ATHENA
Access to cultural heritage
networks across Europe

ECP-2005-CULT-038099

ATHENA

Recommendation for integrating Digital resources present in museums in Europeana

Deliverable number	*D4.3*
Dissemination level	*Public*
Delivery date	*30th April 2010*
Status	*In progress*
Author(s)	*Marie-VéroniqueLeroi (MCC/DREST),* *Johann Holland (Michael Culture Aisbl)* *StéphaneCagnot (Dédale)*

*e***Content***plus*

---

1 OJ L 79, 24.3.2005, p. 1.

Table of Contents

1. Executive summary

This deliverable is part of ATHENA Workpackage 4 (WP4) and addresses European museums. It aims to present all our recommendations for integrating Digital resources present in musuems into Europeana.The deliverable is structured as follows:

- *Executive Summary*: A short summary of the deliverable.

- *Introduction*: Explaining the context of the whole work package in which the deliverable stands, the objectives of the task that the deliverable relates to, the audience for these recommendations and the skills needed to understand and to apply them.

- *General Context*: Presenting all the basic knowledge necessary to have in mind before reading the recommendations. In that part are notably presented the current technological environment, the museum needs, what a datamodel is, a typology of terminology resources, and the connexion between the datamodel and the terminology.

- *Recommendations*: Addressing European museums about terminology management. This part helps museums conceive a terminology, make it interoperable and link it to a network of terminologies in order to enable and to improve the digital resources retrievability on Europeana.

- *Conclusions and perspectives*: A synthesis of the main results and several perspectives for the museums which aim to make their digital resources the more exploitable and retrievable again and again.

2. Introduction

Here we provide the reader with information about what context this deliverable belongs to, and what it is for.

2.1. Context and objectives

2.1.1. Athena WP4

Athena is part of the constellation of projects contributing to Europeana. Its general aim is to help the integration of European museums' digital resources into the Europeana portal. In order to reach that goal, Athena is divided into several workpackages which deal with different important topics. The workpackage 4 (WP4) is dedicated to the terminology management with two strong focuses on multilingualism and SKOS(Simplified Knowledge Organisation System).

All along the project, a WP4 work group (WG4) has led surveys and experimentations in order to get results and to phrase guidelines and recommendations for the museums. WP4 has already produced two deliverables (D4.1 Inventory of resources[1] and D4.2 SKOS guidelines[2]), and has organised two workshops (one in Budapest[3] about the process and issues, and one in Paris[4] about a benchmark of tools for SKOSification and terminology management). All that activity has been supported by a communautarianWiki[5] on which are available the results and all that has been produced to get them.

Now the ATHENA project comes to its end and the WP4 is expected to conclude the WG4 activity by providing final recommendations about terminology management to be addressed to all the museums which intend to make their digital resources available on Europeana. This final deliverable D4.3 is then a general synthesis of the work done during the project within WP4 dedicated to terminology and multilingualism.

2.1.2. Recipients

As a set of recommendations, this deliverable is dedicated to museums that are expected to make their digital resources retrievable on Europeana. We have defined these

---

[1] You can find D4.1:
- in pdf version at: http://www.athenaeurope.org/getFile.php?id=398
- in updated wiki version at:
http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources

[2] You can find D4.2:
- in pdf version at: http://www.athenaeurope.org/getFile.php?id=684
- in updated wiki version at: http://www.athenaeurope.org/athenawiki/index.php/Guidelines

[3] All the documentation about Budapest workshop is available at:
http://www.athenaeurope.org/athenawiki/index.php/Documents#WG4_Second_meeting_.28Budapest.29

[4] All the documentation about Paris workshop is available at:
http://www.athenaeurope.org/athenawiki/index.php/Documents#WG4_Technical_workshop_.28Paris.29

[5] http://www.athenaeurope.org/athenawiki/

recommendations by taking into account the reality of their specific technical and economic situation.We can sum up these specificities through three major elements we develop just below: a gap of skills, a lack of financial means, and a misknowledge of the technological environment.

2.1.2.1. A gap of skills

First of all, we know that there is a gap between the skills of museum people about terminology management and the usual skills required in the technical expert fields of Information Engineering and Linguistics. Ideally, any reader with no expert background in terminology management should be able to understand our recommendations. However, because of the high degree of technicity of the topic, some basic knowledge might be necessary to handle for a good and useful understanding of the recommendations. Thus we have decided to make the recommendations the more easy to understand that we could, and to deliver in the next part[1] of this document a synthesis of all the basic knowledge that the reader is expected to have in mind for the follow-up.

2.1.2.2. A lack of financial means

Then, we are also aware of the critical economic situation in cultural institutions. We cannot occult how much any change in terminology use may have a significant impact on the financial and human state of any structure since a costly effort is then expected to be made. Thus we have defined the recommendations taking into account economic difficulties and constraints so that the museums may be able to handle all the required operations in terminology management in the perspective of Europeana. This deliverable, as it is dedicated to non-expert readers, participates to that idea since museums should not call an external competency to read, to understand and to apply what it is recommended in.

2.1.2.3. A misknowledge of the technological environment

Finally we emphasize the fact that a lot of museums do not seem sensibilized to the current technological environment. The unavoidable evolution of the Web has a strong impact on the way howinstitutions have to manage their data. This evolution occurs under the influence of the different new technologies, norms and standards in use,andit is particularly co-dependent of the evolution of the massive Web usages. Because we consider that this misknowledge may produce important difficulties and misunderstandings, we give below[2]a presentation of some key points featuring the current technological environement and its evolution.

---

[1] see the part 3 "General Context"

[2] see the part 3.1 "Technological environment"

2.2. Methodology

In order to deliver our final recommendations in the most effective way, we have shaped the follow-up in three major parts. The first one called "General Context" presents the whole basic knowledge the reader is expected to become familiar with before reading the recommendations. The second part called "Recommendations" phrases in a practical manner all the advices we address to the museums in the field of terminology management. And the third part "Conclusions and Perspectives" present different initiatives to which museums could have an interest to actively participate because these initiativesare respecting, even implementing, the recommendations.

3. General context

In this part we present the general context of a museum dealing with its own collections. This basic knowledge will help the reader to understand our general approach and really benefit from our recommendations. In order to optimize their intelligibility by museum people, we have chosen to employ a "you-style" phrasing as if the reader is effectively someone working in a museum with no specific skill in Information Engineering and Linguistics. So by now let us say: "you are a museum representative". Moreover, at any step of the presentation, we use and enrich our recommendations with some examples, so that the information shall appear more concrete.

The following information has been gathered and consolidated all along the WP4 activity. Here we synthesize the basic knowledge in five sub-parts, and each of these sub-parts answers a simple question:
- What technological reality you are working in?
- What do we call "terminology management" in your case of museums?
- What is a datamodel in relationship with terminology management?
- What are the different types of terminology you can use?
- How a terminology and a datamodel are connected?

3.1. Technological environment

3.1.1. Social Web

Nowadays you are certainly aware of, even familiar with, the so-called Social Web or Web 2.0. As an evolution of the primarWeb, the Web 2.0 has permitted the emergence of networks of people who are meeting and instantly exchanging online on different platforms like Facebook, Twitter or LinkedIn. After having offered an access to information spread around the world, the Web has allowed new kinds of social relationships. Moreover, the new offered functionalities have enabled all Web users to produce themselves the so-called User Generated Contents (UGC). A new era of information has then appeared in which the information does not come only from editors, but from a mix of heterogeneous sources. For cultural institutions a new scope to interact with users/visitors is now possible through these technologies.

3.1.2. Semantic Web

Then these last years a new trend has appeared: the Semantic Web, also known as Web 3.0. This new version of the Web is the new environment your digital resources will be exploited in.Now they are living in a world of connected pieces of knowledge more than on a network of pieces of information. Roughly speaking, yesterday your digital resources were simply and blindly connected, today their relations with the network can have an explicit meaning.The hyperlink is becoming semantic.

More technicallcy, as we presented it in the deliverable D4.2, the Semantic Web (part of Web 3.0) is "the Web of data with meaning in the sense that a computer program can learn enough

about what the data means to process it"[1]. It provides "a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by World Wide Web Consortium (W3C) with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. It was proposed by World Wide Web inventor Tim Berners-Lee"[2].

If you want more technical information about Semantic Web, please see in the D4.2 the dedicated section.

3.1.3. Linked Open Data

In the world of Semantic Web, a new "philosophy" is getting to become the reference you should know when you want to link your digital resources with the ones already available online.Thisinitiative is called Linked Open Data (LOD). Europeana in order to apply the goals defined in the strategic plan 2011-2015 considers that LOD is critical for the success of its cultural policy[3].

For Europeana, **Linked Open Data** is:
• A technology to combine the many pieces of information we get from data providers.
• A way to share that data with other parties.
• A way to give users the best possible search experience.

From a general point of view, LOD participates to the evolution of the Web which is then no longer a flat list of data but a structured access to all the available resources. If you conform your own data to the LOD norms, you will be able to easily make your data visible through mobile applications, and to benefit from the whole datacloud of URIs in which are already networked some reference resources like DBPedia. For example, if you map your normalised LOD data to DBPedia, you are sure to map them to all the existing other referenceresources in the same time.

Linked Open Data addresses a set of rules, tools and recommendations to the content providers (like museums). Among all of this, most of all you can keep in mind that all the data you want to provide to Europeana have to be named and linked. Our recommendations below help you to complete these required actions before the ingestion of your data on that platform.

3.1.4. Formats

In order to be part of the Linked Data 'cloud' and use Semantic Web technologies the terminology of an institution has to be in compliant format. When you want to represent or

---

[1] http://www.w3.org/People/Berners-Lee/Weaving/glossary.html

[2] http://www.uen.org/core/edtech/glossary.shtml#S

[3] https://version1.europeana.eu/c/document_library/get_file?uuid=374c381f-a48b-4cf0-bbde-172cf03672a2&groupId=10602

model your terminology, and to exploit it on the Web, you have to use a format standard. The most commonly used format standards are SKOS, OWL, RDF, and XML. Some of them can be combined, and some of them can be wrapped by others. Using a format standard will result in the metadata, expressed with your terminology, being effectively represented in a way the Web technologies can recognize and interpret.

Below are brief descriptions of these format standards with the aim of a better understanding of their connections.

XML[1]

XML (Extensible Markup Language) is a set of rules for encoding documents in machine-readable form. It is defined in the XML 1.0 Specification produced by the W3C, and several other related specifications, all free to use open standards.

XML's design goals emphasize simplicity, generality, and usability over the Internet. It is a textual data format, with strong support via Unicode for the languages and scripts of the world. Although XML's design focuses on documents, it is widely used for the representation of arbitrary data structures, for example in web services.

There are many programming interfaces that software developers may use to access XML data, and several schema systems designed to aid in the definition of XML-based languages.

RDF[2]

The Resource Description Framework (RDF) is a family of W3C specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modelling of information that is implemented in web resources, using a variety of syntax formats.

The RDF data model is based upon the idea of making statements about resources (in particular Web resources) in the form of triples. Triples are the expressions of statements about resources which are presented as subject-predicate-object expressions. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object.

The RDF specification is based on the XML encoding.

OWL[3]

---

[1] http://en.wikipedia.org/wiki/XML

[2] http://en.wikipedia.org/wiki/Resource_Description_Framework

[3] http://en.wikipedia.org/wiki/Web_Ontology_Language

The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. The languages are characterised by formal semantics and RDF/XML-based serializations for the Semantic Web. OWL is endorsed by the World Wide Web Consortium and has attracted academic, medical and commercial interest.

In October 2007, a new W3C working group was started to extend OWL with several new features as proposed in the OWL 1.1 member submission. This new version, called OWL 2, soon found its way into semantic editors such as Protégé and semantic reasoners such as Pellet, RacerPro and FaCT++. W3C announced the new version on 27 October 2009.

The OWL family contains many species, serializations, syntaxes and specifications with similar names. This may be confusing unless a consistent approach is adopted. OWL and OWL2 will be used to refer to the 2004 and 2009 specifications, respectively. Full species names will be used, including specification version (for example, OWL2 EL). When referring more generally, OWL Family will be used.

OWL is based on the RDF specification.

SKOS

In this set of formats, SKOS is more and more required by web services. Europeana for instance has decided to format in SKOS all the metadata they harvest for a homogeneous and effective exploitation of the resources, of the data and their related descriptions. SKOS is based on the RDF specification and enable a migration towards OWL ontologies.

SKOS is not a formal knowledge representation language since literally a formal knowledge is expressed as sets of axioms and facts which are the main features of a formal ontology. SKOS is rather used for modeling controlled vocabularies such as thesauri or classifications which are of a different nature than ontologies. The ideas or meanings described by thesauri or other kinds of terminology are referred to as "concepts" even if from the ontological point of view a concept is defined in a different way.

The next section defines more precisely what SKOS is and what its features are.

3.2. Museums case

You are a European museum and you intend to make your collections available on Europeana. Now, some of your collections are natively digital and some others are not. Many European projects such as Minerva or MICHAEL have raised the importance of the digitisation as much for the access by the general public all over Europe than the cultural heritage preservation. Europeana, as the digital European library gives only access to digital elements. This is the step of digitization which enables you to integrate your physical collection into your digital one. But this is not enough. After the digitization, you need to manage your documentation (description of objects, references...) and transform it into a set of metadata related to all your digital objects. So that you finally have a collection of digital resources for managing your collections in your own database or system and then make them available to Europeana.
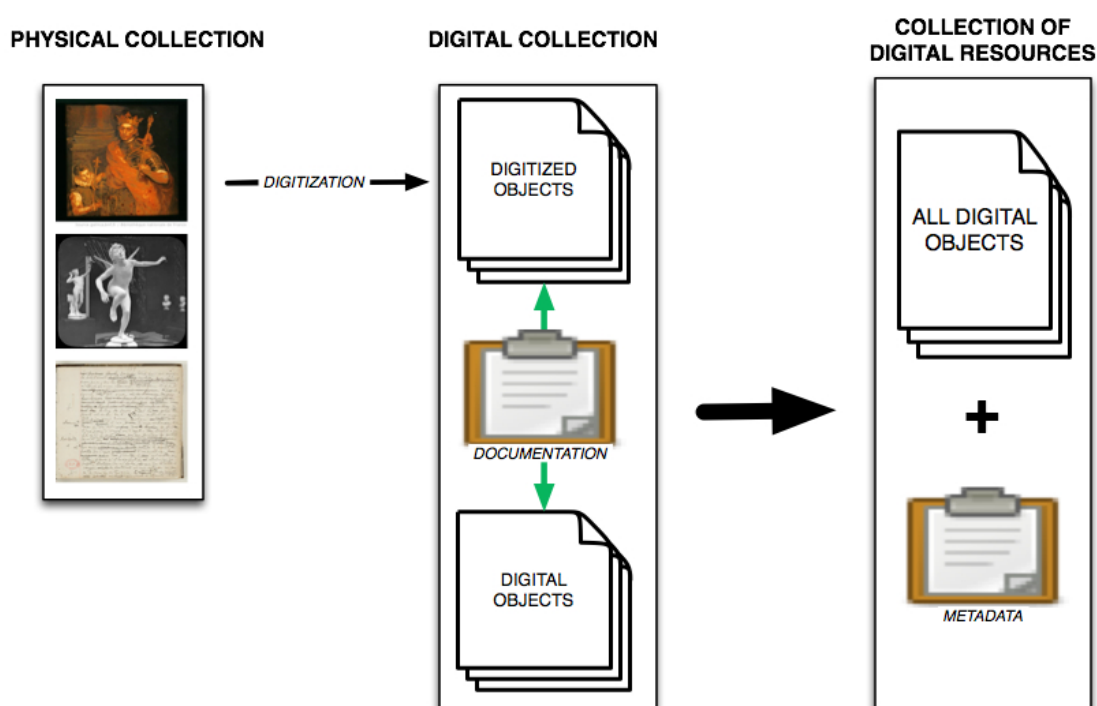
Ex: You have collections of paints, sculptures and manuscripts hold by your museum. In order to make these collections available on the Web, and especially on Europeana, you digitize them by producing photographs, 3D renderings, OCR[1]-generated texts, and by fulfilling digital notices for the complete catalog. All these elements are your digital resources intended to be available on the Web.

---

1 OCR: Optical Character Recognition

More precisely, you intend to make your collection of digital resources available and retrievable on Europeana. Basically, when you want to reach such a goal, first you have to prepare your digital resources. Indeed your resources are not natively compliant with the Europeana's requirements. Athena project aims to help you specifically at this stage by offering you tools, guidelines and recommendations.
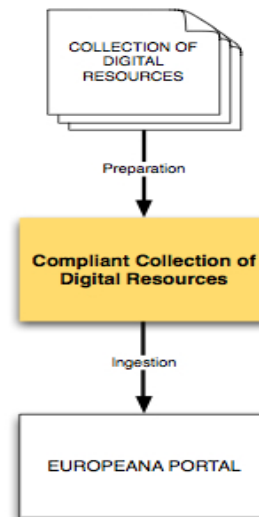
Then, in order to prepare your digital resources for ingestion in Europeana, you have to take care about two general aspects. The first aspect is technical, it consists in guaranteeing the access to every digital resource as an object among a collection of objects. The second aspect is semantic, it consists in exploiting the content of any digital resource as a meaningful element in a collection. This specific aspect of semantic re-use of digital resources is developed further in the document.

Ex: In your collection of paints you notably have a Greco's piece of work: Saint Louis, Roi de France. *You want to make compliant in particular one digital photograph[1] of that work with Europeana requirements: you are going to prepare that picture both technically and semantically.*

---

[1] http://gallica.bnf.fr/ark:/12148/btv1b6904167m/f1.item.r=greco.langEN
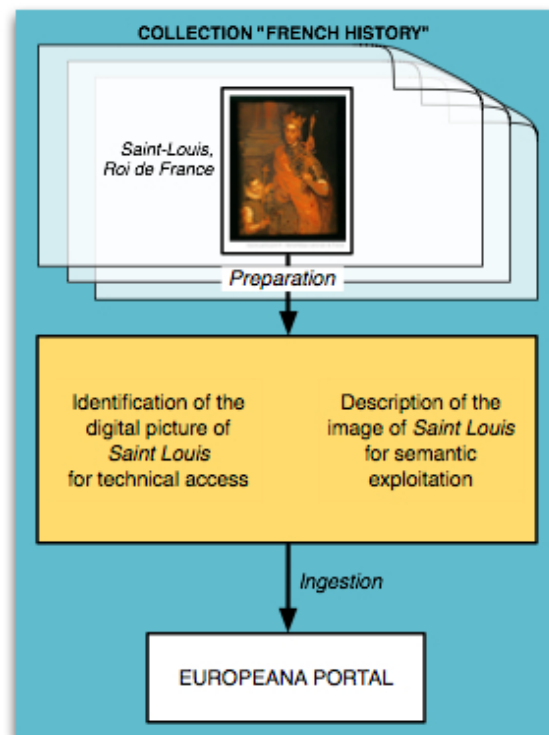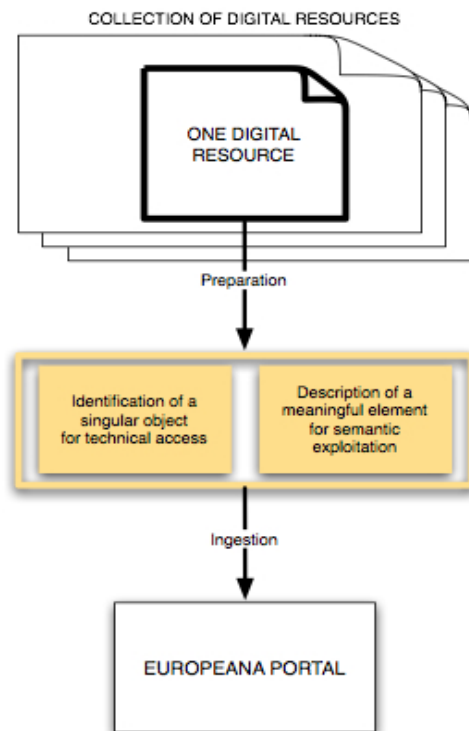
Technical preparation

As you can read in the parallel recommendations[1] jointly provided by Athena WP3 and WP7, the technical preparation requires identifiers and referencing of the collection items. Indeed, the technical access to the digital resource on Europeana implies its identification as a singular object among a mass of items, and its cataloguing as an element of the collection.

You can refer to the documents and guides elaborated by the WP3 of Athena for more information on this technical preparation[2].

The identification name system you can use depends on the type of works you manage, and on your administrative constraints (e.g. if your information system is based on ARK[3] instead of PURL[4], a subscription step in the

Ex: You are technically preparing the digital photograph of Saint Louis. *It means you register the digital file in your file management system. Your identification system based on ARK provides a unique identifier for that picture :*

http://gallica.bnf.fr/ark:/12148/btv1b69 04167m

Then you use this identifier when you employ your datamodel LIDO in which are referenced your collection (French History), the work title (Saint-Louis, Roi des Français), its type (Paint), and the related classification as well (My Italian Paints).

---

[1] See:

- For LIDO datamodel: http://www.athenaeurope.org/getFile.php?id=539
- For Persistent identifiers: http://www.athenaeurope.org/getFile.php?id=772

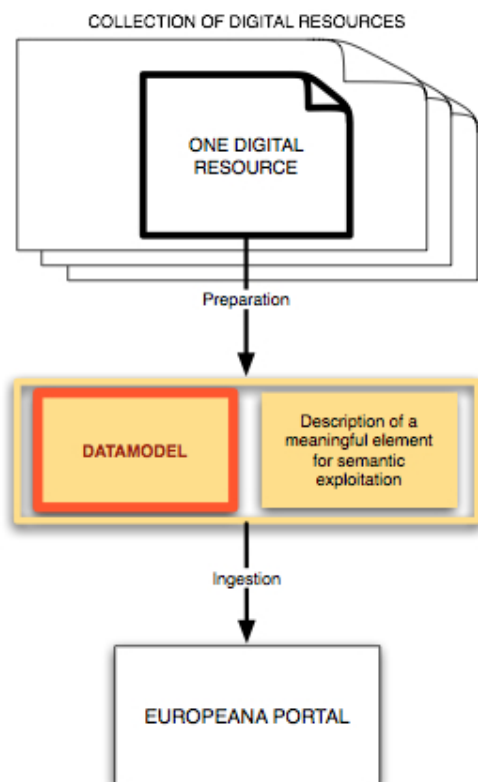[2] http://www.athenaeurope.org/index.php/en/149/athena-deliverables-and-documents

[3] ARK: Archival Resource Key

[4] PURL: Persistent Uniform Resource Locators

identification process impacts the identifier name).

Such an identification is used by the **datamodel** which enables you to declare any item of your collection as a singular element.
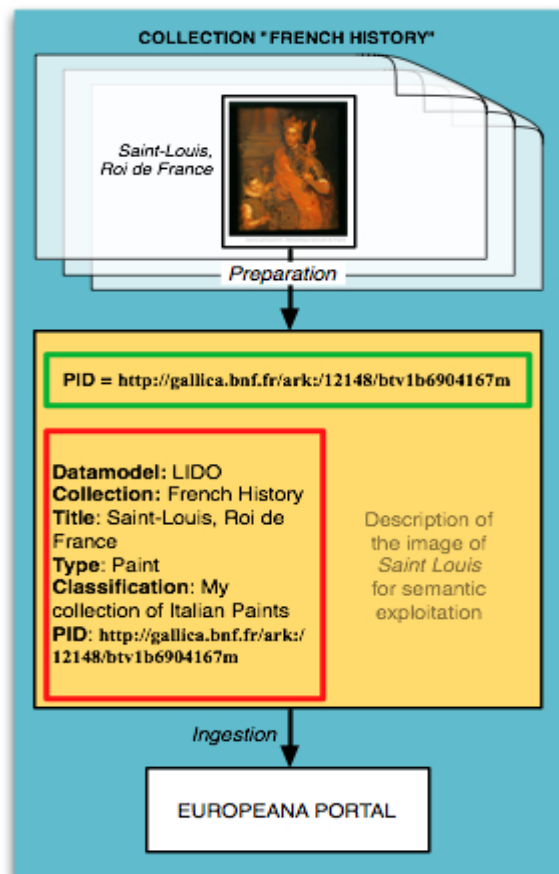

Semantic preparation

The semantic preparation requires a description of the digital resource as a meaningful element in a collection organized according to an editorial point of view. It implies the use of a **terminology** enabling you to feature your digital resource thanks to terms, even concepts and relations in order to contextualise them.

We call **terminology management** all the activity consisting in the handling of the semantic description of the digital resources.

Ex: You are going to semantically prepare your digital picture of Saint Louis. *You want to join metadata to that file in order to express that the author is Il Greco, that it has been painted in 16th century in Italy, and that it belongs to the artistic period called "Renaissance". So that You use your usual terminology to fulfil correctly your metadata schema by giving the value "Il Greco" in the field Author, "16th century" in the field Creation Date, "Italy" in the field Creation Country and "Renaissance" in the field Artistic Period.*
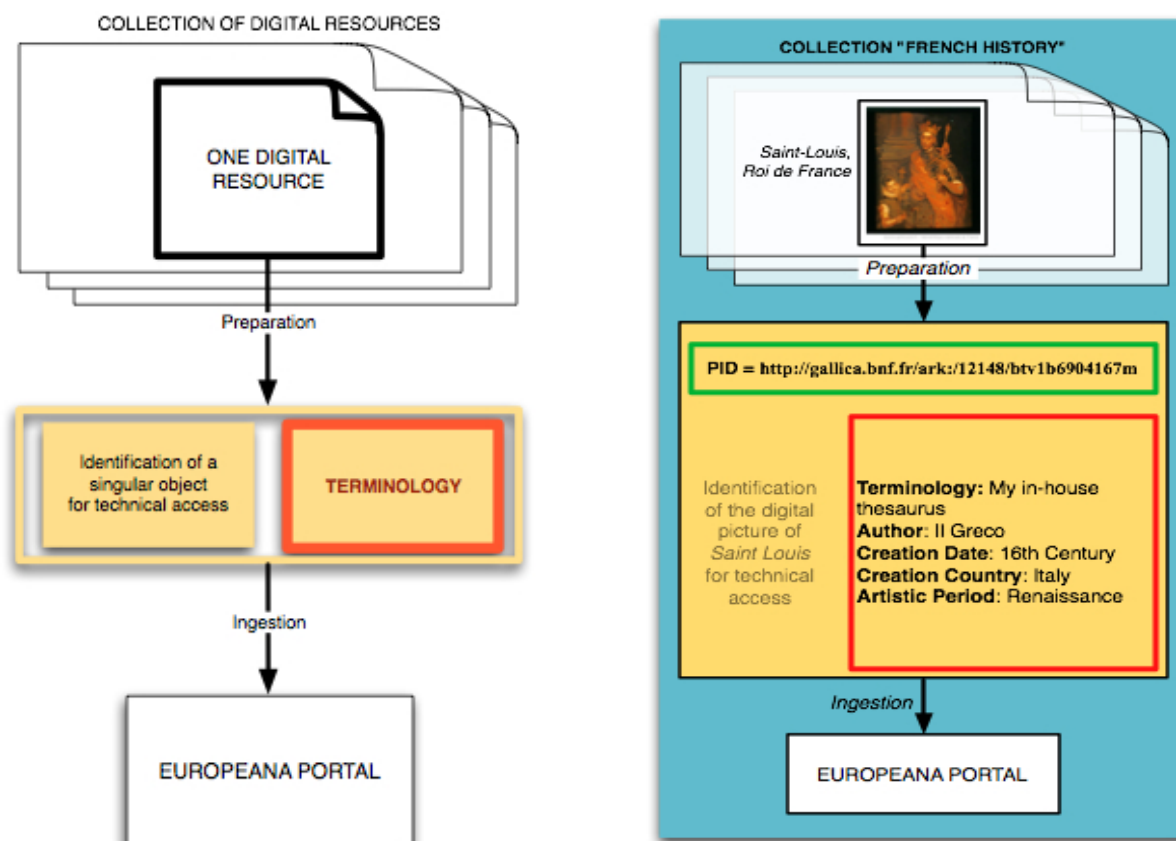
3.3. Datamodel presentation

Here we provide the reader with general presentation of what a datamodel is and what especially LIDO is. All these coming information come from the documents produced in Athena by WP3. This information is given here for the general context and understanding of the deliverable but you are invited to refer to the publications of the WP3[1] for further details.

3.3.1. General presentation

As it is introduced in WP3 report "Standards landscape for European museums, archives, libraries[2]", we can consider that a datamodel in general helps identifying a collection object by giving a core set of informations.The Dublin Core (DC) format is the most commonly used. Namely, 9 out of the 15 DC elements are commonly used for describing an object.

---

[1] http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents

[2] See the report "Standards landscape for European museums, archives, libraries" that you can directly donwload at:
http://www.athenaeurope.org/getFile.php?id=435

These elementsare:
- **Title:** The name (or names) of the object or collection.
- **Creator**: The name of the organisation or individual who originally created the object or collection.
- **Publisher:**The name of the organisation that makesthe object or the collection publicly available.
- **Date**: The date on which the standard wasoriginally published.
- **Identifier:**A number or other identifier under which object (or collection) is published or a URL which points to the definition of the object (or collection).
- **Rights**: Whether rights restrictions, e.g. patents,apply to the object (or collection).
- **Description**: A textual description explaining the object (or collection) and its usage.
- **Subject**: Keywords that identify the nature of the object (or collection).
- **Relation**: Other objects (or collections) that this object (or collection) relates to, and associated websites.

The Dublin Core is a simple metadata element set intended to facilitate discovery of electronic resources. Elements can be grouped into those having data on: Content – Coverage, Description, Type, Relation, Source, Subject, Title; Intellectual Property – Contributor, Creator, Publisher, Rights; Instantiation – Date, Format, Identifier, Language. Its use has been mandated by several governments in Europe (e.g. UK) and throughout the world (e.g. Australia).

3.3.2. LIDO

Among all the existing standards of datamodel, we particularly recommend LIDO (Light Information Describing Objects)to the European Museums.There are four main reasons.

First of all, this datamodel has been defined by Athena WP3 specifically for the museums. Mixing elements coming from Spectrum, MuseumDat and DC, LIDO takes into account the specificities of your situation.

Then LIDO is already mapped with the Europeanadatamodel ESE (Europeana Semantic Elements) and available on the data ingestion platform (Athena Ingester).So if your datamodel is mapped to LIDO you do not have to worry about the compliancy with Europeana today.
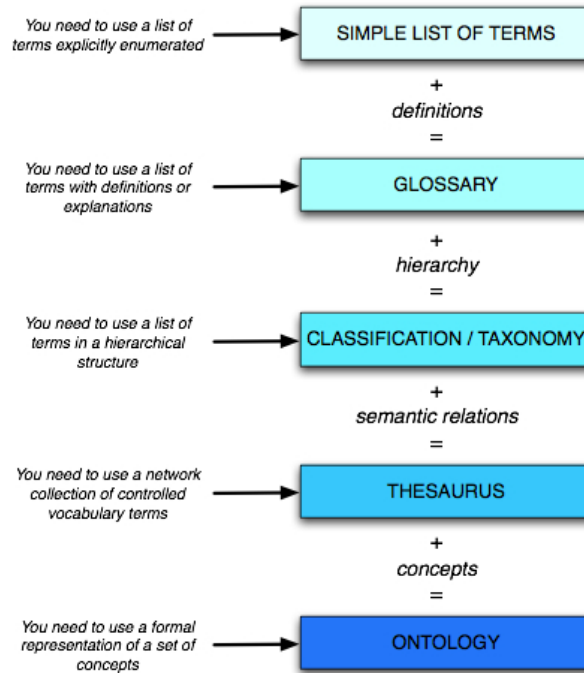
Moreover, LIDO offers more possibilities than Dublin Core to describe efficiently your digital objects since it is conceived as a set of classes gathering fields. These classes are: Object Classifications, Object Identifications, Events, Relations, Administrative Metadata.
A description of object organised into structured classes (such as LIDO datamodel) rather than a flat and linear description (such as Dublin Core) allows a better exploitation of these descriptions by a human user and a machine (search engine, database, ...) as well.

Finally, LIDO with its classes will be easy to map with the next Europeanadatamodel. Indeed Europeana is currently releasing a new datamodel, EDM (EuropeanaDataModel) which will progressively replace ESE. EDM offers a class-based structure which is close to the structure of LIDO and perfectly compliant with the Linked Open Data. If you already use LIDO to be compliant with ESE today, tomorrow the transition with EDM will be easy to be done.

3.4. Types of terminology resources

For the semantic description of your digital resources, different types of terminology are available. We have presented them in detail in our first deliverable D4.1[1]. Here we propose a very schematic graph as a short and synoptic reminder.

Recommendation for integrating Digital resources present in museums in Europeana

ATHENA
Access to cultural heritage
networks across Europe

3.5. Connexionterminology ←→datamodel

In this deliverable we focus on the semantic aspect, thus we provide recommendations about terminology management rather than data description. However the two aspects are not totally separate, there is a connexion in-between we precise here a bit.

This connexion enables you to link the element semantic description to the object technical identification. Indeed the datamodel can transport the semantic descriptions if these descriptions are compliant with its features, *id est* the actual structure of the datamodel (e.g. for my semantic description I need a field "Place" which is or is not in the datamodel set of fields).

The connexion ensures the compliancy of these descriptions with the datamodel.

Ex: Since you want to provide the type of the work Saint-Louis in your datamodel, you have to connect the LIDO field of description Type with your related list of terms structured as following:

- *Cultural Heritage*
  - *Art*
    - ***Painting** (as descriptor in English), Peinture (as equivalent descriptor in French), Dipinto (as equivalent descriptor in Italian)*
    - *Scuplture (as descriptor in English), Sculpture (as equivalent descriptor in French), Scultura (as equivalent descriptor in Italian)*
    - *Manuscript (as descriptor in English), Manuscrit (as equivalent descriptor in French), Manoscritto (as equivalent descriptor in Italian)*

From the theoretical point of view, such a connexion is the link between your grammar (the datamodel) and your vocabulary (the terminology).And from the point of view of the implementation, it means:whateverthe input format of your terminology, there are formats to make your semantic descriptions exploitable by an engine. Moreover, the specifications of the Semantic Web intend to lower this clear separation between data and terminology.

This general context we have proposed must help you understand and apply the recommendations given in the next part. Since these recommendations concern only the terminology management, now we have to focus on the specific part of the semantic exploitation of your description.
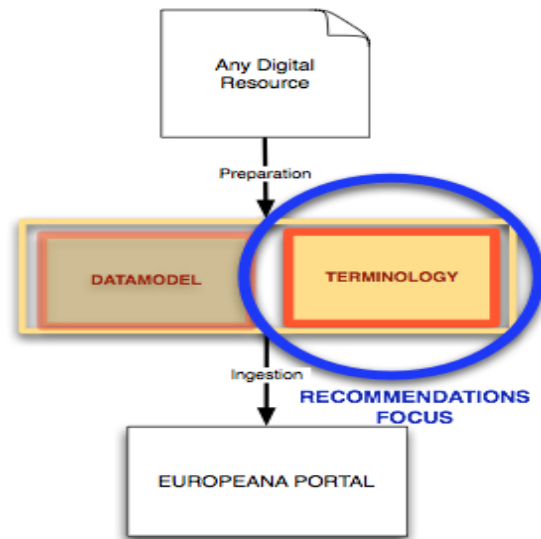
4. Recommendations

4.1. Recommendations purpose

4.1.1. Compliancy and optimization

As we notified in the introduction, our recommendations take into account the recipients' point of view, that is, your background, objectives and interests as a museum representative.

First, regarding your **background**, as we have already said previously, we aim to state our recommendations in a manner **intelligible by non-experts**. In the rest of this deliverable,we keep phrasing things in the same way so that you shall understand and apply what we propose to you.

Then, regarding the **objectives**, our recommendations shall enable you **to be compliant** with Europeana requirements. But that objective is very minimal since, when you look at these official requirements, today you just need to make your data compliant with Semantic Webstandards in orderto fit with the portail constraints for the semantic exploitation of your digital resources descriptions.So you could use OWL[1], which is the most formal language to implement the semantic Web, to format your data in a concern of interoperability, but in fact it may not be necessary or relevant in your case. SKOS[2] is a more "economical" solution since its functionalities cover most of your needs and its use does not require as much costly technical skills as OWL does.

Thus finally we also look at your **interests** by writing these recommendations. Indeed you can do more than SKOSifyingyour terminology (e.g. transforming your terminology into SKOS) in our context without spending too much money and time. There are several "simple" operations you can do on your terminology which will certainly improve the semantic exploitation of your digital resources on Europeana, right now, but most of all in the future. In this spirit, all our recommendations must help you **optimize their retrievability**by a Semantic Search Engine on Europeana. Europeana so far has developed a prototype for a semantic search engine that will be integrated in the main user interface in the future.

4.1.2. Approach in three steps

We have structured our set of recommendations in three steps in order to simplify their presentation and their understanding. Each of this step brings elements for making your terminology compliant with and optimized for the semantic-exploitation requirements of Europeana. Even if they are presented along a linear process, we strongly recommend you not to hesitate to use them in an iterative and more "anarchical" manner if you need it. Do not stay blocked at a step too long, and try to keep acting in any way.

---

[1] OWL: Web Ontology Language

[2] SKOS: Simple Knowledge Organisation System

The first step is about the **conception** of your terminology. So to say, at this stage you manage your terminology "internally" in order to make a thesaurus in a "human" perspective. We present you different operations you can do on your own to build a new terminology or to adapt the one you already use in order to optimize your digital resources descriptions on Europeana. These operations have to be done in priority since they determine the two other steps.

Then the second step consists in making your terminology **interoperable**. Now you bring your terminology out of the museum when you SKOSify it by taking into account the machine perspective. This is the specific part about the connection which enables you to make the relationship between the datamodel and your semantic descriptions. For the time being, SKOS is required by Europeana. Thus we particularly focus on that specific format.

Finally we address you our last recommendationsas they concern the **networking** of your terminology with others. At this last stage you think of being visible in Europe in a network perspective by integrating your terminology in a network of SKOSified Thesauri. This will be considered as the third and last step of our set of recommendations.
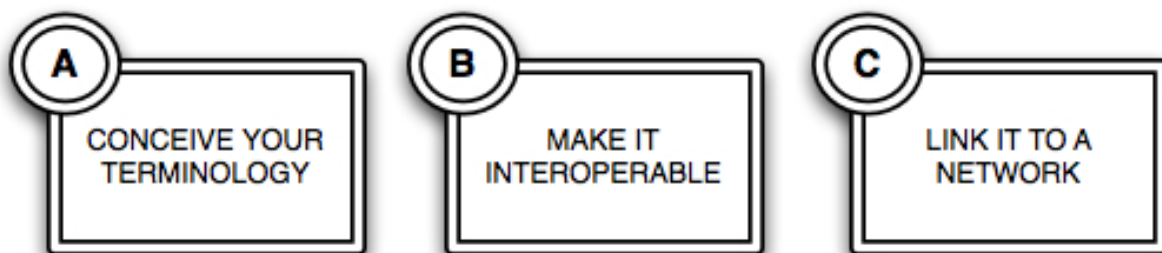
4.2. Conceive your terminology

The conception of your terminology constitutes the framework of all the rest. It determines the operations you shall do later when you will make your terminology interoperablewith other resources,and when you will link it to a network of terminologies.

There is aterminology form that we refer to in that part as an "ideal" means to optimize your semantic descriptions of your digital resources. This terminology form is: **a domain-specific, multilingual and user-oriented thesaurus.** The closer to the ideal form your terminology is, the more optimized the exploitation of your semantic descriptions on Europeana will be. Thus we advice you to try to fit the more you can with such an ideal form by applying our recommendations of conception and refinement we present step by step right now.

4.2.1. Methodology

Here is a task list that you can follow step by step if you want to conceive from scratch a terminology for the semantic description of your digital resources. If you already have an in-house terminology, this is not worthy to trash it, but we recommend you to check it and to refine it if necessary. For that purpose you can use our task list as a quality process.

Even if the task list is presented just below along a linear process, it is really important to use it in an iterative manner. Such a work requires a bit of patience and time that you should consider as an investment. The more precisely your terminology is defined, the more important your return on investment will be. So do not hesitate to switch from a part to another one, to use them by iterations.

The different tasks we are going to detail are:
- A1:Define your collection domain(s)
- A2:Identify your users'expectations (aboutyour semantic descriptions)
- A3:Define your connection with the datamodel
- A4:Choose the terms for thesemanticdescription of your digital resources
- A5:Organise your terms into a thesaurus structure
- A6:Find equivalent terms in other languages
- A7:Implement your thesaurus

Recommendation for integrating Digital resources
present in museums in Europeana

ATHENA
Access to cultural heritage
networks across Europe

A1	**DEFINE YOUR COLLECTION DOMAIN(S)**
	Actions:
	First of all, define your collection domain(s) by answering the following questions:
	• Is there a general domain that your whole collection of items belong to? (e.g. archeology, art, science…)
	• Can you divide your items into several specific sub-domains? (e.g. for the general domain "art": "paints", "sculptures", "cinema","literature")
	Purpose:
	The objective of this first step is to prepare the choice of your terms of description (step A4: Choose your terms). The more specific your domains are, the more precise and non ambiguous your terms will be.
	Prefer a domain-specialization at this step, and later create bridges between specialized thesauri (step A5: organise your terms into a thesaurus structure). Rather than trying to constitute one big thesaurus for all the areas to deal with, we recommend you keeping and enrichingyour existing specific-domain-thesauri without broadening them to other domains. It sounds better to add new thesauri to cover new domains, and to set up "bridges"between the thesauri if the retrieval issue on Europeana requires a cross-domain browsing.A "bridge" in that context consists of mapping terms from different mini-thesauri thanks to relations.
	You can consider that this task is over when, after the step A4: Choose your terms, in your lists there is no more ambiguous term which could belong to several separate domains.
	For example
	If you intend to describe an organ as music instrument and to make a terminology about musicology, and if moreover you aim atdescribing that organ as a religious movable, refer to diffrent micro-thesauri about "musicalinstruments" on one side, and "religious movables" on the other side, instead of mixing terms of these different domains with the ones directly related to musical instruments. So that you will finally have twocollection domains at least to take into account: "musicology" and "religion". And at least two sub-domains: "musical instrument" and "religious movables".

Methods and tools:

As a first step to define your general domain, you can consult HEREIN and MICHAEL which propose a very large typology.

Then, to go deeper in that definition, you can see how the project MIMO has structured several subdomains in its peculiar domain of musical vocabular[1].

---

[1] http://incipioinfodoc.archimed.fr/Idesia/home.aspx?INSTANCE=MIMO&THES=IFD_MIMO_CLASSIF&VIEW=DEFAULT&FORM=0&ACTIVE=TRUE

A2	**IDENTIFY YOUR USERS' EXPECTATIONS**
	Actions: Identify for all the users of your semantic descriptions whichexpectations they have by using them.You can answer the following questions: • Which kind of people shall use your semantic descriptions? (e.g. art amateurs, specialist academics, scientific students) • What are the main interests of these users in accessing to your collections? (e.g. entertainment, research) • Are these people expecting expert descriptions? Which terms are they using to query your collection domains? • In regards with these expectations, which kind of licence are you ready to allow? Do you agree to allow a professional free use of your future thesaurus?
	Purpose: The objective of this step is to prepare the choice of your terms of description (task A4: choose your terms). The idea is to understand what terms the users will spontaneously search in a Web request, and what other terms they should use if they want to be very accurate. Think your terminology as general-user-oriented to fit with the expected request. Because Europeana is a portal for accessing to the collections data, earlier the point of view of who is accessing is taken into account, the more efficient the portal should be. Most of time, requests will not be expressed by professionals, but by the general public. It means the Europeana terminology datacloud does comply with what the general public is "functionally" expecting. If your candidate thesaurus has natively been designed in the same perspective, we can guess that it will bring some relevant results on the portal. Thus we recommend you designing thesauri by considering the skills, habits and expectancies of the general users thenthe professionals as well. It means both two approaches can be considered in the meantime as complementary: 1/ the "bottom- up" approach consists in starting from the needs and habits of the professionals to determine the terminology; 2/ the "top-down" approach on the contrary in coming from the specificities of the access and research by the general users. Regarding the licence for your terminology use, you have to know your legal environment before any choice. What kind of use your institutions is used to allow and under which conditions? It is important to face the point right now even if you will effectively declare the licence at step C1: Definition of the metadata describing your terminology. You can refine this step and consider it as completed when, after the step A4: Choose your terms, a set of test-users can say for each term what type of precise request it refers to or by analysing the query words that have been used to search your collections.

For example

For a collection about underwater archaeology, you may have identified at least two different types of users' expectations:

- General request about spectacular discoveries (e.g. sunk boats, antic ruins felt down in the see): namely to know locations, dates of discoveries, native ages of what has been found, people who have discovered
- Accurate scientific search about investigations: namely to know contexts of the missions, protocols of discovery and datation, hypotheses and arguments

So you better know which kind of informationyour terminology has to provide.

Then, since you consider that the use of your own descriptorsfor the description of other collections can foster the visibility of your digital resources,you can decide to licence your terminology under a Creative Commons Licence CC-By:-Share alike[1](the user can use your descriptors without any modification and with mentioning your institution as author of the terminology).

Methods and tools:

In order to know better the users' expectations, you can at least make two simple surveys and mix their results:

- First a survey consisting in asking the visitors of your physical collections and exhibitions (e.g. ask them if they would be interested in virtual tools for a use in the museum or anywhere else online; and for those who are interested, ask them which kind of request they could do)
- A second survey from the statistics of your own museum website (e.g. what term are often used to access to your collection pages)

---

[1] http://creativecommons.org/licenses/

A3	**DEFINE YOUR CONNECTION WITH THE DATAMODEL**

Actions:

As we explained upper, you have to manage the connection of your terminology with your datamodel. You can make this definition by following two steps:

1. For each kind of query you have identified in step A2: Identify your users' expectations, deduce now which type of information your set of descriptions must contain to provide the user with a relevant answer.
2. In your datamodel, find the description fields corresponding to the kinds of information your users will need.

Purpose:

The objective is to connect your terminology with the datamodel by mapping the terminology model defined within the datamodel and the one you need for satisfying your users' expectations.

Indeed, the first goal is to define all the kinds of semantic information that a search engine could use to give relevant answers to the users' queries about the content of the digital resource. The more your terminology modelcovers the users possible queries, the more relevant and reliable the user query results will be. And the second goal consists of the mapping of that terminology model with the terminology model defined in the datamodel.

Here we invite you not to restrict yourself at this stage by thinking too much to SKOS possibilities or limitations.For instance, SKOS is not designed for the complete description of people so an information such as birth date or death date couldn't be well modelled with the core of SKOS.Anyway, if your users' requests may require these dates, foresee to have a list for that kind of information. You will see below (step B1: Evaluate how far SKOS is compliant with your terminology features) how to manage such peculiarities.

You can consider that this task is over when all kinds of information necessary to provide has a correspondent field in the terminology model defined in the datamodel.

For Example

For instance, in the context of "underwater archaeology", if your users are expecting to find, through the semantic descriptions of your collection items, the locations, the dates of discoveries, and the protocols of discovery of what has been found, your terminology model must use the fields "Place Information", "Date Information", "Acquisition Information"; so that all your digital resources will be precisely retrievable and the semantic results reliable.

Methods and tools:

In order to help you make this connection more easily, you can find a *mapping sheet* in annex of this deliverable.

Among the existing ISO norms, you can look at *BS8723: Structured Vocabularies for Information Retrieval*and also follow the norm ISO 25964-1 Thesauri and interoperability with other vocabularies: Thesaurus for information retrieval (which will be published in 2011) because it expresses more precisely the link betweenthe collections' management and the vocabularies and the technical implementation of thesauri with SKOS.

As a matter of fact, it can be useful to note that the tool xTreedeveloped by Digicult in Germany takes into account this new norm even if you can apply it only if your terminology is already SKOSified.

A4	**CHOOSE YOUR TERMS**

Actions:

Choose your terms for each sub-domain you have defined previously (step A1: Define your collection domain(s)), and deeper for each field of your terminology model. In every case, try to have a middle way between the high precision of an expert/professional vocabulary and the general use by the users you have identified (step A2: Identify your users's expectations).

1. First of all, for each datamodel field you have selected, look at the vocabularies which are proposed either by default in the datamodel, or listed in a repositoryyou already know and consult. Namely, look at their relevance for your users' expectations, and check how relevant it is for you to use and modify them (especially check under which licence conditions you can use and modify them).Afterwards you can decide:
    a. Which vocabularies you will use directly with no change;
    b. Which vocabularies you intend to use directly but after having modified and adapted them.
2. Then, in the case you want to (and you can) modify existing vocabularies and then to use them, follow the respective adaption processes related to all of these vocabularies (do not forget to check the rights conditions).
3. Finally, for each field you don't have terms yet, choose them by your own:
    a. Either by finding a unique compromise term which is an authority (it means the expert term corresponds exactly to the generally-used term);
    b. Or if you cannot find a compromise, by using two different terms as synomyns. You will define the relation in-between later at the step B3: Define with precision the labels expressing concepts(e.g. the expert term will be the descriptor or preferred term and the general term will be an alternative one).
4. Particular cases:
    a. For more precision, in the case ofcompound terms try as much as possible to get a root term which the compound terms will be related to. A compound term is a term made up of two or more words; "comparative literature" is a compound term for example. It will be better to have a term "literature" that will be the broader term of "comparative literature".
    b. Do not avoid terms because they are obsolete or forbidden forms, they may be used in a query and it is important to keep track of the history of your vocabulary.

Purpose:

The objective is to find the best set of terms for the semantic description of your digital resources. The more your set contains expert terms usable by your general users, the more useful and relevant for Web access and retrieval your descriptions will be.

The use of existing vocabularies like Getty thesauri can strongly help the

retrievability of your objects on Europeana. However you have to check which of them are really relevant for your needs. Some of the existing vocabularies are free to use, even enrichable, but some others are not. A case by case checking is then necessary.

If you decide to use an existing vocabulary by modifying it, you have to follow the official adaption process. For all the cases if you do not find a relevant vocabulary for your specific needs, we invite you to create the terms by your own following the existing norms for conceiving a thesaurus (see the norms references given in A7). If you do so, do not forget that obsolete and forbidden forms can be used as terms in a query. So that it can be interesting to have them in your set of terms even if you will not make them really visible at the end (at the step A5: organise your terms into a thesaurus structure and B3: Define with precision the labels expressing concepts).

You can consider that this task is over when after this task a set of test-users can say for each term what it refers to and how their test queries are fully satisfied.

For example

You are using LIDO as a datamodel and have identified the description fields "Place Information" and "Acquisition Information" as mandatory for your needs about locations and protocols of underwater discoveries. So you are now choosing the related terms for these two fields.

When you use LIDO, you have the possibility to use the Getty vocabularies. And among all the vocabularies proposed by Getty, you notice the one about the geographic names (TGN: Thesaurus of Geographic Names) which could be relevant for your needs about locations. After having looked at its content, you decide to use it in order to describe the locations of the underwater discoveries of your collection objects.

However you donot find among the by-default vocabularies proposed within LIDO a relevant one for the descriptions of the scientific protocols used for discovering. So for that field you decide to create your own list of terms: "Archive analysis", "Accurate underwater exploration of a zone", "Fortune"…

Methods and tools:

If you look for existing vocabularies, we advice you to search in the Athena WP4 inventory of resources available at:

http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources

Here you can directly find Getty Vocabularies:

http://www.getty.edu/research/tools/vocabularies/index.html

Recommendation for integrating Digital resources
present in museums in Europeana

ATHENA
Access to cultural heritage
networks across Europe

A5	**ORGANISE YOUR TERMS INTO A THESAURUS STRUCTURE**
	Actions: Structure your terms set by: 1. Creating one micro-thesaurus (list of terms) for each domain or sub-domain you have defined at step A1:Define your collection domain(s). 2. Gathering terms from these micro-thesauri out of the constraint of domain belongingness, but in a transversal or cross-domain approach (thematic approach) 3. Making relations between the lists of terms according to the model of a network of micro-thesauri: "narrower", "broader", and "associated to"
	Purpose: The objective is to transform your simple lists of terms into a thesaurus, that is, a structured network of micro-thesauri. Indeed, among all existing types of terminology resources, we recommend the use of **thesauri** for museums in order to make their collections available to Europeana. First, this type of terminology is quite easy to SKOSify as the SKOS format is intended – even mainly designed – to handle thesauri, so that it can technically comply with the main requirement of Europeana ingestion process. Then, if we compare with the other terminology types, thesaurus features a good mix of richness and usability. Moreover, museums are generally already usingthis kind of terminology rather than ontology or classifications[1]to describe their collections in a well-structured manner (47% of the 149 terminologies we have listed in our initial survey are thesaurus[2]). Without forgetting that regular relations of equivalence and association are particularly relevant for multilinguality. So, even if Europeana is moving toward the ontologies, we keep recommending thesauri to the museums since it appears as a good "middle way" between controlled vocabulary (particularly appreciated in the museums) and ontologies (especially powerful for the retrieval). A thesaurus can be defined as "a networked collection of controlled vocabulary terms". Thesauri allow the connection of terms using several types of relationships which can be hierarchical, associative, equivalence or definition. This means that a thesaurus uses associative relationships in addition to parent-child relationships. A parent-child relationship is expressed by a Broader Term (BT) /Narrower Term (NT) feature. Associative relationships in a thesaurus such as "Related Term" (RT) (e.g. term A is related to term B) are used to express relationships that are neither hierarchical nor equivalent. Equivalence is expressed by the USE (e.g. preferred term)/ Used For (UF) (e.g. non-preferred term). Additional information such as definition or remark can be included in a Scope Note (SN). The equivalence relationship is especially useful within multilingual thesauri. Since the structure of a thesaurus is carried by links of hierarchy and of association

---

[1] To get an overview of the kinds of terminologies, look at the previous part 3.4 Types of terminology resources

[2] cf. D4.1 results

between micro-thesauri, we recommend you to multiply the links in order to improve your terminology exploitation later. The more you have relations between your lists of terms, the more efficiently a search engine will return results to queries. In other words, we advice you to benefit from the thesaurus structure the most you can.

If you respect both the hierarchical and the transversal approaches of the Thesaurus structure, later you will SKOSify your terminology quite easily. SKOS will allow you to manage an arborescence of lists of terms related to different concepts (ConceptScheme tree), and a non-hierarchical set of non-exclusive groups of terms coming from these lists (Collectionsfor thematic grouping).

You can consider that this step is over when there is no more term out of a list, and no more list of terms totally independent of the others.

For example

You have a set of terms describing objects of architecture. Among all the descriptors in your thesaurus you have "monument" and "habitation". For each of them you create one specific micro-thesaurus. In the micro-thesaurus about monuments you have several terms like "palace" which is a type of monument. And in the micro-thesaurus about habitations you have several terms like "apartment" which is a type of habitation. So terms like "monument", "habitation", "apartment" and "palace" are linked by hierarchical relations. But you can also consider that a transversal group of terms around the theme of "building" can be proposed. In this group you could have "palace" and "apartment" which do not belong to the same hierarchy of terms but can though be connected in a thematic group of terms.

Methods and tools:

At the publication date of this deliverable, there is no specific tool guiding the logical design of your thesaurus. So if your vocabulary management tool does not provide this help, you will need to organise it with your daily tools (spreadsheets, ...).

You can have a look on the GEMET, General Multilingual Environmental Thesaurus[1], which propose you different ways to browse the thesaurus: thematic, alphabetic or hierarchical listings.

---

[1] GEMET: http://www.eionet.europa.eu/gemet

A6	**FIND EQUIVALENT TERMS IN OTHER LANGUAGES**

Actions:

According to your institutional situation, your terminology must be multilingual or not. Whatever you have to do regarding multilingualism, we advice you to try to make your thesaurus at least bilingual. So after having chosen your terms in your native language and structured your thesaurus, do the same in at least one other language. To do so, we recommend you:

1. To identify the language(s)in which you want or have to propose your descriptions.
2. **Not** to **proceed to a literal translation of** your terms in the identified languages previously.
3. But to find, for your collection domain(s), existing vocabularies in the language you are interested in and point outthe ones you can consider as equivalent to yours.
4. To ask experts of the domain and speakers of the foreign language to help you find the missing equivalent terms and to validatethe whole choice of equivalent terms.
5. To considerthese equivalent terms as associated terms until you precise their multilingual relationships at step B3: Define with precision the labels expressing concepts.

Purpose:

We strongly recommend you to foresee multilingualism right now even if in your institution this is not internally mandatory. Some institutions are legally mandated to propose multilingual descriptions (e.g. in Belgium in two languages), others technically have to (e.g. in the case of non latin alphabets such as Cyrillic or Greek alphabets). But for the others, even if they do not have this political or technical need, we can say they have in fact a need for visibility. Thus we consider that today every museum has to propose multilingual descriptions (at least in two languages). The Europeana's Strategic Plan 2011-2015[1] gives a strong focus on multilingualism: Europeana is developing a set of tools for providing a multilingual user interface for the access of your collections so if your terminology is included in Europeana with your multilingual terms it would definitely help the multilinguality.

The objective of this step is to find the best set of terms for the semantic description of your digital resources in at least one other language, but without any literal translation. Indeed in this case direct non expert translation produces mistakes or meaning-slidings.Sometimes this is due to "false friends", other times to the fact that the term in your native language is generally used abroad. The more you find equivalent terms by expert in the foreign language you are

---

[1] http://version1.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602

Recommendation for integrating Digital resources
present in museums in Europeana

ATHENA
Access to cultural heritage
networks across Europe

interested in, the more exact your equivalence will be.

Here the difficulties are to use a peer-wise approach instead of translation one; to manage very specific terms without any direct equivalent term; to reach 100% multlilingualism; to prevent yourself to use English as a pivot-language as much as possible since the ambition of the EU is to foster a real multilinguality as one of its main cultural characteristics.

You can consider that this task is over when you have for each major descriptor of your terminology at least one equivalent term in another language.

For example

FACET THEMATIQUES (FR)	FACET THEMATIQUES (NL)	FACET THEMATIQUES (EN)
transport	transport	transport
Transportsurterre	**transport over land**	**transport over land**

The example above comes froma thesaurus (thematic keywords of RMCA, Belgium) that was used for the Athena Thesaurus.

In the example, the term "transport over land" is used for the English term and the Dutch one as well. It is possible that there is no appropriate term for this concept in Dutch or it is possible that this English term is acknowledged in Dutch rather than its Dutch literal equivalent.

Like in the example, in some cases, the use of terms in a language that is not the original one may be necessary either because the common use acknowledges the use of this term or because there is no proper equivalent in the expected language. These terms are known as "coin" terms. In this case, beware of providing context or use information on the term (please refer to A7: Implement your thesaurus and B7: Ensure the documentation of the concepts).

Methods and tools:

You can consult a repository of free-to-use terminologies on the Athena Wiki where the resources are classified by domainsand languages :

http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources

You can also consult the norm *ISO 5964: 1985* since it notably deals with the transposition of a monolingual thesaurus to a multilingual one. However, this norm does not take into account the technological reality. It helps you to make it on the paper without considering a technical implementation thanks to a tool. That particularity can lead to contradictions later when you implement the mapping of equivalent terms. So we advice you just to refer to it for the core design of your terminology and keeping in mind that semantic Web technologies (definition of RDF format, the SKOS W3C recommendation, ...) may have solved some of the issues pointed out in this norm. The upcoming norm ISO 25964-1 that we already mentioned will address better the design of a monolingual or multilingual with consideration to the technological reality.

A7	**IMPLEMENT YOUR THESAURUS**
	Actions:
	Since you have just conceived your thesaurus structure, chosen your terms and found equivalent terms in different langages, you have now to technically make the thesaurus by:
	• Refining your general conception and preparing the implementation by consulting some standards which have been elaborated to provide guidance for the elaboration of thesaurus:
	o Three already finalized standards*: ISO 2788:1986: + ISO 5964: 1985 + ANSI/NISO Z39.19-2003*
	o **But most of all: *BS8723: Structured Vocabularies for Information Retrieval*and the upcoming*ISO 25964-1 - Thesauri and interoperability with other vocabularies: Thesaurus for information retrieval***
	• Using your in-house thesaurus or collections management tool, or if there is not any terminology management part in your collections management tool, using a spreadsheet tool (such as Microsoft Excel or Calc from Open Office) to practically declare and organize the lists of terms and the transversal groups.
	Purpose:
	The objective is to effectively build the thesaurus you have previously conceived. If your conception is satisfying, the technical concretisation will be quick and easy. Before trying to technically make your thesaurus, we recommend you to consult standards giving guidance for elaboration of such terminology. Indeed the work of ISO is a good guidance to implement your thesaurus.
	If the 3 following standards *ISO 2788:1986: + ISO 5964: 1985 + ANSI/NISO Z39.19-2003* are finalized and interesting to know when you want to conceive precisely a thesaurus, we recommend you the latest ones.
	*-BS8723: Structured Vocabularies for Information Retrieval:*This standard, which is a British adaption of the ISO 2788, intends to take into account every kind of terminology, not only thesauri, and focuses also on the interoperability between vocabularies. It takes into account the connection between terminologies and collections & objects, in the perspective of a SKOSification.
	- ISO 25964: Thesauri and Interoperability with other Vocabularies. This norm is divided in two parts: the first part on "Thesaurus for Information retrieval" will be published in 2011. The second part about "Interoperability with other vocabularies" will be published in 2012. This norm gives an update of the previous norms on thesauri (ISO2799 and ISO5964) for their design but also some technical specifications for thesaurus design and maintenance softwares. Some recommendations for interchange formats and protocols are available. The UML[1]

---

[1] UML: Unified Modeling Language

diagram presenting the general design of a thesaurus and its implementation defined by this norm is included in the annexes.

We make our recommendations according to the recommendations of these standards.

Among all the exiting tools we have identified during our benchmark[1], no one is really dedicated to the implementation of a new thesaurus. Ideally, in the perspective of the SKOSification (especially step B2: Roughly SKOSify your thesaurus), you should directly use at this very step an XML editor in which you could already format your terminology in RDF. However you can make it more easily by using a spreadsheet tool andthen converting it in an XML.

XML is not mandatory here, but your terminology will be in a more standard form than in a spreadsheet. The first interest of XML is thatyou are making a first step for your terminology SKOSification. The second one is that the arborescence displayof XML (for instance in a Web browser) helps to see in one sight how your thesaurus is structured.

Anyway, even if we can say that the previous steps did not require very specific knowledge in Information Engineering, this very one requires for the first time technical skills.

For example

You have a thesaurus about architecture in which there are two micro-thesauri:one about monuments, and another one about habitations. In your "monument" list of terms, you have for example "palace", "triumphalarch","therms"… And in the "habitation" list you can have "apartment", "hut", "house", "squat"… Finally, your transversal group of termson the theme of "buildings" gathers "palace", and "apartment".

In order to implement such a thesaurus, you use OpenOffice as a spreadsheet software. Your main sheet is called "Architecture Thesaurus". In the first column you have the micro-thesauri names ("monument", "habitation"). In the second the related terms which are in hierarchical relation.

Sub-domains	Terms
Monument	
	Palace
	Triumphal Arch
	Therms
Habitation	
	Apartment
	Hut
	House
	Squat

Then, in order to declare the transversal grouping of terms related to the theme "buildings", you create a new sheet in your spreadsheet entitled "buildings" in which the first column gives the terms and the second the source micro-thesauri.

---

[1] http://www.athenaeurope.org/athenawiki/index.php/Benchmark

Terms	Source micro-thesauri
Appartment	Habitation
Palace	Monument

Methods and tools:

If you do not have an in-house thesaurus management tool wich enables you to implement a thesaurus from scratch and convert it in XML, we advice you to use a spreadsheet toolsuch as OpenOffice.

It is a free tool whose functionalities are adapted to organise terms according to both hierarchical and transversal approaches. And you can export your file data into an XML conversion thanks to the function Save As.

The use of a thesaurus implies a few issues. The main difficulty is that at this point you are not manipulating concepts but terms. This is a difficulty because in the expanding world of Semantic Web, concepts are now better exploited than terms because they are considered as independent of the language in use. On the contrary, terms are relative to the language.

So the question is: How to keep using a thesaurus without preventing the exploitation of concepts by the Semantic Web machine?We will see below that the solution consists in the second step, when you make your thesaurus interoperable.So far we were at the term-level which determines the following step. We go now to the concept-level by leaving the floor of the language. This abstraction enables multilingualism even more efficiently.

4.3.Make it interoperable

After having conceived your terminology the closer to the ideal form you could, you have to make it interoperable.
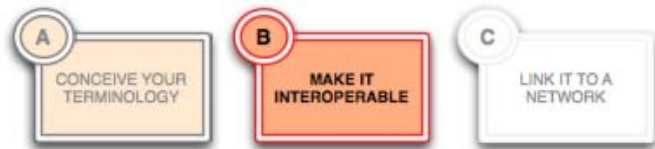
The conception step aimed at producing a thesaurus from the internal point of view of the museum in a human perspective. Now we recommend you to metaphorically go out of the museum, and to take into account the machine concern by SKOSifying your thesaurus.

4.3.1. Benefits from using SKOS

RDFS[1] and OWL are the languages that have been formally defined for knowledge representation. SKOS is one language among this formal languages' family. The major difference is that SKOS has been designed to model every type of controlled vocabulary. It can be used to represent a thesaurus as well as a classification or a subject headings list. Then it is a good compromise for the institutions who are using these types of resources, and who are willing to be compliant with the Semantic Web technologies without developing sophisticated ontologies.

The SKOS data model is consistent with the formal ontology language OWL. Therefore the migration from a SKOS version of a terminology towards a formal ontology in OWL can be handled without major difficulties.

Since the SKOS model is very simple, but still complete enough, the implementation of a SKOS version has a low cost for migration. As we made the distinction in the D4.2 introduction, SKOS is not a formal knowledge representation. But for an institution managing simple list of terms, or classifications and thesauri in the best case, it would be extremely costly and time consuming to develop a formal ontology perfectly compliant with Semantic Web technologies (using OWL for example). Therefore SKOS provides a structure based on classes and properties which give a powerful data model for migrating and porting these terminologies towards Semantic Web technologies.

Institutions must keep in mind that the adoption of the SKOS model is not a total replacement of the data model in use in the institution but a format for publishing and reusing their terminology and for ensuring the portability of this terminology for a semantic interoperability. Indeed usually knowledge organization systems (KOS), e.g. controlled vocabularies and thesauri, are used for indexing, and then porting these KOS into SKOS would enable the use of these indexing KOS for retrieval as well.

---

[1] RDFS: Resource Description Format Schema

However SKOS may not be the appropriate language for every type of controlled vocabulary. For instance, authority files which usually provide a list of persons cannot be migrated to a SKOS version properly since the scope of this type of terminology is real persons and not concepts. Another point is that the SKOS semantic relations properties cannot really apply to authority files since a person cannot be related to another one with hierarchical (narrower/broader) or associative (related) links.

If we would like to sum up all the reasons for you to use SKOS as a format for expressing your descriptions, we would remind:

- First SKOS is particularly well adapted to multilingual terminologies.
- Then SKOSification is an economical way to get to the conceptual level without employing an ontology. You can benefit from SKOS by migrating your thesaurus to a simili-ontology with a minimum of time and financial costs.
- Finally it is important to consider that SKOS is evolving and it will be easier and easier to customize it thanks to new SKOS classes you will be able to define by your own or the ones that will be proposed with the evolution of this standard.

Anyway, if you are considering to SKOSify your thesaurus, you have to change a bit your perspective. So far you were dealing with terms.Now you have to manage **concepts** since the Semantic Web in a multilingualism perspective requires concepts rather than terms to exploit. You also need to be distanced from your professional framework and make explicit a knowledge that you can easily infer considering your professional background.

4.3.2. Methodology

In fact when you SKOSify your thesaurus, you are technically applying the connection you have defined previously at step A3: Define your connection with the datamodel. SKOS is the format we advice you to use among those which fit today with Europeanarequirements for the ingestion and the exploitation of your digital resources.

As aformat it enables the mapping of terminology models. And in our case the two terminology models are:

- Your own terminology model (that you have defined in step A5: Organise your terms into a thesaurus structure)
- The official terminology model designed in your datamodel (in LIDO if you have chosen it)
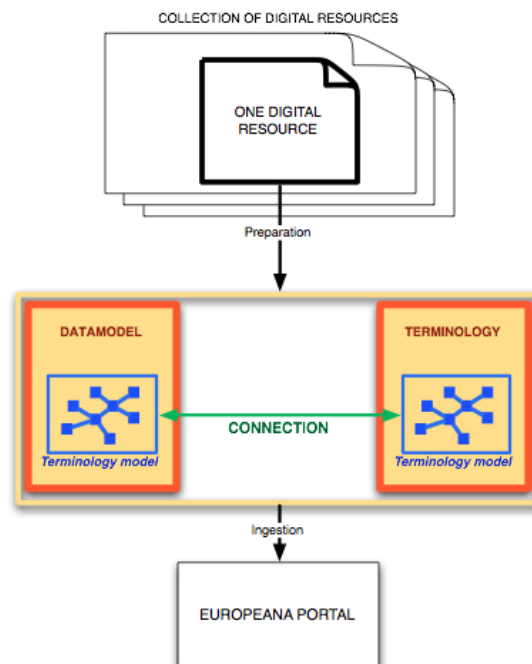
So this second big step is about the conversion of your thesaurus into a SKOS version. To complete the implementation, a set of actions can be done[1]. Most of them concern the mapping of terminology models. As it has been done for the first step, we present them as a linear process for reasons of readability. But of course, these tasks are iterative as well.

The different tasks we are going to detail are:
- B1:Evaluate how far SKOS is compliant with your terminology features
- B2:Roughly SKOSify your terminology
- B3:Define with precision the labels expressing concepts
- B4: Identify your concepts and validate the structure
- B5:Ensure the documentation of concepts
- B6:Map your concepts
- B7:Map your(multilingual) terms
- B8: Validateyour SKOSification

---

[1] Some of these actions come from the guidelines proposed in the deliverable D4.2 in which you can find more precisions

B1	**EVALUATE HOW FAR SKOS IS COMPLIANT WITH YOUR TERMINOLOGY FEATURES**
	Actions: As a prolog of the technical SKOSification of your terminology, you must check if SKOS is fully appropriate to your terminology features. Your terminology has been designed for satisfying the users' needs you have identified at step A2: Identify your users' expectations. But SKOS may be uncompliant with some of these needs. 1. Check if in your thesaurus you have only independent descriptors (concepts or sub-domain names). If yes, SKOS is not necessary, an RDF representation can be enough. 2. Check if in your thesaurus you have a list of people names; If yes, you will need to specifically use FOAF in addition to SKOS. 3. Check if in your thesaurus you have a list of location names; If yes, you will need to specifically use SKOS paying attention to the hierarchy to be defined (geographical information versus political information) 4. Check if in your thesaurus you have a list of Institution names; If yes, you shall need to specifically use Vcard[1]/hCard[2] and FOAF in addition to SKOS. 5. Check if in your thesaurus subject lists you have different terms which differ from others only by the gender or the number. In this case you need to precise the gender or number relation between terms, you can use SKOS-XL as an extension of SKOS.

---

[1] Vcard : http://www.w3.org/TR/vcard-rdf/

[2] hCard : http://microformats.org/wiki/hcard

Purpose:

The objective is to SKOSify your thesaurus, that is, to make your terminology interoperable with a datamodel like LIDO, and consequently with Europeana. But before starting any procedure for converting a terminology into SKOS, you must have checked how far SKOS is the appropriate format considering the features of your terminology. In the case of authority files for instance, SKOS may not be the most appropriate format. We have listed five different cases in which SKOS has to be mixed with other formats.

- Semantic relations: Canthe descriptors (then concepts) of the terminology be linked together via semantic relations? =>if the terminology only contains independent descriptors without any semantic relations, a SKOS modelization is not absolutely necessary, an RDF representation may be more convenient.

- People names: Is your terminology dealing with objects and abstract things that could be assimilated to concepts? Is the terminology dealing with persons? =>if the terminology is dealing with persons and not objects or abstract things, a standard like FOAF (Friend Of A Friend) FOAF : http://www.foaf-project.org would be more apropriate. If the terminology is dealing with both of them, a mix of FOAF and SKOS could be interesting.

- Location names: Is your terminology dealing with locations names? =>if the terminology is dealing with location names and not objects or abstract things, SKOS simple RDF can be used to model it.

- Institution names: Is your terminology dealing with Institution names? =>if the terminology is dealing with Institution names and not objects or abstract things, a standard like Vcards/hcards would be more apropriate. If the terminology is dealing with both of them, a mix of VCARDS/HCARDS[1]and SKOS or OWL could be interesting.

- Gender and number relations: Is your terminology dealing with terms which differ the ones from the others by gender or number? =>if the terminology is dealing with gender and/or number versions of terms, you can use the SKOS extension: SKOS-XL. Indeed SKOS-XL enables you to precise such relations between terms. More generally, SKOS-XL is useful when you want to link concepts and lexical resources by providing information about terms from the general language, out of the specialities.

---

[1] VcARD/hCARD is a microformat dedicated to the publication of contact details of people, organizations or places.

For example

- In a terminology on architecture, suppose you have a term "stained glass" as part of a religious building. If you provide the equivalent term in French "vitrail", it may be relevant to provide also the plural form "vitraux" for query concerns.

If you evaluate that the distinction between singular or plural form of a term and then a label is relevant for your terminology, you may use the SKOS-XL extension in order to provide specific information on each label rather than general information on a concept expressed by different labels.

-If you intend to model an authority list with authors or people's names, you can have a look on the VIAF[1]terminology.

As this authority file is the result of the mapping of several terminologies from various institutions (mainly libraries), FOAF and SKOS format are combined. The website provide a multilingual display and the possibility to view the results of a query in RDF.

Methods and tools:

You may have a look on the website of the W3C for getting more details on SKOS and the SKOS-XL extension:

http://www.w3.org/TR/skos-reference/skos-xl.htm

Please also refer to the use cases and requirements on SKOS defined by the W3C:

http://www.w3.org/TR/skos-ucr

---

[1] VIAF, Virtual Internation Authority File: http://viaf.org/

B2	**ROUGHLY SKOSIFY YOUR THESAURUS**

Actions:

In these recommendations we are dealing with "subjects" mainly, that is, we are just considering the conversion into SKOS of your terminology without taking into account the particular cases we have listed at step B1 (B1: Evaluate how far SKOS is compliant with your terminology features).

In order to make a first SKOSification of your thesaurus, you can proceed according to 4 different ways:

1. If you have a collection management tool proposing a function of export in SKOS, use it directly.
2. If you do not have such an in-house tool, and haved implemented your thesaurus in a spreadsheet, you can use the tool XL2XML,or XLTaxonomy, and their by-defaut style sheet.
3. If you prefer defining yourself the style sheet enabling the SKOSification of your XML file, you can use a tool such as Annocultor.
4. Or if you prefer to build your style sheet from scratch by analysing your XML source file, do it on the paper by your own.

Purpose:

SKOSification is a conversion process requiring a stylesheet in which the conversion rules are written to transform the thesaurus into a structure of concepts. The tricky point of this process is specifically the definition / implementation of the style sheet. You can do it by different ways, but in any case this step requires a strong technical skill. Moreover, the more resource you spend at this step, the more guarantee about the SKOSification quality you will have. Thus we advice you to proceed the more expensive procedure if you can, because it is really worthy in regards with the next steps.

Just below we present you four ways to define and implement the SKOSification style sheet. They are listed from the most economic to the most costly.

1/ First of all, if you have an in-house terminology management tool, check if there is a conversion function to SKOS. If yes, just use it and export the result into an XML file. In this case the style sheet is already defined inside the tool. You do not need any specific skill to manage it.

2/ Another method to SKOSifiy your thesaurus without spending too much resource consists of using the XL2XML tool[1]. Indeed this tool has a predefined style sheet and its use is precisely presented in online guidelines. However it requires to work on a spreadsheet you have to configure in a perfect match with what the tool expect as input. So if your thesaurus has a simple structure, this way

---

[1] https://www.seegrid.csiro.au/wiki/bin/view/Siss/ExcelToRdfTool#XL2XML_conversion_tool

can be an economic and efficient solution. But it can miss flexibility. Another tool which provide the same kind of features is XLTaxonomy[1]. This tool enables you to load a spreadsheet where the terms are organised in a precise way and proceed with the simple conversion into SKOS.

3/ The third possibility consists in using a specific tool. Annocultor[2] was developed to SKOSify many terminologies with the same structure on the row. In this case you can manage the style sheet according to your thesaurus features. This tool has now evolved and may not be relevant for the SKOSification of a single terminology. However this method requires a strong technical expertise since you have to use command lines. But if this style sheet is correctly defined, the tool then generates alone the SKOSified XML file.

4/ The last possibility is the most costly in resource because in this case you do everything by your own: defining the style sheet, and writing the SKOSified XML file. Nevertheless, this solution is the one which offers the highest guarantee of adequation with your thesaurus features, and of usability of the final file at the next steps. To do so, start from the XML file generated at step A7: Implement your thesaurus.

For example

Suppose your terminology is structured in a spreadsheet as follow:

Sub-domains	Terms
Monument	
	Palace
	Triumphal Arch
	Therms
Habitation	
	Apartment
	Hut
	House
	Squat

You can use a tool such as XL-Taxonomy which can convert into SKOS easily your terminology. With this tool you can define which term will be the broader and narrower concepts of your terminology.

[1] http://www.modelfutures.com/file_download/16/xlTaxonomy.zip

[2] http://annocultor.eu/
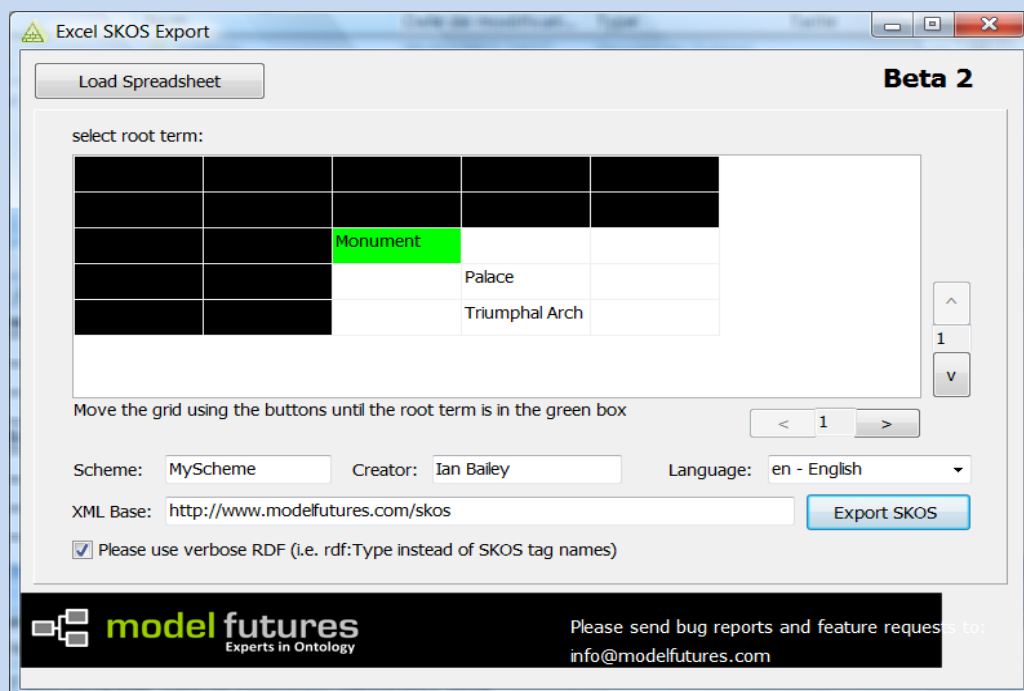
You may need to check the SKOS output produced by the tool and if needed you may convert it inthe SKOS/RDF form that can be exploited by your tools. The tool can produce either an XML or SKOS file.

Methods and tools:

The Athena Format is the format that Athena Thesaurus is expressed with. This format is here proposed to the museums who want to map later their own terminology with Athena Thesaurus (or to use the Athena Thesaurus as the starting point for the construction of their own terminology). As a SKOS-based format, the Athena Format guarantees to the museums that their descriptions respect the relative Europeana requirement regarding SKOS.

Find here the Athena Format:

http://www.athenaeurope.org/athenawiki/index.php/References

B3	**DEFINE WITH PRECISION THE LABELS EXPRESSING CONCEPTS**
	Actions:
	After having identified your concepts and mapped them with the datamodel, you have to precisely define the labels which express these concepts. To complete such a definition, here are our recommendations:
	- Preferred labels must be unique within a concept scheme - Each concept must be expressed with one preferred label per language (mandatory) - Avoid the concatenation of several words for a same label - Privilege the use of the lemma for the preferred label and possibly the other labels - Privilege the typography in use by convention in the languages involved - Avoid repeating the same information in different ways (e.g. symmetric and inverse properties)
	All these recommendations are detailed right below.

Purpose:

The objective is to define your concepts labels with precision.

Preferred labels must be unique within a concept scheme

As it is required by the SKOS data model, no two concepts from a same concept scheme should have the same preferred label in a given language. However as natural languages are highly polysemous and full of homographs, the SKOS data model does not forbid that one concept can have two same preferred labels in two different languages.

Each concept must be expressed with one preferred label per language (mandatory)

As we saw above, the SKOS data model does not forbid the absence of preferred label, but labels are meant to help the understanding and refining the meaning of a concept. This is especially true in a multilingual context and it is helpful for purposes of administration and maintenance. Therefore we recommend using one preferred label per language. It is important to note that this also means that is not possible to have several preferred labels in the same language.

Avoid the concatenation of several words for a same label

In order to get the most accurate description, we recommend avoiding several values as a preferred term. For example, double concepts such as "dwelling/houses" must be considered as two different concepts that are linked by a semantic relation. The use of scope notes can help to reinforce the closeness of these two concepts. The link between the two terms must be defined in order to provide the best description. We can state that "dwelling" and "houses" are synonyms; then the double concepts can be modelled as follows:

Dwelling: preferred label and houses: alternative label

Another possibility in the case of double concepts is to model the two concepts as related concepts.

Privilege the use of the lemma for the preferred label and possibly the other labels

The preferred label should consist in a single word term or a compound words term in natural language. This means that no artificial word or code must be used to label a concept. Such code must be defined using the skos:notation property. The lemma of a word represents its canonical form. We strongly recommend this form of terms to be used as preferred label. For instance, in English or in French, the usual form of a lemma in the case of nouns is the singular for the number and the masculine for the gender.

Privilege the typography in use by convention in the languages involved

The labels should respect the typographical rules that are usually in use in the languages of the labels. For instance, in English all the words referring to a language or nationality starts with an upper-case character whereas in French, these words will be in lower case characters. Thus we recommend respecting the conventions that are in use for each language involved. Any exception to this guideline must be documented via documentation properties of the model.

For verbal forms, infinitive forms will be privileged. Thus the forms of terms should be based on the conventions in the languages involved. If the concept is only expressed with labels in specific forms that do not correspond to the lemma, this must be documented via the documentation properties (skos:note, skos:changeNote, skos:editorialNote or skos:historyNote). In the case of compound terms, if possible, the addition of adjectives or verbs to a noun phrase should be limited. In the same spirit, the use of articles and prepositions should be avoided in order not to extend the length of the label. From the computing systems point of view, these guidelines can help the efficiency of a retrieval system.

Avoid the duplication of information

The SKOS data model consists of classes and properties as we saw above. Meanings are to be deduced by an efficient use of these properties. As some of the properties available in the SKOS model are proposed as pairs (inverse or symmetric), this supposes that the use of one property implies the opposite or the reverse. Therefore it is better to avoid duplication and not to repeat the same information in different ways. SKOS terminologies are processed by machines. So the less redundant information there is, the faster the results of a query can be retrieved. The main properties to pay attention to in order to avoid duplication of information are:

- *Inverse properties*: The use of the skos:broader or skos:narrower property implies the inverse meaning. Asserting that A has a broader concept B implies that B has a narrower concept A. This is true also for the skos:broaderTransitive and skos:narrowerTransitive property.
- *Symmetric properties*: The skos:related property is symmetric then if an assertion that A is related to B is made, there is no need to make the following assertion, B is related to A.

For example

You have roughly SKOSified your terminology about architecture.

You may have different terms which are equivalent to express a concept but you have to define a preferred term (as it is usually the case in a thesaurus) and keep only one preferred term per language.

- If you have a concept scheme (group of concepts) on Architecture and that you have "apartment" and "flat" as a narrow concept of "habitation" then you have to specify which is the preferred label. Then you can express these terms as follows (represented here as Turtle[1] format) :

Ex:apartment rdf:type skos:Concept;

skos:prefLabel « apartment »@en ;

  skos:altLabel « flat »@en ;

  skos:prefLabel « appartement »@fr.

« flat » is defined as an alternative label and apartment is the preferred label in English.

- If there are compound terms in your terminology, try as much a possible to decompose them in order to get to a simple form.

Your terminology has a concept "Musical instrument" in order to define with precision the labels of your concepts, you decompose this concept into two concepts: "Music" and "Instrument"

Ex:music rdf:type skos:Concept
skos:prefLabel « music »@en;
skos :prefLabel « musique »@fr.
ex :music skos :narrower ex :intrument.

Methods and tools:

Since your thesaurus has already been roughly SKOSified, it can be open and modified in the online tool xTree. This tool is interesting because it helps you to "write" labels in SKOS through a Web user interface, and it implements the draft of normISO 25964-1.

Other editing tools such as Protege[2] (with the plugin SKOSed) or ThManager[3]allow you to manage the labels used to express the concepts of your terminology. Please refer to the Benchmark section of the Athena Wiki for an up to

---

[1] Turtle Terse RDF Triple : http://www.w3.org/TeamSubmission/turtle/

[2] You can find Protégé and SKOSed at: http://protege.stanford.edu/

[3] You can find ThManager at: http://thmanager.sourceforge.net/

	date version of the existing tools for SKOS editing: http://www.athenaeurope.org/athenawiki/index.php/Benchmark

B4	**IDENTIFY YOUR CONCEPTS**
	Actions: Since you have refined your SKOSified version of your thesaurus by precising the labels, you can now go further by technically identifying your concepts and map them with the datamodel. To do so we advice you to follow the 5 stars scheme proposed by Tim Berners-Lee[1]: ✴ make your stuff available on the Web (whatever format) under an open license ✴✴ make it available as structured data (e.g., Excel instead of image scan of a table) ✴✴✴ use non-proprietary formats (e.g., CSV[2] instead of Excel) ✴✴✴✴ use URIs to identify things, so that people can point at your stuff ✴✴✴✴✴ link your data to other data to provide context

	Purpose:
	
	The W3C define two main steps to proceed to the identification of concepts:
	
	   - Creating (or reusing) a Uniform Resource Identifier (URI) to uniquely identify the concept
	   - Asserting in RDF using the rdf:type property that the resource identified by this URI is of type skos:Concept
	
	Use of a Persistent Identifying System for the definition of the URIs
	
	As we described them above, we recommend the use of standards for the identification of the concepts. Indeed, as the identification of concepts is achieved with the definition of HTTP URIs, these URI must be declared to persistent identification systems such as PURL which is normalised. This will also be of a great benefit since it is location-independent, e.g. if the terminology is moved from one location (housing server) to another, the URIs identifying the concepts of this terminology will not have to be modified.
	
	Use of non-explicit URIs
	
	It is highly recommended to use non-explicit URIs[3] in order to avoid the reuse of a same URI for identifying two different concepts. Indeed as natural languages are by definition ambiguous and polysemous, it is possible that two different concepts

---

[1] http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/

[2] CSV: Comma-separated values

[3] A non-explicit URI is a URI whose the name transmits no semantic information; a series of numbers and special characters with no meaning can constitute a good non-explicit URI.

might have two similar labels. The use of explicit URIs supposes that the choice of one specific natural language has been made during the definition or the migration of the terminology which cannot be convenient in a multilingual context.

For example

Suppose your terminology is hosted and managed by your institution but used by several other institutions. You have to define your identifiers so they can state the origin of the concepts (domain name) but also being flexible enough so the other institutions donot have to make any modification if your identification system change. It is better to use non explicit URIs in order to avoid the ambiguity of natural languages.

The Bibliothèque Nationale de France (BnF), the French National Library, for example is using the ARK persistent identifiers system (see details below).

Here is an example of URI with ARK from the BnF:

http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11931420f

Methods and tools:

Different systems for Persistent Identifiers are in use. Here some information of these main systems:

PURL: A PURL (Persistent Uniform Resource Locators) consists of a URL; instead of pointing directly to the location of a digital object, the PURL points to a resolver, which looks up the appropriate URL for that resource and returns it to the client as an HTTP redirect, which then proceeds as normal to retrieve the resource. PURLs are compatible with other document identification standards such as the URN.

URN: The URN (Uniform Resource Name) is designed to describe an *identity* rather than a *location*; for example, a URN may contain an ISBN (International Standard Book Number, used as a unique, commercial book identifier).

NBN: National Bibliography Numbers (NBNs) is a URN namespace used solely by national libraries, in order to identify deposited publications which lack an identifier, or to reference descriptive metadata (cataloguing) that describe the resources. These can be used either for objects with a digital representation, or for objects that are solely physical, in which case available bibliographic data is provided instead.

ARK: The Archival Resource Key (ARK) is a URL scheme developed at the US National Library of Medicine and maintained by the California Digital Library.

ARKs are designed to identify objects of any type – both digital and physical objects.The ARK scheme encourages semantically opaque identifiers for core objects. Unlike an ordinary URL, an ARK is used to retrieve three things: the object itself, its metadata, and a commitment statement from its current provider.

Open URL: An OpenURL contains resource metadata encoded within a URL and is designed to support mediated linking between information resources and library services. This standard is not *primarily* designed as a persistent identifier/resolver but is described as a metadata transport protocol.

DOI: The Digital Object Identifier (DOI) is an indirect identifier for electronic documents based on Handle resolvers[1]. According to the International DOI Foundation (IDF), formed in October 1997 to be responsible for governance of the DOI System, it is a 'mechanism for permanent identification of digital content'.

We can see from these short introductions that some of these standards are more adapted to specific field (for instance, URN and NBN are more adapted for the libraries), however standards such as PURL or DOI could be used for definition of URIs.

You can also have a look on the booklet "Persistent identifiers: Recommendations for institutions"[2] elaborated by the WP3 of ATHENA.

---

[1] Handle was a previous system for identification and references of resources.

[2] http://www.athenaeurope.org/getFile.php?id=779

B5	**MAP YOUR CONCEPTS**

Actions:

During the SKOSification of your terminology, you can map your concepts by defining semantic relations in-between. We recommend you to provide precision about these relations:

- Provide precision even if the concepts you want to link are not in an immediate hierarchical relation
- Avoid the mix of hierarchical and associative relations to ensure the consistency of the semantic relations

Purpose:

The objective is to start the auto-documentation of your terminology through its SKOSification. This task is also important to avoid possible ambiguities.

Non-immediate hierarchical relations

In some cases, semantic relations between concepts have to be described with precision in order to avoid a loss of meaning or information and also avoid designing information which will not make any sense. For example the skos:broaderTransitive/skos:narrowerTransitive pair of properties allows to describe with precision relations between concepts when two levels of hierarchy are impacted. Then the use of these transitive properties is preferred in order to assert a non-immediate hierarchical relationship between two concepts. However there is a possibility to use an extension to the SKOS data model in order to remove the symmetry of a property if this creates confusion in the meaning of the concepts.

Consistency of the semantic relations

In order to ensure consistency, mixing hierarchical relationships with associative ones should be avoided. For example, a concept A cannot be related to another concept B if this concept A is the narrower concept of a concept C. Therefore a special attention must be paid when designing the semantic relations between concepts.

For example

Each terminology is designed for a specific purpose. As presented in the A section of recommendations you have conceived your terminology to answer your own needs. Considering this, you may have to pay attention to the structure of the terminology where you intend to map your concepts with.

For example if your terminology has a concept scheme (group of concepts) on Music with a generic concept "musical instrument" and another concept schemeon Religion with a concept "organ" you can proceed with different mappings (represented here as Turtle[1] format) :

ex:musical instrument rdf:type skos:Concept;

skos:prefLabel « musical instrument »@en.

skos:broader ex:music

ex:organ rdf:type skos:Concept;

  skos:prefLabel « organ »@en.

  skos :broader ex :religion

➜ ex :organ skos :broader ex :musical instrument

As you can define an organ as a musical instrument, you can proceed with a mapping to the concept of "musical instrument". This concept will then be related to two different concept schemes.

This mapping of concepts allow you to avoid the repetition of concepts.

Methods and tools:

At the moment, there is no free tool available for helping and guiding this mapping process.Though this mapping process is mainly intellectual and rely on a human expert validation.

Proprietary tools such as ITM3 (Mondeca) propose a mapping user interface with a complete validation workflow.

You can have a look on the Benchmark led within the WP4 of Athena as the list of tools keep on evolving:

http://www.athenaeurope.org/athenawiki/index.php/Benchmark

---

[1] Turtle Terse RDF Triple : http://www.w3.org/TeamSubmission/turtle/

B6	**MAP YOUR TERMS**

Actions:

At the step A5 (A5: Organise your terms into a thesaurus structure), when you were conceiving your thesaurus before thinking to its SKOSification, you already made a first mapping of (groups of) terms. You have implemented this mapping when you technically set your thesaurus up at step A7: Implement your thesaurus. Now we propose you to refine and to improve this mapping of terms thanks to the possibilities that SKOS offers. We recommend you to do it first with the terms in your native language by:

- Making explicit the semantic relations between labels
- Documenting in scope notes any change of terms in your thesaurus

After proceeding with a refinement of your terms mapping, you can now define and implement the mapping between the equivalent terms you identified at step A6 (A6: Identify equivalent terms). Here we recommend you:

- To provide for each concept an equivalent label in the languages involved in your terminology
- To use the same system of language tags for defining the language of label

Purpose:

The objective is to refine and to improve your mapping of terms since your mapping ot terminology models has been done before, and particularly to enable multilingualism by expressing the semantic equivalence of terms in different languages. The more linked (equivalent) terms you have, the more exploitable (in different languages) your thesaurus will be.

Provide for each concept an equivalent label in the languages involved in your terminology

Special attention must be paid to the multilingual labels expressing the concepts. These multilingual labels must be defined in the correct way in the different languages of the terminology so that the equivalencies can be computed from the SKOS representation of concepts.

Use the same system of language tags for defining the language of label

There are several systems which are normalized and equivalent: for example the three tags "en", "en-GB" or "en-Latn" are different language tag systems referring to one language which is the English from Great Britain in Latin alphabet. In the case of terminology where different languages of different alphabet are involved, the tag system "language-alphabet" (for example "en-Latn") may be useful for providing more precision. We recommend using the same system of tags for every

language attribute of the terminology. In the case where a specific language tags system is not required, we recommend the use of the language systems defined in ISO 639-11where the language tags are coded on two letters in lower case.

For example

For example, if your terminology on Architecture has a concept "dwelling" where "dwelling" and "houses" is the alternative label and if your terminology is bilingual French and English, you will then have to provide the equivalencies of these concept labels in French.

ex:dwelling rdf:type skos:Concept
  skos:prefLabel "dwelling"@en;
  skos:prefLabel "habitation"@fr;
skos:altLabel "houses"@en;

skos:altLabel "maisons"@fr;

You can refer to existing terminology to enrich your terminology or rely on the work of an expert from the domain to get the exact terms to express your concepts.

Methods and tools:

As for the mapping of concepts, there is no specific tool for helping and guiding the mapping of terms. Though you might find some very specific tools developed by professionals of Natural Language Processing (NLP) for multilingual alignment. Europeana has set a repository of tools and lexical resources developed and used in the framework of NLP[1] and there is also an incubator for open source code for these technologies[2]

These tools help you to extract automatically similar terms on the basis of their lemma. This can be useful for a first automatic extraction that could be refined by the expert afterwards.

You can of course use the SKOS editors (ThManager or SKOSed-Protege, ...) to proceed with the mapping of terms once the conceptual work is done.

---

[1] http://europeanalabs.eu/wiki/WP2LanguageResources

[2] http://incubator.apache.org/opennlp/index.html

B7	**ENSURE THE DOCUMENTATION OF CONCEPTS**
	Actions: Here the step of documentation consists of giving information about the changes appearing through time by making a separation between the concepts and the labels. We advice you: • To provide documentation for each change that may occur to a concept and its labels • To provide as much as possible documentation to concepts with scope notes
	Purpose: **Provide documentation for each change that may occur to a concept and its labels** The SKOS data model provides number of documentation properties in order to refine the meaning of a concept or keep track of the changes on the label(s) of a concept and/or its meaning. For the purposes of administration and maintenance of the terminology, each change must be reported in the SKOSified terminology using change notes (skos:changeNote) or editorial notes (skos:editorialNote). **Provide as much as possible documentation to concepts with scope notes** As mentioned above, documentation on concepts helps to refine the meaning of a concept. The use of scope notes (skos:scopeNote) can be very helpful in enabling a better understanding of the concepts with contextual information. Examples may also be provided via skos:example property. Documentation of concepts is especially needed in the case of homographs/homonyms in the same language or different languages for the labels expressing the concept. Then scope notes and examples can provide the user with a semantic disambiguation. In order to make your documentation, you can use more or less precise notes whichare proposed in SKOS format: • 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*can be used to provide general documentation on a concept. All the

other types are specializations of this general property.

The *skos:changeNote* and *editorialNote* are mainly useful for the purpose of administration and maintenance. The *skos:definition*, *skos:example*, *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.

For example

The use of notes can help to keep track of the history of a concept or to give details on a concept. For example, if you have a concept "gothic art" in your terminology on Architecture, you can have a scope note introduced by the property skos:definition where you can state that gothic art appeared during the second half of the medieval period.

You can also make explicit a knowledge that is implicit for you by giving definitions for your concepts.

Methods and tools:

You can proceed with the already mentioned SKOS editors to provide documentation of your concepts and their terms.

B8	**VALIDATE YOUR SKOSIFICATION**
	Actions: SKOSification is a process of conversion of your thesaurus elements into a specific format. It means that the conversion is supported by rules, and that the result of such a process must be syntactically correct in regards with the format "grammar". Thus you have to check at the end if the SKOSified version of your thesaurus is correct or not. This step consists of the the validation of concepts and labels mapping, and of the respect of SKOS formalism. To do so we advice you to use the Webservice Pool Party.
	Purpose: The W3C offers on line a validation tool but it doesnot take into account the latest version of the SKOS model[1]. Pool Party, a thesaurus management system, offers online SKOS services[2] for converting and checking the consistency of your SKOS thesaurus. From a technical point of view, in order to check the consistency of your converted terminology to the SKOS model, we recommend using the online web service Pool Party. Pool Party offers a free online tool for validating SKOS files that may be already online or stored on your local repositories. This tool checks the consistency of the SKOSified terminology according to the following points which refer to our guidelines: - **Valid URIs**: the tool checks if there is not any unauthorised character in the URI. Although if an URI is used twice for identifying two different concepts, there will not be any alert or warning. - **Missing language tags**: the tool checks if all the labels and notes have a language tag - **Missing labels**: the tool checks that each concept has at least one preferred label. - **Loose concepts**: all the concepts that are isolated and not linked to other concepts are pointed out as loose concepts - **Disjoint OWL classes**: some elements of the SKOS model are compliant with OWL elements then the tool checks the eventual consistency with OWL elements that may be in the SKOSified terminology - **Consistent use of labels**: the rules for the use of labels are checked by the tool in order to avoid the use of a same label as a preferred label and alternative or hidden label, and to avoid the use of two preferred labels in a same language, ... - **Consistent usage of mapping properties**: the tool checks the consistency in the mapping relations.

---

[1] http://www.w3.org/2004/02/skos/validation

[2] http://demo.semantic-web.at:8080/SkosServices/index

- **Consistent usage of semantic relations**: the tool checks that there is no mix between hierarchical and associative semantic relationships.

From the content point of view, only the administrators and users of the terminology can validate the final migration of the terminology into SKOS format at least for an initial transformation process.Indeed they will be able to confirm or modify the general design of the terminology and its semantic relations according to the indexing and retrieval efficiency.

For example

Here is the output of the SKOS validator of Pool Party:

All the main sections are checked in green: in this case the terminology is well-SKOSified.

Methods and tools:

The editors such as Protege-SKOSed proceed with a first rough parsing of the terminology before allowing the editing however this is just a primary parsing. To be sure that your terminology is well-skosified you will need to use tools such as Pool Party.

You can find information and use online Pool Party at: http://poolparty.punkt.at/

4.4. Link it to a network

After having made interoperable your
terminology thanks to its
SKOSification, now we advice you to
link your terminology to a network of
resources.
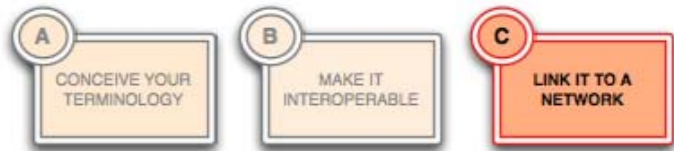
Indeed, the more your terminology is linked to others, the more its terms are retrievable by a
Semantic Search Engine. In order to help you link your terminology with others, we propose
you to follow a 4-step process:

- C1: Definition of metadata on your terminology
- C2: Identification of other resources for mapping
- C3: Mapping with other resources
- C4: Validation of the interoperability

Once again, even if our recommendations are presented along a linear process, you would
better follow them iteratively.

ATHENA
Access to cultural heritage
networks across Europe

C1	**DEFINITION OF METADATA ON YOUR TERMINOLOGY**

Actions:

Before effectively linking your terminology to a network of resources, we recommend you to make a documentation of your terminology as a whole by defining metadata on it. There is not specific metadata schema you could use, but we guess that a Dublin Core extended could be a good start. Here are the information the metadata on your terminology should provide:

- Terminology name
- Owner
- Domains of description
- Languages
- Contributors
- Creation date
- Modifications dates
- Terminology type (e.g. thesaurus)
- Licence
- Status (e.g. draft or published)
- Norms (e.g. RDF/SKOS, RDF/OWL)
- Kind of structure (e.g. flat list of terms, strictly in arborescence, mix of arborescence and transversal groups)

Purpose:

Here the first step consists of describing the terminology as a whole in order to identify it as a unique and precise element connected to a network of resources. All the metadata on your terminology are expected to to be input in a specific database related to a repository of resources.

Among all the tools dedicated to terminology management, some enable the edition of metadata. For instance ThManager enables you, even requires, to define information about the terminology before its use. This tool generates a metadata file in relationship with a repository database.

Rights issue

Because you aim at linking your own thesaurus to a network of other terminologies, you have to face the issue of rights. In order to map or to duplicate external terms, the other source terminologies must be free for use. And on the contrary, if you want a bilateral mapping, your own terms must be free fore use too. We recommend you to place your thesaurus under a Creative Commons Licence like CC-by-share alike, or CCo(universal public domain).

Tags

In order to declare your metadata, we recommend you:

- To wrap all of your metadata in <rdf:Description> tags.
- To use DC tags for defining the title (<dc:title>), the identifier (<dc:identifier>), the creator (<dc:creator>), the contributors (<dc:contributor>), the format (dc:format>), the languages (<dc:language>), the description (<dc:description> and the status (<dc:status>).
- To use DCterms tags for precising the norms in use like SKOS and RDF (<dcterms:conformsTo>), indicating dates of creation and release (<dcterms:created> and <dcterms:issued>), and declaring rights elements (<dcterms:RightOwner> and <dcterms:license>)
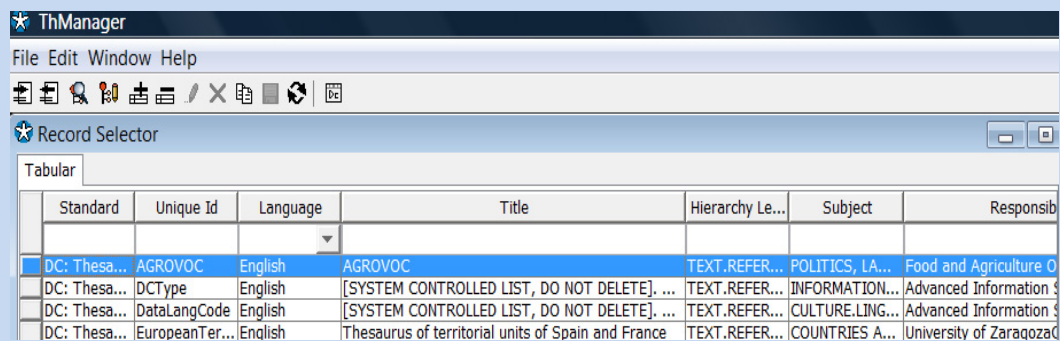
For example

You can see how the Athena Thesaurus is documented by looking at the header of this page :

http://www.athenaeurope.org/athenawiki/index.php/RDF/XML_version

ThManager[1] is an editing tool for SKOS thesauri which allow to registrate in an internal database several thesauri.

Here a screenshot of the registry of ThManager:

Methods and tools:

The Dublin Core is the basic format for providing metadata on a resource. Moreover it will enable the harvesting via interchange protocol such as OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting)[2] then your terminology will be visible to be part of an existing registry.

The Linked Heritage project, which will be presented further, will provide a terminology registry where the institution will be able to declare their terminology and their needs.

---

[1] http://thmanager.sourceforge.net

[2] OAI-PMH: http://www.openarchives.org/OAI/openarchivesprotocol.html

C2	**IDENTIFICATION OF RESOURCES FOR MAPPING**

	Actions:
	
	 Before linking your terminology to others, you need to identify those which can be interesting for mapping.
	
	 To identify such relevant terminologies, we recommend you:
	
	 • To browse terminology repositories by using two criteria:
	    o Domain of description linkable with yours
	    o Languages you are interested in for mapping of equivalent terms
	 • To check the rights for use of each of these identified terminologies
	 • To identify in the terminology you can map the (groups or lists of) terms you are interested in

	Purpose: To link your thesaurus to a network of resources implies first that you want to benefit from the semantic exploitability of all the relations you are going to implement between your internal terms and external others proposed in different terminologies.Then it notably implies you are considering to also benefit from multilingualism by mapping equivalent terms in different languages. So we can say there are at least two main complementary key reasons for looking for other terminologies: 1. To find terms potentially belonging to the same domains than yours 2. To find terms potentially equivalent to yours in other languages To achieve such goals, a repository of terminologies appears as a very useful solution. Thus we have started a repository available online at: http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources This is a result of the inventory of resources we made during the Athena project. This repository is dedicated to free of use terminologies from European museums. Other repositories can also be useful Europeana datacloud[1] or DBPedia[2].
	For example Your terminology has a list of places' names and you would like to map your own list with some reference terminology. You can have a look on some terminology repositoriesto see which resource you could map your concepts with. You can have a look on the datacloud of Europeana which are the terminologies already mapped and used for search and retrieval:

---

[1] http://eculture.cs.vu.nl/europeana/www/datacloud.html

[2] http://wiki.dbpedia.org/OnlineAccess

	http://eculture.cs.vu.nl/europeana/www/datacloud.html The Thesaurus of Geographic Names (TGN)[1] from the Getty and Geonames[2] are major resources for places's names. Then you can start proceeding with the mapping of your locations' names with those of the TGN and Geonames.
	Methods and tools: You can look our inventory of free-for-use resources at: http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources You can also look at the Europeanadatacloud: http://eculture.cs.vu.nl/europeana/www/datacloud.html Or the LOD datacloud of DBPedia: http://wiki.dbpedia.org/About The mapping of your concepts with a resource such as DBpedia ensure you the enrichment of your terminology since this is the RDF version of the articles available on Wikipedia.

---

[1] TGN-Getty : http://www.getty.edu/research/tools/vocabularies/tgn/

[2] Geonames : http://www.geonames.org/

C3	**MAPPING WITH OTHER RESOURCES**

Actions:

First of all map your thesaurus concepts with external ones by reproducing now what you did at step B3: Define with precision the labels expressing concepts.

Then map your thesaurus terms with external ones reproducing now what you did at step B6: Map your terms. If it is possible and useful, duplicate terms in your own thesaurus.

Purpose:

The recommendations of the part B (B: Make your terminology interoperable) proposed an internalmapping of your thesaurus concepts and terms. Now, in this part C, we invite you to do the same with concepts and terms which belong to other terminologies.

So if we compare with what we recommended at steps B3 (B3: Define with precision the labels expressing concepts) and B6 (B6: Map your terms), only a few differences happen.

Regarding the concept mapping, the main difference is that you now need to get one identifier for each terminology.We consider that the root of your URI naming system should be used as the identifier of your terminology.

If during your search of terms for mapping, you have noted sets of terms you would like to have in your terminology rather than having a mapping with them, you can enrich your thesaurus by integrating them. The integration of terms may be interesting if you intend to give an online access to your terminology so users can browse it. Check you have the right to do so (i.e. if the source terminologies licence allow the duplication with no condition). Then express them in your own format.

In order to implement the mapping of terms and concepts between different terminologies, you can use the Athena Format which is based on the SKOS format.

For example

Each terminology is designed for a specific purpose. As presented in the A section of recommendations you have conceived your terminology to answer to your own needs. Considering this, you may have to pay attention to the structure of the terminology where you intend to map your concepts with.

For example if your terminology is about Architecture and that you plan to map it with a thesaurus on Environment you have to pay attention to the structure of these two terminologies and define the degree of matching.

If you map the concept of "cultural heritage" from your terminology with the same concept from the GEMET thesaurus, you will be able to enrich the information induced by your concept thanks to the mapping link and its inference.

cultural heritage

Definition:
The inherited body of beliefs, customs, artistic activity and knowledge that has been transmitted by ancestors. (Source: RF

broader terms
- culture (society)

narrower terms
- art
- cultural goods
- historical evolution
- literature
- music
- natural heritage
- restoration
- traditional culture

related terms
- world heritage site

Scope note:
scope note is not available

Groups:
INFORMATION, EDUCATION, CULTURE, ENVIRONMENTAL AWARENESS

Themes:
social aspects, population

Other relations:

Has exact match

AGROVOC: Cultural heritage

EuroVoc: cultural heritage

Wikipedia article
Cultural heritage

Arabic:	تراث ثقافي
Basque:	kultur ondare
Bulgarian:	Културно наследство
Catalan:	patrimoni cultural
Chinese:	文化遗产
Czech:	dědictví kulturní
Danish:	kulturarv
Dutch:	cultureel erfgoed
English (US):	cultural heritage
Estonian:	kultuuripärand
Finnish:	kulttuuriperintö
French:	patrimoine culturel
German:	Kulturerbe
Greek:	πολιτιστική κληρονομιά
Hungarian:	kulturális örökség
Irish:	oidhreacht chultúrtha
Italian:	patrimonio culturale
Latvian:	kultūras mantojums
Lithuanian:	kultūros paveldas
Maltese:	wirt kulturali
Norwegian:	kulturarv
Polish:	dziedzictwo kulturowe
Portuguese:	património cultural
Romanian:	patrimoniu cultural
Russian:	культурное наследие
Slovak:	kultúrne dedičstvo
Slovenian:	kulturna dediščina
Spanish:	patrimonio cultural
Swedish:	kulturarv
Turkish:	kültür mirası

This screenshot represents the concept of "cultural heritage" in the GEMET thesaurus. The mapping could be of a real benefit since this concept is already mapped with other terminology resources such as Agrovoc[1] or EuroVoc[2] and also to the corresponding article in Wikipedia.

This mapping can also help enriching your own terminology namely regarding the multilingualism since the GEMET provides terms in 30 languages to express its concepts

---

[1] Agrovoc : http://aims.fao.org/website/AGROVOC-Thesaurus/sub

[2] EuroVoc : http://eurovoc.europa.eu/

Methods and tools:

Athena Format is presented in details at:

http://www.athenaeurope.org/athenawiki/index.php/References

The URI is a crucial element when you intend to map your terminology with other ones and to make it part of the Linked Open Data.

As for the mapping of concepts and terms presented in the B section (B5 and B6), there is no free open source tool for guiding the mapping process but you can have a look on some proprietary tools if necessary. Please refer to the Benchmark section of the Athena Wiki:

http://www.athenaeurope.org/athenawiki/index.php/Benchmark#Tools

C4	**VALIDATION OF THE INTEROPERABILITY**
	Actions:
	Here the validation process looks like the one presented at step B8: Validate your SKOSification. Proceed exactly like in step B8.
	- Check and validate your SKOSified terminology once the mapping is done
	- Check the ingestion of your collections and their quality once ingested
	- Check the interoperability of your terminology through a semantic search engine
	Purpose:
	The purpose of this step is to validate the whole process of designing a terminology, making it interoperable and linking it to other vocabularies.
	This step will also enable the multilingualism through the interoperability regardless of the languages available in your terminology.
	For example
	The semantic SearchLab[1] developped within the Europeana initiative is a work in progress but it could allow you to check some queries on your terminology and the inferences enabled by the mapping.
	If your terminology is well-structured, skosified and linked with other resources, a query on "Léonard de Vinci" can give the following results:

[1] Europeana Thought lab: http://www.europeana.eu/portal/thought-lab.html

The results are presented as clusters according to the status (author, represented figure, …) and the objects indexed with "Léonard de Vinci" or "Leonardo Da Vinci" are brought as results regardless of the language of the query.

Methods and tools:

A free tool such as Pearltrees[1] can help you to have an overview of your terminology by providing you a graphical interface.

The best way to validate the interoperability is to check the syntax and the consistency of your terminology and test it thanks to simple and complex queries.

---

[1] Pearltrees : http://www.pearltrees.com

5. Conclusions and perspectives

All the recommendations we have phrased in the previous pages depend on the current state of the art. In the future new resources, norms, standards, guidelines and tools will appear to help you manage your terminology in the context of LOD.

We cannot anticipate on what will happen and when. However we can right now draw some tangible perspectives that we think you should keep aware of. From our current point of view, we consider there are at least three important initiatives to follow: the collaborative Athena Thesaurus, the work on multilingual terminologies in the frame of the new Best Practice Network Linked Heritage that will continue and enlarge the work done by Athena, and the next version of Europeana (v2).

5.1. Athena Thesaurus

5.1.1. Presentation

We call Athena Thesaurus the thesaurus produced and updated by all contributors during and after the project. As a thesaurus, the Athena Thesaurus is a network of controlled vocabularies, that is, an amount of terms organised by domains of description and structured thanks to bridges in-between.

This Athena Thesaurus is:
- **SKOSified**: The Athena Thesaurus is already SKOSified; it fits with Europeana requirements; so it can be directly used for description by institutions in case
- **Free of rights**: Any institution can use it as it likes without paying any fee; hence an institution which enrich the Athena Thesaurus by terms coming from its own terminology must check if it has rights to do so for free distribution and modification
- **Evolving**: We are considering to enable a collaborative workflow to produce and update the Athena Thesaurus; a specific interface with moderation process can be imagined; The Linked Heritage project will proceed with the evolution of the Athena Thesaurus within the Linked Heritage Thesaurus.
- **Available online**: We can imagine a Web service helping an institution to use the Athena Thesaurus online for description; of course this terminology will be downloadable for a use offline. The terminology management platform that will be developed in the framework of the Linked Heritage project will provide such a Web service.
- **Mappable**: We consider to enable the mapping through a Web service of terminologies with the Athena Thesaurus; to do so, there are a few requirements: 1/ the terminology must be syntactically and semantically valid; 2/ it must be well-SKOSified. While these requirements are not satisfied, the mapping would not be possible.

The Athena Thesaurus is an on-going resource available online at:

http://www.athenaeurope.org/athenawiki/index.php/Athena_Thesaurus

5.1.2. Athena Thesaurus v1

It is important to keep in mind that the ATHENA Thesaurus in its current version is at a draft status since it was mainly created for the purpose of testing in the experimental framework. So at this current stage, this thesaurus does not intend to be a standard.

Considering the properties of each of the selected resources, the elaboration of this first version of the ATHENA Thesaurus was done in several mapping steps.

The mapping was performed from the most general resource to the most specific. Then a first mapping between the Michael Subjects thesaurus and the RMCA keywords thesaurus was done before mapping this version with the PICO thesaurus which is the most specific.
The Hungarian monolingual thesaurus KÖZTAURUSZ was then mapped with the result of this first mapping.

The approach adopted to build this version of the thesaurus consisted in merging the non-published resources and make mapping links to the published one. We have considered each of the source terminology as a concept scheme.

In order to provide a thematic organisation of the concepts, and as designed in the PICO thesaurus, we set four thematic collections, namely "who", "what", "where" and "when".

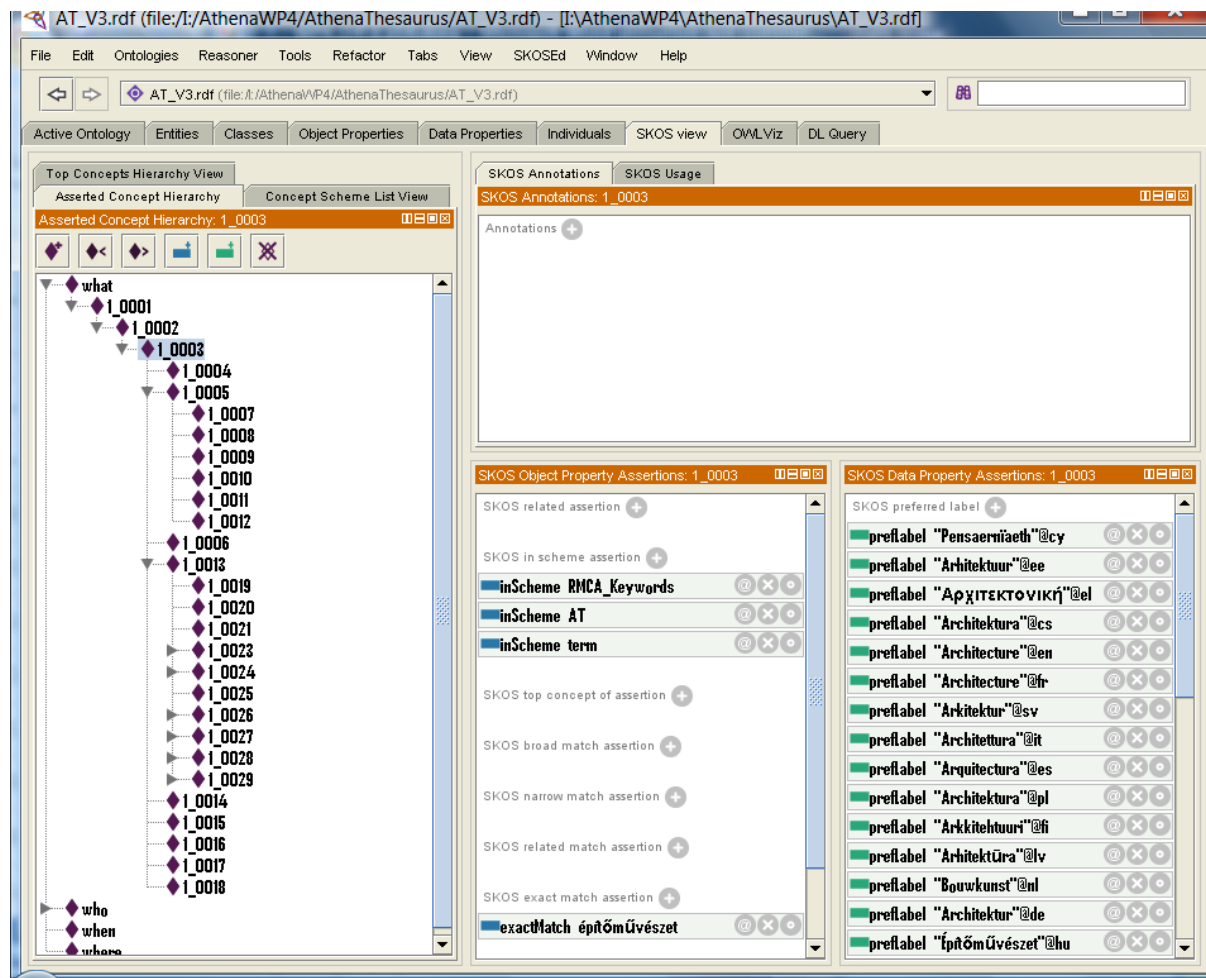Here follows a screenshot of the ATHENA Thesaurus:

Figure1: Preview of the ATHENA Thesaurus in SKOSed

The mapping of these four resources was done manually for the purpose of testing. As our benchmark on terminology and dedicated tools is still ongoing, tools for automatic or semi-automatic mapping will be studied.

The mapping process helped making the connections between very different resources and also enabled the multilinguality; the Athena Thesaurus has now concepts expressed in 5 languages: English, Italian, French, Dutch and Hungarian.

In order to use the SKOSed tool, which is a plug-in to Protege (tool dedicated to the management of OWL ontologies), the URIs have been set in an explicit form. As we recommend in our guidelines. A next step for the elaboration of this thesaurus will be to choose a Persistent Identifier System (described in the Guidelines section) and define a sustainable way to identify the concepts.

5.2. Europeana achievements

Europeana and its development relies on the outcomes of several European projects which are thematic (Athena for the European museums, APEnet for the archives, ...) or specific (EFG European Film Gateway, MIMO Musical Instruments Museums Online, ...)

Besides these projects, the core of Europeana has been developed within two specific projects: Europeana version1[1] and Europeana Connect[2].

Europeana Version 1 has provided the general framework for the integration and coordination of all the content providers to Europeana. Europeana Version 2 will start from August 2011 and will focus on Europeana content and Linked Open Data.

Europeana Connect which started in May 2009 has produced some main outputs[3] regarding terminology and multilingualism :

- EuMDR, Europeana Metadata Registry: this registry will provide details and documentation about the metadata scheme in use from the different content providers of Europeana. This achievement for metadata scheme will be reused within the Linked Heritage project for developing a terminology registry.

- Europeana Language Resources Repository: this repository gathers information and links to Open Source languages Resources and Tools. These resources can be lexical or terminological and multilingual or monolingual. You can access this repository through this URL http://europeanalabs.eu/wiki/WP2LanguageResources

The Semantic Search prototype that we already mentioned is one of the main development of Europeana requiring terminology and interoperability. You can access this semantic search engine via this URL: Europeana Thought lab : http://www.europeana.eu/portal/thought-lab.html

5.3.Linked Heritage

In Athena WP4 activity, thanks to an experiment we have raised some issues and consequently make an effort to specify an ideal tool for terminology management specifically dedicated to non-expert users[4].

This results will be taken into account in the new CIP ICT PSP project: **Linked Heritage.** This project is starting right now, and addresses the coordination of Standards and Technologies for the enrichment of Europeana. Among all of the WP, one(Work Package 3) willbededicated to terminology management and multilingualism. In fact, this WP aims to take advantage of ATHENA WG4 activity, and to implement a prototype of an integrated software platform for terminology management. The WP will also benefit from the work done on the ATHENA Thesaurus in order to continue its completion and extend it to the other domains as Linked Heritage is a cross-domain project.

---

[1] Europeana version 1: http://www.version1.europeana.eu/web/europeana-project/home

[2] Europeana Connect: http://europeanaconnect.eu

[3] Europeana Connect results: http://europeanaconnect.eu/results-and-resources.php

[4] see D4.2, Conclusion part

Moreover, as our benchmark helped us to identify relevant tools and structures in regards with terminology management, possible solutions and developments are to be investigated in order to adapt the ATHENA Ingester for terminology mapping or integrating components of the xTree tool. In both cases, collaboration with partners of the Linked Heritage project will be reinforced in order to propose a sustainable solution.

The main outcomes of this WP within Linked Heritage are a state of the art of terminologies in use in European institutions, a definition of functional needs regarding terminology, the development of a terminology registry and the development of a terminology management platform.

The WP3 of Linked Heritage will then bring together all the efforts led within Athena WP4 and Europeana achievements to reach these objectives and provide a sustainable solution for terminology management.

6. Annexes

6.1. Mapping sheet for connection with the SKOS datamodel

My Terminology *(Ex: Architecture)*	Thesaurus	Athena Format	Explanation
Micro-Thesaurus *Ex:Architecture*	*Micro-thesaurus* *Ex:Architecture*	*skos:ConceptScheme (class)* *skos:hasTopConcept(property)* *Ex:Architecture*	*If your terminology has a micro thesaurus on Architecture, you can describe it as a concept Scheme according to the SKOS model.*
Group of terms *Ex:buildings*	*Thesaurus Array* *Ex:buildings*	*Skos:Collections (class)* *Ex:buildings*	*If your terminology has thematic or other specific groups of terms, the SKOS Collections class allows you to reproduce these groups of terms.*
term level N *ex: Monument*	*Descriptor, vedette* *ex: Monumento* *Non-descriptor* *Used For (UF)* *ex: construction*	*Concept* *- Preferred label* *skos:preflabel* *ex: Monumento @it* *- Alternative label* *skos:altlabel* *ex: construction* *- Hidden label* *Skos :hiddenLabel* *ex :hut*	*The descriptors/terms of your terminology can be defined as prefererred, alternative or hidden label*

term level N-1 *ex: Palace*	*Narrower term (NT)* *ex: Palace*	*skos:narrower* *ex: Palace*	*The Narrower term of your terminology can be defined as a skos :narrower concept*
Term level N+1 *ex:Architecture*	*Broader term (BT)* *ex:Architecture*	*skos:broader* *ex:Architecture*	*The Broader term of your terminology can be defined as a skos :broader concept*
Term level N *ex:Building*	*Related term (RT)* *ex:Building*	*skos :related* *ex:Building*	*The Related term of your terminology can be defined as a skos :related concept*
Notes	*Notes* *Scope note* *Definition* *Date of entry* *...*	*Notes* *skos :scopeNote* *skos :definition* *skos :historyNote* *…*	*The SKOS model gives you a large choice of notes that can be easily transposed from your terminology if this is a thesaurus.*

This mapping sheet intends to help you seeing quickly to which SKOS feature the structure of your terminoly may refer to.

6.2. Acronyms

ARK: Archival Resource Key
CSV: Comma-Separated Values
DC: Dublin Core
DOI: Digital Object Identifier
EDM: Europeana Datamodel
EFG: European Film Gateway
ESE: Europeana Semantic Elements
FOAF: Friend Of A Friend
GEMET: General Multilingual Environmental Thesaurus
ISO: International Standard Organisation
LIDO: Light Information Describing Objects
LOD: Linked Open Data
MIMO: Musical Instruments Museums Online
NBN: National Bibliography Numbers
OAI-PMH:Open Archive Initiative – Protocol for Metadata Harvesting
OCR: Optical Character Recognition
OWL: Web OntologyLanguage
PURL: Persistent Uniform Resource Locators
RDF: Resource Description Framework
RDFS: RDF Schema
SKOS: Simple Knowledge Organisation System
UGC: User Generated Content
UML: Unified Modeling Language
URI: Uniform Resource Identifier
URN: Uniform Resource Name

6.3. W3C: World Wide Web Consortium UML Diagram presented in the norm ISO 25964-1