Linked Heritage, Terminologies & Multilingualism

The development of a Terminology Management Platform

Roxanne Wyns
Royal Museums of Art and History (Belgium)
r.wyns@kmkg-mrah.be

Table of content

I/   The Linked Heritage project

II/   Work Package 3: Terminologies & Multilingualism

III/   Development of the Terminology Management Platform

IV/   The semantic web, SKOS & recommendations for improved interoperability

V/   Linked Heritage semantic mapping experiment

VI/   Perspective

I/ The Linked Heritage project – Coordination of standards and technologies for the enrichment of Europeana

- Linked Heritage at a glance
- Linked Heritage strengths

Linked Heritage at a glance

❑ **Best Practice Network started on 1st of April 2011**

❑ **38 partners from 24 countries and 5 external contributors (such as Russia & Israel)**

❑ **Objectives**

- To contribute large quantities of new content to Europeana (3 millions objects) – cross domain

- To experiment with the enhancement of the quality of both new and existing Europeana content

- To demonstrate improved search, retrieval and use of Europeana content

- For the first time, the private publishing sector is involved to provide content to Europeana

Linked Heritage at a glance

Linked Heritage Work Packages (WP)

WP1 Project Management and Coordination [ICCU]

WP2 Linking Cultural Heritage Information [CT]

WP3 Terminology [KMKG & MCC]

WP4 Public Private Partnership [Editeur]

WP5 Technical Integration [NTUA]

WP6 Coordination of Content [CyI]

WP7 Dissemination & Training [UniPD & ICCU]

Linked Heritage at a glance

Linked Heritage thematic working groups (TWG)

WP2   Linking Cultural Heritage Information

WP3   Terminology

WP4   Public-Private Partnership

WP7   E-learning

Open to both consortium partners and external people

Linked Heritage strengths

- **LIDO:** Interchange standard format elaborated within ATHENA, widely adopted across Europe, allows to transfer rich data (WP2)

- **PIDs:** Essential for the sustainability of Europeana (WP2)

- **Terminology:** Experimental work on semantic interoperability of multilingual terminologies (WP3)

- **Private-public partnership:** Unique case in the EU project landscape (WP4)

- **MINT:** Aggregation tool for the integration of different standards and harvesting formats, focus on semantic interoperability (WP5)

- **Millions of items to Europeana:** Linked Heritage contribution to Europeana (WP6)

- **Dissemination and training:** At national and international level (WP7)

Linked Heritage strengths

Linked Heritage capitalises on the lessons learnt in previous projects to innovate in new projects

II/ Work Package 3: Terminologies & Multilingualism

- – Organisation & collaboration
- – WP3 objectives

Organisation & collaboration

- **WP leaders:** KMKG & MCC

  – Coordination of the work, deliverables, training and dissemination

- **Technical partners:** Univ. of Savoie, IST, DigiCult, NTUA

  – Development of the Terminology Management Platform (TMP)

- **Thematic Working Group (TWG):** + 50 members (internal & some external)

  – To contribute to WP3 documents & deliverables

  – To create, edit, map thesauri using the TMP

  – To test and provide feedback on the functionalities of the TMP

WP3 general objectives

❑ WP3 will develop a prototype of a **Terminology Management Platform (TMP)** to allow a collaborative creation of a network of multilingual cross-domain thesauri and controlled vocabularies

❑ WP3 will experiment in the creation of a **network of multilingual cross-domain Linked Heritage Thesauri** using existing or within the project developed tools

❑ WP3 aims to **reduce the gap in terminology management skills** that exists in cultural heritage institutions and the private sector by providing not only tools, but also **guidelines, recommendations and training (materials)**

III/ Development of the Terminology Management Platform (TMP)

– Identification of needs in terminology management

– From theory to practice

– Architectural schema of the TMP

– Prototype version (www.culture-terminology.org)

Identification of needs in terminology management

Athena and Linked Heritage defined that

- Many museums use an **in-house non-standard terminology**

- Reason: **Costs** to use a reference terminology (e.g. AAT) or **specific needs** (language, domain...)

- Often these vocabularies are **only exportable in CSV**, not in XML and not in a standard interoperability schema like SKOS

- However, these vocabularies could be very useful for:

  - **Alignment of concepts** used in content brought together from different sources and domains

  - **Multilingual access**

  - **Improved search** and **retrieval** of content

From theory to practice

How to create a network of interlinked multilingual vocabularies?

Functional needs identified by Athena WP4 - Benchmark

1. Registration of a terminology in a repository

2. SKOSification of a terminology

3. Search and navigation into a network of vocabularies

4. Mapping of the terminology to another terminology resource

5. Enrichment of a thesaurus

6. Collaborative moderation of an update of the terminology

From theory to practice

WP3 is developing a prototype of a Terminology Management Platform to deal with these needs

The TMP is

❑ 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 out easily which actions to do according to his/her needs

❑ To be **flexible** enough to be adapted to new standards and versions

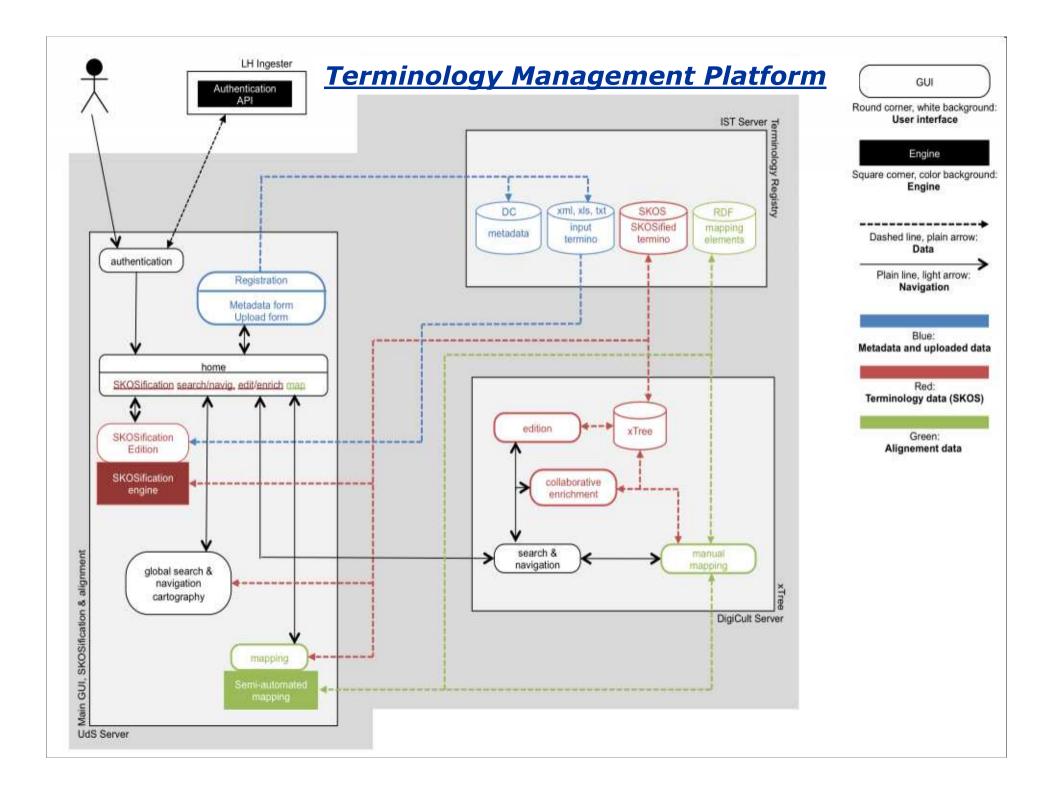# Architectural schema of the TMP

Architectural schema of the TMP

- Authentication API (NTUA)

- Terminology registry (IST server)

- Main GUI, SKOSification & alignment module (UdS server)

- Editing & mapping module (DigiCult-xTree server)

Beta version on

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

Terminology Management Platform

Legend (right panel):

- GUI — Round corner, white background: **User interface**
- Engine — Square corner, color background: **Engine**
- Dashed line, plain arrow: **Data**
- Plain line, light arrow: **Navigation**
- Blue: **Metadata and uploaded data**
- Red: **Terminology data (SKOS)**
- Green: **Alignement data**

Diagram labels:

- LH Ingester — Authentication API
- IST Server — Terminology Registry
  - DC metadata
  - xml, xls, txt input termino
  - SKOS SKOSified termino
  - RDF mapping elements
- UdS Server — Main GUI, SKOSification & alignment
  - authentication
  - Registration — Metadata form / Upload form
  - home — SKOSification search/navig. edit/enrich map
  - SKOSification Edition
  - SKOSification engine
  - global search & navigation cartography
  - mapping
  - Semi-automated mapping
- DigiCult Server — xTree
  - edition
  - xTree
  - collaborative enrichment
  - search & navigation
  - manual mapping

www.culture-terminology.org

Microsoft Outlook We...

LINKeD HeRITAGe

Coordination
of standards and technologies
for the enrichment
of EUROPEANA

New   Edit   Map   Search / browse

Welcome on the Culture Terminology Website, an ongoing effort under the linkedHeritage project.

This website aims to provide a service that allow any cultural institution to import and store his thesaurus.

This service will also allow Institution to map this thesaurus to others thesaurus already imported in the service.

Thesaurus contained in the Culture Terminology will be publicly exposed.

This website is actually an alpha release for testing purpose.

Until this service is under alpha version, be aware that datas you put now can be deleted at anytime.
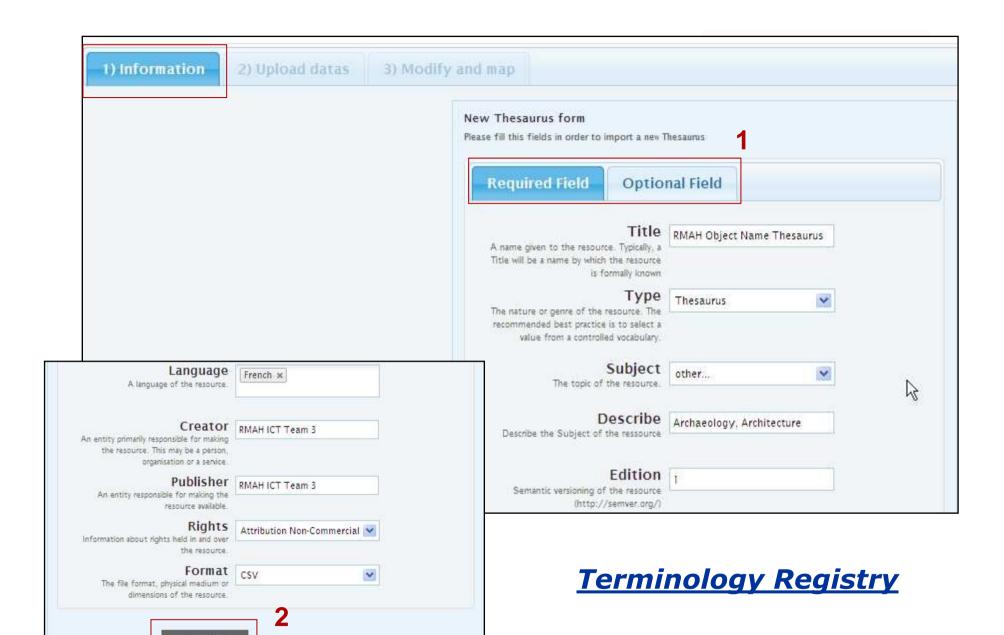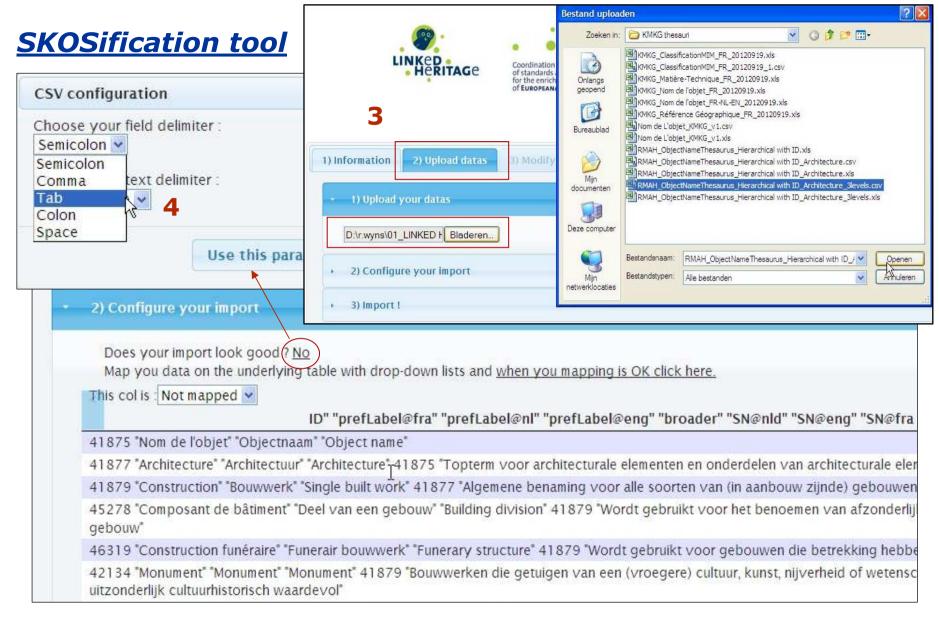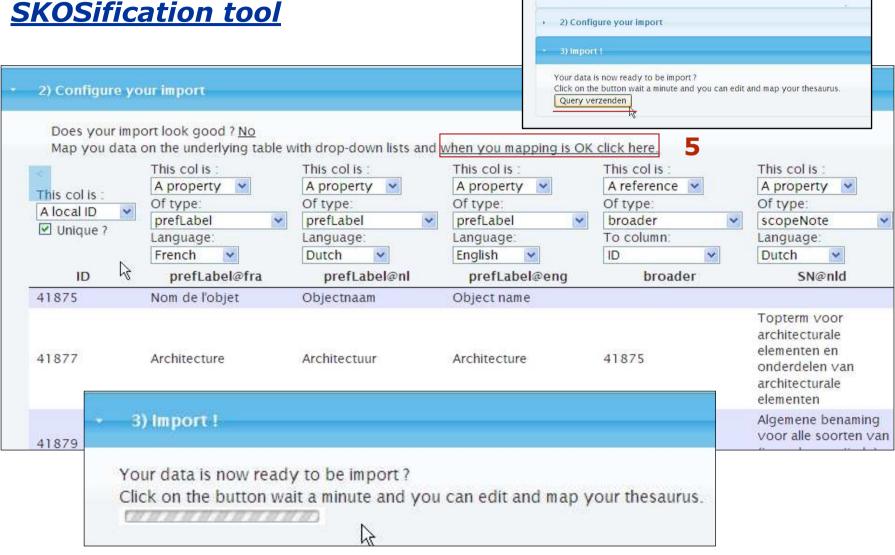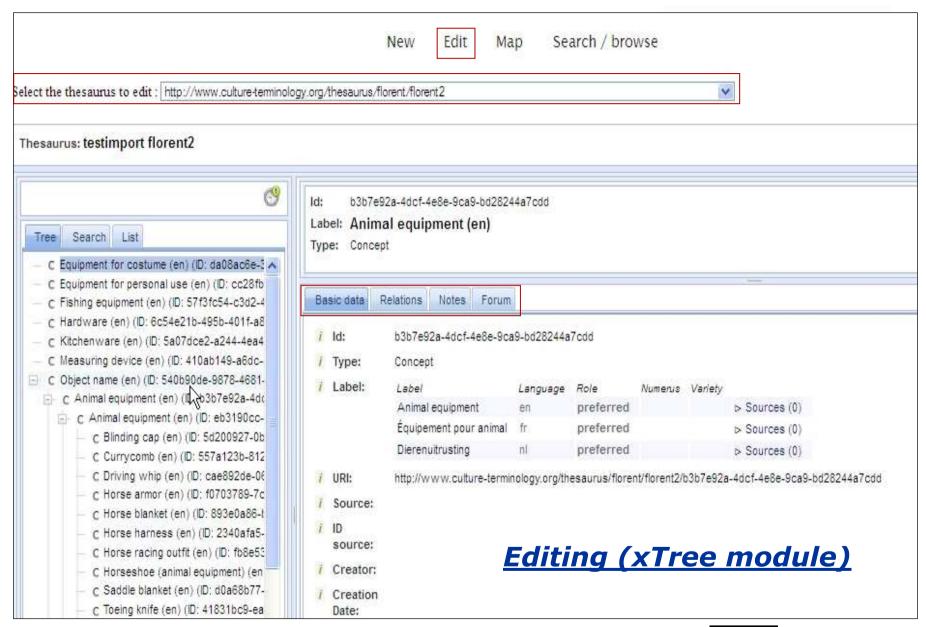
Stay tuned this situation will change soon !

A project of **LINKeD HeRITAGe**   Co-funded by the

1) Information 2) Upload datas 3) Modify and map

New Thesaurus form
Please fill this fields in order to import a new Thesaurus

1

Required Field **Optional Field**

Title
A name given to the resource. Typically, a Title will be a name by which the resource is formally known
RMAH Object Name Thesaurus

Type
The nature or genre of the resource. The recommended best practice is to select a value from a controlled vocabulary.
Thesaurus

Subject
The topic of the resource.
other...

Describe
Describe the Subject of the ressource
Archaeology, Architecture

Edition
Semantic versioning of the resource (http://semver.org/)
1

Language
A language of the resource.
French ×

Creator
An entity primarily responsible for making the resource. This may be a person, organisation or a service.
RMAH ICT Team 3

Publisher
An entity responsible for making the resource available.
RMAH ICT Team 3

Rights
Information about rights held in and over the resource.
Attribution Non-Commercial

Format
The file format, physical medium or dimensions of the resource.
CSV

2

Import file

Terminology Registry

SKOSification tool

SKOSification tool

New   Edit   Map   Search / browse

Select the thesaurus to edit : http://www.culture-terminology.org/thesaurus/florent/florent2

Thesaurus: **testimport florent2**

Tree   Search   List

- C Equipment for costume (en) (ID: da08ac6e-3
- C Equipment for personal use (en) (ID: cc28fb
- C Fishing equipment (en) (ID: 57f3fc54-c3d2-4
- C Hardware (en) (ID: 6c54e21b-495b-401f-a8
- C Kitchenware (en) (ID: 5a07dce2-a244-4ea4
- C Measuring device (en) (ID: 410ab149-a6dc-
- C Object name (en) (ID: 540b90de-9878-4681-
  - C Animal equipment (en) (ID: b3b7e92a-4d
    - C Animal equipment (en) (ID: eb3190cc-
      - C Blinding cap (en) (ID: 5d200927-0b
      - C Currycomb (en) (ID: 557a123b-812
      - C Driving whip (en) (ID: cae892de-06
      - C Horse armor (en) (ID: f0703789-7c
      - C Horse blanket (en) (ID: 893e0a86-
      - C Horse harness (en) (ID: 2340afa5-
      - C Horse racing outfit (en) (ID: fb8e53
      - C Horseshoe (animal equipment) (en
      - C Saddle blanket (en) (ID: d0a68b77-
      - C Toeing knife (en) (ID: 41831bc9-ea

Id: b3b7e92a-4dcf-4e8e-9ca9-bd28244a7cdd
Label: Animal equipment (en)
Type: Concept

Basic data   Relations   Notes   Forum

i	Id:	b3b7e92a-4dcf-4e8e-9ca9-bd28244a7cdd				
i	Type:	Concept				
i	Label:	Label	Language	Role	Numerus	Variety
		Animal equipment	en	preferred		▷ Sources (0)
		Équipement pour animal	fr	preferred		▷ Sources (0)
		Dierenuitrusting	nl	preferred		▷ Sources (0)

i URI: http://www.culture-terminology.org/thesaurus/florent/florent2/b3b7e92a-4dcf-4e8e-9ca9-bd28244a7cdd

i Source:

i ID source:

i Creator:

i Creation Date:

Editing (xTree module)

Graphical mapping

IV/ The semantic web, SKOS & recommendations for improved interoperability

- The Semantic web and Linked Data

- Semantic data formats

- SKOS tutorial

The Semantic web and Linked Data

The Semantic web or Web 3.0

❑ Concept introduced by Sir Tim Berners-Lee (W3C)

❑ Proposed as the solution for the current problems in sharing and retrieving relevant data on the current Web where

– Content is not well structured, has inexplicit semantics, is not interoperable (HTML, URLs to link)

– Expressive questions cannot be asked by the user

– Multiple data queries, human interpretation and knowledge is needed to retrieve relevant and "complete" results

*"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." (Tim Berners-Lee)*

The Semantic web and Linked Data

The semantic web is an extension of the current Web

❑ It includes semantic information (context and meaning!) in web pages

❑ This meaning allows both people and machines to better interpret the data

❑ It creates links so that a person or machine can explore the web of "related" data via these links

These links are at the heart of the Semantic web and are needed for integration and reasoning of data on the Web = **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." (Sir Tim Berners-Lee)*

The Semantic web and Linked Data

Vocabularies play an important role in the Semantic Web and Linked Data world

❑ They are the basic building blocks for linking data

❑ They help with the interpretation and integration of data between different datasets

❑ And so may lead to the discovery of new relationships between information expressed in a different natural language

Semantic data formats

The **Simple Knowledge Organisation System** or **SKOS** format is generally used to share, group, align and link vocabularies in an interoperable environment

First a general understanding of some **other semantic web formats** is required to better understand the SKOS format

- XML

- RDF

- OWL

- SKOS

All of these are **W3C standards**

Semantic data formats

Extensible Mark-up Language (XML)

❑ A set of rules for encoding documents in machine-readable format

❑ Simple, general, usable to exchange data on the internet

```
− <lido:lido>
    <lido:lidoRecID lido:type="local">RMAH-94399-FR</lido:lidoRecID>
  − <lido:descriptiveMetadata xml:lang="fr">
    − <lido:objectClassificationWrap>
      − <lido:objectWorkTypeWrap>
        − <lido:objectWorkType>
            <lido:conceptID lido:type="MuseumPlusThiID">47324</lido:conceptID>
            <lido:term>Likembe</lido:term>
          </lido:objectWorkType>
        − <lido:objectWorkType>
            <lido:term xml:lang="en">musical instruments</lido:term>
          </lido:objectWorkType>
        </lido:objectWorkTypeWrap>
```

Semantic data formats

Resource Description Framework (RDF)

❑ Forms the basis of Semantic web technologies

❑ Universal language to describe the characteristics of resource on the web

❑ Using XML for syntax and URIs for naming

❑ Makes statements about resources in the form of **subject-predicate-object triples**

❑ RDF triples provides a labelled connection using URIs to make it possible to link data with one another

❑ In this way a machine is able to find the semantic relations between data

Semantic data formats

- ❑ **The different parts of a triple are**

  - – **Subject** – the thing being described

  - – **Predicate** – a trait, aspect, or property of the thing, which expresses a relationship between the subject and object

  - – **Object** – the thing that is the value of the predicate (trait, aspect or property) of the object thing

- ❑ **So in the statement "The Kiss was created by Gustav Klimt"**

  - – **Subject** – The Kiss

  - – **Predicate** – Created by

  - – **Object** – Gustav Klimt

- ❑ **In terms of representation:**

  - – **Subject** – must be a URI

  - – **Predicate** – must be a URI

  - – **Object** – may be a URI or a constant value or "literal" (e.g. "oil on canvas")

Semantic data formats

Web Ontology Language (OWL)

- A family of knowledge representation languages for authoring ontologies

- Characterised by formal semantics and RDF/XML-based serializations

- Provides a more expressive language than SKOS to enhance the exchange of information

The word "ontology" is usually used for a more complex and formal representation of a set of concepts, whereas "vocabulary" is used for less complex and formal lists of terms

But for the semantic representation of the later, a simpler formal language will often do...

Semantic data formats

Simple Knowledge Organisation System (SKOS)

❏ Solution for converting a "classic" thesaurus or vocabulary managed in a local database into a terminology compliant with the 4 Linked Data principles being

  – 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 semantic information

  – Include links to other URIs, so that they can discover more things

❏ Based on the RDF specification, enables a migration towards OWL

❏ Ideal compromise for modelling controlled vocabularies such as thesauri and classifications

❏ "Low-cost" migration path to Semantic web interoperability

Semantic data formats

- Structured according to the **ISO 25964 norm** which is dedicated to thesauri and interoperability with other vocabularies and is divided in two parts

  - **Part 1:** Thesauri for information retrieval

  - **Part 2:** Interoperability with other vocabularies

- The **ISO norm and SKOS model are useful guidelines** to follow when conceiving or adapting a terminology as a preparative step towards later interoperability

Semantic data formats

<u>S</u>omething <u>K</u>ool <u>O</u>riginal and <u>S</u>exy?

❑ More and more required by web services

❑ The format recommended and used by **Europeana** for sharing terminology resources

❑ Captures much of the similarity of organisation systems such as thesauri and classifications

❑ Comprehensible format for the non-expert user to map to

❑ Adopted by Linked Heritage WP3 as the central pivot format for the resources uploaded, stored, edited and mapped in the TMP

❑ Important role in the linking data process, but can also be of great use in Europeana (multilingual access, navigation tool…)

SKOS tutorial

SKOS classes

❑ **Concept (*skos:Concept*)**

- An idea, notion or unit of thought

- Unique → 1 concept, can be expressed with multiple lexical labels

❑ **Concept scheme (*skos:ConceptScheme*)**

- Groups of concepts, micro-thesaurus

- The same concept can be part of more than one concept scheme

❑ **Collection (*skos:Collection*)**

- Thematic group of concepts

- Different from concept scheme → the thesaurus as a whole could be considered as a concept scheme where several thematic groups of concepts could be designed as collections

SKOS tutorial

Identifiers

❑ Each concept must be identified in a unique way in order to avoid any ambiguity

❑ Use HTTP URIs to identify concepts according to the RDF, Semantic web and Linked Data principles

❑ Identifiers are introduced as **rdf:resource** properties for

- Each new concept being introduced

- Semantic relations or mappings to other concepts

```
<rdf:Description rdf:about="http://iaaa.cps.unizar.es/thesaurus/HYDROBIOLOGY">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:related rdf:resource="http://iaaa.cps.unizar.es/thesaurus/AQUACULTURE"/>
    <skos:prefLabel xml:lang="fr">HYDROBIOLOGIE</skos:prefLabel>
```

SKOS tutorial

Labels

❑ Distinction between the concept itself and the terms used to express this concept

❑ Terms referring to a concept are expressed via lexical labels

❑ Allows you to have a term in any natural language with or without Latin characters

❑ SKOS defines 3 types of lexical labels

- **Preferred Label**

- **Alternative label**

- **Hidden Label**

SKOS tutorial

❑ **Preferred label (*skos:preflabel*)**

- Preferred lexical label assigned to a concept for a specific natural language

- Only one preferred label is allowed in the same language

❑ **Alternative label (*skos:altlabel*)**

- For synonyms, different spellings and acronyms of the preferred label

- For each preferred label you can have many alternative labels

❑ **Hidden label (*skos:hiddenlabel*)**

- For mentioning misspellings or obsolete forms of a term (UF-Used For)

- Not visible, useful for retrieval

All can be provided in different languages by using language tags with the *xml:lang* attribute (e.g. *skos:prefLable*@*en*)

SKOS tutorial

Documentation properties

❑ Different types of notes can be used to provide more information on the concept (e.g. ***skos:note***, ***skos:definition***…)

❑ Can be provided in different languages by using language tags with the *xml:lang* attribute (e.g. ***skos:note@en***)

SKOS tutorial

Semantic relationships and mapping

- ❑ Feature that enables **interoperability between vocabularies**

- ❑ Consisting of semantic relationships and mapping properties to connect between different concepts

- ❑ 2 semantic relationships play a crucial role in defining the concepts

  - **Hierarchical relationship**

  - **Associative relationship**

SKOS tutorial

□ **Hierarchical properties**

- *skos:broader* : The concept has a more general meaning

- *skos:narrower* : The concept has a more specific meaning

- The concept can have more than one broader and more than one narrower concept

- To be used for direct hierarchical links between 2 concepts

- For non-immediate links, use transitive properties *skos:BroaderTransitive* and *skos:narowwerTransitive*

□ **Associative properties**

- *skos:related* : For associative links between 2 concepts without hierarchical relation

- The *skos:related* property is symmetric

- Associative and semantic relations should not be mixed

SKOS tutorial

Interoperability

❑ In this context it is mainly about **mapping** concepts of one vocabulary to the concepts of another, thus **creating a network of interlinked terminologies**

❑ A mapping is a relationship between a concept in one vocabulary and one or more concepts in another

❑ There are 3 mapping types defined

– **Equivalence**

– **Hierarchical**

– **Associative**

SKOS tutorial

❑ **Equivalence**

- A concept in the terminology 'A' is **equivalent** to a concept in terminology 'B'

- E.g. "***Architecture@en***" of terminology 'A' is **equivalent** "***Architectura@it***" of terminology 'B'

- The equivalence mapping is expressed by

  - ▪ **Use (USE):** For specifying the preferred term
  - ▪ **Used For (UF):** For specifying the non-preferred term

SKOS tutorial

❑ **Hierarchical**

– A concept in the terminology 'A' is **narrower or broader** than a concept in terminology 'B'

– E.g. "***Buildings@en***" of the terminology 'A' can be mapped as **narrower** in **hierarchical relation** "***Architecture@en***" from the terminology 'B'

– The hierarchical mapping relation is expresses by

▪ **Broader term (BT):** This relationship comprises broader concepts with a more generic meaning in relation to a narrower concept that has a more specific meaning

▪ **Narrower term (NT):** This relationship comprises narrower concepts with a more specific meaning in relation to a broader concept that has a more generic meaning

SKOS tutorial

❑ **Associative**

- A concept in terminology 'A' is **related to a concept** in the terminology 'B'.

- E.g. "***Villas@en***" in the terminology 'A' is **related** to the concept "***Houses@en***" in the terminology 'B'.

- The associative mapping is expressed by

  ▪ **Related Term (RT):** This relationship comprises related concepts and see-also-references

SKOS tutorial

Diagram of the different mapping cases

Case	Diagrammatic Representation
1. Equivalence The diagram implies equivalent sets. Circle A and B overlap. Example: ancient monuments USE **monuments** (A) (B) monuments UF ancient monuments (B) (A)	A = B
2. Hierarchical The diagram implies class inclusion Example: **mammals** NT **dogs**	A B
3. Associative The diagram implies semantic overlap, ie. there is and element of meaning common to both terms Example: **gold** RT **money**	A B

Source: A. d'Andrea, Introduction to SKOS. Power Point Presentation 2009-07-16, Rome)
www.athenaeurope.org/getFile.php?id=287

V/ Linked Heritage semantic mapping experiment

– Creating a network of object name thesauri

– Editing tool of the TMP (xTree module)

– Mapping concepts (xTree module)

Creating a network of object name thesauri

Semantic mapping experiment

❑ Creation of a network of LH terminologies

❑ Using the TMP in an early stage

- Work on object name thesauri

- Start small – having the network grow over time

- British museum object name thesaurus + 2 partner terminologies imported in TMP

Status

- Semantic mapping: 4 terminologies & 5000 concepts mapped together, 4 languages (FR, EN, IT, NL)

- New LH partners invited to join in second stage using the TMP prototype

Editing tool of the TMP
(xTree module)

Open a vocabulary

Editing tool of the TMP
(xTree module)

Search for a concept (Tree, Search, List)

Editing tool of the TMP (xTree module)

Basic data

Editing tool of the TMP
(xTree module)

Relations to other concepts

Editing tool of the TMP
(xTree module)

Mappings

Editing tool of the TMP
(xTree module)

Notes

Editing tool of the TMP
(xTree module)

Collaboration platform - Forum

Editing tool of the TMP (xTree module)

Edit a terminology: Add a new concept

Editing tool of the TMP (xTree module)

Edit a terminology: Enrich a concept

Editing tool of the TMP (xTree module)

Edit a terminology – Add multiple lexical labels...

Mapping concepts (xTree module)

Semantic mappings between vocabularies: Open multiple vocabularies

Mapping concepts (xTree module)

Select concept from own vocabulary

Mapping concepts (xTree module)

Search for matching concept in other vocabulary

Mapping concepts (xTree module)

When finding a matching concept, copy URI

Mapping concepts (xTree module)

LINKED HERITAGE

To map, add URI to own vocabulary

Concept "Piece of Furniture" (KMKG) is mapped to concept "Mobilia" (ICCD)

VI/ Perspective

Perspective

- ❑ Continue the collaborative development of Linked Heritage terminologies

    – Enlargement of the network of interlinked object name thesauri

    – LIDO terminologies

        ▪ Events http://terminology.lido-schema.org/eventType

        ▪ Actor roles

    – Start work on new domains & subject thesauri (e.g. Geography)

- ❑ Finalise the TMP prototype (D3.3, Month 24-March 2013)

- ❑ Organise training workshops and provide additional manuals

- ❑ Work towards a productive version of the TMP → AthenaPlus

QUESTIONS?

THANK YOU!

Roxanne Wyns
Royal Museums of Art and History (Belgium)
r.wyns@kmkg-mrah.be

Linked Heritage website: *http://www.linkedheritage.eu*