YOUR TERMINOLOGY
AS A PART OF THE SEMANTIC WEB
RECOMMENDATIONS
FOR DESIGN AND MANAGEMENT

YOUR TERMINOLOGY AS PART OF THE SEMANTIC WEB: RECOMMENDATIONS FOR DESIGN AND MANAGEMENT

edited by
Athena WP4
«Terminology and multilingualism»
and
Linked Heritage WP3
«Terminology and multilingualism»

texts by
Marie-Véronique Leroi,
Ministry of Culture and Communication (France)
Johann Holland,
Michael Culture Aisbl (Belgium)
Stéphane Cagnot,
Dédale (France)

Full results available at http://www.athenaeurope.org/athenawiki

Linked Heritage & Athena general co-ordinator
Rossella Caffo

design
mt milani,
geo graphic sdf

printed
in the month of november 2011
by Repro Stampa Ind. Grafica, Villa Adriana Tivoli (Roma)

Linked Heritage logo
Susan Hazan

web version
http://www.linkedheritage.org/index.php?en/181/publications

table of contents

Foreword

The aim of the Linked Heritage project is to support cultural institutions in providing object data for publication in Europeana.

A way of supporting them is by providing instruments and tools helping them in enlarging their knowledge on digitisation issues.

This booklet containing recommendations for design and management of terminologies, produced in the framework of Linked Heritage, represents another piece to be added to the results already reached during the Athena project, when four booklets were produced dealing with standards, GIS, persistent identifiers and LIDO.

Such tools are not only useful to our Best Practice network but to cultural institutions in general, considering that the previous booklets were downloaded 10,000 times in a few months.

“Your Terminology as part of the semantic web” represents a continuity between the two projects. Started during Athena in the framework of the activities of the working group “Terminology and Multilingualism” (WP4), it was

completed during Linked Heritage by the thematic working group dealing with the same topics (WP3).

Therefore, the experience of two Best Practice networks is merged into this practical guide that will lead the readers into the challenging world of terminologies.

Rossella Caffo
Linked Heritage Project Coordinator

Foreword

The volume of the digital cultural resources is growing very rapidly thanks to the investment of national, regional, European and international digitization initiatives.

Further, the development of portals and thematic applications is encouraging more and more researchers, students and cultural lovers in general to approach the digital cultural heritage as new form of knowledge.

However, the potential of the digital technologies could be locked because of the limits of accessing such huge amount of data.

Access can be limited by the bandwidth of the connectivity or by the actual computing power of the involved servers and workstations. The e-infrastructures born in the last years to support the European research are coping with these limitations. Examples of them are the National Research and Educational Networks, connected at European level by Géant, which provide high speed connectivity and which are progressively opening their facilities to the cultural sector. Other relevant examples are the National Grid Initiatives together with the European Grid Initiative, which provide share computing and storage capability to the various domain of the research, including the digital cultural heritage.

Access, though, have another major and more subtle enemy: the linguistic barriers. One of the most famous richness of Europe is represented by its languages, but this risks often to be perceived as a difficulty instead of an opportunity. As long as the digital domain of the cultural heritage is widening, the issue of preserving the multilingual characteristic of the cultural knowledge is becoming a priority.

As a starting point, multilingualism means to leave the freedom to choose which language to use for the description of a cultural object. Multilingualism therefore means from one hand to describe a cultural object with the language of the people who know that object, either because they studied it, or because the object belongs to the civilization that speaks that language, or because the memory institutions who holds that object use that language.
From the other hand, multilingualism means that a researcher can find that cultural object by searching in his/her own language and can understand the description of that cultural object even if it was written in a different language.
To achieve these basic objectives, the multilingual terminologies are fundamental and it is for this reason that this booklet represents a very important instrument for everybody who is approaching the multilingual dimension of the digital cultural heritage.

Antonella Fresa
Linked Heritage Technical Coordinator

Introduction

Collections management usually implies the use of one or several terminology resources. Then to give a complete access to your collections, you have to provide metadata on your digital collections and your terminologies as well. Making your terminology as a part of the Semantic Web is the lead to improve the visibility of your collections and objects at a national and international level.

This publication aims in priority at people working in European museums, experts or non-experts in Information Engineering and/or Linguistics, who have an interest on terminology and multilingualism.

It proposes recommendations helping you to manage your terminology, to make it compliant and optimized for improving multilinguality on Europeana and to make it as a part of the Semantic Web following.

The more you respect them, the more you improve the future retrievability of your digital collections online.

1. Context and objectives

1.1 From Athena WP4 to Linked Heritage WP3

Linked Heritage is a Best Practice network within the ICT-PSP funding Program. Enhancing the aggregation of new content and its retrievability through Europeana are the main objectives of this project. Linked Heritage is a legacy of the Athena project. Athena was also part of the constellation of projects contributing to Europeana. It started in November 2008 and ended in April 2011. Its general aim was to help the integration of European museums' digital resources into the Europeana portal. Among the different workpackages (WP) of the Athena project, the WP4 was dedicated to the terminology management with two strong focuses on multilingualism and SKOS (Simplified Knowledge Organisation System).

Athena WP4 produced three deliverables (D4.1 Inventory of resources[1], and D4.2 SKOS guidelines[2] and D4.3 Final recommendations[3]) on the basis of surveys, experimentations and the organisation of workshops.

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 You can find D4.3: in pdf version at: http://www.athenaeurope.org/getFile.php?id=1029 in updated wiki version at: http://www.athenaeurope.org/athenawiki/index.php/Recommendations_Introduction

All that activity has been supported by a communautarian Wiki[4] on which are available the results and all that has been produced to get them.

The Athena WP4 dedicated to terminology and multilingualism has concluded its 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.

The Linked Heritage WP3 as a legacy of the Athena WP4 is building on this work and this publication is then a synthesis of the recommendations and guidelines phrased within Athena.

4 http://www.athenaeurope.org/athenawiki/Documentation From workshops : http://www.athenaeurope.org/athenawiki/index.php/Documents

1.2 Recipients

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

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.

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 publication, 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.

1.2.3 A misknowledge of the technological environment

Finally we emphasize the fact that a lot of museums do not have up-to-date and precise vision of the current technological environment. The unavoidable evolution of the Web has a strong impact on the way how institutions have to manage their data. This evolution occurs under the influence of the different new technologies, norms and standards in use, and it is particularly co-dependent of the evolution of the massive Web usage. Because we consider that this misknowledge may produce important difficulties and misunderstandings, we give below a presentation of some key points featuring the current technological environment and its evolution.

1.3 Technological environment

1.3.1. 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 primar Web, 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.

1.3.2. 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 technically, 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"[5]. 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"[6].

5 http://www.w3.org/People/Berners-Lee/Weaving/glossary.html

6 http://www.uen.org/core/edtech/glossary.shtml#S

1.3.3. 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. This initiative 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[7].

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.

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

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 reference resources 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, first 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.

1.3.4. 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 Europeana datamodel 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 Europeana datamodel. Indeed Europeana is currently releasing a new datamodel, EDM (Europeana Data Model) 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.

1.3.5. 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 a compliant format.

When you want to represent or 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

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

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

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 respectively to the 2004 and 2009 specifications. 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.

2. Recommendations purpose

2.1 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 publication, 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 Web standards in order to fit with the portal constraints for the semantic exploitation of your digital resources descriptions. So you could use OWL (Web Ontology Language) , which is the most formal and complex 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 (Simple Knowledge Organisation System) 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 SKOSifying your 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 to 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.

2.2 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.

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 recommendations as 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.

3. 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 interoperable with other resources, and when you will link it to a network of terminologies.

3.1 Methodology

Here is a task list that you can follow step by step if you need 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 (about your semantic descriptions)
- A3: Define your connection with the datamodel
- A4: Choose the terms for the semantic description of your digital resources
- A5: Organise your terms into a thesaurus structure
- A6: Find equivalent terms in other languages
- A7: Implement your thesaurus

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. archaeology, 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 enriching your 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 micro-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.

A1 DEFINE YOUR COLLECTION DOMAIN(S)

Example	If you intend to describe an organ as music instrument and to make a terminology about musicology, and if moreover you aim at describing that organ as a religious movable, refer to different micro-thesauri about "musical instruments" 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 two collection domains at least to take into account: "musicology" and "religion". And at least two sub-domains: "musical instrument" and "religious movables".
Methods & tools	As a first step to define your general domain, you can consult HEREIN (http://european-heritage.net) and MICHAEL (http://www.michael-culture.eu) 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[8].

8 http://incipioinfodoc.archimed.fr/Idesia/home.aspx?INSTANCE=MIMO&THES=IFD_MIMO_CLASSIF&VIEW=DEFAULT&FORM=o&ACTIVE=TRUE

A2 IDENTIFY YOUR USERS' EXPECTATIONS

Actions	Identify for all the users of your semantic descriptions which expectations 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 objects and the collections data, the earlier the point of view of who is accessing is taken into account, the more efficient the portal would be. Most of the time, requests will not be expressed by professionals, but by the general public. It means the Europeana 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.

A2 IDENTIFY YOUR USERS' EXPECTATIONS

Purpose	Thus we recommend you designing thesauri by considering the skills, habits and expectations of the general users then the 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 is 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.
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

A2 IDENTIFY YOUR USERS' EXPECTATIONS

Example	• 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 information your terminology has to provide. Then, since you consider that the use of your own descriptors for 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[9] (the user can use your descriptors without any modification and with mentioning your institution as author of the terminology).
Methods & 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 terms are often used to access to your collection pages).

9 http://creativecommons.org/licenses/

A3 DEFINE YOUR CONNECTION WITH THE DATAMODEL

Actions	As we explained above, 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 kind 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 model covers 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.

A3 DEFINE YOUR CONNECTION WITH THE DATAMODEL

Purpose	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.
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 & tools	In order to help you make this connection more easily, you can find a mapping sheet in annex. Among the existing ISO norms, you can look at BS8723: Structured Vocabularies for Information Retrievaland also follow the norm ISO 25964-1 Thesauri and interoperability with other vocabularies: Thesaurus for information retrieval (which has been published in 2011) because it expresses more precisely the link between the 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 xTree developed 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[5] 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 repository you 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);

Actions	*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).
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).

A4 CHOOSE YOUR TERMS

Purpose	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.
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 do not 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 & 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

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 using this kind of terminology rather than ontology or classifications[10] to describe their collections in a well-structured manner (47% of the 149 terminologies we have listed in our initial survey are thesaurus[11]).

10 To get an overview of the kinds of terminologies, look at the D4.1, part 3.4 «Types of terminology resources», or online at: http://www.athenaeurope.org/athenawiki/index.php/Terminology_management#Types_of_terminology_resources

11 see D4.1 results.

A5 ORGANISE YOUR TERMS INTO A THESAURUS STRUCTURE

Purpose

Without forgetting that regular relations of equivalence and association are particularly relevant for multilingualism. 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 between micro-thesauri, we recommend you to multiply the links in order to improve your terminology exploitation later.

A5 ORGANISE YOUR TERMS INTO A THESAURUS STRUCTURE

Purpose	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 (Collections for 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.
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.

A5 ORGANISE YOUR TERMS INTO A THESAURUS STRUCTURE

Methods & tools

For the time being, 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[12], which propose you different ways to browse the thesaurus: thematic, alphabetic or hierarchical listings.

12 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: • To identify the language(s) in which you want or have to propose your descriptions. • Not to proceed to a literal translation of your terms in the identified languages previously. • But to find, for your collection domain(s), existing vocabularies in the language you are interested in and point out the ones you can consider as equivalent to yours. • To ask experts of the domain and speakers of the foreign language to help you find the missing equivalent terms and to validate the whole choice of equivalent terms. • To consider these 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).

A6 FIND EQUIVALENT TERMS IN OTHER LANGUAGES

Purpose

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[13] 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 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.

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

A6 FIND EQUIVALENT TERMS IN OTHER LANGUAGES

Purpose

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.

Example

FACET THEMATIQUES (F)	FACET THEMATIQUES (NL)	FACET THEMATIQUES (E)
transport	transport	transport
transport sur terre	transport over land	transport over land

The example above comes from a 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 & tools

You can consult a repository of free-to-use terminologies on the Athena Wiki where the resources are classified by domains and languages : http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources

A6 FIND EQUIVALENT TERMS IN OTHER LANGUAGES

Methods & tools

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 technologies may have solved some of the issues pointed out in this norm. The latest norm ISO 25964-1 that we already mentioned will address better the design of a monolingual or multilingual terminology 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: – Three already finalized standards: ISO 2788:1986: + ISO 5964: 1985 + ANSI/NISO Z39.19-2003 – 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.

A7 IMPLEMENT YOUR THESAURUS

Purpose

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" has been 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 (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.

A7 IMPLEMENT YOUR THESAURUS

Purpose	Among all the existing tools we have identified during our benchmark[14], none 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 and then 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 that you are making a first step for your terminology SKOSification. The second one is that the arborescence display of 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.
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", "triumphal arch","therms"... and in the "habitation" list you can have "apartment", "hut", "house", "squat"... Finally, your transversal group of terms, on the theme of "buildings", gathers "palace", and "apartment".

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

A7 IMPLEMENT YOUR THESAURUS

Example

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.

Terms	Source micro-thesauri
Apartment	Habitation
Palace	Monument

Methods & 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 tool such as OpenOffice.

It is a free tool which 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. Make your terminology 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.

4.1 Benefits from using SKOS

RDFS (Resource Description Format Schema) 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.

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.

4.2 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 Europeana requirements for the ingestion and the exploitation of your digital resources.

As a format 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 . 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: Validate your SKOSification

B1 EVALUATE HOW FAR SKOS IS COMPLIANT WITH YOUR TERMINOLOGY FEATURES

Actions	As a prologue 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. • 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. • 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. • 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) • Check if in your thesaurus you have a list of Institution names; If yes, you shall need to specifically use Vcard[15]/hCard[16] and FOAF in addition to SKOS. • 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.
Purpose	The objective is to SKOSify your thesaurus, that is, to make your terminology interoperable with a datamodel like LIDO, and consequently with Europeana.

15 Vcard : http://www.w3.org/TR/vcard-rdf/
16 hCard : http://microformats.org/wiki/hcard

B1 EVALUATE HOW FAR SKOS IS COMPLIANT WITH YOUR TERMINOLOGY FEATURES

Purpose

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: Can the 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.

B1 EVALUATE HOW FAR SKOS IS COMPLIANT WITH YOUR TERMINOLOGY FEATURES

Purpose	• 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 Vcard/hCard would be more apropriate. If the terminology is dealing with both of them, a mix of Vcard/ hCard and SKOS or OWL could be interesting. (VcARD/hCARD is a microformat dedicated to the publication of contact details of people, organizations or places.) • 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.
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.

B1 EVALUATE HOW FAR SKOS IS COMPLIANT WITH YOUR TERMINOLOGY FEATURES

Example	• If you intend to model an authority list with authors or people's names, you can have a look on the VIAF (Virtual International Authority File)[17] 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 & 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

17 VIAF, Virtual International 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: • If you have a collection management tool proposing a function of export in SKOS, use it directly. • If you do not have such an in-house tool, and have implemented your thesaurus in a spreadsheet, you can use the tool XL2XML,or XLTaxonomy, and their by-default style sheet. • If you prefer defining yourself the style sheet enabling the SKOSification of your XML file, you can use a tool such as Annocultor. • 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.

B2 ROUGHLY SKOSIFY YOUR THESAURUS

Purpose

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 SKOSify your thesaurus without spending too much resource consists of using the XL2XML tool[18]. 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 can be an economic and efficient solution. But it can miss flexibility. Another tool which provide the same kind of features is XLTaxonomy[19]. 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.

18 https://www.seegrid.csiro.au/wiki/bin/view/Siss/ExcelToRdfTool#XL2XML_conversion_tool
19 http://www.modelfutures.com/file_download/16/xlTaxonomy.zip

B2 ROUGHLY SKOSIFY YOUR THESAURUS

Purpose

3. The third possibility consists in using a specific tool. Annocultor[20] 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.

Example

Suppose your terminology is structured in a spreadsheet as follow:

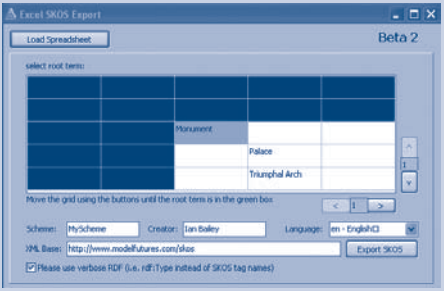
Sub-domains	Terms
Monument	
	Palace
	Triumphal Arch
	Therms
Habitation	
	Apartment
	Hut
	House
	Squat

20 http://annocultor.eu/

B2 ROUGHLY SKOSIFY YOUR THESAURUS

Example

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.



You may need to check the SKOS output produced by the tool and if needed you may convert it in the SKOS/RDF form that can be exploited by your tools. The tool can produce either an XML or SKOS file.

Methods & 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.

B3 DEFINE WITH PRECISION THE LABELS EXPRESSING CONCEPTS

Purpose

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 it 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.

B3 DEFINE WITH PRECISION THE LABELS EXPRESSING CONCEPTS

Purpose

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 commonly 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.

B3 DEFINE WITH PRECISION THE LABELS EXPRESSING CONCEPTS

Purpose

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.

B3 DEFINE WITH PRECISION THE LABELS EXPRESSING CONCEPTS

Purpose	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.
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[21] 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 as possible to decompose them in order to get to a simple form.

21 Turtle Terse RDF Triple : http://www.w3.org/TeamSubmission/turtle/

B3 DEFINE WITH PRECISION THE LABELS EXPRESSING CONCEPTS

Example	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 & 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 norm ISO 25964-1.
	Other editing tools such as Protégé[22] (with the plugin SKOSed) or ThManager[23] 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 date version of the existing tools for SKOS editing: http://www.athenaeurope.org/athenawiki/index.php/Benchmark

22 You can find Protégé and SKOSed at: http://protege.stanford.edu/
23 You can find ThManager at: http://thmanager.sourceforge.net/

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[24]: * 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 Comma-separated values) 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.

24 http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/

Purpose	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 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 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. 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.
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 do not 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).

Example	Here is an example of URI with ARK from the BnF: http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11931420f
Methods & tools	Different systems for Persistent Identifiers are in use. Here some information on 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.

Methods & tools

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 (Handle was a previous system for identification and references of resources). 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 contents'.

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 refer to the booklet "Persistent identifiers: Recommendations for institutions"[25] elaborated by the WP3 of Athena.

25 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.

B5 MAP YOUR CONCEPTS

Purpose	**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.
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 scheme on Religion with a concept "organ" you can proceed with different mappings (represented here as Turtle[26] 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".

26 Turtle Terse RDF Triple: http://www.w3.org/TeamSubmission/turtle/

Example	This concept will then be related to two different concept schemes. This mapping of concepts allow you to avoid the repetition of concepts.
Methods & 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 glance at the Benchmark led within the WP4 of Athena as the list of tools keep on evolving: http://www.athenaeurope.org/athenawiki/index.php/Benchmark

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 of 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.

B6 MAP YOUR TERMS

Purpose	**Provide for each concept an equivalent label in the languages involved in the 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 equivalences 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 639-11 where the language tags are coded on two letters in lower case.
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.

Example	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 an 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 & 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)[27] for multilingual alignment. Europeana has set a repository of tools and lexical resources developed and used in the framework of NLP and there is also an incubator for open source code for these technologies[28]. 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-Protégé, ...) to proceed with the mapping of terms once the conceptual work is done.

27 http://europeanalabs.eu/wiki/WP2LanguageResources
28 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.

B7 ENSURE THE DOCUMENTATION OF CONCEPTS

Purpose	In order to make your documentation, you can use more or less precise notes which are 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.
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 & 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 in the validation of concepts and labels mapping, and in 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 does not take into account the latest version of the SKOS model[29]. Pool Party, a thesaurus management system, offers online SKOS services[30] 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 unauthorized character in the URI. Although if an URI is used twice for identifying two different concepts, there will not be any alert or warning.

29 http://www.w3.org/2004/02/skos/validation
30 http://demo.semantic-web.at:8080/SkosServices/index

B8 VALIDATE YOUR SKOSIFICATION

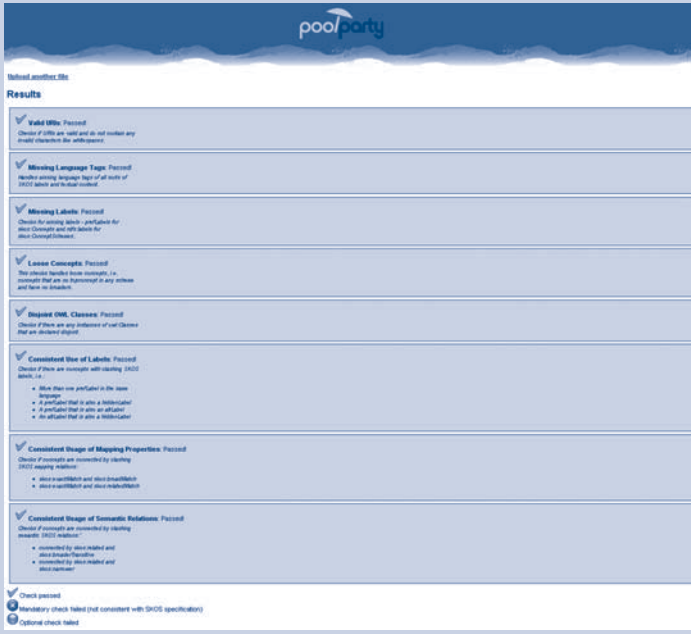
Purpose	**Missing language tags**: the tool checks if all the labels and notes have a language tag **Missing labels**: the tool checks that each concept has at least one preferred label. **Loose concepts**: all the concepts that are isolated and not linked to other concepts are pointed out as loose concepts **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. 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.

B8 VALIDATE YOUR SKOSIFICATION

Example

Beside 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 & tools

The editors such as Protégé-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/

5. Link your terminology 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.

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 a 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 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.

C1 DEFINITION OF METADATA ON YOUR TERMINOLOGY

Purpose

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 for 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

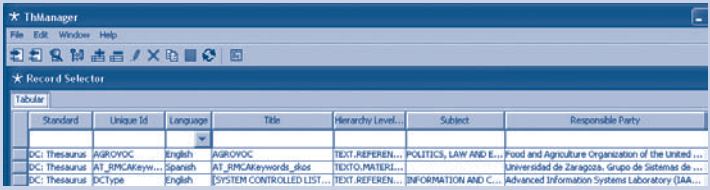
In order to declare your metadata, we recommend you:
• To wrap all of your metadata in <rdf:Description> tags.
• To use DC tags for defining the title (<dc:title>), the identifier (<dc: identifier>), the creator (<dc:creator>), the contributors (<dc: contributor>), the format (dc:format>), the languages (<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>)

C1 DEFINITION OF METADATA ON YOUR TERMINOLOGY

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[31] is an editing tool for SKOS thesauri which allows to registrate in an internal database several thesauri.
Here a screenshot of the registry of ThManager:



Methods & 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)[32] 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.

31 http://thmanager.sourceforge.net
32 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: – Domain of description linkable with yours – 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: • To find terms potentially belonging to the same domains than yours • To find terms potentially equivalent to yours in other languages • To achieve such goals, a repository of terminologies appears as a very useful solution.

C2 IDENTIFICATION OF RESOURCES FOR MAPPING

Purpose	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[33] or DBPedia[34].
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 repositories to 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: http://eculture.cs.vu.nl/europeana/www/datacloud.html The Thesaurus of Geographic Names (TGN)[35] from the Getty and Geonames[36] are major resources for places' names. Then you can start proceeding with the mapping of your locations' names with those of the TGN and Geonames.

33 http://eculture.cs.vu.nl/europeana/www/datacloud.html
34 http://wiki.dbpedia.org/OnlineAccess
35 TGN-Getty: http://www.getty.edu/research/tools/vocabularies/tgn/
36 36 Geonames: http://www.geonames.org/

C2 IDENTIFICATION OF RESOURCES FOR MAPPING

Methods & 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.

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 internal mapping 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.

C3 MAPPING WITH OTHER RESOURCES

Purpose	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.
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. 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 is an editing tool for SKOS thesauri which allow to registrate in an internal database several thesauri.
	The screenshot beside of ThManager registry represents the concept of "cultural heritage" in the GEMET thesaurus.

C3 MAPPING WITH OTHER RESOURCES

Example

The mapping could be of a real benefit since this concept is already mapped with other terminology resources such as Agrovoc[37] or EuroVoc[38] and also to the corresponding article in Wikipedia.

37 Agrovoc: http://aims.fao.org/website/AGROVOC-Thesaurus/sub
38 EuroVoc: http://eurovoc.europa.eu/

C3 MAPPING WITH OTHER RESOURCES

Methods & 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.
Example	The semantic SearchLab[39] 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: 

C4 VALIDATION OF THE INTEROPERABILITY

Example	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 & tools	A free tool such as Pearltrees[40] 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.

39 Europeana Thought lab: http://www.europeana.eu/portal/thought-lab.html
40 Pearltrees: http://www.pearltrees.com

6. Conclusion

All the recommendations we have phrased in this publication 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.

The Linked heritage project is the legacy of the Athena Project. Athena was an eContent plus project aiming at providing objects and collections from European museums to Europeana. The Athena WP4 was dedicated to terminology and multilingualism. The Linked Heritage Workpackage 3 takes into account the results of the Athena WP4 activity.

Thanks to the experiment led within Athena WP4, some issues have been raised and consequently we have made an effort to build the technical specifications for an ideal tool for terminology management specifically dedicated to non-expert users.

The Linked Heritage WP3 builds on the legacy of the Athena WP4 activity to implement a prototype of an integrated software platform for terminology management.

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.

7. Annexes

7.1 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
OWL: Web Ontology Language
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
VIAF: Virtual Information Authority File

7.2 References

Here you can find a structured reminder of all the references that have been mentionned above. They are organised in 5 categories:

- Repositories
- Vocabularies
- Datamodels and formats
- Recommendations / guidelines
- Tools

7.2.1 Repositories

D4.1 Inventory of resources:
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

DBPedia: http://wiki.dbpedia.org/OnlineAccess

Europeana datacloud: http://eculture.cs.vu.nl/europeana/www/datacloud.html

7.2.2 Vocabularies

Agrovoc: http://aims.fao.org/website/AGROVOC-Thesaurus/sub
Eurovoc: http://eurovoc.europa.eu/
GEMET: http://www.eionet.europa.eu/gemet
Geonames: http://www.geonames.org/
Getty Vocabularies: http://www.getty.edu/research/tools/vocabularies/index.html

HEREIN: http://thesaurus.european-heritage.net/herein/thesaurus
MICHAEL: http://www.michael-culture.org
MIMO terminology: http://incipioinfodoc.archimed.fr/Idesia/home.aspx?INSTANCE=MIMO&THES=IFD_MIMO_CLASSIF&VIEW=DEFAULT&FORM=0&ACTIVE=TRUE
TGN-Getty: http://www.getty.edu/research/tools/vocabularies/tgn/

7.2.3 Datamodels and formats

LIDO: http://www.athenaeurope.org/index.php?en/112/news/21/lido-the-harvesting-format-used-within-athena
Athena Format: http://www.athenaeurope.org/athenawiki/index.php/References#Athena_Format
SKOS reference: http://www.w3.org/TR/2009/REC-skos-reference-20090818/
SKOS wiki: http://www.w3.org/2001/sw/wiki/SKOS/Datasets

7.2.4 Recommendations / guidelines

D4.2 Guidelines for SKOSification:
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

Persistent identifiers: **Recommendations for institutions**: http://www.athenaeurope.org/getFile.php?id=779

Norms:
BS8723: Structured Vocabularies for Information Retrieval
ISO 25964-1 Thesauri and interoperability with other vocabularies: Thesaurus for information retrieval

7.2.5 Tools

Our benchmark results about tools are available at: http://www.athenaeurope.org/athenawiki/index.php/Benchmark#Tools

Among the whole set of tools:
Annocultor: http://annocultor.eu/
ITM3: http://www.mondeca.com/Products/TM
Pearltrees: http://www.pearltrees.com
Pool Party: http://poolparty.punkt.at/
Protégé + SKOSed: http://protege.stanford.edu/
ThManager: http://thmanager.sourceforge.net/
XL2XML: https://www.seegrid.csiro.au/wiki/bin/view/Siss/ExcelToRdfTool#XL2XML_conversion_tool
XLTaxonomy: http://www.modelfutures.com/file_download/16/xlTaxonomy.zip
xTree: http://xtree.digicult-museen.de/

7.3 Mapping Sheet

The mapping sheet intends to help you see quickly to which SKOS feature the structure of your terminology may refer to.

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 preferred, alternative or hidden label

My terminology (ex: Architecture)	Thesaurus	Athena Format	Explanation
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.

Michael
Culture
Association

European Network

providing new
experiences of
European digital cultural heritage

www.michael-culture.eu

Education and Culture DG
Culture Programme

Parmigianino, *La schiava turca* - Parma, Pinacoteca Nazionale

Michael Culture Association

A not-for-profit organisation, founded in April 2007 under Belgian law.

• Supports the Minerva network of European professionals working on digital cultural heritage.
It federates professionals and experts through support and development of the transmission and exchange of information, via publications, conferences and events.
It has members from all over Europe, including public agencies, cultural institutions and private organisations.

• Participates in major European initiatives
Michael Culture Association aims at enhancing the digitization and promotion of European cultural heritage.
It is participating in the ongoing construction of Europeana, the European Digital Library.
The association is a member of the executive committee of the EDL foundation and is also a partner in major European projects.

• Promotes and develops the Michael portal
The European catalogue of online digitized cultural collections offers access to over 10.000 collections of high quality cultural content from archives, museums, libraries etc, bringing together national initiatives from across Europe.

JOIN Michael Culture Association

www.michael-culture.eu

contact
info@michael-culture.eu

With the support of the Culture Programme of the European Commission

Education and Culture DG
Culture Programme

minerva Minerva Knowledge Base Digitising Content Together
MEDCULT
Michael Multilingual Inventory of Cultural Heritage in Europe
ATHENA Access to cultural heritage networks across Europe
LINKED HERITAGE
DC NET Digital Cultural Heritage Network
INDICATE
europeana project group member

in the same series