQL is part of the Semantic Web Stack since it is the normalised protocol for querying data available in RDF format. The SPARQL query language is very powerful as it can retrieve all data from an RDF database or very precise information according to a complex query. Web services providing data compliant with Semantic Web technologies are expected to provide a SPARQL endpoint.

---

<sup>46</sup> SPARQL : <http://www.w3.org/TR/rdf-sparql-query/>

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)  
http://dbpedia.org

Query Text  
select distinct ?Concept where {[ ] a ?Concept} LIMIT 100

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: HTML (The CXML output is disabled, see [details](#))

Execution timeout: 30000 milliseconds (values less than 1000 are ignored)

Options:  Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

Figure 30 : SPARQL endpoint from DBPedia (<http://dbpedia.org/sparql>)

The screenshot above presents the SPARQL endpoint provided for Dbpedia<sup>47</sup>, the Semantic Web version of Wikipedia. Options for the display of the results are proposed.

The query is composed of the verb “SELECT” , the variable “?Concept” and the clause “WHERE” is presented between the brackets {}. This query basically lists all the pages with different URI. The result must be limited to 100 (LIMIT 100).

Here are the results provided by this query :

<sup>47</sup> See <http://dbpedia.org/>

| Concept   |
|---|
| <a href="http://www.w3.org/2002/07/owl#Thing">http://www.w3.org/2002/07/owl#Thing</a>   |
| <a href="http://dbpedia.org/ontology/Agent">http://dbpedia.org/ontology/Agent</a>   |
| <a href="http://dbpedia.org/ontology/Band">http://dbpedia.org/ontology/Band</a>   |
| <a href="http://dbpedia.org/ontology/Organisation">http://dbpedia.org/ontology/Organisation</a>   |
| <a href="http://schema.org/MusicGroup">http://schema.org/MusicGroup</a>   |
| <a href="http://schema.org/Organization">http://schema.org/Organization</a>   |
| <a href="http://dbpedia.org/class/yago/MusicalGroupsDisestablishedIn1986">http://dbpedia.org/class/yago/MusicalGroupsDisestablishedIn1986</a> |
| <a href="http://dbpedia.org/class/yago/DanceBand108249960">http://dbpedia.org/class/yago/DanceBand108249960</a>                               |
| <a href="http://dbpedia.org/class/yago/Organization108008335">http://dbpedia.org/class/yago/Organization108008335</a>                         |
| <a href="http://dbpedia.org/class/yago/RockGroup108250501">http://dbpedia.org/class/yago/RockGroup108250501</a>                               |
| <a href="http://dbpedia.org/class/yago/YagoLegalActor">http://dbpedia.org/class/yago/YagoLegalActor</a>                                       |
| <a href="http://umbel.org/umbel/rc/Band_MusicGroup">http://umbel.org/umbel/rc/Band_MusicGroup</a>   |
| <a href="http://dbpedia.org/class/yago/Abstraction100002137">http://dbpedia.org/class/yago/Abstraction100002137</a>                           |
| <a href="http://dbpedia.org/class/yago/BritishPunkRockGroups">http://dbpedia.org/class/yago/BritishPunkRockGroups</a>                         |
| <a href="http://dbpedia.org/class/yago/Group100031264">http://dbpedia.org/class/yago/Group100031264</a>                                       |
| <a href="http://dbpedia.org/class/yago/MusicalGroupsEstablishedIn1981">http://dbpedia.org/class/yago/MusicalGroupsEstablishedIn1981</a>       |
| <a href="http://dbpedia.org/class/yago/MusicalOrganization108246613">http://dbpedia.org/class/yago/MusicalOrganization108246613</a>           |
| <a href="http://dbpedia.org/class/yago/SocialGroup107950920">http://dbpedia.org/class/yago/SocialGroup107950920</a>                           |
| <a href="http://dbpedia.org/class/yago/YagoLegalActorGeo">http://dbpedia.org/class/yago/YagoLegalActorGeo</a>                                 |

Figure 31 : Results of a SPARQL query

➤ **The top of the cake**

The upper part of the stack still has to be implemented or at least normalised. The unifying logic and proof layers intend to enhance the usability and usefulness of the Web and its interconnected resources. The top part of the stack is less technical and is more related to usability and institutional aspects. **Unifying logic** implies that a logical reasoning is implemented so that inference of new facts and check of consistency is possible. The **proof layer** would explain the logical reasoning steps. The cryptography is intended to protect the RDF data. For example RDF data can be protected thanks to a digital signature. This would ensure the reliability of the data and help enabling trustfulness towards the data and the links that are made between these data. The user interface & applications layer is the final step to achieve the vision of the Semantic Web. For the time being, tools and data are made available in a machine-understandable format but there are not really fully Semantic-Web-based applications dedicated to human users. Semantic Web implies that structured data is interconnected with other data and that meaning could be inferred by these links; if an institution makes available its data as Linked Data and spurious links are made from these data, this could attempt the image of the institution and its trustworthiness. This is why these parts of the cake are more institutional and related to the use and production of data than technical.

### 3.3 Online material



We present here the online documentation that will be integrated in the static pages of the TMP. This documentation consists of a basic 'help online' feature, which would point to the detailed documentation presented above; a FAQ section; a section dedicated to licenses and finally some video tutorials.


#### 3.3.1 Help online

The help online would briefly give some hints on how to proceed with the actions described below within the TMP. Here we present a first draft of what could become an online helpdesk<sup>48</sup>. This idea will evolve on the basis of the user feedback gathered during upcoming training sessions.

##### *How to create a concept ?*

There are two ways to proceed with the creation of a concept within a terminology. First you have to edit the terminology. Then you choose one of these methods:

Method 1 : You add a new concept using the  icon and then you move it to the relevant hierarchy using the  icon that appears when you mouse over the concept that needs to be moved.

Method 2 : You select an existing concept, and under the selected one a child concept can be added. Click the icon  to add a child concept.

##### *How to make a mapping ?*

A 'mapping' refers to a link between two concepts from two different terminologies (e.g. two different concept schemes). The concepts that you want to map together have to be equal (skos :exactMatch) or similar (skos :closeMatch) or have a more general or specific relation (skos :broadMatch / skos :narrowMatch).

First you have to open the terminologies where each of the concepts to be mapped are. If you have a precise idea of the concepts to be mapped, you can upload your terminology on the left and search for the concept to be mapped on the right. It is possible to search through metadata and concepts as well.


Once you have defined which are the concepts to be mapped together you select them in the two trees of concepts and the dialog box presented below will ask you to define the type of link that should be made between these two concepts.



##### *How to modify the hierarchy of my terminology ?*

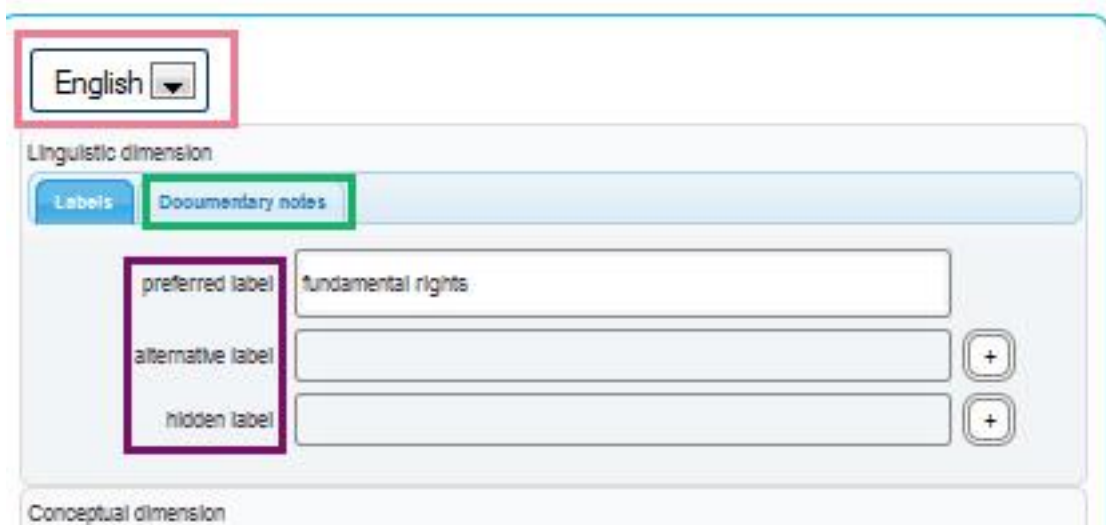
<sup>48</sup> You could add here in a footnote who would be staffing this online helpdesk (e.g. MCC, KMKG, others, ...)



You open your terminology and search for the concept to be moved. Once selected, the concept can be moved using the  icon. If the concept you wish to move contains children concepts, they will all be moved with their parent concept.

### *How to provide a translation and/or additional documentation to my concepts ?*

First you edit the terminology and select the concept that is going to be enriched with translations and/or documentation. Then on the right part of the screen, a drop-down list (framed in red on the figure below) is available to change the language of the description so labels in the relevant language can be added. Labels and documentation of concepts belong to the linguistic dimension of the concept. Therefore this information is language dependant. You have to check that the language that is set is really the language that will be used for adding new labels or new notes.



The screenshot shows the 'Linguistic dimension' section of the TMP interface. At the top, there is a dropdown menu set to 'English', which is highlighted with a red box. Below this, there are two tabs: 'Labels' and 'Documentary notes'. The 'Documentary notes' tab is selected and highlighted with a green box. Underneath, there is a table with three rows: 'preferred label', 'alternative label', and 'hidden label'. The 'preferred label' field contains the text 'fundamental rights'. To the right of the 'alternative label' and 'hidden label' fields, there are plus signs in circles, indicating that new labels can be added. Below the linguistic dimension section, the 'Conceptual dimension' section is partially visible.

### **3.3.2 FAQ (Frequently Asked Question)**

It has become very common for any web service to propose a Frequently Asked Questions page to the users who might need quick and sharp information on which features/services the website offers. In the case of the TMP, the FAQ section could cover two levels :

- questions and answers about the Semantic Web
- questions and answers about the TMP

Here is a first version of these FAQs which will evolve according to the user feedback and the questions they might ask.

### *What is the TMP?*

The TMP is a set of tools that has been developed within the Linked Heritage and AthenaPlus project to provide an intuitive interface for managing a terminology using Semantic Web technologies. TMP is free to use but registration is necessary for creating, SKOSifying, editing, mapping or publishing terminologies.



### *What is a terminology?*

Under the name “terminology” we consider any kind of vocabulary. In the framework of the TMP, we consider simple lists of terms, thesauri, classifications, ontologies, subject headings, and any kind of controlled vocabularies as a terminology.

### *Who is it for?*

The TMP is dedicated to professionals from the GLAM sector (Galleries, Libraries, Archives and Museums) but is open to any person or organisation willing to create, manage, share and publish their terminologies.

### *What is the Semantic Web?*

The semantic Web is an extension of the current Web that intends to propose structured and meaningful content that can be understood by a machine. A machine that can understand the meaning of data will be able to process better these data and propose extended results for a simple query.

### *What is Linked Data?*

Linked Data is the principle of interconnecting data available in a Semantic Web compliant format. It is this interconnection between data that brings additional meaning to these data.

### *I am not a professional but I would like to use some terminologies via the TMP, can I?*

Yes you can! TMP is open to any user. You can create your own terminology and view all the terminologies that are public. You can make suggestions to improve a terminology, or get in contact with the administrator or owner of the terminology in the TMP.

### *Can I trust the data available on the TMP?*

The fact that the majority of TMP users belongs to an institution, helps the owners and administrators to evaluate the suggestions (e.g. for modification) that are made on their terminologies and it contributes to ensure the trustworthiness of the data stored on the TMP.

### *Why should I do Semantic Web stuff?*

Because it is the web of tomorrow with powerful machines and technologies that can understand content available on the Web. Providing your terminologies in a format compliant with the Semantic Web will help machines and users understand better your knowledge and expertise on a domain. The metadata on your objects and collections will then not just remain simple metadata, but structured and meaningful metadata. Semantic Web technologies can also help you enhancing the multilingualism of your terminologies and collection registration.

### *What is a concept?*

A concept is an abstract idea. It is expressed by terms that may exist in different languages. The notion of concept is crucial from the perspective of the Semantic Web because the Semantic Web implies that each resource is unique and unambiguous but natural languages are a natural source of ambiguity. For example you can get the abstract idea behind the following terms “school”, “education”, and “teacher”. The concept is independent from the natural language in which it is expressed; for example the term “school” in English and the term “école” in French express the same concept and the definition “lugar para estudiar” in Spanish can be used to define this same concept.

### *What is the difference between a term and a concept?*

A term is a label that is used to express a concept. There could be preferred terms, alternative terms or hidden terms, all referring to one concept. The concept is language independent whereas the term is fixed in a natural language. Here is an example in SKOS of a Concept and the preferred labels used to express this concept.

```
<rdf:Description
rdf:about="http://www.athenaeurope.org/athenawiki/AthenaThesaurus/Michael_Subjects#Education">
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
<skos:inScheme
rdf:resource="http://www.athenaeurope.org/athenawiki/AthenaThesaurus/Michael_Subjects"/>
<skos:prefLabel xml:lang="cs">Vzdělávání</skos:prefLabel>
<skos:prefLabel xml:lang="ee">Haridus</skos:prefLabel>
<skos:prefLabel xml:lang="en">Education</skos:prefLabel>
<skos:prefLabel xml:lang="fr">Éducation</skos:prefLabel>
<skos:prefLabel xml:lang="it">Pedagogia</skos:prefLabel>
<skos:prefLabel xml:lang="cy">Addysg</skos:prefLabel>
<skos:prefLabel xml:lang="fi">Kasvatustiede</skos:prefLabel>
<skos:prefLabel xml:lang="sv">Utbildning</skos:prefLabel>
<skos:prefLabel xml:lang="el">Εκπαίδευση</skos:prefLabel>
<skos:prefLabel xml:lang="nl">Onderwijs</skos:prefLabel>
<skos:prefLabel xml:lang="lv">Pedagoģija</skos:prefLabel>
<skos:prefLabel xml:lang="hu">Pedagógia</skos:prefLabel>
<skos:prefLabel xml:lang="bg">Педагоγика</skos:prefLabel>
<skos:prefLabel xml:lang="pl">Edukacja</skos:prefLabel>
<skos:prefLabel xml:lang="sk">Vzdelávanie</skos:prefLabel>
<skos:prefLabel xml:lang="es">Educación</skos:prefLabel>
</rdf:Description>
```

### *I do not have concepts in my thesaurus, can I still do Semantic Web stuff?*

Yes you can! The purpose of the TMP is to bring any kind of terminology to the Semantic Web. Therefore you can import your thesaurus without any problem. The TMP is here to facilitate the process but the conceptual work on how to bring your thesaurus descriptors and non-descriptors into concepts or terms, is up to you and depends on your specific needs. Even if your terminology is not structured, TMP provides you the toolset necessary to structure, edit or reorganise your terminology.

### **3.3.3 Video tutorials**

Some specific functionalities will also be illustrated with video tutorials. These videos will be produced using the Camvideo tool which allows recording screen activity as a video file.

Repeating an action seen on a video can be hundred times more efficient than hundred lines of written explanation.

The recording of the videos is foreseen for the following functionalities :

- Register a terminology
- Prepare a CSV export for SKOSification
- SKOSify a terminology
- Create a concept
- Change the hierarchy of a concept
- Add labels or documentary notes to concepts
- Make a mapping between two concepts
- Edit a mapping link
- Export in SKOS from TMP?

These fonctionnalités have been considered as the most useful and complex to understand without seeing how it is done. A 20 seconds video could be more efficient than a long sentence to help the user.

Once the videos will be ready, they will be aligned with the content of the “Online help”. The list of possible video tutorials is not fixed and can evolve according to the questions and user feedbacks.

All the videos will also be put online on the Michael Culture association’s Vimeo channel<sup>49</sup>.

---

<sup>49</sup> <http://vimeo.com/user14659022>

## 4 DISSEMINATION

So far we presented the documentation available for the TMP online as contextual help and as a detailed documentation.

In this section we present the dissemination material that is available about the TMP and more widely the work on terminology. We also present briefly here dissemination material that is going to be prepared in order to foster the use of the TMP, in- and outside of the AthenaPlus project consortium.

The dissemination of the TMP is not just about getting the most users possible, but mainly to get the most of users feedback in order to improve the tool and extend the context and domains where it can be used.

As the developments of the TMP started within the Linked Heritage project, a multitude of training and dissemination material has already been prepared in order to guide the cultural institutions in providing their terminologies according to Semantic Web standards. These materials were mainly presentations with a general introduction of the Semantic Web and a presentation of the tool.

The Linked Heritage work package dedicated to the dissemination of the project has also contributed to promoting the work done on terminology and the TMP.

This is for example the case with an interview summarizing in five minutes the stakes, uses and needs, and how Linked Heritage dealt with terminology. This video has been produced by the Michael Culture Association and can be reached from the following link : <http://vimeo.com/58878997>

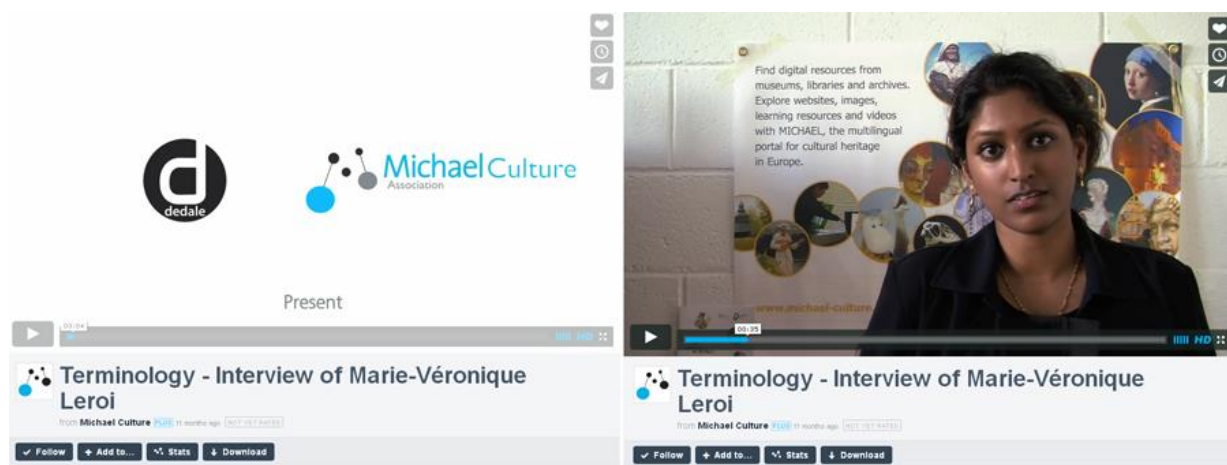
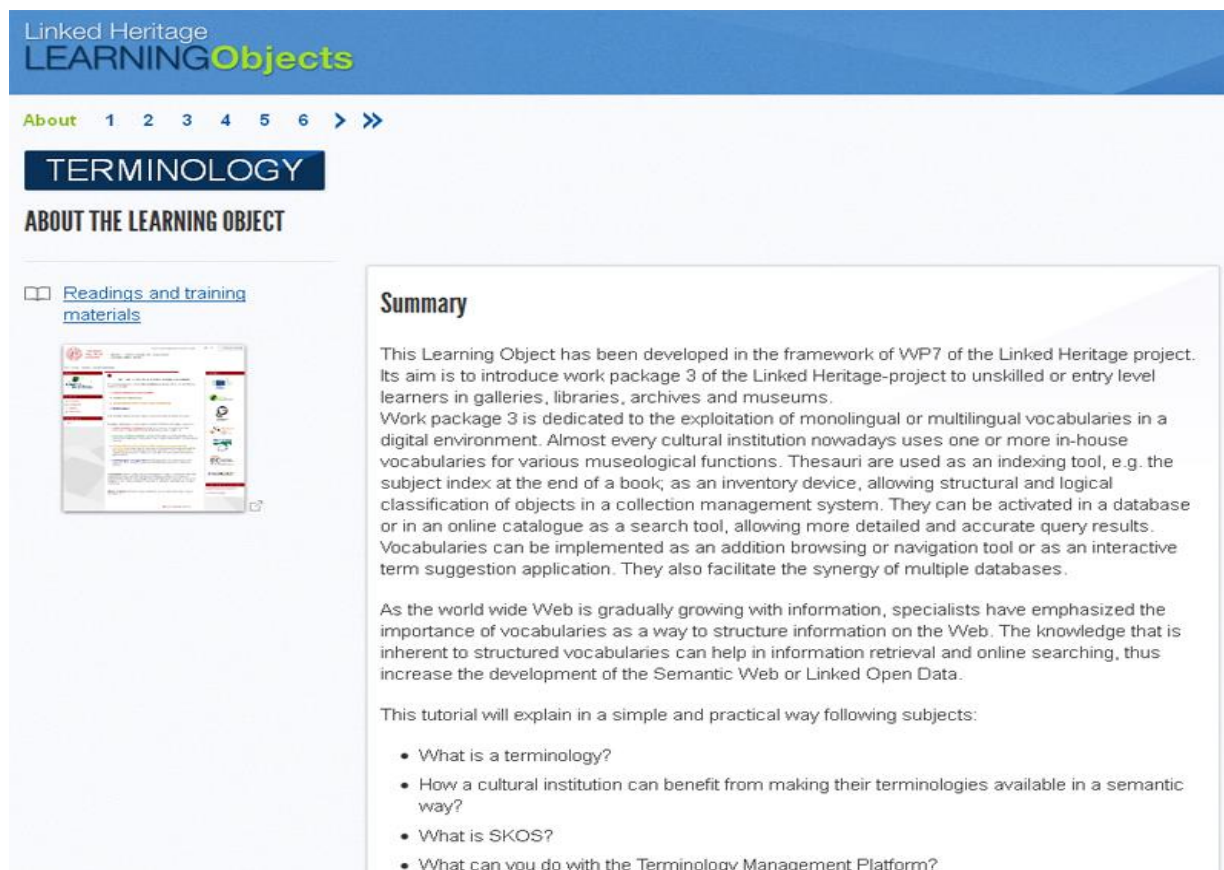


Figure 32 : Terminology interview

This same dissemination work package has also published a learning object on terminology. A learning object is a collection of content and information with a learning objective. The learning object developed in cooperation with the University of Padova introduces the notion of terminology and what could be the tasks related to terminology management: <http://linkedheritage.cab.unipd.it/training/LO-07/en/00.html>.

The terminology learning object has the following sections :

- Terminology and terminology management
- Semantic Web : How to join?
- SKOS, Simplified Knowledge Organisation System
- TMP, Terminology Management Platform
- Literature and Glossary




Linked Heritage  
**LEARNING**Objects

About 1 2 3 4 5 6 > >>

## TERMINOLOGY

### ABOUT THE LEARNING OBJECT

[Readings and training materials](#)



#### Summary

This Learning Object has been developed in the framework of WP7 of the Linked Heritage project. Its aim is to introduce work package 3 of the Linked Heritage-project to unskilled or entry level learners in galleries, libraries, archives and museums.

Work package 3 is dedicated to the exploitation of monolingual or multilingual vocabularies in a digital environment. Almost every cultural institution nowadays uses one or more in-house vocabularies for various museological functions. Thesauri are used as an indexing tool, e.g. the subject index at the end of a book; as an inventory device, allowing structural and logical classification of objects in a collection management system. They can be activated in a database or in an online catalogue as a search tool, allowing more detailed and accurate query results. Vocabularies can be implemented as an addition browsing or navigation tool or as an interactive term suggestion application. They also facilitate the synergy of multiple databases.

As the world wide Web is gradually growing with information, specialists have emphasized the importance of vocabularies as a way to structure information on the Web. The knowledge that is inherent to structured vocabularies can help in information retrieval and online searching, thus increase the development of the Semantic Web or Linked Open Data.

This tutorial will explain in a simple and practical way following subjects:

- What is a terminology?
- How a cultural institution can benefit from making their terminologies available in a semantic way?
- What is SKOS?
- What can you do with the Terminology Management Platform?

**Figure 33 : Terminology Learning Object summary**

The video tutorials that will be made for the TMP could also be added to the learning object to update it.

These materials were mainly aiming at disseminating the TMP to the project partners. But the TMP is intended to all cultural professionals and could provide an interesting software solution to most of the European and Europeana-related projects dealing with terminologies and multilingualism.

This is the reason why an open seminar on terminology and multilingualism was organised in Paris the 18<sup>th</sup> of April 2013. Europeana and other European projects were invited to present their own needs, user experiences and tools for terminology management. The seminar was a good opportunity to go beyond the boundaries of the project for disseminating the best practices and the use of the TMP.

Presentations that were made during this seminar are accessible online :

<http://www.michael-culture.eu/blog/seminar-multilingualism-and-terminology-paris-18-04-2013.html>

This was so far the lessons learnt from the Linked Heritage project. Within the AthenaPlus project where some of the partners heard about the TMP before the beginning of the project, the continuing dissemination of the TMP would require the preparation of a leaflet or a card (such as a postcard) presenting the tool and giving the URL from where it can be accessed. This will be discussed with the WP7 in charge of the dissemination.

Other dissemination material, that could also be training material, would be the preparation of a quick user manual with screenshots and short explanations.

## 5 TRAINING SESSIONS AND WORKSHOPS

From the beginning of the Athena Plus project, two training sessions have been organised in conjunction with the plenary meetings. It appears that this format of training is convenient for a general introduction to Semantic Web concepts and a general presentation of the tool. But this format is not the best in order to receive real user feedback.

| Nature of event          | Place  | Date                         | Number of participants |
|--------------------------|--------|------------------------------|------------------------|
| Technical session on TMP | Rome   | 18 <sup>th</sup> March 2013  | + - 80                 |
| Training session         | Berlin | 8 <sup>th</sup> October 2013 | 26                     |

During the Linked Heritage project, the same kind of organisational concept was proposed except for one event in Paris in September 2012 where the morning was dedicated to the general introduction to Semantic Web, general introduction to the tool and the afternoon was set up as a hands-on session. This format was interesting as it brought valuable feedback on the use of the TMP to the developing team.

Another workshop organised within the Linked Heritage project as well has been very instructive. This workshop was organised in Brussels by the Belgian partner PACKED vzw, Centre of Expertise in Digital Heritage. The participants were not direct partners of the project but museums and institutions managing collections and terminologies and collaborating with the project partner PACKED. It appeared during this workshop that all the institutions were not using the same collections management system and that for some institutions bringing their terminologies into the TMP would require more preparatory work. An interesting use case was the case of the Adlib proprietary CMS which can format CSV export only in a specific way. Many institutions are using this tool as a collection management system, therefore using a tool such as Google Refine to preprocess this exported CSV file before importing it in the TMP could be useful for many users. The future idea of integrating this kind of tools in the TMP in order to facilitate the SKOSification and accommodate more TMP users, has emerged from this use case.

Considering the valuable feedback brought by institutions working on a daily basis and using their terminologies, it seems to be more appropriate and beneficial to organise smaller training sessions that would be more efficient to also put in practice the theoretical aspects presented as an introduction. This could be also very useful to improve the usability of the tool.

Further trainings will then be organised along the plenary meetings but changing the format, and organised more as workshops where small groups can be formed to test and use specific functionalities and doing specific tasks on terminologies so we can have a valuable feedback.

Beside these training sessions, workshops in the same range of the one organised in Brussels could be organised in some countries.

We have selected the following countries on the basis of the results of the survey presented in the *D4.1 First release GLAM sector reference terminologies*<sup>50</sup>. Here we present the list of the future pilot training sessions to be organised in cooperation with each partner country :<sup>51</sup>

| Country | Place    | Partner                   | Possible Period |
|---------|----------|---------------------------|-----------------|
| Italy   | Rome     | ICCU                      | April 2014      |
| Hungary | Budapest | PIM, National Gallery     | June 2014       |
| France  | Paris    | MCC, Unisav, MSH Grenoble | September 2014  |
| Germany | Berlin   | SPK, UNIMAR               | March 2015      |

These pilot sessions will help refining and improving the tool but also the documentation and training material that comes along with the tool. The TMP manual will be then published in a second version.

<sup>50</sup> First release GLAM sector reference terminologies : <http://www.athenaplus.eu/getFile.php?id=187>

<sup>51</sup> This planning is only a proposal as it has to be validated with each organising partner.



The first release of the TMP has been launched in December 2013 and the second release is expected for January 2015, this is why it is important to organise at least three hands on pilot sessions between these two releases.

#### ➤ **Evaluation**

Evaluation is a crucial aspect when developing a software. Its usability, its functionalities have to be evaluated in order to offer a quality service. An evaluation survey has been defined in order to assess the different aspects of the TMP (see in the annexes). This evaluation survey will be distributed and enriched for each training sessions. The evaluation feedback will be taken into account for implementing the second release of the TMP in January but also a technical roadmap for the future developments of the TMP beyond the Athena Plus project.

For the time being, the purpose of the workshops and trainings organised was to improve the tool and tackle in the best way the professional context of the terminology managers and providers. For this second phase we should combine the pedagogic aspect of these training sessions with a user focus group approach.

#### ➤ **Collaboration**

We need to define and develop some proof of concepts at institution, national and Europeana level to convince the institutions to follow the lead. Partners will be given the floor to propose their own proofs of concept so others can note the benefits. For example, other European projects have expressed their interest in using a tool like the TMP. This is the case for the Siera<sup>52</sup> project or Europeana Photography<sup>53</sup> who already imported their terminologies into the TMP.

The TMP can also play a role in the optimisation of search results in Europeana. Enriched vocabularies will guarantee multilingual queries and richer information to users looking for information on the Europeana portal.

In the AthenaPlus project content is delivered to Europeana through the MINT tool. Providers export a LIDO-schema from their CMS, upload this schema in MINT and then map the (rich) LIDO elements to the (simple) Europeana format EDM.

Projects such as PartagePlus and EU Fashion have domain-specific vocabularies in SKOS, which are linked to the LIDO format in the MINT-tool. The selected concepts from the enriched vocabulary are then mapped to EDM, where they are fully exploited.

The TMP will also deliver enriched vocabularies. However, it is not possible to enrich LIDO in the MINT-tool with a *meta-thesaurus*, i.e. a SKOS/RDF file containing many mapped concepts from different terminologies.

This is why we decided with the technical partners (UniSav, NTUA and Unimar) to create a set of micro-thesauri. These micro-thesauri will be enriched by the AthenaPlus partners (e.g. with translations) and uploaded in the TMP. A SKOS/RDF version of the thesauri will be used to enrich the LIDO metadata schema, where they can be mapped to EDM.

A Thesaurus Working Group will define which LIDO-supported metadata will be considered for enrichment (plenary meeting Bucharest, March 2014).

#### ➤ **Evolution of the TMP**

TMP as any software will evolve according to its users and their feedback. The evaluation presented above and set during the trainings will be crucial for defining the priorities for the future technical developments.

WP4 in cooperation with Europeana will be attentive to the evolution and correspondence work done for SKOS and the ISO 25964<sup>54</sup> in order to take this into account as a possible evolution of the TMP.

---

<sup>52</sup> SIERA : <http://sina.birzeit.edu/SIERA/featured-news/news-events/birzeit-university-wins-fp7-funded-project-siera/>

SIERA is a FP7 research project in the field of multilingual and multicultural knowledge sharing technologies.

<sup>53</sup> Europeana Photography : <http://www.europeana-photography.eu/> EuropeanaPhotography (European ancient photographic vintage repositories of digitized pictures of historical quality) is a EU-funded digitization project aimed at enriching Europeana with masterpieces of early photography provided by a public-private partner consortium.

<sup>54</sup> See Section 3.2.5 SKOS and ISO 25964 in this deliverable.

Solutions for simplifying the import of any terminology into the TMP are also investigated. The integration of tools such as Google Refine<sup>55</sup> could be indeed very useful to prepare the import and the SKOSification of a terminology.

---

<sup>55</sup> See more on Google Refine on page 27 of this deliverable

## 6 CONCLUSIONS

The trainings and workshops organised so far within the Linked Heritage project and the AthenaPlus project allowed adapting the documentation to be provided for the TMP. As the (technical) world of the Semantic Web was very new for most of the partners, the documentation has been focused on the theoretical aspects and the vision of the Semantic Web. This documentation does not mean to be exclusive and is complementary with the existing recommendations and guidelines for managing multilingual terminologies.

The major difficulty encountered during a presentation of the work done for terminologies is to convince the institutions of the benefits they may get in moving towards the Semantic Web.

For some institutions publishing their terminologies as Semantic Web compliant information sources is logical, but making links with other resources is not obvious. A pedagogic effort on showing the real benefit from publishing and linking data is really the priority for the next steps of this work package.

As mentioned above, the trainings and workshops will involve a more restricted number of participants so it can be more efficient and easier to proceed with the evaluation of the tool.

This is why we propose to organise small and focussed training sessions so we could have a more precise feedback and get a better knowledge of the environment, the needs and conditions the institutions may have. In particular a better knowledge of the collection management system used in the cultural institutions among the Athena plus partners during these training sessions would help for sure answering their expectations and improving and extending the functionalities proposed by the TMP.

## 7 APPENDIX 1: REFERENCES

- Linked Data : <http://www.w3.org/DesignIssues/LinkedData.html>
- RDF : <http://www.w3.org/RDF>
- RDF-Schema : <http://www.w3.org/TR/rdf-schema>
- « Persistent identifiers: Recommendations for institutions » : <http://www.athenaeurope.org/getFile.php?id=779>
- « Correspondence between ISO 25964 and SKOS/SKOS-XL Models » : [http://www.niso.org/apps/group\\_public/download.php/9627/Correspondence%20ISO25964-SKOSXL-MADS-2012-10-21.pdf](http://www.niso.org/apps/group_public/download.php/9627/Correspondence%20ISO25964-SKOSXL-MADS-2012-10-21.pdf)
- « Your terminology as a part of the semantic web recommendations for design and management » : <http://www.linkedheritage.eu/getFile.php?id=244>
- SKOS Simple Knowledge Organization System Primer, W3C Working Group Note 18 August 2009 : <http://www.w3.org/TR/skos-primer>
- SKOS Simple Knowledge Organization System Reference, W3C Recommendation 18 August 2009: <http://www.w3.org/TR/skos-reference>

## **8 APPENDIX 2: DEFINITION OF TERMS AND ABBREVIATIONS**

ARK : Archival Resource Key

CSV : Comma-Separated Values

DOI : Digital Object Identifier

DTD : Document Type Definition

GLAM sector : Galleries, Libraries, Archives and Museums

HTML: HyperText Mark Language

HTTP: HyperText Transfer Protocol

ISO : International Organisation for Standardisation

OWL : Web Ontology Language

PURL : Persistent URL

RDF : Resource Description Framework

RIF : Rule Interchange Format

SKOS : Simple Knowledge Organisation System

SPARQL : SPARQL Protocol and RDF Query Language

TMP : Terminology Management Platform

URI : Uniform Resource Identifier

URL : Uniform Resource Locator

URN : Uniform Resource Name

UTF-8 : Unicode Transformation Format 8 bits

UTF-16 : Unicode Transformation Format 16 bits

W3C : WorldWide Web Consortium

Web 3.0 : Semantic Web, Linked Data

XML : eXtensible Markup Language

## 9 ANNEXES

### 9.1 Evaluation survey for the TMP

TMP Evaluation survey  
<http://www.culture-terminology.org>

#### Access to the TMP

- Did you have any difficulty in accessing the TMP through its URL : <http://www.culture-terminology.org> ?  
If yes, please describe the problem.

- Did you have any difficulty in creating an account? If yes, please describe the problem.

- Did you have any difficulty in connecting to the TMP? If yes, please describe the problem.

#### User interface

- Is the proposed user interface adapted to your needs? If not, please explain why :

- Yes
- No



- Is the interface intuitive and effective enough? If not, please explain why :

- Yes
- No

- Is the documentation available detailed enough? If not, please list what is missing :

- Yes
- No

- Does the actual user interface need improvements? If yes, please detail.

- Yes
- No

### Use of the TMP

- Did you have any difficulty with any functionality proposed by the TMP? If yes, please detail.

- Yes
- No

- Is the tool adapted to your needs and your use? If not, why?

- Yes
- No

- Which improvement(s) should be done for a better use of the tool?

- Did you face any operating error while you were using the tool? If yes, which were these errors? Were they frequent? If possible, include a screenshot of the error message (if any)

- Yes
- No

- Which is the functionality of the TMP that you needed and used the most?

- SKOSification
- Editing
- Mapping
- Registry/Search
- Other

- What would be new functionalities that you might need ? Please describe them below.

### User experience

- The TMP and its documentation helped you acquire a better understanding of the Semantic Web and its general approach

- Excellent
- Good
- Average
- Poor

- The TMP and its documentation helped you to acquire a better knowledge of SKOS

- Excellent
- Good
- Average
- Poor

- Which elements would you like to deepen ?

**9.2 Presentation made for the training session in Berlin, the 8<sup>th</sup> of October 2013**



---

**AthenaPlus**  
**WP 4**  
**Terminologies and semantic enrichment**

**Training on the Terminology Management Platform**

**Marie-Véronique Leroi & Eva Coudyzer**

**AthenaPlus Training**



Berlin, AthenaPlus Training  
(2<sup>nd</sup> plenary meeting)

07/10/2013

**1. Terminology?**

**2. Semantic Web**

**3. Terminology Management Platform (TMP)**

**4. Practical exercises**

## Terminology?

### Who?

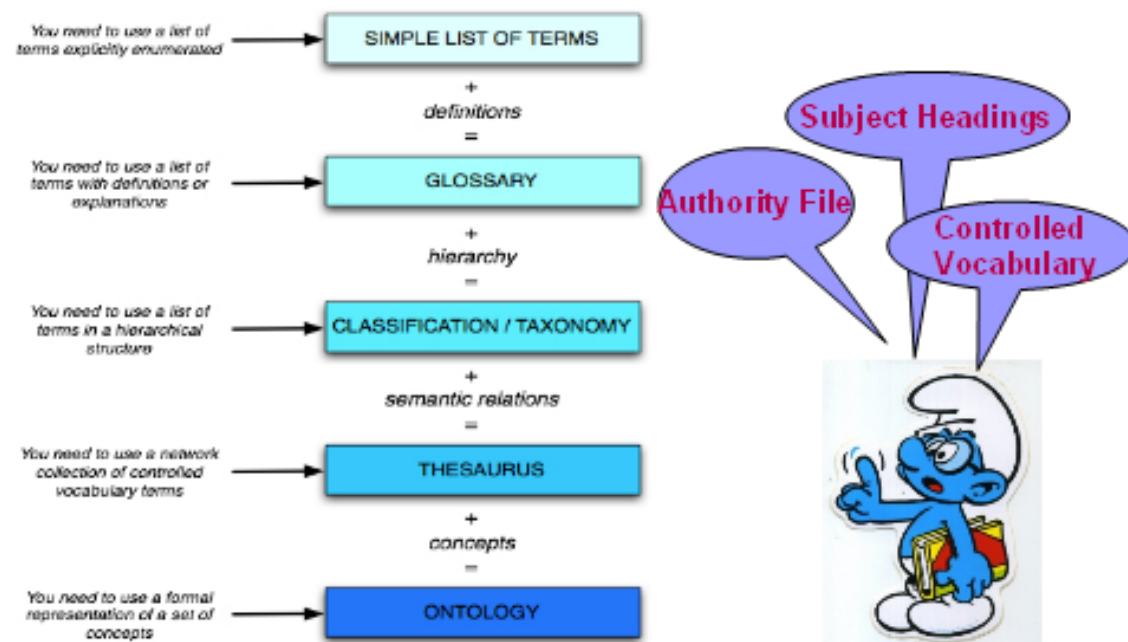
- Galleries, libraries, archives and museums (GLAM) use monolingual or multilingual terminologies.

### Definitions

- Aitchison et al. (2000): "a vocabulary of controlled indexing language, formally organized so that a priori relationships between concepts are made explicit".
- Hodge (2000): "Knowledge Organization Systems are used to organize materials for the purpose of retrieval and to manage a collection. A KOS serves as a bridge between the user's information need and the material in the collection. With it, the user should be able to identify an object of interest without prior knowledge of its existence. Whether through browsing or direct searching, whether through themes on a web page or a site search engine, the KOS guides the user through a discovery process".



## Terminology?







## Some examples : Getty Art & Architecture Thesaurus (NL)



## Why do GLAM-s ectors use thesauri?

### Use as :

- Indexing tool
- Search tool / query formulation support feature
- Interactive term suggestion
- Browse and navigation tool
- To facilitate combination of multiple databases or unified access to multiple databases

### Role :

- Thesauri have been a primary dimension of the research and development interests of experts in information retrieval, online searching, user interface design, knowledge organization in general, indexing and abstracting, cataloguing and classification, and information search behavior studies (Ali Shiri)

### What?

- Introduced in 2001 by Tim Berners-Lee
- Collaborative effort led by W3C
- Evolution of the Web of documents

**Principle?** share and reuse data on the Web (semantic interoperability)

**Why?** better search results, irrespective of language

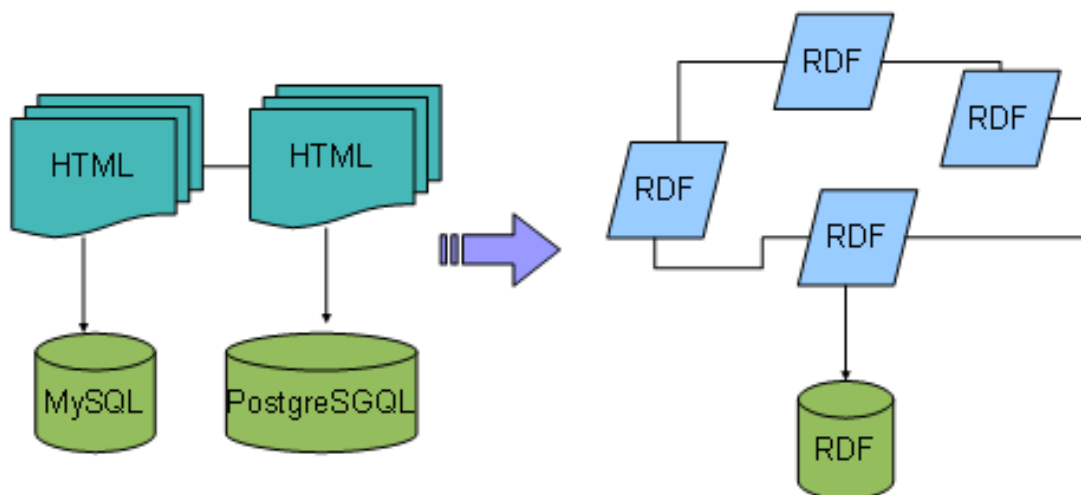
**How?** By automatically linking "separate" data on the Web (linked data); by inter-thesauri mapping

**Tools?** Allow machines to understand the meaning, or "semantics", of information on the Web

Thesauri can serve as a basis for developing domain-specific ontologies



## Web of data



Web of Documents

Web of Data

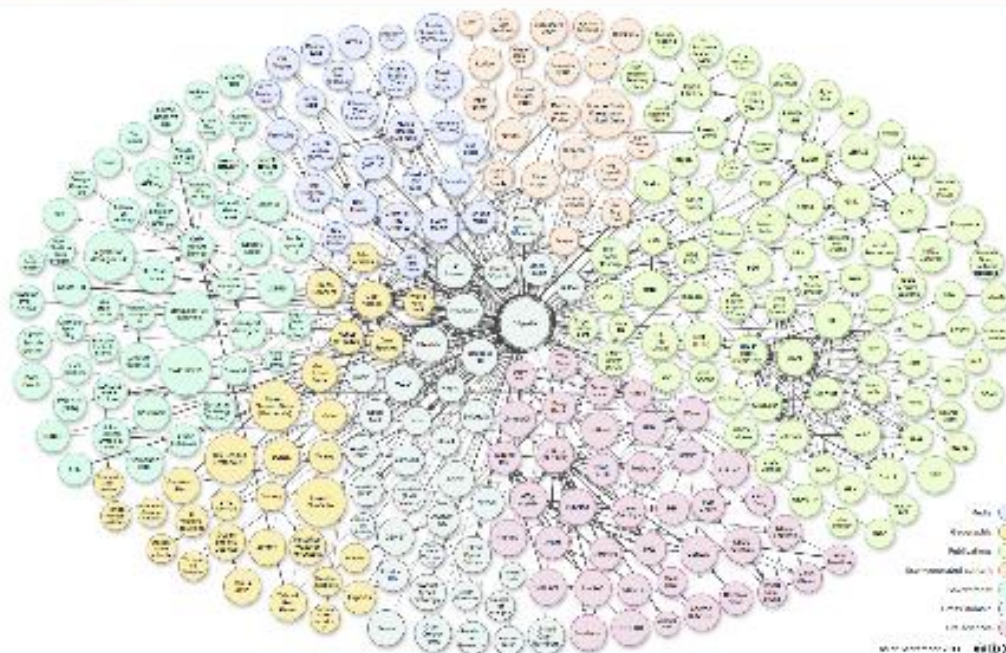


### Semantic Web

*"the Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it" (Tim Berners-Lee)*

### Linked Data

*"The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data." (Tim Berners-Lee)*







## Why joining the Semantic Web?

- **Keep providing the same quality of service as a public institution and improve it**
- **Give access to structured and normalised cultural data**
- **Be part of the Linked Open Data Cloud**
- **Enhance the semantic interoperability at national and international level**
- **Enhance the visibility of your collections and professional knowledge**
- **Help the user in getting optimised results to specific requests**
- **Be a trusted source of cultural data**



Berlin, AthenaPlus Training  
(2<sup>nd</sup> plenary meeting)

07/10/2013



## Athena/Linked Heritage Booklet

"TERMINOLOGY" COVERS SEVERAL TYPES OF VOCABULARIES: LEXICON, THESAURUS, CONTROLLED VOCABULARY, ONTOLOGY... THE TYPE OF VOCABULARY IS HIGHLY CONNECTED TO ITS FUNCTION (INDEXATION, SEARCH AND RETRIEVAL, TRANSLATION...). SKIPPIFYING YOUR TERMINOLOGY IS A CRUCIAL STEP TO SHARE IT AND TO CONNECT IT TO OTHER KNOWLEDGE RESOURCES AND TO IMPROVE VISIBILITY AND USE AT EUROPEAN LEVEL.  
[www.athena-plus.europa.org/athena2013](http://www.athena-plus.europa.org/athena2013)

<http://www.linkedheritage.org> | [info@linkedheritage.org](mailto:info@linkedheritage.org)



YOUR TERMINOLOGY

AS A PART OF THE SEMANTIC WEB

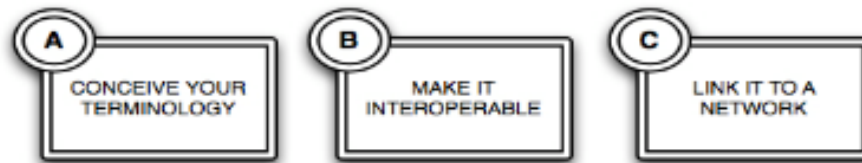
RECOMMENDATIONS

FOR DESIGN AND MANAGEMENT



Berlin, AthenaPlus Training  
(2<sup>nd</sup> plenary meeting)

07/10/2013



**A : What is the best way to create a terminology?**

**Create or adapt an existing terminology**

**B : What format do I use to make my terminology interoperable?**

**Technical implementation of the created or adapted terminology  
(in the perspective of the Semantic Web)**

**C : How can I link my terminology to other terminologies?**

**Linking of the SKOSified terminology with other resources (in the  
perspective of Linked Open Data)**



**Conceive a terminology : Create or adapt an existing terminology**

- **Define your collection domain(s)**
- **Identify your users' expectations (about your semantic descriptions)**
- **Define your connection with the data model**
- **Choose the terms for the semantic description of your digital resources**
- **Organise your terms into a thesaurus structure**
- **Find equivalent terms in other languages**
- **Implement your thesaurus**





□ **Make the terminology interoperable : Technical implementation ( => Semantic Web)**

- **SKOS: Simple Knowledge Organization System**
- **Standard created by the World Wide Web Consortium**
- **Allows flexible and machine-readable encoding of thesauri / other controlled vocabularies**
- **Within this standard thesaurus relationships can be encoded in such a way as to be used by various web services and web research systems**
- **A large number of terminologies will be brought in SKOS, to flexibly and consistently function within the semantic web environment**



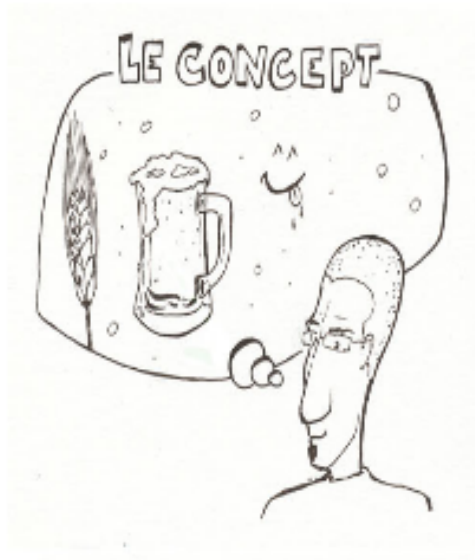
□ **Link the terminology to a network : Linking with other resources (=> Linked Data)**

- **Links several different thesauri**
- **Expand search functionalities through federated searching of multiple controlled vocabularies and linked data sources**
- **Allow for the integration of thesauri into many web-based search engines and services**
- **Provide semantically rich visualisation of thesauri and links between and among thesauri**
- **Facilitate multilingual information access and retrieval**
- **Provide easy access to thesauri for indexing and information representation purposes (Ali Shir)**





## Term vs Concept : The concept



### □ The concept

- A concept is defined as a unit of thought
- Concepts can often be expressed in a variety of different ways. They exist in the mind as abstract entities independent of terms used to express them.
- There is no notion of language for the concept but only for the terms expressing the concept

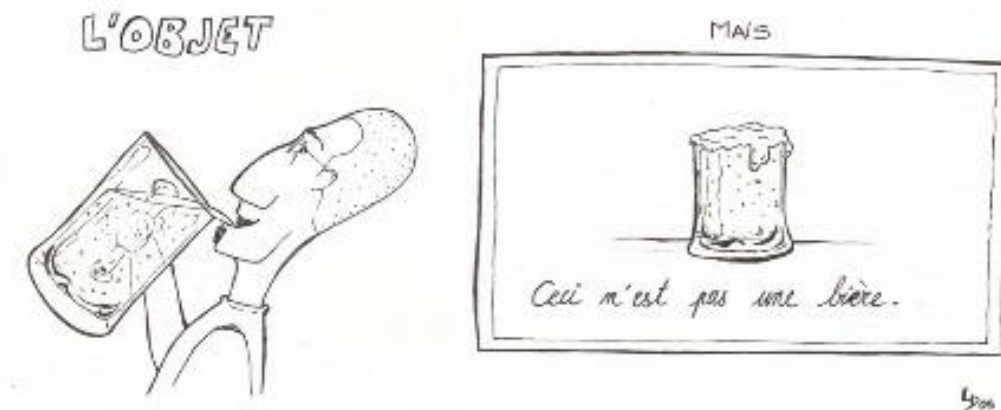
## Term vs Concept : The term



### □ The term

- A term is defined as a word or phrase used to label a concept
- Examples : hops, leaven, pint, ...
- Thesaurus terms can be either preferred terms or non-preferred terms (descriptors / non-descriptors)

## Term vs Concept : the object



### - Rdf : Resource Description Framework

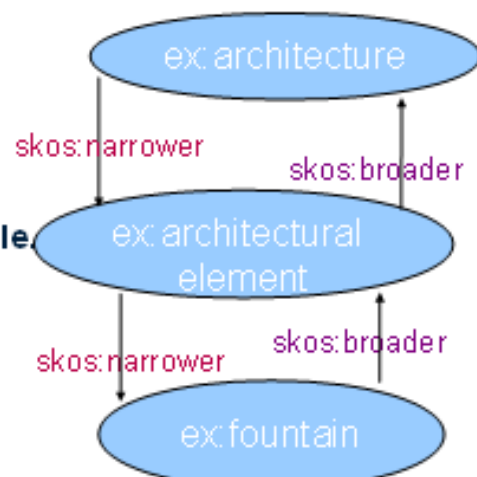


- Each information is structured as a triple

- Subjects and objects have an URI

- URI = Unique identifier

- URI = one of the Semantic Web & Linked Data principle



- Simplified Knowledge Organisation System (SKOS)
- Recommendation of the W3C since August 2009
- Interoperable format for thesaurus
- Descriptors became SKOS concepts expressed with labels as preferred terms (non-descriptors = non-preferred terms = alternative labels)
- Ideal compromise between a thesaurus and an ontology (RDF/XML)



- Concept : unit of thought (*skos:Concept*)
- Concept scheme : groups of concepts, micro-thesaurus (*skos:ConceptScheme*)
- Collection : thematic groups of concepts (*skos:Collection*)
- Concepts are expressed with labels :
  - Preferred Label : *skos:prefLabel*
  - Alternative label : *skos:altLabel*
  - Hidden Label : *skos:hiddenLabel*
- Semantic relations :
  - Hierarchical : *skos:broader* / *skos:narrower*
  - Associative : *skos:related*



- 
- « **Controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms** » (ISO 25964-1)
  - Norm ISO 25964-1 published in August 2011 : **Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval**
  - **Second part on Interoperability published in 2012**
  - **Revision of the ISO 2788, ISO 5964 and BS8723**
  - **Thesaurus oriented and evolution of SKOS**



### SKOS example

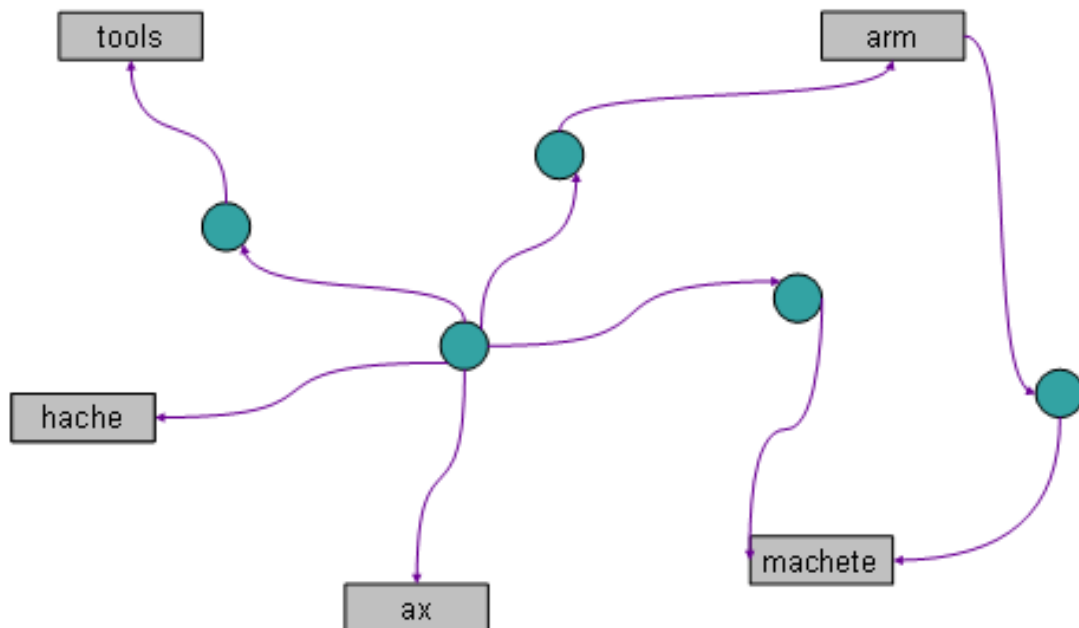
#### - Tools

- Hammer
- Saw
- Ax



#### - Arm

- Firearm
  - Gun
  - Pistol
- Cold steel
  - Sword
  - Ax
  - Crossbow

### SKOS example



## SKOS example

## SKOS output

```
<j.0:prefLabel xml:lang="en">arm</j.0:prefLabel>
<j.0:prefLabel xml:lang="fr">Change me</j.0:prefLabel>
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
<j.0:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/MVL/
/machete"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/MVL/m
achete/fe7c33c5-6225-4f9a-b1a3-2a4a27c6737a">
  <j.0:broader rdf:resource="http://www.culture-terminology.org/thesaurus/MVL/
machete/bdf87820-4754-4034-a8f8-b7eac11a20fc"/>
  <j.0:prefLabel xml:lang="en">axe</j.0:prefLabel>
  <j.0:prefLabel xml:lang="fr">Change me</j.0:prefLabel>
  <j.0:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/MVL/
/machete"/>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
</rdf:Description>
<rdf:Description rdf:about="http://www.culture-terminology.org/thesaurus/MVL/m
achete/cb658cbe-0c44-4584-9da5-1c28476fed27">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <j.0:inScheme rdf:resource="http://www.culture-terminology.org/thesaurus/MVL/
/machete"/>
```



## How to connect and build bridges in-between multilingual vocabularies?

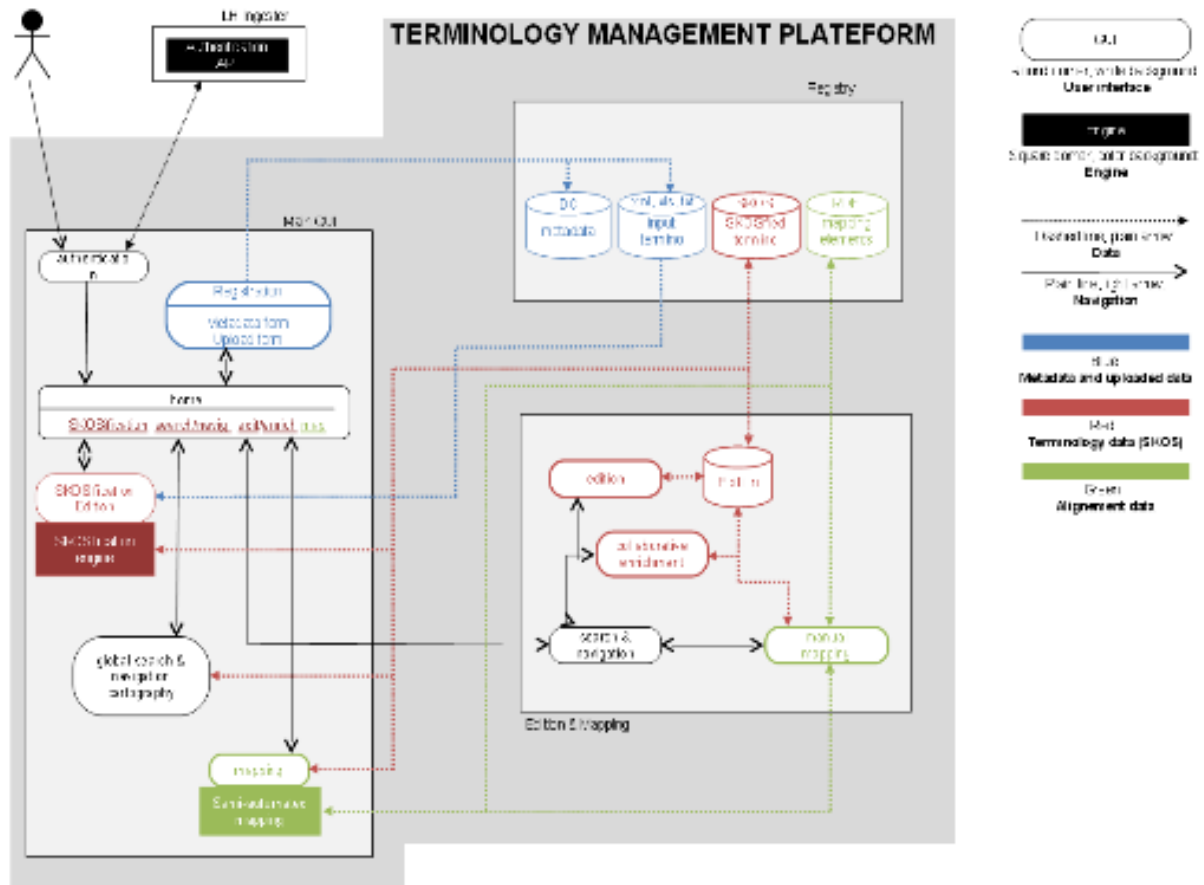
### Functional needs identified by Athena WP4: Benchmark

- Registration of a terminology in a repository
- SKOSification of a terminology
- Search and navigation into a network of vocabularies
- Mapping of the terminology with a thesaurus
- Enrichment of a thesaurus
- Collaborative moderation of an update of the thesaurus



- **To be a web service: For collaborative work online ;**
- **To have a user-friendly GUI: Adapted for a non-expert use in European museums, libraries and archives**
- **To combine open-source components: Such a service must stay independent of proprietary codes and formats**
- **To be logically structured with an intuitive Workflow: The user must find which actions to do according to his/her needs**
- **To be flexible enough to be adapted to new standards: What if SKOS is updated in a new version or evolving towards an ontology description?**





### Use cases

- Register an existing terminology > Import
- Register a new terminology > Create
- Modify/update the metadata of the terminology
- Create a terminology from different terminologies
- Edit/enrich a terminology
- Search through the terminology
- Mapping of terminologies



- **Add a new concept**
- **Modify a concept:**
  - **Modify the labels**
  - **Modify/add definitions and notes (scope notes skos:scopeNotes, editorial notes, skos:editorialNote, history note skos:historyNote)**
  - **Add a language for labels and/or notes**
  - **Change the hierarchy of a concept, change the semantic relations of the concept**
  - **Add an illustration to a concept with its metadata**
  - **Merge two concepts**
- **Delete a concept**



- **Mapping between terminologies in the same language**
- **Mapping between terminologies in different languages**
- **Mapping between terminologies with the same granularity (precision in the coverage of a domain)**
- **Mapping between terminologies with different granularity (for example mapping of a specific thesaurus to a general one for the top concepts only)**



| Role         | Actions                 |                  |              |                       |                    |                           |               |                     |                    |
|--------------|-------------------------|------------------|--------------|-----------------------|--------------------|---------------------------|---------------|---------------------|--------------------|
|              | View Public terminology | Make suggestions | CRUD Concept | CRUD outgoing 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                       |                  |              |                       |                    |                           |               |                     |                    |

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

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



## User management on TMP

---

- **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 some 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. So 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.



Berlin, AthenaPlus Training  
(2<sup>nd</sup> plenary meeting)

07/10/2013



## Terminology Management Platform (TMP)

---

<http://www.culture-terminology.org>



Berlin, AthenaPlus Training  
(2<sup>nd</sup> plenary meeting)

07/10/2013



[Home](#)

TMP (Terminology Management Platform) is a "tool box" for creating, editing and managing thesaurus, classifications, subject headings, ontology and any other kind of terminology. This platform is mainly dedicated to cultural institutions from any sector (libraries Archives and Museums) who hold or are willing to create terminologies

You can register your terminology and proceed easily to the SKOSification (conversion into SKOS) and/or to the mapping of your terminology with another one

**Search public terminology**

**Connect to TMP**

Email

Password

[Sign in!](#)   [Forgot your password?](#)



A project of Co-funded by the and the



[Home](#) > [search public terminologies](#)

Show  entries Search

| Title                         | Subject         | Languages      | Author         |                                     |
|-------------------------------|-----------------|----------------|----------------|-------------------------------------|
| <a href="#">Being</a>         | Philosophy      | fr             | GR             | <input type="button" value="View"/> |
| <a href="#">Droit</a>         | Law             | fr             | GR             | <input type="button" value="View"/> |
| <a href="#">Fortification</a> | Archaeology     | de, en         | JL             | <input type="button" value="View"/> |
| <a href="#">Furniture</a>     | Furniture       | en, fr         | nwl            | <input type="button" value="View"/> |
| <a href="#">Objects</a>       | Economics       | en, fr, en, fr | Mt             | <input type="button" value="View"/> |
| <a href="#">ObjectsName</a>   | Archaeology     | en, fr, nl     | KMKG           | <input type="button" value="View"/> |
| <a href="#">Palissy</a>       | Furniture       | fr             | MCC-Inventaire | <input type="button" value="View"/> |
| <a href="#">test</a>          | Decorative arts | sk             | qdkdqpf        | <input type="button" value="View"/> |
| <a href="#">titre</a>         | Decorative arts | fr             | fiorent        | <input type="button" value="View"/> |

Showing 1 to 9 of 9 entries Previous Next

A project of Co-funded by the and the



[My home](#)

**Search terminologies**

 Search  

**Create new terminologies**

**Link two terminologies**

Show  entries Search:

| Title                            | Subject         | Languages                              | Author        |  |
|----------------------------------|-----------------|--|---------------|--|
| <a href="#">Michael Subjects</a> | Geography       | en, fr, en, fr, en, fr, en, fr, en, fr | MCA           | <input type="button" value="Edit"/> <input type="button" value="export"/><br><input type="button" value="Delete"/> |
| <a href="#">Objectnaam KIK</a>   | Archaeology     | en, nl                                 | V&Sronique    | <input type="button" value="Edit"/> <input type="button" value="export"/><br><input type="button" value="Delete"/> |
| <a href="#">Ornamental Art</a>   | Decorative arts | fr                                     | Julia Kostova | <input type="button" value="Edit"/> <input type="button" value="export"/><br><input type="button" value="Delete"/> |



[My home](#) > Terminology creation

Create a terminology from scratch. Fill in general information (metadata) on your terminology (\* mandatory fields)

|             |                      |                   |                      |
|-------------|----------------------|-------------------|----------------------|
| Title *     | <input type="text"/> | Alternative Title | <input type="text"/> |
| Subject *   | Select a value...    | Description       | <input type="text"/> |
| Languages * | Select language...   | Creation Date     | <input type="text"/> |
| Edition *   | <input type="text"/> | Date of issue     | <input type="text"/> |
| Creator *   | <input type="text"/> | Contributor       | <input type="text"/> |
| Publisher * | <input type="text"/> | Identifier        | <input type="text"/> |
| Rights *    | Select a value       | Source            | <input type="text"/> |
|             |                      | Amount of Concept | Select a value       |

1) Upload your terminology

2) Configure your import

Does your import look good? **No**

Map you data on the underlying table with drop-down lists and [when you mapping is OK click here.](#)

| This col is : | This col is :         | This col is :       | This col is :     | This col is : | This col is :   | This col is : | This col is : | This col is : | This col is : |
|---------------|-----------------------|---------------------|-------------------|---------------|---|---------------|---------------|---------------|---------------|
| Not mapped    | Not mapped            | Not mapped          | Not mapped        | Not mapped    | Not mapped  | Not mapped    | Not mapped    | Not mapped    | Not mapped    |
| ID            | prefLabel@fra         | prefLabel@nl        | prefLabel@eng     | broader       | SN@nl   | SN@en         | SN@fr         |               |               |
| 41875         | Nom de l'objet        | Objectnaam          | Object name       |               |   |               |               |               |               |
| 41877         | Architecture          | Architectuur        | Architecture      | 41875         | Toeterm voor architecturale elementen en onderdelen van architecturale elementen    |               |               |               |               |
| 41879         | Construction          | Bouwwerk            | Single built work | 41877         | Algemene benaming voor alle soorten van (in aanbouw zijnde) gebouwen                |               |               |               |               |
| 45278         | Composant de bâtiment | Deel van een gebouw | Building division | 41879         | Wordt gebruikt voor het benoemen van afzonderlijke kamers of ruimten van een gebouw |               |               |               |               |

2) Configure your import

Does your import look good? **No**

Map you data on the underlying table with drop-down lists and [when you mapping is OK click here.](#)

| This col is : | This col is :  | This col is : | This col is :     | This col is : | This col is : | This col is : |
|---------------|----------------|---------------|-------------------|---------------|---------------|---------------|
| A local ID    | A property     | A property    | A property        | A reference   | A property    | Not mapped    |
| Of type:      | Of type:       | Of type:      | Of type:          | Of type:      | Of type:      |               |
| prefLabel     | prefLabel      | prefLabel     | prefLabel         | broader       | Select one    |               |
| Language:     | Language:      | Language:     | To column:        |               |               |               |
| French        | Dutch          | English       | ID                |               |               |               |
| ID            | prefLabel@fra  | prefLabel@nl  | prefLabel@eng     | broader       |               | SN@en         |
| 41875         | Nom de l'objet | Objectnaam    | Object name       |               |               |               |
| 41877         | Architecture   | Architectuur  | Architecture      | 41875         |               |               |
| 41879         | Construction   | Bouwwerk      | Single built work | 41877         |               |               |

Select one

- Concept
- altLabel
- prefLabel
- scopeNote
- ConceptScheme
- Collection
- OrderedCollection
- changeNote
- definition
- editorialNote
- example
- hiddenLabel
- historyNote
- notation
- note





[My home](#) > [Edit the terminologie](#)

**Edit metadata**

Title :  - Author :

*Instrument de musique*

- Change me
- Équipement de peinture
- Équipement d'impression
- Nom de l'objet
  - Instrument de musique*
  - Moyen de communication
  - Arme et armure
  - Jout et équipement de sport
  - Forme d'objet et objet dérivé
  - Matériel visuel
  - Dispositif de fixation et fermoir
  - Moyen de transport
  - Article de toilette et instrument médical
  - Moyen de punition et d'instruction
  - Mobilier
    - Meuble
    - Accessoire de meuble

Langue de dimension : Français

Labels / Documentation

preferred label :

alternative label :

hidden label :

---

Conceptual dimension

has broader :

has narrower :



[My home](#) > [Map terminologies](#)

Title : P100%Jal hodourus - Author : mibac

Euphonia

- when
- who
  - Mairias
    - Companies related to
      - Families
      - Foreign institutions
      - Performing art companies
      - Enterprises
        - Constitutional bodies and bodies with constitutional relevance
          - Senate of the Republic
          - National Council for Economic and Labour
          - Chamber of Deputies
          - Presidency of the Republic
          - Parliament
            - Italian Government**
            - Council of Ministers
            - State Council
            - High Council of the Judiciary
            - Constitutional Court
      - Academies (Institutions)
      - Independent Authorities
      - Museums
      - Ministries, Departments, General Government Bodies
      - Libraries, multimedia centres and documentation centres
      - Historic and Security Bodies
      - Regional and Local Authorities
      - Religious Bodies
      - Publishing houses
      - Education and research institutions
      - Educational and consultancy organisations

Title : Euphonia - Author :

- types of photograph : practice
- photographic techniques
- keywords
  - health
  - society
  - religion and belief
  - science and technology
  - landscapes
  - leisure
    - national government
      - national government**
      - local government
    - parliament
    - political party
    - politician
    - regional government
    - public administration
    - local government
    - minister
    - people of state
  - lifestyle and leisure
  - culture, world and peace
  - sport
  - human interest
  - education
  - labour

Select the type of the link

close/Cancel

Labels / Notes / Relations

preferred label :

alternative label :

hidden label :

Labels / Notes / Relations

preferred label :

alternative label :

hidden label :



- **TMP released as a demo version within the Linked Heritage project**
- **The development of the TMP continue within the AthenaPlus project (WP4 terminology)**
- **Connection between the TMP and the Mint tool is under preparation**
- **Next prototype release expected in december 2013, second release in February 2015**
- **The work on terminologies initiated within LH will be continued within AthenaPlus.**



Berlin, AthenaPlus Training  
(2<sup>nd</sup> plenary meeting)

07/10/2013



## Contact

---

**Thank you for your attention !**

**Marie-Véronique Leroi (MCC)**  
**marie-veronique.leroi@culture.gouv.fr**

**Eva Coudyzer (KMKG)**  
**e.coudyzer@kmsg-mrah.be**

[www.athenaplus.eu](http://www.athenaplus.eu)

[Twitter |https://twitter.com/AthenaPlusEU](https://twitter.com/AthenaPlusEU)  
[LinkedIn |http://Minkd.in/19XwW0z](http://Minkd.in/19XwW0z)



Berlin, AthenaPlus Training  
(2<sup>nd</sup> plenary meeting)

07/10/2013