



VOTRE TERMINOLOGIE

COMME ÉLÉMENT DU WEB SÉMANTIQUE

RECOMMANDATIONS

DE CONCEPTION ET DE MANAGEMENT



VOTRE TERMINOLOGIE

COMME UN ÉLÉMENT WEB SÉMATIQUE

RECOMMANDATIONS

DE CONCEPTION ET DE GESTION

Votre terminologie
comme élément du web sémantique:
Recommandations
de conception et de management

édité par les groupes de
travail ATHENA WP4 et Linked
Heritage WP3 « Terminologie
et multilinguisme »

textes de
Marie-Véronique Leroi,
Ministère de la Culture et de
la Communication (France)
Johann Holland, Michael
Culture Aisbl (Belgique)
Stéphane Cagnot, Dédale
(France)

L'ensemble des résultats est
disponible à l'adresse [http://
www.athenaeurope.org/
athenawiki](http://www.athenaeurope.org/athenawiki)

Linked Heritage et ATHENA
Coordinateur général
Rossella Caffo

Coordinateur technique
Antonella Fresa

design
mt milani
logo Athena
Susan Hazan

version Web [http://www.
athenaeurope.org/index.
php?en/110/promotional-
material](http://www.athenaeurope.org/index.php?en/110/promotional-material)

Ce document est sous licence
Creative Commons de type
Attribution Pas d'Utilisation
Commerciale Partage à
l'Identique (CC-BY-NC-SA)
[http://creativecommons.org/
licenses/by-nc-sa/3.0/](http://creativecommons.org/licenses/by-nc-sa/3.0/)



Table des matières

Avant-propos	6
1. Contexte et objectifs	11
1.1 D'Athena WP4 à Linked Heritage WP3	11
1.2 Destinataires	13
1.2.1 Un manque de compétences	13
1.2.2 Un manque de moyens financiers	14
1.2.3 Une méconnaissance de l'environnement technologique	14
1.3 Environnement technologique	15
1.3.1 Web social	15
1.3.2 Web sémantique	15
1.3.3 Web des données Ouvertes	16
1.3.4 LIDO	18
1.3.5 Formats	19
2. Objectif et structure des recommandations	23
2.1 Optimisation et comptabilité	23
2.2 Approche en trois étapes	25
3. Concevoir sa terminologie	27
3.1 Methodologie	27
Etapes:	
A1: Définissez le(s) domaine(s) de votre collection	29
A2: Identifiez les attentes de vos utilisateurs	31
A3: Définissez votre lien avec le modèle de données	35
A4: Choisissez les termes pour la description sémantique de vos ressources numériques	37
A5: Organisez vos termes selon une structure de thésaurus	41
A6: Trouvez des termes équivalents dans d'autres langues	45
A7: Produisez techniquement votre thésaurus	49
4. Rendre sa terminologie intéropérable	54
4.1 Avantages de SKOS	54
4.2 Méthodologie	57
Etapes:	
B1: Évaluez la compatibilité de SKOS avec les caractéristiques de votre terminologie	59
B2: SKOSifiez sommairement votre terminologie	63
B3: Définissez précisément les labels exprimant les concepts	67
B4: Identifiez vos concepts et validez la structure	72
B5: Mappez vos concepts	76
B6: Mappez vos termes (multilingues)	79
B7: Vérifiez que vos concepts sont documentés	82
B8: Validez votre SKOSification	84
5. Lier sa terminologie à un réseau	87
Etapes:	
C1: Définissez les métadonnées de votre terminologie	88
C2: Identifiez les autres ressources avec lesquelles mapper votre terminologie	91
C3: Mappez votre terminologie avec les autres ressources	94
C4: Validez l'interopérabilité	98
6. Conclusion	100
7. Annexes	102
7.1 Acronymes	102
7.2 References	103
7.2.1 Répertoires dépôts	103
7.2.2 Vocabulaires	103
7.2.3 Modèles de données et format	104
7.2.4 Recommandations / directives	104
7.2.5 Outils	105
7.3 Tableau de correspondance	105

Avant-propos

L'objectif du projet Linked Heritage est d'aider les institutions culturelles à fournir les données de leurs objets, pour les publier sur Europeana.

Une manière de les aider consiste à fournir des instruments et des outils leur servant à approfondir leur connaissance des enjeux propres à la numérisation.

Ce livret qui contient des recommandations pour concevoir et gérer des terminologies, et qui a été réalisé dans le cadre de Linked Heritage, représente un élément de plus à ajouter aux résultats du projet ATHENA durant lequel ont été rédigés quatre livrets portant sur les standards, les systèmes d'information géographique (GIS), les identifiants persistants et LIDO.

Il s'avère que de tels outils ne sont pas utiles qu'à notre réseau de bonnes pratiques, mais plus généralement à toutes les institutions culturelles, comme en témoigne le fait que les livrets précédents ont été en quelques mois téléchargés plus de 10 000 fois.

« Votre terminologie comme élément du Web Sémantique » marque la continuité entre les deux projets. Débuté pendant ATHENA dans le cadre des activités du groupe de travail « Terminologie et multilinguisme » (WP4), ce livret a ensuite été complété durant Linked Heritage par le groupe de travail thématique étudiant les mêmes sujets (WP3).

L'expérience des deux réseaux de bonnes pratiques est donc réunie dans ce guide pratique qui conduira les lecteurs dans le monde difficile des terminologies.

Rossella Caffo

Coordinateur général du projet Linked

Avant-Propos

Grâce aux diverses initiatives de numérisation, que ce soit à l'échelle régionale, nationale, régionale, européenne ou internationale, le nombre de ressources culturelles numériques augmente actuellement de façon très rapide.

De plus, le développement de portails et d'applications thématiques encourage de plus en plus de chercheurs, d'étudiants et plus généralement d'amoureux de la culture à considérer le patrimoine culturel numérique comme une nouvelle forme de connaissance. Toutefois, le potentiel des technologies numériques pourrait être gâché tant il est encore difficile d'accéder à une telle masse d'informations.

L'accès peut être limité par l'état actuel de la bande passante des réseaux ou par celui de la puissance de calcul des serveurs et des postes de travail utilisés. Les e-infrastructures qui sont apparues ces dernières années pour soutenir la recherche européenne subissent ces contraintes. C'est le cas par exemple pour les Réseaux Nationaux de la Recherche et de l'Education (National Research and Educational Networks), connectés au niveau européen par Géant, ou pour les Grilles Nationales (National Grid Initiatives) associées à l'initiative européenne des grilles (European Grid Initiative), qui fournissent de la connexion à haut débit, et qui partagent de la puissance de calcul et toute une série de moyens facilitant l'accès au secteur culturel.

Sans compter qu'il existe un autre ennemi majeur à l'accès, et qui est plus subtil encore : les barrières linguistiques.

Une des plus célèbres richesses de l'Europe réside dans sa diversité linguistique qui pourtant est plus souvent perçue comme une difficulté que comme une opportunité. Le développement de la composante numérique du patrimoine culturel conduit à faire de la préservation de cette caractéristique multilingue un enjeu prioritaire.

En premier lieu, le multilinguisme implique qu'on est libre de choisir quelle langue utiliser pour décrire un objet culturel. Ainsi, le multilinguisme signifie d'une part que pour décrire un objet il est possible d'utiliser la langue de ceux qui connaissent cet objet, soit parce qu'ils l'étudient, soit parce que cet objet appartient à une civilisation qui parle cette langue, soit parce que les institutions patrimoniales qui détiennent cet objet utilisent cette langue. D'autre part, le multilinguisme signifie qu'un chercheur peut retrouver cet objet culturel en formulant des requêtes dans sa propre langue et peut comprendre la description de cet objet même si celle-ci a été écrite dans une autre langue que la sienne.

1. Contexte et objectifs

Pour atteindre ces objectifs essentiels, les terminologies multilingues sont fondamentales et c'est pourquoi ce livret constitue un outil très important pour quiconque souhaite aborder la dimension multilingue du patrimoine culturel numérique.

L'expérience de deux Réseaux de bonnes pratiques (Best Practice Networks), Athena et Linked Heritage, est réunie dans ce guide pratique qui mènera les lecteurs au sein d'un monde fascinant et aux enjeux réels.

Antonella Fresa

Coordnatrice technique du projet Linked Heritage

- 1 Vous pouvez trouver le D4.1 : en version pdf à l'adresse : <http://www.athenaeurope.org/getFile.php?id=398> en version wiki à jour à l'adresse : http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources
- 2 Vous pouvez trouver le D4.2 : en version pdf à l'adresse : <http://www.athenaeurope.org/getFile.php?id=684> en version wiki à jour à l'adresse : <http://www.athenaeurope.org/athenawiki/index.php/Guidelines>
- 3 Toute la documentation associée aux workshops est disponible à l'adresse : <http://www.athenaeurope.org/athenawiki/index.php/Documents>
- 4 <http://www.athenaeurope.org/athenawiki/>

1.1 D'Athens WP4 à Linked Heritage WP3

Linked Heritage est un Réseau de bonnes pratiques (Best Practice Network), financé par le programme européen ICT-PSP. Il a pour objectifs principaux d'améliorer l'ingestion de nouveaux contenus et leur accessibilité sur le portail Europeana. Il est le fruit du projet Athena qui faisait également partie de la constellation de projets gravitant autour d'Europeana. Athena avait débuté en novembre 2008 et s'est achevé en avril 2011. Son objectif général était d'aider à l'intégration dans le portail Europeana de toutes les ressources numériques dont disposent les musées européens. Parmi les différents lots de travail (WP) du projet Athena, le WP4 portait spécifiquement sur les questions de gestion de terminologies avec deux accents mis sur le multilinguisme et le format SKOS (Simple Knowledge Organisation System – Système simple d'organisation des connaissances).

Le lot de travail WP4 d'Athens a produit trois livrables (D4.1 Inventory of resources¹ ; D4.2 SKOS guidelines² ; D4.3 Final recommendations³) rédigés sur la base d'études, d'expérimentations et d'organisation d'ateliers⁴.

L'ensemble des résultats de toute cette activité, ainsi que tout ce qui a été produit et imaginé pour les obtenir, est disponible sur le Wiki communautaire⁵ du groupe de travail.

Le lot de travail WP4 d'Athena, dédié aux questions de terminologie et de multilinguisme, a mis un terme à son activité en adressant au final à tous les musées qui souhaitent rendre disponibles leurs ressources numériques sur Europeana un ensemble de recommandations portant sur la gestion de terminologies.

En tant que continuation du WP4 d'Athena, le lot de travail WP3 du projet Linked Heritage s'appuie sur tout ce travail, et en ce sens, la présente publication constitue une synthèse des recommandations et directives formulées dans le cadre d'Athena.

1.2 Destinataires

En tant qu'ensemble de recommandations, cette publication s'adresse aux musées qui envisagent de rendre disponibles sur Europeana leurs ressources numériques. La réalité de la situation économique et technique spécifique aux musées a été prise en compte dans l'élaboration de ces recommandations. Trois constats principaux permettent de résumer les particularités en question : un manque de compétences, un manque de moyens financiers, et une méconnaissance de l'environnement technologique. Ces trois points sont développés ci-dessous.

1.2.1. Un manque de compétences

Il existe tout d'abord un fossé entre les compétences en gestion de terminologie que possèdent en général les membres des musées et celles techniques et expertes habituellement requises par les domaines de l'Ingénierie des Connaissances et de la Linguistique. Idéalement, n'importe quel lecteur dépourvu d'un tel bagage expert en gestion de terminologies devrait être en mesure de comprendre les recommandations formulées dans ce guide. Néanmoins, le haut degré de technicité du sujet nécessite quelques connaissances fondamentales à maîtriser pour faciliter leur compréhension et leur utilisation. C'est pourquoi il a été décidé de rendre ces recommandations les plus compréhensibles possibles et de fournir dans le même temps quelques unes de ces connaissances fondamentales.

1.2.2 Un manque de moyens financiers

Il faut ensuite signaler combien la situation économique des institutions culturelles est critique. Tout changement dans l'utilisation de terminologies peut avoir un impact significatif sur l'équilibre humain et financier d'une structure puisqu'il implique un coût. Ces difficultés et contraintes économiques ont été prises en compte dans l'élaboration des recommandations pour que les musées puissent effectivement suivre toutes les opérations de gestion de terminologies requises pour la bonne exploitation de leurs ressources numériques par Europeana. La présente publication dédiée aux lecteurs non experts participe de cette idée dans la mesure où elle devrait permettre aux musées de ne pas avoir à faire appel à une compétence externe pour lire, comprendre et appliquer ce qui y est recommandé.

1.2.3 Une méconnaissance de l'environnement technologique

Enfin il est important de noter que beaucoup de musées n'ont pas de vision précise et actualisée de leur environnement technologique. L'évolution inévitable du Web a un impact fort sur la façon dont les institutions gèrent leurs données. Cette évolution se produit sous l'influence des différentes nouvelles technologies, normes et standards en usage, et s'avère particulièrement co-dépendante de l'évolution de l'usage massif du Web. Dans la mesure où cette méconnaissance peut produire difficultés et mécompréhensions, la suite de cette partie présente quelques unes des notions clefs qui caractérisent l'environnement technologique actuel et son évolution.

1.3 Environnement technologique

1.3.1. Web Social

Vous avez certainement entendu parler de ce que l'on a coutume d'appeler le Web Social ou Web 2.0. Peut-être même en êtes-vous familier. Le Web 2.0 est une évolution du Web originel, et a permis l'apparition de réseaux d'internautes qui s'y rencontrent et y discutent en ligne de manière instantanée sur différentes plateformes comme Facebook, Twitter ou LinkedIn. Après avoir offert un accès à toute l'information disséminée à travers le monde, le Web a rendu possibles de nouvelles formes de sociabilité. De plus, les nouvelles fonctionnalités ainsi offertes par le Web ont permis aux internautes de produire et de publier eux-mêmes des contenus, connus généralement sous l'appellation : User Generated Contents (UGC – contenus générés par l'utilisateur). Une nouvelle ère d'information s'est ainsi ouverte dans laquelle l'information ne provient pas seulement des éditeurs, mais d'un mélange de sources hétérogènes. Pour les institutions culturelles, un nouvel horizon d'interaction avec les internautes / visiteurs s'ouvre maintenant grâce à ces technologies.

1.3.2. Web Sémantique

Ces dernières années une nouvelle tendance s'est affirmée : celle du Web Sémantique, aussi appelé Web 3.0. Cette nouvelle version du Web forme le nouvel environnement dans lequel vos ressources numériques seront exploitées. Elles vivent désormais dans un monde où sont connectés entre eux des éléments de connaissance, et non plus simplement dans un réseau d'informations.

Pour le dire rapidement, vos ressources numériques étaient hier connectées par des liens simples et aveugles, alors qu'aujourd'hui ces liens avec le réseau peuvent avoir un sens explicite. En d'autres termes, l'hyperlien est devenu sémantique.

D'un point de vue plus technique, le Web Sémantique (qui n'est en fait qu'une partie du Web 3.0) est « le Web des données qui portent un sens, c'est-à-dire des données qu'un programme informatique peut traiter en fonction du peu qu'il peut en apprendre sur leur sens. »⁵. Il fournit « un cadre commun qui permet le partage et la réutilisation des données sans tenir compte des frontières entre les mondes des applications, des entreprises et des communautés. Il s'agit d'un effort collaboratif conduit par le World Wide Web Consortium (W3C) et avec la participation d'un grand nombre de chercheurs et de partenaires industriels. Il s'appuie sur le Resource Description Framework (RDF) qui intègre une variété d'applications utilisant le XML pour la syntaxe et les URIs pour la nomenclature. Il a été proposé par l'inventeur du World Wide Web, Tim Berners-Lee. »⁶

1.3.3. Web des données ouvertes

Dans le monde du Web Sémantique, une nouvelle « philosophie » est en passe de devenir une référence qu'il est bon de connaître dès lors qu'on souhaite lier ses ressources numériques avec celles qui sont déjà disponibles en ligne. Cette initiative s'appelle le Web des données ouvertes, ou Linked Open Data (LOD). Pour poursuivre ses ambitions définies dans le plan stratégique 2011-2015, Europeana considère que LOD est essentiel au succès de sa politique culturelle⁷.

En effet pour Europeana, le Web des données ouvertes est :

- Une technologie qui permet d'articuler tous les éléments d'information provenant des fournisseurs de données.
- Un moyen de partager ces données avec d'autres parties.
- Une manière d'offrir aux utilisateurs la meilleure expérience de recherche possible.

D'un point de vue général, LOD participe à l'évolution du Web qui passe ainsi d'une liste à plat de données à un accès structuré à toutes les ressources disponibles. Si vous conformez vos propres données aux normes du LOD, vous serez en mesure de les rendre visibles sur des applications mobiles, et de bénéficier des avantages du nuage complet d'URIs dans lequel sont d'ores et déjà mises en réseau des ressources de référence comme DBPedia. Par exemple, si vous mappez à DBPedia vos données normalisées selon les LOD, vous êtes par conséquent sûr(e) de les mapper dans le même temps avec toutes les autres ressources de référence qui existent.

Le Web des données ouvertes propose aux fournisseurs de contenus (les musées par exemple) un jeu de règles, d'outils et de recommandations. De tout ce jeu vous pouvez dans un premier temps garder en mémoire que toutes les données que vous souhaitez fournir à Europeana doivent être nommées et liées. Pour cela, les recommandations de ce livret vous aideront à accomplir les actions nécessaires avant l'envoi de vos données sur la plateforme technique d'Europeana.

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

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

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

1.3.4. LIDO

Parmi tous les standards de modèles de données qui existent, nous recommandons particulièrement LIDO (Light Information Describing Objects) aux musées européens. Il y a à cela quatre raisons principales.

Tout d'abord, ce modèle de données a été spécifiquement défini pour les musées par le lot de travail WP3 d'Athena.

Il s'agit d'un mélange d'éléments en provenance de Spectrum, MuseumDat et DC, et qui ainsi tient compte des spécificités de votre situation.

Ensuite il est important de noter que LIDO est déjà mappé au modèle de données d'Europeana, ESE (Europeana Semantic Elements), et qu'il est disponible sur la plateforme d'ingestion (Athena Ingester). Du coup si votre modèle de données est mappé avec LIDO, vous n'avez pas à vous préoccuper de sa compatibilité avec la version actuelle d'Europeana.

De plus, LIDO offre plus de possibilités que Dublin Core pour décrire efficacement vos objets numériques puisqu'il est conçu comme un ensemble de classes regroupant des champs. Ces classes sont : Object Classifications (classifications d'objets), Object Identifications (Identifications d'objets), Events (événements), Relations, Administrative Metadata (métadonnées administratives). Bien mieux qu'une description linéaire et sans hiérarchie (à l'instar de Dublin Core), une description d'objet qui est organisée en classes structurées (comme c'est le cas pour LIDO) permet une meilleure exploitation de cette description par un être humain aussi bien que par une machine (moteur de recherche, base de données).

Enfin, LIDO et ses classes seront plus faciles à mapper avec le prochain modèle de données d'Europeana. En effet Europeana est en train de finaliser un nouveau modèle de données, EDM (Europeana Data Model) qui progressivement remplacera ESE. EDM offre une structure de classes proche de celle de LIDO et parfaitement compatible avec le Web des données ouvertes. Si aujourd'hui vous utilisez déjà LIDO pour être compatible avec ESE, demain la transition avec EDM sera facile à opérer.

1.3.5. Formats

Pour faire partie du nuage des données ouvertes et utiliser les technologies du Web Sémantique, la terminologie d'une institution doit être exprimée en un format compatible. Dès que vous voulez représenter ou modéliser votre terminologie, et l'exploiter sur le Web, vous devez utiliser un format standard. Les plus communément utilisés sont SKOS, OWL, RDF, et XML. Certains d'entre eux peuvent être associés, certains peuvent également être encapsulés par d'autres. L'utilisation d'un format standard aura pour résultat de rendre les métadonnées que vous avez exprimées avec votre terminologie, effectivement représentées de façon à ce que les technologies du Web puissent les reconnaître et les interpréter. Ci-après se trouvent de brèves descriptions de ces formats standards et qui ont été rédigées dans le but de rendre leurs liens plus intelligibles.

XML :

XML (Extensible Markup Language) consiste en un ensemble de règles pour encoder tout document dans une forme lisible par une machine. Il est défini dans le document de spécification XML 1.0 produit par le W3C, ainsi que dans d'autres spécifications en lien avec celle-ci, toutes libres d'utilisation.

Le XML a été conçu pour mettre en avant la simplicité, la généralité et la facilité d'usage sur Internet. C'est un format de données textuelles s'appuyant fortement sur l'Unicode pour gérer les différentes langues et écritures du monde. Bien qu'il ait été conçu spécifiquement pour le document, le XML est largement utilisé pour représenter des structures de données arbitraires, par exemple celles des services Web. Il existe de nombreuses interfaces de programmation que les développeurs de logiciels peuvent utiliser pour accéder aux données XML, ainsi que plusieurs systèmes graphiques pour aider à définir schématiquement des langages s'appuyant sur le XML.

RDF :

Le Resource Description Framework (RDF) est une famille de spécifications du W3C qui avait été conçue à l'origine comme un modèle de données pour métadonnées. Il en est venu à être utilisé comme une méthode générale pour la description conceptuelle ou la modélisation d'information, et qu'on retrouve dans les ressources Web, utilisant pour ce faire divers formats de syntaxe.

Le modèle de données RDF est conçu pour déclarer des énoncés sur les ressources (en particulier celles du Web) sous la forme de triplets. Les triplets sont le moyen d'exprimer

ces énoncés qui se présentent comme des expressions de type sujet-prédicat-objet. Le sujet y indique la ressource, et le prédicat indique quant à lui les traits ou aspects de la ressource tout en exprimant une relation entre le sujet et l'objet.

La spécification en RDF est à base de XML.

OWL :

Le Web Ontology Language (OWL) est une famille de langages de représentation de connaissances pour la création d'ontologies. Les langages en question sont caractérisés par de la sémantique formelle et des arrangements sous formes de séries de connaissances, exprimés en RDF à base d'XML, et destinés au Web Sémantique. OWL est soutenu par la W3C et a attiré l'intérêt des mondes universitaire, médical et commercial. En octobre 2007, un nouveau groupe de travail du W3C a commencé d'étendre OWL à de nouvelles possibilités telles que proposées dans la soumission OWL 1.1. Cette nouvelle version appelée OWL 2 se trouva vite prise en compte par des logiciels d'édition sémantique tels que Protégé et de contrôle d'inférence comme Pellet, RacerPro and FaCT++. Le W3C a annoncé le 27 octobre 2009 une nouvelle version à venir.

La famille OWL contient plusieurs types, sérialisations, syntaxes et spécifications portant des noms qui se ressemblent. A moins d'adopter une approche stricte et poussée, des confusions sont possibles. OWL et OWL2 seront utilisés pour faire respectivement référence aux spécifications de 2004 et 2009. Les noms complets de types seront utilisés, incluant la version de spécification (par exemple OWL2 EL).

Dans les cas où la référence sera plus générale, l'expression « famille OWL » sera plus adéquate.
OWL s'appuie sur la spécification RDF.

SKOS :

Parmi les formats présentés dans cette partie, SKOS est de plus en plus demandé par les services Web. Europeana par exemple a décidé de formater en SKOS toutes les métadonnées moissonnées par leur service pour fournir une exploitation homogène et efficace des ressources cataloguées, des données présentes et de leurs descriptions. SKOS s'appuie sur la spécification RDF et permet de migrer vers des ontologies OWL. SKOS n'est pas à strictement parler un langage de représentation de connaissances formelles, puisque une connaissance formelle est littéralement exprimée sous la forme de groupes d'axiomes et de faits qui composent les caractéristiques principales d'une ontologie formelle. SKOS est plutôt utilisé pour modéliser des vocabulaires contrôlés tels que les thésaurus ou les classifications qui sont d'une nature différente de celle de l'ontologie. Les idées et les significations qui sont décrites par les thésaurus et les autres types de terminologie sont souvent considérées comme des « concepts » même si d'un point de vue ontologique un concept est défini autrement.

2. Objectif et structure des recommandations

2.1 Optimisation et compatibilité

Comme ce fut indiqué dans l'introduction, nos recommandations adoptent le point de vue de ceux qui en sont les destinataires, c'est-à-dire que sont pris en compte le bagage technologique, les objectifs et les intérêts propres aux musées.

Tout d'abord, en ce qui concerne le bagage technologique, nous avons cherché jusqu'ici à exprimer nos recommandations d'une manière qui soit intelligible par des non- experts du domaine. Dans la suite de cette publication nous continuerons de formuler les choses de la même manière pour que vous puissiez comprendre et appliquer ce qui vous est proposé.

Ensuite, pour ce qui est des objectifs muséaux, nos recommandations doivent vous permettre d'être compatible avec les contraintes et attentes d'Europeana. Ceci dit cet objectif est vraiment minimal puisque l'exigence principale ne consiste qu'à rendre vos données compatibles avec les standards du Web Sémantique, afin de correspondre aux contraintes du portail en vue de l'exploitation sémantique des descriptions de vos ressources numériques. Si bien que vous pourriez utiliser OWL (Web Ontology Language), qui est actuellement le langage le plus formel et le plus complexe pour une exploitation de type Web

Sémantique, afin de formater vos données de manière à ce qu'elles soient interopérables, mais à vrai dire il n'est pas sûr que ce soit nécessaire ni pertinent dans votre cas. SKOS (Simple Knowledge Organisation System) est une solution plus « économique » puisque ses fonctionnalités couvrent la plupart de vos besoins et son utilisation n'exige pas autant de compétences techniques coûteuses que celle de OWL. C'est pourquoi au final nous avons tenu à défendre vos intérêts en écrivant ces recommandations. En effet dans le contexte que nous décrivons vous pouvez faire plus que SKOSifier votre terminologie sans dépenser trop d'argent et de temps. Il y a plusieurs opérations « simples » que vous pouvez appliquer à votre terminologie et qui amélioreront certainement l'exploitation sémantique de vos ressources numériques sur Europeana, et ce aujourd'hui même, mais surtout à l'avenir. Dans cet esprit, toutes nos recommandations doivent vous aider à optimiser la prédisposition de vos données à être efficacement trouvables par un moteur de recherche sémantique. Pour information, Europeana a pour l'heure développé un prototype de moteur sémantique qui sera dans le futur intégré dans l'interface utilisateur du portail.

2.2 Approche en trois étapes

Nous avons structuré nos recommandations en trois étapes afin de simplifier leur présentation et leur compréhension.

Chacune de ces étapes apporte des éléments pour rendre votre terminologie compatible avec les exigences d'Europeana et optimisée pour son exploitation sémantique. Même si elles sont présentées sous la forme d'une séquence linéaire, nous vous recommandons fortement de ne pas hésiter à les suivre de manière itérative voire « anarchique » si nécessaire. Ne restez pas bloqué(e) à une étape trop longtemps, et de manière générale essayez de rester actif(ve) de quelque façon que ce soit.

La première étape concerne la conception de votre terminologie. Pour le dire ainsi, c'est comme si à ce point vous gériez votre terminologie de manière « interne » (à votre institution) en vue de fabriquer un thésaurus pour un usage « humain ». Nous vous présenterons différentes opérations que vous pouvez réaliser par vous-même afin de construire une nouvelle terminologie ou d'adapter celle que vous utilisez déjà, opérations qui optimiseront les descriptions de vos ressources numériques sur Europeana. Ces opérations doivent être réalisées en priorité car elles déterminent les deux autres grandes étapes.

La deuxième étape quant à elle consiste à rendre votre terminologie interopérable. A ce point de la démarche, en la SKOSifiant pour qu'elle soit efficacement exploitable par la « machine », c'est comme si vous tourniez votre terminologie vers le monde « environnant » votre musée. Il s'agit ici de la partie spécifiquement consacrée à la mise en relation de vos descriptions sémantiques avec le modèle de données. Pour l'heure SKOS est fortement demandé par Europeana. C'est pourquoi nous avons mis l'accent sur ce format en particulier.

Enfin les dernières recommandations que nous vous adressons portent sur la mise en réseau de votre terminologie avec d'autres. Lors de cette dernière étape, votre préoccupation consiste à être visible en Europe dans une logique de réseau en intégrant votre terminologie dans un réseau déjà constitué de thésaurus SKOSifiés. Il s'agira là de la troisième et dernière étape de nos recommandations.

3. Concevoir sa terminologie

La conception de votre terminologie est au fondement de tout le reste. Elle détermine les opérations que vous devrez accomplir ensuite quand vous rendrez votre terminologie interopérable avec d'autres ressources, et quand vous l'intègrerez à un réseau de terminologies.

3.1 Méthodologie

Voici une liste de tâches que vous pouvez suivre pas à pas si vous devez concevoir ex nihilo la terminologie qui servira à la description de vos ressources numériques. Si vous avez déjà une terminologie « maison », nous vous recommandons de la vérifier et de l'améliorer si besoin. En ce sens vous pouvez utiliser notre liste de tâches comme un guide pour évaluer la qualité de votre terminologie. Même si cette liste est présentée comme une séquence linéaire de tâches, il est très important de l'utiliser de manière itérative. Un travail comme celui-ci demande un peu de patience et de temps, ce que vous devriez voir comme un investissement. Plus votre terminologie est précisément définie, plus votre retour sur investissement sera important. N'hésitez donc pas à passer d'une partie à une autre de ces recommandations, à les utiliser par itérations.

Les différentes tâches que nous détaillerons par la suite sont :

A1 : Définissez le(s) domaine(s) de votre collection

A2 : Identifiez les attentes de vos utilisateurs

A3 : Définissez votre lien avec le modèle de données

A4 : Choisissez les termes pour la description sémantique de vos ressources numériques

A5 : Organisez vos termes selon une structure de thésaurus

A6 : Trouvez des termes équivalents dans d'autres langues

A7 : Produisez techniquement votre thésaurus

A1

DÉFINISSEZ LE(S) DOMAINE(S) DE VOTRE COLLECTION

Actions

Avant toute chose, définissez le(s) domaine(s) de votre collection en répondant aux questions suivantes :

- Existe-t-il un domaine général auquel votre collection d'objets appartient dans sa totalité ? (ex : archéologie, art, science...)
- Pouvez-vous répartir vos objets en différents sous-domaines spécifiques ? (ex. pour le domaine général « art » : « peintures », « sculptures », « cinéma », « littérature »)

Objectifs

L'objectif de cette première étape est de préparer le choix de vos termes de description (étape A4 : Choisissez les termes pour la description sémantique de vos ressources numériques). Plus vos domaines sont spécifiques, plus vos termes seront précis et dépourvus d'ambiguïté.

Privilégiez dès cette étape une spécialisation par domaines précis, et plus tard créez des ponts entre tous les thésaurus spécialisés que vous aurez (étape A5 : Organisez vos termes selon une structure de thésaurus). Plutôt que d'essayer de construire un seul gros thésaurus pour tous les domaines à couvrir, nous vous recommandons de conserver et d'enrichir vos thésaurus spécifiques à des domaines donnés sans les étendre à d'autres domaines. Il est plus judicieux d'ajouter de nouveaux thésaurus pour couvrir de nouveaux domaines, et d'établir ensuite des ponts entre les thésaurus puisqu'il s'avère que le dispositif de recherche d'Europeana s'appuie sur un traitement des descriptions transversal aux domaines. Un « pont » dans ce contexte consiste à mapper entre eux des termes en provenance de micro-thésaurus à l'aide de relations.

Vous pouvez considérer que cette tâche est terminée quand, après la tâche A4 : Choisissez les termes..., ne se trouvent plus dans vos listes de descripteurs de termes ambigus susceptibles d'appartenir à plusieurs domaines distincts.

DÉFINISSEZ LE(S) DOMAINE(S) DE VOTRE COLLECTION

Exemple

Si vous envisagez de décrire un orgue en tant qu'instrument de musique, et de construire une terminologie sur la musicologie, et si en plus de cela vous cherchez à décrire un orgue en tant que meuble religieux, référez à différents micro-thésaurus sur « les instruments de musique » d'un côté, et « les meubles religieux » de l'autre, au lieu de mélanger les termes de ces domaines différents. Si bien que vous aurez au final à prendre en compte au moins deux domaines pour votre collection : « musicologie » et « religion ». Et au moins deux sous-domaines : « instruments de musique » et « meubles religieux ».

Méthodes et outils

8 http://incipiainfodoc.archimed.fr/Idesia/home.aspx?INSTANCE=MIMO&THES=IFD_MIMO_CLASSIF&VIEW=DEFAULT&FORM=0&ACTIVE=TRUE

Au moment de commencer à définir votre domaine général, vous pouvez consulter HEREIN et MICHAEL qui proposent une très large typologie. Ensuite, pour aller plus en profondeur dans la définition, vous pouvez voir comment le projet MIMO a structuré plusieurs sous-domaines appartenant à son domaine spécifique du vocabulaire musical⁸.

IDENTIFIEZ LES ATTENTES DE VOS UTILISATEURS

Actions

Pour tous les utilisateurs de vos descriptions sémantiques, identifiez les attentes qu'ils peuvent avoir en les utilisant. Pour cela vous pouvez répondre aux questions suivantes :

- Quels types de personnes sont susceptibles d'utiliser vos descriptions sémantiques ? (ex : amateurs d'art, universitaires spécialistes, étudiants en sciences)
- Quelles sont leurs motivations principales à consulter vos collections ? (ex : divertissement, recherche)
- Ces personnes s'attendent-elles à des descriptions expertes ? Quels termes utilisent-elles lorsqu'elles interrogent vos domaines de collection ?
- En fonction de ces attentes, quel type de licence êtes-vous prêt(e) à leur accorder ? Êtes-vous d'accord pour autoriser un usage professionnel libre de droits du thésaurus que vous êtes en train de construire ?

Objectifs

L'objectif de cette étape est de préparer le choix de vos termes de description (étape A4 : Choisissez les termes pour la description sémantique de vos ressources numériques). L'idée ici est de comprendre quels termes les utilisateurs emploieront spontanément lors d'une requête sur le Web par moteur de recherche, et quels autres termes il leur faudra utiliser s'ils veulent être plus précis.

Concevez votre terminologie en la pensant d'abord pour les utilisateurs les plus généraux afin qu'elle corresponde à leurs attentes. Puisque Europeana est un portail d'accès aux objets et aux données des collections, plus le point de vue de ceux qui consultent est pris en compte en amont, et plus le portail sera efficace. La plupart du temps, les requêtes ne sont pas exprimées par des professionnels, mais par le grand-public. Cela veut dire que le nuage de données d'Europeana doit correspondre à ce que le grand public en attend d'un point de vue « fonctionnel ».

IDENTIFIEZ LES ATTENTES DE VOS UTILISATEURS

Si le thésaurus que vous construisez est à son origine conçu dans cette perspective-là, nous pouvons imaginer qu'il permettra au portail de fournir quelques résultats pertinents.

C'est pourquoi nous vous recommandons de concevoir vos thesaurus en tenant compte des compétences, habitudes et attentes du grand public aussi bien que des professionnels. Ce qui signifie que deux approches peuvent être adoptées dans leur complémentarité :

1/ l'approche « bottom-up » consiste à partir des besoins et habitudes des professionnels pour définir la terminologie ;

2/ l'approche « top-down » consiste au contraire à s'intéresser en priorité aux particularités du grand public quant aux questions d'accès et de recherche. En ce qui concerne la licence d'utilisation de votre terminologie, vous devez connaître le cadre légal avant de faire votre choix. Quel type d'utilisation votre institution a l'habitude d'autoriser et sous quelles conditions ? Il est important de traiter la question dès maintenant même si le choix de licence effectif ne se fera qu'à l'étape C1 : Définissez les métadonnées de votre terminologie.

Vous pouvez reprendre cette étape et la considérer comme achevée quand, après l'étape A4 : Choisissez les termes..., un groupe d'utilisateurs test peut dire précisément pour chaque terme choisi à quelle type de requête il répond, ou quand l'analyse des mots-clés de requête utilisés vous conforte dans le choix de vos termes.

IDENTIFIEZ LES ATTENTES DE VOS UTILISATEURS

Exemple

Dans le cas d'une collection d'éléments d'archéologie sous-marine, vous avez pu identifier chez vos utilisateurs au moins deux types d'attentes différents :

- Des requêtes générales à propos de découvertes spectaculaires (ex : épaves de bateaux, ruines antiques ayant sombré au fond d'une mer) : pour connaître les lieux, les dates de découverte, l'âge de ce qui a été trouvé, les personnes ayant fait les découvertes.
- Des recherches scientifiques très poussées concernant les enquêtes menées : pour connaître le contexte des missions scientifiques, les protocoles de découverte et de datation, les hypothèses et arguments.

Vous connaissez donc mieux quel type d'information doit fournir votre terminologie. Du coup, puisque vous considérez que l'utilisation de vos descripteurs pour décrire d'autres collections peut améliorer la visibilité de vos propres ressources numériques, vous pouvez décider de mettre votre terminologie sous licence Creative Commons CC-BY:-Share alike⁹ : l'utilisateur peut se servir de vos descripteurs sans aucune modification et en mentionnant votre institution comme auteur de la terminologie.

⁹ <http://creativecommons.org/licenses/>

Méthodes et outils

Afin de mieux connaître les attentes des utilisateurs, vous pouvez réaliser au moins deux études simples et croiser leurs résultats :

- D'abord une étude questionnant les visiteurs de vos collections matérielles et de vos expositions (ex : demandez-leur s'ils seraient intéressés par des outils virtuels à utiliser dans le musée ou en ligne à partir de n'importe quel point d'accès ; et aux intéressés, demandez quel type de requêtes ils formuleraient)
- Une autre étude réalisée à partir des données statistiques du site Web de votre institution (ex : quel terme en particulier est souvent utilisé pour accéder aux pages de la collection)

IDENTIFIEZ LES ATTENTES DE VOS UTILISATEURS

Concevez votre terminologie en la pensant d'abord pour les utilisateurs les plus généraux afin qu'elle corresponde à leurs attentes. Puisque Europeana est un portail d'accès aux objets et aux données des collections, plus le point de vue de ceux qui consultent est pris en compte en amont, et plus le portail sera efficace. La plupart du temps, les requêtes ne sont pas exprimées par des professionnels, mais par le grand-public. Cela

DÉFINISSEZ VOTRE LIEN AVEC LE MODÈLE DE DONNÉES

Actions

Comme nous l'avons expliqué plus haut, vous devez gérer le lien qui unit votre terminologie à votre modèle de données. Pour définir ce lien, vous pouvez procéder en deux temps :

1/ pour chaque type de requête que vous avez identifié à l'étape A2 : Identifiez les attentes de vos utilisateurs, déduisez maintenant quel genre d'information vos descripteurs doivent couvrir pour proposer à l'utilisateur une réponse pertinente.

2/ dans votre modèle de données, notez les champs de description qui correspondent aux genres d'information dont ont besoin vos utilisateurs.

Objectifs

L'objectif est de connecter votre terminologie au modèle de données en mappant d'un côté le modèle de la terminologie qui est défini dans le modèle de données, avec de l'autre côté le modèle nécessaire à la satisfaction de vos utilisateurs et de leurs attentes. En effet, le premier enjeu consiste à définir tous les types d'information sémantique qu'un moteur de recherche serait susceptible d'exploiter pour fournir des réponses pertinentes aux requêtes que formulent les utilisateurs qui souhaitent en savoir plus sur le contenu d'une ressource numérique. Plus le modèle de votre terminologie couvre les requêtes possibles, plus les résultats de ces requêtes seront pertinents et fiables. Et le deuxième enjeu consiste à mapper ce modèle de terminologie avec celui défini dans le modèle de données. A ce point nous vous invitons à ne pas trop vous contraindre en anticipant trop sur les possibilités et les limites de SKOS. Par exemple, SKOS n'est pas conçu pour une description complète de personnes, si bien qu'une information comme la date de naissance ou de décès serait difficile à modéliser avec les caractéristiques cœur de SKOS.

DÉFINISSEZ VOTRE LIEN AVEC LE MODÈLE DE DONNÉES

Quoi qu'il en soit, si les requêtes de vos utilisateurs exigent ces dates, prévoyez d'avoir une liste à part pour ce type d'information. Vous verrez plus bas (étape B1 : Évaluez la compatibilité de SKOS avec les caractéristiques de votre terminologie) comment gérer ce cas particulier. Vous pouvez considérer que cette tâche est achevée quand chaque type d'information qu'il faut fournir a un champ correspondant dans le modèle de terminologie qui est lui défini dans le modèle de données.

Exemple

Dans le contexte de « l'archéologie sous-marine », si les utilisateurs s'attendent à obtenir, grâce aux descriptions sémantiques de vos objets, les lieux et les dates de découverte, et les protocoles de découverte de ce qui a été trouvé, votre modèle de terminologie dans le modèle de données LIDO doit utiliser les champs « Place Information », « Date Information », « Acquisition Information » ; ainsi pour ce genre de requêtes toutes les ressources numériques proposées le seront précisément avec des résultats sémantiques fiables.

Méthodes et outils

Pour vous aider à réaliser ce lien plus facilement, un tableau de mise en relation (« mapping sheet ») est disponible en annexe. Parmi les normes ISO existantes, vous pouvez consulter BS8723: Structured Vocabularies for Information Retrieval, et suivre également la norme ISO 25964-1 Thesauri and interoperability with other vocabularies: Thesaurus for information retrieval (qui a été publiée en 2011) parce qu'elle exprime plus clairement d'une part le lien entre la gestion des collections et les vocabulaires, et d'autre part la mise en œuvre technique de thésaurus avec SKOS. Il peut être aussi utile de noter que l'outil xTree développé par Digicult en Allemagne tient compte de cette nouvelle norme bien que vous ne puissiez utiliser l'outil que si votre terminologie est déjà SKOSifiée.

CHOISISSEZ LES TERMES POUR LA DESCRIPTION DE VOS RESSOURCES

Actions

Choisissez vos termes pour chaque sous-domaine que vous avez défini précédemment (étape A1 : Définissez le(s) domaine(s) de votre collection), et plus loin pour chaque champ de votre modèle de terminologie. Dans chaque cas, essayez d'emprunter une voie intermédiaire entre la haute précision d'un vocabulaire expert / professionnel et l'utilisation générale faite par les utilisateurs que vous avez identifiés (étape A2 : Identifiez les attentes de vos utilisateurs).

1. Avant toute chose, pour chaque champ que vous avez sélectionné dans le modèle de données, jetez un œil aux vocabulaires qui sont proposés soit par défaut dans le modèle de données, soit référencés dans un répertoire que vous connaissez déjà et pouvez consulter. Plus précisément, évaluez leurs pertinences respectives au regard des attentes de vos utilisateurs, et voyez s'il est intéressant pour vous d'en utiliser certains, voire de les modifier (en particulier, vérifiez sous quelles conditions les licences vous le permettent). Suite à cela vous pouvez décider :

- a/ quels vocabulaires vous utiliserez directement sans modification ;
- b/ quels vocabulaires vous pensez utiliser mais seulement après les avoir modifiés et adaptés.

2. Puis, dans le cas où vous souhaitez (et pouvez) modifier des vocabulaires existants pour ensuite les utiliser, suivez les procédures de modification et d'adaptation relatives à tous ces vocabulaires (n'oubliez pas de vérifiez les droits d'utilisation).

3. Enfin, pour chaque champ n'ayant pas encore de termes suite à ces deux premières actions, choisissez-les librement :

CHOISISSEZ LES TERMES POUR LA DESCRIPTION DE VOS RESSOURCES

a/ soit en trouvant le terme qui à la fois fasse autorité pour les experts et corresponde aussi à ce que le grand public attend ;

b/ soit, si vous ne pouvez trouver le compromis unique, en utilisant comme synonymes deux termes différents. Vous définirez leur relation de synonymie plus tard à l'étape B3 : Définissez précisément les labels exprimant les concepts (ex : le terme expert sera alors le descripteur ou le terme préféré, et le terme général en sera une alternative)

4. Cas particuliers :

a/ pour plus de précision, dans le cas des termes composés, essayez autant que possible d'avoir un terme « racine » avec lequel les termes composés seront liés. Un mot composé est un terme composé de deux mots ou plus ; « littérature comparée » est par exemple un terme composé. Il sera préférable d'avoir un terme « littérature » qui en sera le terme parent (ou « plus large »).

b/ N'évitez pas forcément les termes qui vous paraissent avoir des formes obsolètes ou oubliées, ils pourraient eux aussi être utilisés dans une requête, et de plus il est important de garder une trace de l'histoire de votre vocabulaire.

Objectifs

L'objectif est de trouver le meilleur ensemble de termes qui serviront à la description sémantique de vos ressources numériques. Plus votre ensemble contient des termes experts utilisables par le grand public, plus vos descriptions seront utiles et pertinentes et faciliteront le traitement par un moteur de recherche de vos ressources.

En utilisant des vocabulaires de référence comme les thésaurus du Getty, vous aiderez Europeana à exploiter vos descriptions et à améliorer par conséquent l'accès à vos ressources. Cependant vous devez vérifier lesquels

CHOISISSEZ LES TERMES POUR LA DESCRIPTION DE VOS RESSOURCES

de ces vocabulaires sont réellement pertinents au vu de vos besoins. Certains sont libres d'utilisation, voire modifiables, mais d'autres ne le sont pas. Un contrôle au cas par cas est en fait nécessaire.

Si vous décidez d'utiliser un vocabulaire de référence en le modifiant, vous devez suivre la procédure officielle permettant son adaptation. Pour tous les cas où vous ne trouvez pas de vocabulaire pertinent au vu de vos besoins, nous vous invitons à créer vous-mêmes les termes en suivant les normes existantes dédiées à la conception de thésaurus (voir les références de normes données à l'étape A7). Si vous le faites, n'oubliez pas que les formes obsolètes ou oubliées peuvent être utilisées dans des requêtes par mots-clefs. Il peut donc être intéressant de les avoir dans votre ensemble de termes même si finalement vous ne les rendrez pas forcément visibles (cela se décide aux étapes A5 : Organisez vos termes selon une structure de thésaurus et B3 : Définissez précisément les labels exprimant les concepts).

Vous pouvez considérer que cette tâche est achevée quand un groupe d'utilisateurs test peut dire pour chaque terme à quoi il se réfère et en vérifiant comment leurs requêtes test sont pleinement satisfaites.

qui a été trouvé, votre modèle de terminologie dans le modèle de données LIDO doit utiliser les champs « Place Information », « Date Information », « Acquisition Information » ; ainsi pour ce genre de requêtes toutes les ressources numériques proposées le seront précisément avec des résultats sémantiques fiables.

Exemple

Vous utilisez LIDO comme modèle de données et vous y avez identifié les champs de description « Place Information » et « Acquisition Information » comme indispensables pour couvrir vos besoins de descriptions des lieux et protocoles de découvertes sous-marines.

CHOISISSEZ LES TERMES POUR LA DESCRIPTION DE VOS RESSOURCES

Vous arrivez donc au point de choisir les termes afférents à ces deux champs.

Quand vous utilisez LIDO, vous avez la possibilité d'utiliser les vocabulaires du Getty. Et parmi eux, vous remarquez celui qui traite des noms géographiques (TGN: Thesaurus of Geographic Names) et qui pourrait être pertinent dans votre cas. Après y avoir jeté un œil, vous décidez effectivement de l'utiliser en vue de décrire les lieux où ont été découverts les objets de votre collection d'archéologie sous-marine.

Par ailleurs vous ne trouvez pas parmi les vocabulaires proposés par défaut avec LIDO, un qui soit pertinent pour décrire les protocoles scientifiques de découvertes. Par conséquent vous décidez pour ce champ de créer par votre propre liste de termes : « Analyse d'archive », « Exploration précise d'une zone sous-marine », « Hasard »...

Méthodes et outils

Si vous cherchez des vocabulaires de référence, nous vous conseillons de chercher dans l'inventaire de ressources réalisé par le lot de travail WP4 d'Athena à l'adresse : http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources
Vous pouvez aussi accéder directement aux vocabulaires du Getty : <http://www.getty.edu/research/tools/vocabularies/index.html>

ORGANISEZ VOS TERMES SELON UNE STRUCTURE DE THESAURUS

Actions

Structurez votre ensemble de termes :

1. En créant un micro-thésaurus (liste de termes) pour chaque domaine ou sous-domaine que vous avez identifié à l'étape A1 : Définissez le(s) domaine(s) de votre collection.
2. En rassemblant certains termes de ces micro-thésaurus sans qu'ils appartiennent forcément aux mêmes domaines, mais en privilégiant une approche thématique et transversale aux domaines.
3. En reliant les listes de termes sur le modèle d'un réseau de micro-thésaurus à l'aide des relations : « parent » (Broader), « enfant » (Narrower), « associé à » (Related to).

Objectifs

L'objectif est de transformer vos listes de termes en thésaurus, c'est-à-dire en un réseau structuré de micro-thésaurus. Car en effet, parmi tous les types de terminologies qui existent, nous recommandons aux musées qui souhaitent rendre leurs collections visibles sur Europeana d'utiliser des thésaurus. Premièrement, ce type de terminologie est assez facile à SKOSifier puisque le format SKOS a été prévu – voire spécialement conçu – pour le traitement de thésaurus, afin d'être techniquement compatible avec les contraintes techniques imposées par la procédure d'ingestion sur la plateforme Europeana. Si on le compare ensuite aux autres types de terminologies, le thésaurus se présente comme un bon compromis entre terminologie riche et terminologie facile à utiliser. De plus, en règle générale les musées utilisent déjà ce type de terminologie, plutôt que des ontologies ou des classifications¹⁰ pour décrire leurs collections d'une manière bien structurée (47% des 149 terminologies que nous avons répertoriées dans notre étude initiale sont des thésaurus¹¹). Sans compter sur le fait que les relations ordinaires d'équivalence et d'association de termes à la base

¹⁰ Pour avoir un aperçu des types de terminologies, jetez un œil au D4.1, partie 3.4 «Types of terminology resources», ou en ligne à l'adresse : http://www.athenaeurope.org/athenawiki/index.php/Terminology_management#Types_of_terminology_resources

¹¹ cf. les résultats du D4.1

A5 ORGANISEZ VOS TERMES SELON UNE STRUCTURE DE THESAURUS

des thésaurus sont particulièrement pertinentes pour gérer le multilinguisme. Par conséquent, même si Europeana se penche actuellement sur les ontologies, nous continuons de penser qu'il est préférable pour les musées d'utiliser des thésaurus qui sont un bon compromis entre les vocabulaires contrôlés (spécialement appréciés dans les musées) et les ontologies (qui sont plus puissantes pour répondre aux requêtes des moteurs de recherche).

Un thésaurus peut être défini comme « un réseau de vocabulaires contrôlés ». Les thésaurus permettent de connecter des termes à l'aide de plusieurs types de relations qui peuvent être hiérarchiques, associatives, d'équivalence ou de définition. Cela signifie qu'un thésaurus ajoute des relations d'association à des relations parents-enfants. Une relation parent-enfant est exprimée par une fonction Broader Term (BT) / Narrower Term (NT). Les relations d'association dans un thésaurus du type Related Term (RT) (ex : le terme A est associé au terme B) sont employées pour exprimer des relations qui ne sont ni hiérarchiques ni d'équivalence. L'équivalence est exprimée à l'aide de la fonction de relation USE (ex : terme préféré) / Used For (UF) (ex : terme non préféré). Toute information additionnelle comme une définition ou une remarque peut être fournie dans une Scope Note (SN). La relation d'équivalence est tout spécialement utile dans le cas des thésaurus multilingues.

Dans la mesure où la structure d'un thésaurus s'appuie sur des liens de hiérarchie et d'association entre micro-thésaurus, nous vous recommandons de multiplier les liens afin d'améliorer l'exploitation ultérieure de votre terminologie. Plus vous avez de relations entre vos listes de termes et plus un moteur de recherche sera

A5 ORGANISEZ VOS TERMES SELON UNE STRUCTURE DE THESAURUS

capable d'en tirer des résultats efficaces. Dit autrement, nous vous conseillons de tirer parti au maximum de ce que permet la structure d'un thésaurus. Si vous observez de manière conjointe les deux approches transversale et hiérarchique qu'autorise cette structure, la SKOSification à venir de votre terminologie en sera grandement facilitée. SKOS vous permettra de gérer aussi bien une arborescence de listes de termes liés à différents concepts (ConceptScheme tree), qu'un ensemble non hiérarchisé de groupes de termes venant de ces mêmes listes mais non exclusifs les uns des autres (Collections for thematic grouping).

Vous pouvez considérer que cette tâche est achevée quand il n'existe plus aucun terme isolé de toute liste et quand chaque liste de termes est liée à au moins une autre liste.

Exemple

Vous avez un ensemble de termes pour décrire des objets d'architecture. Parmi tous les descripteurs de votre thésaurus se trouvent « monument » et « habitation ». Pour chacun de ces deux descripteurs vous créez un micro-thésaurus spécifique. Dans celui consacré aux monuments se trouvent plusieurs termes dont « palais » qui est un type de monument. Et dans le micro-thésaurus portant sur les habitations se trouvent plusieurs termes dont « appartement » qui est un type d'habitation. Les termes comme « monument », « habitation », « appartement » et « palais » sont donc liés entre eux par des relations de hiérarchie. Mais vous pouvez proposer aussi un groupe transversal de termes portant sur la thématique « bâtiment ». Dans ce groupe pourraient alors figurer « palais » et « appartement » qui sans appartenir à la même hiérarchie de termes peuvent pourtant être connectés dans un même groupe thématique.

A5 ORGANISEZ VOS TERMES SELON UNE STRUCTURE DE THESAURUS

Méthodes et outils

A l'heure actuelle, il n'existe aucun outil dédié à la conception logique d'un thésaurus. Donc si votre propre outil de gestion de terminologies ne fournit pas ce service, vous aurez besoin de structurer votre thésaurus à l'aide des outils de bureautique les plus quotidiens (les tableurs informatiques par exemple).

Par ailleurs vous pouvez jeter un œil au GEMET, General Multilingual Environmental Thesaurus¹², qui présente différentes manières de parcourir un thésaurus : thématique, alphabétique ou hiérarchique.

¹² GEMET: <http://www.eionet.europa.eu/gemet>

A6 TROUVEZ DES TERMES ÉQUIVALENTS DANS D'AUTRES LANGUES

Actions

Suivant votre situation institutionnelle, votre terminologie doit être multilingue ou monolingue. Quelles que soient vos obligations concernant le multilinguisme, nous vous conseillons d'utiliser un thésaurus comprenant au minimum deux langues. Pour cela, après avoir choisi vos termes dans votre langue maternelle et structuré votre thésaurus, faites-en de même avec au moins une autre langue. En ce sens nous vous recommandons :

1. D'identifier les langues dans lesquelles vous voulez ou devez proposer des descriptions.
2. De ne pas procéder à une traduction littérale de vos termes dans les langues en question.
3. Mais de trouver, pour votre domaine de collection, des vocabulaires existant dans les langues qui vous intéressent et de repérer ceux que vous pouvez considérer comme équivalents au vôtre.
4. De demander à des experts du domaine qui parlent nativement les langues en question de vous aider à trouver les équivalences de termes qui vous manquent et de valider tous les choix.
5. De gérer provisoirement ces termes équivalents comme étant des termes simplement associés en attendant que vous précisiez leur relation multilingue à l'étape B3 : Définissez précisément les labels.

Objectifs

Nous vous recommandons très fortement de prévoir le multilinguisme dès cette étape de la conception de votre terminologie, même si cela n'est pas exigé dans le cas de votre institution. Certaines ont l'obligation légale de proposer des descriptions multilingues (ex : en deux langues minimum en Belgique), d'autres y sont contraints techniquement (ex : dans le cas des alphabets non latins comme les alphabets grec ou cyrillique). Mais pour toutes les autres institutions, bien qu'elles n'aient pas obligatoirement ce besoin politique

A6 TROUVEZ DES TERMES ÉQUIVALENTS DANS D'AUTRES LANGUES

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

ou technique, nous pouvons dire qu'elles ont en fait un besoin de visibilité. C'est pourquoi il serait profitable dans le cadre de la société d'information actuelle que chaque musée fournisse des descriptions multilingues (dans au moins deux langues). Dans le plan stratégique d'Europeana 2011-2015¹³ l'accent est mis sur le multilinguisme : Europeana est en train de développer un ensemble d'outils qui seront au service d'une interface d'accès multilingue, et il y va de votre intérêt de fournir au service des descriptions multilingues qui amélioreront la visibilité de vos collections.

L'objectif de cette étape est de trouver le meilleur ensemble de termes servant à la description sémantique de vos collections numériques dans au moins une autre langue, mais sans procéder à une traduction littérale. En effet la traduction directe et non experte produit des erreurs et des glissements de sens. Cela est dû parfois à la présence de « faux amis », d'autres fois au fait que le terme issu de votre langue maternelle est généralement utilisé à l'étranger aussi mais sans que vous le sachiez. Plus vous trouvez les termes équivalents par l'intermédiaire d'un expert du domaine parlant la langue étrangère qui vous intéresse, et plus vos équivalences seront exactes.

Ici apparaissent quelques difficultés comme d'utiliser une approche pair à pair à la place de la traduction directe ; de gérer des termes très spécialisés sans qu'il existe d'équivalent direct ; d'atteindre 100% de multilinguisme ; de vous empêcher autant que possible d'utiliser l'anglais comme langue pivot puisque l'ambition de l'Union Européenne est justement d'encourager une réelle diversité de langues et d'en faire un de ces traits culturels essentiels.

A6 TROUVEZ DES TERMES ÉQUIVALENTS DANS D'AUTRES LANGUES

Vous pouvez considérer que cette tâche est accomplie dès lors que chaque descripteur principal de votre terminologie possède bien au moins un terme équivalent dans une autre langue.

Exemple

FACET THÉMATIQUES (F)	FACET THÉMATIQUES (D)	FACET THÉMATIQUES (N)
transport	transport	transport
transport over land	transport over land	transport over land

L'exemple ci-dessus provient d'un thésaurus réel (thematic keywords du RMCA, Belgique) qui a servi à bâtir le Thésaurus Athena. Le terme « transport over land » qui y est utilisé pour le descripteur anglais est aussi utilisé pour la version en néerlandais. Ce cas se présente car il est possible qu'il n'y ait aucun terme approprié pour désigner ce concept en néerlandais, ou alors il est possible également que ce terme anglais soit reconnu et utilisé en néerlandais plutôt que son équivalent littéral.

Comme en témoigne cet exemple, dans certains cas l'utilisation de termes dans une autre langue que celle attendue peut s'avérer nécessaire car l'usage commun reconnaît ce terme dans cette langue-là ou parce qu'il n'existe pas de véritable équivalent dans la langue voulue. Ces termes sont appelés « coin terms ». Si ce cas se présente à vous, faites l'effort de renseigner le contexte ou d'utiliser l'information disponible sur le terme (veuillez vous référer à l'étape A7 : Produisez techniquement votre thésaurus et B7 : Vérifiez que vos concepts sont documentés).

Méthodes et outils

Vous pouvez consulter un répertoire de terminologies libres d'utilisation sur le Wiki d'Athena où les ressources sont classées par domaines et par langues :
http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources

Vous pouvez aussi consulter la norme ISO 5964: 1985 car elle traite notamment de la transposition d'un thésaurus monolingue en une version multilingue. Toutefois cette norme ne tient pas compte de la réalité technologique. Elle vous aide à le faire sur le papier sans considérer la mise en œuvre technique par un outil. Cela peut notamment mener ultérieurement à des contradictions quand vous mapperez vos termes équivalents. Nous vous conseillons donc de vous y référer pour bâtir le cœur de votre terminologie tout en gardant à l'esprit que des technologies peuvent de fait résoudre quelques difficultés signalées par la norme. La norme à venir ISO 25964-1 que nous avons déjà mentionnée portera plus précisément sur la conception d'une terminologie monolingue ou multilingue selon ce que permet la réalité technologique du moment.

Actions

Puisque vous venez de concevoir la structure de votre thésaurus, de choisir vos termes et de leur trouver des équivalents dans différentes langues, il vous faut maintenant réaliser techniquement le thésaurus :

- En affinant une dernière fois si besoin votre structure générale et en consultant quelques standards pour préparer la mise en œuvre. Ces standards ont été élaborés pour fournir des conseils de construction de thésaurus :
 - o Trois standards déjà finalisés : ISO 2788:1986 + ISO 5964: 1985 + ANSI/NISO Z39.19-2003
 - o Mais plus que tout autre : BS8723: Structured Vocabularies for Information Retrieval et la norme à venir ISO 25964-1 - Thesauri and interoperability with other vocabularies: Thesaurus for information retrieval
- En utilisant la fonction de gestion de thésaurus de votre outil maison de gestion de collections, ou si une telle fonction n'existe pas, en vous servant d'un tableur (tel que Microsoft Excel ou Calc d'Open Office) pour déclarer et organiser d'un point de vue pratique les listes de termes et les groupes thématiques transversaux.

Objectifs

L'objectif est de construire techniquement le thésaurus que vous avez conçu précédemment. Si la conception a été faite de manière satisfaisante, la réalisation technique sera rapide et facile. Avant de procéder à la fabrication technique de votre thésaurus, nous vous recommandons de consulter certains standards qui donnent des conseils pour élaborer ce type de terminologie. En effet le travail de l'organisme de normalisation ISO est d'une bonne aide pour construire techniquement votre thésaurus.

Si les trois standards suivants ISO 2788:1986 + ISO 5964: 1985 + ANSI/NISO Z39.19-2003 sont finalisés et intéressants à connaître pour concevoir précisément un thésaurus, nous vous recommandons les plus récents.

A7 PRODUISEZ TECHNIQUEMENT VOTRE THESAURUS

- BS8723: Structured Vocabularies for Information Retrieval. Ce standard, qui est une adaptation britannique de l'ISO 2788, cherche à tenir compte de tout type de terminologie, donc non pas seulement des thésaurus, et se concentre également sur l'interopérabilité entre les vocabulaires. Il prend en compte la connexion entre les terminologies et les collections (et leurs objets), dans la perspective d'une SKOSification.

- ISO 25964: Thesauri and Interoperability with other Vocabularies. Cette norme est divisée en deux parties : la première sur « Thesaurus for Information retrieval » sera publiée en 2011. La seconde partie intitulée « Interoperability with other vocabularies » le sera en 2012. Cette norme est une mise à jour des précédentes normes sur les thésaurus (ISO2799 et ISO5964) quant aux méthodes de conception mais offre aussi des spécifications techniques pour la conception de thésaurus et la maintenance de logiciels dédiés. Quelques recommandations de formats d'échange et de protocoles sont disponibles.

Dans les annexes se trouve le diagramme UML (Unified Modeling Language) qui présente la façon générale de concevoir un thésaurus et sa réalisation technique telle que définie par cette norme.

Nous formulons nos recommandations en fonction de celles fournies par ces standards.

Parmi tous les outils que nous avons identifiés durant notre étude sur l'état de l'art¹⁴, aucun n'est véritablement dédié à la réalisation technique d'un nouveau thésaurus. Dans l'idéal, puisqu'il sera question plus tard de SKOSification (particulièrement à l'étape B2 : SKOSifiez sommairement votre terminologie), vous devriez directement utiliser dès maintenant un éditeur XML grâce auquel vous pourriez déjà formater votre terminologie en RDF.

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

A7 PRODUISEZ TECHNIQUEMENT VOTRE THESAURUS

Néanmoins vous pouvez le faire aussi de manière plus simple et plus légère en utilisant un tableur et en convertissant ensuite votre tableau en XML. Le XML n'est pas indispensable ici, mais votre terminologie épousera une forme plus standard que celle d'un classeur. Le premier intérêt du XML est de conduire à faire un premier pas dans la SKOSification de votre terminologie. Le second intérêt réside dans le fait que l'affichage en arborescence en XML (par exemple dans un navigateur Web) aide à voir en un coup d'œil comment votre thésaurus est structuré.

Quoi qu'il en soit, alors que les étapes précédentes ne réclamaient pas de connaissance très pointue en Ingénierie des Connaissances, celle-ci en revanche demande pour la première quelques compétences techniques.

Exemple

Dans votre thésaurus sur l'architecture se trouvent deux micro-thésaurus : un sur les monuments, et un autre sur les habitations. Dans votre liste de termes « monument », apparaissent des termes comme « palais », « arc de triomphe », « thermes »... et dans la liste « habitation » vous pouvez avoir « appartement », « hutte », « maison », « squat »... et pour finir un groupe transversal réunit des termes comme « palais » et « appartement » autour de la thématique « bâtiment ». Pour réaliser techniquement ce thésaurus, vous vous servez d'OpenOffice comme logiciel tableur. Dans le classeur, votre feuille principale est appelée « Thésaurus d'Architecture ». La première colonne recense les noms de micro-thésaurus (« monument », « habitation »). La seconde liste les termes relatifs à ces sous-domaines et qui sont dans une relation de hiérarchie.

PRODUISEZ TECHNIQUEMENT VOTRE THESAURUS

Sub-domaines	Termes
Mouvement	Marcher
	Transmettre
	Donner
Habitation	Appartement
	Maison
	Logement
	Logis

Puis, pour déclarer le rassemblement transversal de termes liés au thème « bâtiment », vous créez une nouvelle feuille dans votre classeur, intitulée « bâtiment », et dans laquelle la première colonne donne les termes et la seconde les micro-thésaurus dont ils sont tirés.

Termes	Thésaurus micro-thésaurus
Marcher	Marcher
Donner	Donner

Méthodes et outils

Si vous n'avez pas d'outil maison de gestion de thésaurus qui vous permette de réaliser techniquement un thésaurus et de le convertir ensuite en XML, nous vous conseillons d'utiliser un tableur comme OpenOffice. C'est un outil libre et gratuit dont les fonctionnalités sont adaptées quand on souhaite organiser des termes que ce soit de manière hiérarchique ou transversale. Et il vous permet grâce à la fonction « Enregistrer Sous » d'exporter vos données en les convertissant en XML. L'utilisation d'un thésaurus soulève quelques difficultés. La première d'entre elles est qu'à ce point vous ne manipulez pas des concepts mais des termes. Il s'agit bien d'une difficulté car dans le monde du Web Sémantique qui ne cesse de s'étendre, les concepts sont désormais mieux exploités que les termes. Cela est dû au fait que les concepts sont considérés comme étant indépendants de la langue utilisée, alors que les termes au contraire lui sont relatifs.

PRODUISEZ TECHNIQUEMENT VOTRE THESAURUS

Du coup se pose la question : comment continuer d'utiliser un thésaurus sans empêcher l'exploitation de concepts par une machine taillée pour le Web Sémantique ? Nous verrons plus bas que la solution réside dans la deuxième grande étape de notre ensemble de recommandations, celle consistant à rendre interopérable votre thésaurus. Nous étions jusqu'ici au niveau du terme, niveau indispensable et qui détermine le reste. Mais désormais nous passons au niveau du concept en nous éloignant des contingences de la langue. L'abstraction que cela requiert est importante à effectuer, notamment parce qu'elle permet de mettre en œuvre plus efficacement le multilinguisme.

4. Rendre sa terminologie interoperable

Une fois que vous avez conçu votre terminologie en la rendant la plus proche possible de la forme idéale que nous préconisons, vous devez la rendre interoperable.

L'étape de conception avait pour objectif de produire un thésaurus à partir du point de vue interne du musée et dans une perspective de traitement « humain ». A partir de maintenant nous vous recommandons de sortir (métaphoriquement) du musée, et de prendre en compte la question du traitement de votre thésaurus par la machine en procédant à la SKOSification de votre terminologie.

4.1 Avantages de SKOS

RDFS (Resource Description Format Schema) et OWL sont des langages qui ont été définis pour la représentation formelle de connaissances. SKOS fait partie de cette famille de langages formels. Mais à la différence des deux autres langages cités, SKOS a été conçu pour modéliser tout type de vocabulaire contrôlé. Il peut être utilisé pour représenter un thésaurus aussi bien qu'une classification ou une liste de rubriques. C'est aussi un bon compromis pour les institutions qui utilisent déjà ces types de ressources terminologiques, et qui cherchent à les rendre compatibles avec les technologies du Web Sémantique sans avoir à développer des ontologies sophistiquées. Le modèle de données

de SKOS est compatible avec le langage formel d'ontologie OWL. Ce qui a pour conséquence de faciliter la migration d'une terminologie exprimée en SKOS vers une version en OWL.

Même si le modèle SKOS est très simple, il reste assez complet, ce qui fait que la SKOSification de votre terminologie et d'éventuelles migrations ultérieures se feront à bas coût. Comme nous l'avons indiqué dans l'introduction du livrable Athena D4.2, SKOS n'est pas à proprement parler un langage de représentation formelle de la connaissance. Mais pour une institution ayant l'habitude de gérer des listes de termes, des classifications ou des thésaurus dans le meilleur des cas, cela prendrait beaucoup de temps et s'avérerait coûteux de développer une ontologie formelle parfaitement adaptée aux technologies du Web Sémantique (en utilisant OWL par exemple). A son échelle, SKOS fournit une structure s'appuyant sur des classes et des propriétés qui composent un modèle de données puissant pour migrer ou porter ce type de ressources terminologiques vers les technologies du Web Sémantique.

Les institutions doivent garder en mémoire que l'utilisation du modèle SKOS ne conduit pas à remplacer totalement les modèles de données qui sont en usage jusque-là.

Il s'agit d'un format pour publier et faciliter la réutilisation d'une terminologie, et pour garantir la portabilité de cette terminologie et son interopérabilité sémantique. En effet, les systèmes d'organisation des connaissances (KOS – Knowledge Organization System), par exemple les vocabulaires contrôlés et les thésaurus, qui sont utilisés pour l'indexation de ressources, pourraient sous une forme SKOSifiée être aussi employés pour améliorer l'accès aux ressources.

Cependant il est possible que SKOS ne soit pas approprié à tous les types de vocabulaire contrôlé. Par exemple les listes d'autorités qui habituellement fournissent des listes de personnes ne peuvent pas être migrées en SKOS de manière adéquate puisque ces terminologies portent sur des personnes réelles et non sur des concepts. Une autre limite est que les propriétés sémantiques et relationnelles de SKOS ne peuvent s'appliquer vraiment à des listes de ce genre car une personne ne peut être liée à une autre selon des liens hiérarchiques (narrower / broader) ou associatifs (related).

Résumons toutes les raisons pour vous d'utiliser SKOS comme format pour exprimer vos descriptions :

- D'abord SKOS est particulièrement bien adapté aux terminologies multilingues.
- Ensuite la SKOSification d'une terminologie est un moyen économique d'accéder au niveau conceptuel sans employer d'ontologie. Vous pouvez profiter de SKOS en migrant votre thésaurus vers une pseudo-ontologie en un minimum de temps et d'argent.
- Enfin l'évolution de SKOS fait qu'il sera dans un avenir proche de plus en plus facile de le personnaliser grâce aux nouvelles classes que

vous pourrez définir par vous-mêmes ou à celles qui seront proposées par d'autres tant que ce standard évoluera.

Quoi qu'il en soit, si vous envisagez de SKOSifier votre thésaurus, vous devez changer légèrement de perspective. Jusqu'ici vous vous occupiez de termes. A partir de maintenant vous devez gérer des concepts que le Web Sémantique dans une perspective multilingue saura mieux exploiter que les termes. Vous devez également sortir de votre cadre professionnel habituel et rendre explicite un savoir que votre bagage professionnel vous permettra de préciser facilement.

4.1 Méthodologie

En fait quand vous SKOSifiez un thésaurus, vous appliquez techniquement la connexion que vous avez définie précédemment à l'étape A3 : Définissez votre lien avec le modèle de données. SKOS est le format que nous vous conseillons d'utiliser parmi tous ceux qui aujourd'hui respectent les contraintes d'Europeana pour l'import et l'exploitation de vos ressources numériques.

En tant que format, SKOS permet le mapping de modèles de terminologie. Et dans notre cas les deux modèles de terminologie sont :

- Votre propre modèle de terminologie (que vous avez défini à l'étape A5: Organisez vos termes selon une structure de thésaurus)
- Le modèle de terminologie officiel défini dans votre modèle de données (dans LIDO par exemple si vous l'avez choisi)

Cette deuxième grande étape est donc celle de la conversion de votre thésaurus en une version SKOSifiée. Pour y parvenir, un ensemble d'actions peuvent être entreprises. La plupart d'entre elles touchent au mapping de modèles de terminologie. Comme cela fut le cas lors de la première grande étape de conception, nous présentons ici ces actions, par souci de lisibilité, comme si elles appartenait à une séquence linéaire stricte. Mais bien sûr ces tâches peuvent tout aussi bien être suivies de manière itérative.

Les différentes tâches que nous allons détailler sont :

- B1 : Évaluez la compatibilité de SKOS avec les caractéristiques de votre terminologie
- B2 : SKOSifiez sommairement votre terminologie
- B3 : Définissez précisément les labels exprimant les concepts
- B4 : Identifiez vos concepts et validez la structure
- B5 : Mappez vos concepts
- B6 : Mappez vos termes (multilingues)
- B7 : Vérifiez que vos concepts sont documentés
- B8 : Validez votre SKOSification

B1 ÉVALUER LA COMPTABILITÉ DE SKOS AVEC LES CARACTÉRISTIQUES DE VOTRE TERMINOLOGIE

Actions

En guise de prologue à la SKOSification technique de votre terminologie, vous devez vérifier si SKOS est complètement approprié à vos caractéristiques. Votre terminologie a été conçue pour satisfaire les besoins utilisateurs que vous avez identifiés à l'étape A2 : Identifiez les attentes de vos utilisateurs. Il se peut que SKOS ne soit pas compatibles avec certains de ces besoins.

1. Vérifiez si dans votre thésaurus vous n'avez que des descripteurs indépendants (concepts ou noms de sous-domaines). Si oui, SKOS n'est pas nécessaire, une représentation en RDF peut suffire.
2. Vérifiez si dans votre thésaurus vous avez une liste de noms de personnes ; si oui, vous aurez besoin d'utiliser FOAF en plus de SKOS.
3. Vérifiez si dans votre thésaurus vous avez une liste de noms de lieux ; si oui, vous aurez besoin d'utiliser SKOS d'une manière un peu spécifique en prenant garde à la définition des hiérarchies (information géographique versus information politique).
4. Vérifiez si dans votre thésaurus vous avez une liste de noms d'institutions ; si oui, vous aurez besoin d'utiliser Vcard¹⁵/hCard¹⁶ en plus de SKOS.
5. Vérifiez si dans vos listes de sujets vous avez des termes qui diffèrent les uns des autres seulement par le genre ou le nombre. Si le cas se présente, il vous faudra préciser la relation de genre ou de nombre entre les termes. Pour cela vous pouvez utiliser une extension de SKOS : SKOS-XL.

¹⁵ Vcard : <http://www.w3.org/TR/vcard-rdf/>

¹⁶ hCard : <http://microformats.org/wiki/hcard>

B1 ÉVALUER LA COMPTABILITÉ DE SKOS AVEC LES CARACTÉRISTIQUES DE VOTRE TERMINOLOGIE

Objectifs

L'objectif est de SKOSifier votre thésaurus, c'est-à-dire de rendre votre terminologie interopérable avec un modèle de données comme LIDO, et par la suite avec Europeana. Mais avant d'entreprendre toute procédure de conversion en SKOS de votre terminologie, vous devez avoir vérifié jusqu'à quel point SKOS est le format approprié compte-tenu des caractéristiques de votre terminologie. Dans le cas des listes d'autorité par exemple, il est probable que SKOS ne soit pas le format le plus approprié. Nous avons listé cinq cas où il est préférable d'articuler SKOS à d'autres formats.

- Relations sémantiques : peut-on lier entre eux les descripteurs (et donc les concepts) de la terminologie à l'aide de relations sémantiques ? => si la terminologie ne contient que des descripteurs indépendants les uns des autres, sans relation sémantique entre eux, une modélisation en SKOS n'est pas absolument nécessaire, une représentation en RDF peut s'avérer plus pratique.
- Noms de personnes : votre terminologie traite-t-elle d'objets ou d'entités abstraites qui pourraient être assimilées à des concepts ? Traite-t-elle de personnes ? => Si la terminologie traite de personnes et non d'objets ou d'entités abstraites, un standard comme FOAF (Friend Of A Friend : <http://www.foaf-project.org>) serait plus approprié. Si la terminologie traite des deux, un mélange de FOAF et de SKOS pourrait être intéressant.
- Noms de lieux : votre terminologie traite-t-elle de noms de lieux ? => Si la terminologie traite de noms de lieux mais pas d'objets ou d'entités abstraites, un simple RDF peut être utilisé pour la modéliser.
- Noms d'institutions : votre terminologie traite-t-elle de noms d'institutions ? => si votre terminologie traite de noms d'institutions et pas d'objets ou d'entités abstraites, un standard comme Vcard/hCard devrait être approprié. Si la terminologie traite des deux, un mélange de Vcard16/hCard17 et SKOS ou OWL pourrait

B1 ÉVALUER LA COMPTABILITÉ DE SKOS AVEC LES CARACTÉRISTIQUES DE VOTRE TERMINOLOGIE

être intéressant. (Vcard/hCARD est un micro-format destiné à la publication d'informations de contact sur les personnes, les organisations et les lieux.)

- Relations de genre et de nombre : votre terminologie traite-t-elle de termes qui diffèrent les uns des autres par le genre ou le nombre ? => Si la terminologie traite de termes dans plusieurs versions de genre ou de nombre, vous pouvez utiliser l'extension SKOS : SKOS-XL. En effet SKOS-XL vous permet de préciser ce type de relations entre les termes. Plus généralement, SKOS-XL est utile pour lier des concepts et des ressources lexicales à l'aide d'informations sur les termes exprimées dans une langue très générale, indépendamment de toute spécialité.

Exemple

Dans une terminologie sur l'architecture en anglais et en français, supposons que vous avez le terme « stained glass » comme élément d'une liste portant sur les édifices religieux. Si vous proposez le terme français équivalent « vitrail », il peut être pertinent de fournir également la forme plurielle « vitraux » car cette forme peut aussi être fréquemment utilisée lors d'une requête en français.

Si vous estimez que la distinction entre la forme singulière ou plurielle d'un terme et son label principal est pertinente dans le cas de votre terminologie, vous pouvez utiliser l'extension SKOS-XL pour associer de l'information spécialisée à chaque label plutôt que de fournir de l'information générale sur un concept exprimé par plusieurs labels possibles.

Si vous avez l'intention de modéliser une liste d'autorités comprenant des noms de personnes ou d'auteurs, vous pouvez jeter un œil sur la terminologie VIAF (Virtual International Authority File¹⁷). Dans la mesure où cette liste d'autorités est le résultat du mapping de plusieurs terminologies en provenance de diverses

17 VIAF, Virtual International Authority File: <http://viaf.org/>

B1 ÉVALUER LA COMPTABILITÉ DE SKOS AVEC LES CARACTÉRISTIQUES DE VOTRE TERMINOLOGIE

institutions (des bibliothèques principalement), FOAF et SKOS ont été articulés pour l'exprimer. Le site Web fournit une interface multilingue et la possibilité d'afficher en RDF les résultats d'une requête.

Méthodes et outils

Vous pouvez regarder sur le site Web du W3C pour obtenir plus d'informations détaillées sur SKOS et son extension SKOS-XL :
<http://www.w3.org/TR/skos-reference/skos-xl.htm>
Veuillez également vous référer aux cas d'usages et aux exigences liés à SKOS et définis par le W3C :
<http://www.w3.org/TR/skos-ucr>

B2 SKOSIFIEZ SOMMAIREMENT VOTRE TERMINOLOGIE

Actions

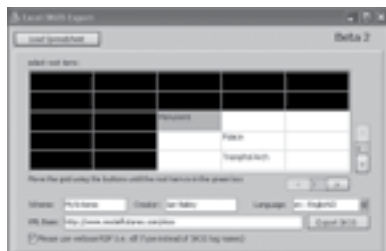
Dans ces recommandations nous parlons principalement de « sujets », c'est-à-dire que nous ne prenons en compte pour l'instant que le cas plus général de SKOSification de votre terminologie sans nous attarder sur les cas particuliers listés à l'étape B1 : Évaluez la compatibilité de SKOS avec les caractéristiques de votre terminologie.

Dans l'optique de réaliser une première SKOSification grossière de votre thésaurus, vous pouvez procéder de 4 manières différentes :

1. Si vous avez un outil de gestion de collections qui présente une fonctionnalité d'export en SKOS, utilisez-la directement.
2. Si vous n'avez pas ce type d'outil ou de fonctionnalité, et que vous avez utilisé un tableur pour concrétiser votre thésaurus, vous pouvez utiliser l'outil XL2XML, ou XLTaxonomy, et les feuilles de style qu'ils offrent par défaut.
3. Si vous préférez définir vous-même la feuille de style permettant la SKOSification de votre fichier XML, vous pouvez vous servir d'un outil comme Annocultor.
4. Ou si vous préférez construire votre feuille de style ex nihilo en procédant à une analyse de votre fichier XML source, faites-le plutôt sur papier avant de traduire le résultat informatiquement.

Objectifs

La SKOSification est une procédure de conversion qui nécessite une feuille de style dans laquelle sont écrites les règles de conversion qui permettent de transformer le thésaurus en une structure de concepts. Le point clef de cette procédure est justement la définition et la réalisation de cette feuille de style. Vous pouvez procéder de différentes manières, mais dans tous les cas cette tâche réclame une compétence technique forte. De plus, plus vous mettez de moyens à réaliser cette étape, et plus vous aurez de garanties quant à la qualité



Il est probable que vous ayez besoin de vérifier le SKOS fourni en sortie de l'outil et si nécessaire de le convertir ensuite dans la forme SKOS/RDF qui sera alors exploitée par vos outils. Cet outil peut produire soit un fichier

Méthodes et outils

Le Format Athena est le format dans lequel est exprimé le Thésaurus Athena. Il est ici proposé aux musées qui souhaitent mapper leurs propres terminologies avec le Thésaurus Athena (ou qui veulent partir du Thésaurus Athena pour construire leur terminologie). En tant que format s'appuyant sur SKOS, le Format Athena garantit aux musées que leurs descriptions respectent bien les exigences techniques d'Europeana concernant SKOS. A cette adresse vous pouvez consulter le Format Athena :
<http://www.athenaeurope.org/athenawiki/index.php/References>

Actions

Une fois que vous avez identifié vos concepts et que vous les avez mappés avec le modèle de données, vous devez définir précisément les labels qui expriment ces concepts. Pour y parvenir, voici nos recommandations :

1. Les « Preferred labels » doivent être uniques au sein d'un schéma conceptuel.
2. Chaque concept doit obligatoirement être exprimé à l'aide d'un « Preferred label ».
3. Évitez de concaténer plusieurs mots pour former un seul et même label.
4. Utilisez de préférence un lemme pour distinguer le « Preferred label » des autres labels s'ils existent.
5. Utilisez de préférence la typographie conventionnelle en usage dans les langues concernées par votre terminologie.
6. Évitez de répéter la même information par des voies différentes (ex : les propriétés inverses et symétriques).

Toutes ces recommandations sont détaillées ci-dessous.

Objectifs

L'objectif de cette tâche est de définir précisément les labels de vos concepts.

- Les « Preferred labels » doivent être uniques au sein d'un schéma conceptuel
 Conformément à ce qui est exigé par le modèle de données de SKOS, deux concepts d'un même schéma conceptuel ne devraient pas avoir le même « Preferred label » dans une langue donnée. Cependant, en raison du fait que les langues naturelles sont hautement polysémiques et pleines d'homonymes, le modèle de données de SKOS n'interdit pas d'avoir pour un concept donné deux fois le même « Preferred label » dans deux langues différentes.

- Chaque concept doit obligatoirement être exprimé à l'aide d'un « Preferred label » par langue
 Dans le même ordre d'idée, le modèle de données de

DÉFINISSEZ PRÉCISEMENT LES LABELS EXPRIMANT LES CONCEPTS

SKOS n'interdit pas l'absence de « Preferred label » pour un concept, car les labels servent surtout à aider la compréhension et à préciser le sens des concepts. Cela est particulièrement vrai dans un contexte multilingue et très utile pour les actions d'administration et de maintenance. Par conséquent nous recommandons d'utiliser un « Preferred label » par langue. Il est important de constater que cela implique également qu'il n'est pas possible d'avoir plusieurs « Preferred labels » pour une seule et même langue.

- Evitez de concaténer plusieurs mots pour former un seul et même label
Pour obtenir la description la plus pointue possible, nous vous recommandons d'éviter l'utilisation de plusieurs valeurs en guise de termes préférés. Par exemple les concepts doubles tels que « demeures / maisons » doivent être considérés comme deux concepts différents et séparés mais liés entre eux par une relation sémantique. L'utilisation de « scope notes » peut alors aider à renforcer la proximité de ces deux concepts. Dans ce cas le lien entre les deux termes doit être clairement défini afin d'avoir au final la meilleure description qui soit. En acceptant que « demeure » et « maison » sont deux synonymes, le double concept de départ peut finalement être modélisé comme suit :
Demeure: preferred label and maison: alternative label
Une autre manière de gérer le cas des doubles concepts consiste à modéliser les deux concepts comme étant des concepts liés entre eux (relation « related to »).

- Utilisez de préférence un lemme pour distinguer les « Preferred labels » des autres labels s'ils existent
Le « Preferred label » devrait consister en un terme de langue naturelle formé d'un mot simple ou de mots composés. Cela signifie qu'il vaut mieux ne pas associer

DÉFINISSEZ PRÉCISEMENT LES LABELS EXPRIMANT LES CONCEPTS

comme label à un concept un mot inventé et inusité ou un code arbitraire. Si l'information que constitue le code est importante, elle peut être portée par la propriété SKOS `skos:notation property`. Le lemme d'un mot indique qu'il s'agit de sa forme canonique. Nous recommandons fortement de former ainsi les termes à utiliser comme « Preferred labels ». Par exemple, en anglais ou en français, la forme habituelle, le « lemme » d'un mot dans le cas des noms est le singulier pour le nombre et le masculin pour le genre (en français).

- Utilisez de préférence la typographie conventionnelle en usage dans les langues concernées par votre terminologie
Les labels que vous utilisez devraient respecter les règles typographiques qui sont en vigueur dans les langues relatives à ces labels. Par exemple, en anglais tous les mots désignant une langue ou une nationalité commencent avec une majuscule tandis qu'en français ces mots n'en ont pas. C'est pourquoi nous recommandons de respecter les conventions en usage pour chaque langue concernée. Toute exception volontaire à cette directive doit être documentée à l'aide des propriétés de documentation du modèle.
Pour toute forme verbale il est préférable de privilégier la forme de l'indicatif. Là encore il est important de choisir pour les termes les formes conventionnelles dans les langues considérées. Si jamais un concept est exprimé seulement par des labels aux formes spécifiques qui ne correspondent pas au lemme, cela doit être documenté à l'aide des propriétés de documentation (`skos:note`, `skos:changeNote`, `skos:editorialNote` ou `skos:historyNote`). Dans le cas des mots composés, si cela est possible, il est préférable de limiter au maximum l'ajout d'adjectifs ou de verbes dans des phrases nominales.

DÉFINISSEZ PRÉCISEMENT LES LABELS EXPRIMANT LES CONCEPTS

Dans le même ordre d'idées, l'utilisation d'articles et de prépositions devrait être évitée pour ne pas augmenter la longueur des labels. Si on se place du point de vue des systèmes informatiques qui traitent les descriptions exprimées par votre terminologie, ces directives concernant les labels peuvent améliorer leur efficacité et rendre vos données beaucoup plus facilement exploitables et accessibles.

- Évitez de dupliquer l'information

Le modèle de données de SKOS est composé de classes et de propriétés telles que celles que nous avons déjà mentionnées plus haut. Les sens des concepts et des relations doivent pouvoir être déduits grâce à un usage efficace des propriétés. Le fait que les propriétés disponibles dans le modèle de SKOS sont proposées sous forme de paires (les propriétés sont inverses ou symétriques) suppose que l'utilisation d'une propriété implique son opposée ou son inverse. Du coup il vaut mieux éviter de dupliquer et de répéter la même information par des voies diverses. Les terminologies SKOSifiées sont traitées par des machines qui sont programmées pour le faire de manière exhaustive. Par conséquent, moins il y a d'information redondante, plus la machine fournira rapidement des résultats à une requête. Les principales propriétés auxquelles il faut accorder de l'attention à ce sujet sont :

- Propriétés inverses : l'utilisation des propriétés `skos:broader` ou `skos:narrower` implique le sens inversé. Déclarer que A a un concept parent B implique que B a un concept enfant A. Cela est également vrai pour les propriétés `skos:broaderTransitive` et `skos:narrowerTransitive`.
- Propriétés symétriques : la propriété `skos:related` est symétrique, donc s'il est déclaré que A est associé à B, il n'y a pas besoin de déclarer que B est associé à A.

DÉFINISSEZ PRÉCISEMENT LES LABELS EXPRIMANT LES CONCEPTS

Exemple

Vous avez sommairement SKOSifié votre terminologie sur l'architecture. Il est probable que vous ayez plusieurs termes équivalents pour exprimer un concept mais vous devez définir un terme favori ou « Preferred term » (comme cela se passe habituellement quand on utilise un thésaurus) et ne garder qu'un seul « Preferred term » par langue.

- Si vous avez un schéma conceptuel (ou groupe de concepts) sur l'architecture et que vous avez en anglais les deux termes « apartment » et « flat » pour exprimer un concept fils de « habitation », vous devez alors spécifier lequel des deux termes est le « Preferred label ». Ce que vous pouvez exprimer comme suit (en utilisant par exemple comme ici le format Turtle²¹ pour représenter cette partie de votre terminologie SKOSifiée) :

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

Méthodes et outils

Puisqu'à ce stade votre thésaurus a déjà été sommairement SKOSifié, il peut être ouvert et modifié dans l'outil en ligne xTree. Cet outil est intéressant parce qu'il vous aide à « écrire » des labels en SKOS grâce à une interface Web orientée utilisateur, et parce qu'en plus il met en œuvre le brouillon de norme ISO 25964-1.

D'autres outils d'édition comme Protégé²² (augmenté du plug-in SKOSed) et ThManager²³ vous permettent de gérer les labels utilisés pour exprimer les concepts de votre terminologie. Veuillez vous référer à la section d'état de l'art du wiki Athena pour avoir une version à jour des outils disponibles pour l'édition de terminologies SKOSifiées :

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

22 Vous pouvez trouver Protégé et SKOSed à l'adresse : <http://protege.stanford.edu/>

23 Vous pouvez trouver ThManager à l'adresse : <http://thmanager.sourceforge.net/>

IDENTIFIEZ VOS CONCEPTS ET VALIDEZ LA STRUCTURE

Actions

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

Puisque vous avez affiné la version SKOSifiée de votre thésaurus en y précisant les labels, vous pouvez désormais aller plus loin en identifiant techniquement vos concepts et en les mappant avec le modèle de données. Pour ce faire nous vous conseillons de suivre le schéma « 5 étoiles » de Tim Berners-Lee²⁴ :

- * rendez vos données disponibles sur le Web (quel que soit leur format) sous une licence ouverte
- ** rendez-les disponibles sous une forme structurée (ex : un tableau Excel à la place d'une image scannée de ce tableau)
- *** utilisez des formats non propriétaires (ex : le CSV - Comma-Separated Values à la place d'Excel)
- **** utilisez des URIs pour identifier les entités, pour que les internautes puissent accéder directement à vos données
- ***** liez vos données à d'autres données afin de fournir des éléments de contexte

Objectifs

Le W3C définit deux étapes principales lorsqu'il s'agit d'identifier des concepts :

- La création (ou la réutilisation) d'une URI (Uniform Resource Identifier – Identifiant uniforme de ressource) pour identifier le concept de manière strictement unique
- La déclaration en RDF, à l'aide de la propriété `rdf:type`, que la ressource identifiée par cette URI est de type `skos:Concept`

Utilisation d'un système d'identification persistante pour définir les URIs

Nous recommandons d'utiliser des standards (décrits plus hauts) pour identifier les concepts. Ensuite, une fois que cette identification a été réalisée grâce à la définition des URIs (protocole `http`), ces dernières doivent être déclarées à des systèmes d'identification pérenne tels que PURL qui est un système normalisé.

Rendre sa terminologie interopérable

IDENTIFIEZ VOS CONCEPTS ET VALIDEZ LA STRUCTURE

L'utilisation de tels systèmes vous offre aussi l'avantage de rester indépendant(e) des adresses de stockage ; par exemple si la terminologie est déplacée d'une adresse (serveur hôte) à une autre, les URIs d'identification des concepts n'ont pour autant pas à être modifiées.

Utilisation d'URIs non explicites

Il est hautement recommandé d'utiliser des URIs non explicites pour éviter de réutiliser une même URI pour identifier deux concepts différents. Comme les langues naturelles sont en effet ambiguës et polysémiques par définition, il est possible que deux concepts aient deux labels similaires. L'utilisation d'URIs explicites suppose qu'une langue naturelle en particulier a été choisie pendant la définition ou la migration de votre terminologie, ce qui n'est pas pratique dans un contexte multilingue. Une URI non explicite étant une URI dont le nom ne transmet aucune information sémantique une série de chiffres ou de caractères spéciaux ne signifiant rien peut constituer une bonne URI non explicite.

Exemple

Supposez que votre terminologie soit techniquement hébergée et gérée par votre institution, mais utilisée par plusieurs autres. Vous devez définir vos identifiants de telle manière qu'apparaisse dans les URIs l'origine des concepts (le nom de domaine) mais aussi en restant suffisamment souple pour que les autres institutions qui les utilisent n'aient pas à faire de modification de leur côté à chaque fois que du vôtre le système d'identification change. Il est préférable d'utiliser des URIs non explicites pour éviter l'ambiguïté des langues naturelles. La Bibliothèque Nationale de France (BnF), par exemple, utilise actuellement le système d'identifiants persistants ARK (voir plus bas pour les détails). Voici un exemple d'URI de la BnF et définie à l'aide du système ARK : <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11931420f>

Rendre sa terminologie interopérable

IDENTIFIEZ VOS CONCEPTS ET VALIDEZ LA STRUCTURE

Méthodes et outils

Plusieurs systèmes d'identifiants persistants sont en usage aujourd'hui. Voici quelques informations concernant ceux qui s'avèrent être les principaux :

PURL : un PURL (Persistent Uniform Resource Locator) est une URL d'un type particulier ; plutôt que de pointer directement vers l'adresse de stockage de l'objet numérique, le PURL pointe vers ce qu'on appelle un « résolveur d'adresse » qui recherche l'URL correcte de cette ressource et la retourne au poste client sous la forme d'une redirection HTTP, qu'il suffit alors de suivre de façon habituelle pour accéder à la ressource. Les PURLs sont compatibles avec d'autres standards d'identification de documents tels que les URN.

URN : l'URN (Uniform Resource Name) est conçu pour décrire une identité plutôt qu'une adresse ; par exemple un URN peut contenir un code ISBN (International Standard Book Number, utilisé comme identifiant commercial unique de livres.)

NBN : un NBN est un espace de nommage de type URN utilisé uniquement par les bibliothèques nationales, afin d'identifier des publications laissées en dépôt et à qui il manque un identifiant, ou afin de référencer des métadonnées de catalogage qui décrivent les ressources. Les NBNs peuvent être utilisés soit dans le cas d'objets ayant une représentation sous forme numérique, soit dans le cas d'objets qui sont seulement physiques et auxquels on a substitué des données bibliographiques.

ARK : l'ARK (Archival Resource Key) est un schéma URL qui a été développé à la Bibliothèque nationale de Médecine des Etats-Unis (US National Library of Medicine) et maintenu par la Bibliothèque Numérique de Californie (California Digital Library). Les ARKs sont conçus pour identifier des objets de tout type – qu'ils

IDENTIFIEZ VOS CONCEPTS ET VALIDEZ LA STRUCTURE

soient sous forme numérique ou physique. Le schéma ARK encourage l'utilisation d'identifiants non porteurs de sens pour les objets les plus au cœur de la collection. Différemment d'une URL ordinaire, un ARK est utilisé pour accéder à trois choses : l'objet lui-même, ses métadonnées, et une déclaration d'engagement de la part du fournisseur actuel.

OpenURL : Une Open-URL contient encodées dans une adresse URL des métadonnées sur la ressource ainsi identifiée, et est conçue pour assurer le chaînage entre des ressources d'information et des services de bibliothèques. Ce standard n'a pas été originellement conçu pour fournir des identifiants ou des « résolveurs d'adresse » persistants mais il est décrit comme un protocole de transport de métadonnées.

DOI : un Digital Object Identifier (DOI) est un identifiant indirect pour documents électroniques s'appuyant sur des « résolveurs d'adresse » de type Handle (Handle était un précédent système pour l'identification et le référencement de ressources). Selon la Fondation Internationale de DOI (IDF - International DOI Foundation), formée en octobre 1997 pour présider à l'évolution du système DOI, il s'agit d'un « mécanisme pour l'identification permanente de contenus numériques ».

Grâce à ces courtes introductions, nous pouvons voir que certains de ces standards sont plus adaptés à un champ en particulier (par exemple l'URN et le NBN sont plus adaptés aux bibliothèques), tandis que d'autres comme PURL ou DOI pourraient servir à la définition d'URLs. Pour toutes ces questions d'identification, vous pouvez également vous reporter au livret (en langue anglaise) « Identifiants Pérennes : Recommandations pour les institutions »²⁵ qui a été élaboré par le lot de travail WP3 d'Athena.

25 <http://www.athenaeurope.org/getFile.php?id=779>

Actions

Au cours de la SKOSification de votre terminologie, vous pouvez mapper vos concepts en définissant entre eux des relations sémantiques. Nous vous recommandons de préciser ces relations :

- Même si les concepts que vous souhaitez lier ne sont pas dans une relation hiérarchique immédiate.
- Tout en évitant de mélanger relations hiérarchiques et associatives et ainsi augmenter la probabilité que vos relations soient bien cohérentes (c'est-à-dire correctes eu égard aux règles qui régissent SKOS)

Objectifs

L'objectif est de profiter de la SKOSification pour commencer dès cette étape l'auto-documentation de votre terminologie. Cette tâche est aussi l'occasion de lever certaines ambiguïtés.

Relations hiérarchiques non immédiates

Dans certains cas, les relations sémantiques entre concepts doivent être décrites avec précision si on veut éviter une perte de sens ou d'information, ou si on souhaite éviter également de concevoir de l'information inintelligible. Par exemple les paires de propriétés `skos:broaderTransitive` / `skos:narrowerTransitive` permettent de décrire précisément des relations entre concepts qui touchent à deux niveaux de hiérarchie. L'utilisation de ces propriétés transitives est alors à privilégier pour déclarer toute relation non immédiatement hiérarchique entre deux concepts. Si la symétrie de telles propriétés rend plus confus le sens des concepts qui sont ainsi mis en relation, il est possible de la bloquer grâce à une extension du modèle de données de SKOS.

Consistance des relations sémantiques

Pour garantir la consistance de votre terminologie SKOSifiée, il est préférable de ne pas mélanger des relations hiérarchiques avec des relations associatives. Par exemple, un concept A ne peut pas être associé à un autre

concept B si ce concept A est déjà le concept-fils enfant (narrower) d'un autre concept C. Il est donc important d'accorder une attention particulière à cette tâche de conception des relations sémantiques entre concepts.

Exemple

Chaque terminologie est conçue dans un but spécifique. Comme ce fut présenté dans la section A des recommandations, c'est pour répondre à vos besoins propres que votre terminologie a été conçue. Dès lors, vous devrez probablement porter votre attention sur la structure de votre terminologie afin de la mapper correctement avec les concepts que vous avez définis à l'aide de SKOS.

Par exemple si votre terminologie en anglais a d'un côté un schéma conceptuel (groupe de concepts) portant sur la Musique et contenant le concept générique « musical instrument » et de l'autre un schéma conceptuel sur la Religion et contenant un concept « organ », vous pouvez réaliser plusieurs mappings différents (représentés ici en format Turtle²⁶) :

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

```
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
```

Dans la mesure où vous pouvez définir un orgue comme un instrument musical, vous pouvez mapper ce concept « organ » à celui de « musical instrument ». Le concept « organ » sera alors associé aux deux schémas conceptuels. Ce mapping précisément vous évite la répétition du concept dans les deux schémas.

Méthodes et outils

Il n'y a pour le moment aucun outil libre et disponible pour vous aider et vous guider dans la procédure de mapping. En effet, cette procédure de mapping est principalement une opération intellectuelle et dépend d'une validation par un expert.

Des outils propriétaires comme ITM3 (Mondeca) proposent des interfaces utilisateurs dédiés au mapping et proposant de bout en bout un suivi d'étapes et des outils de validation.

Vous pouvez jeter un œil à l'état de l'art élaboré par le lot de travail WP4 d'Athena en gardant à l'esprit que la liste d'outils en question continue d'évoluer :

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

Actions

A l'étape A5 : Organisez vos termes selon une structure de thésaurus, alors que vous étiez en train de concevoir votre thésaurus avant de penser à sa SKOSification, vous aviez déjà réalisé un premier mapping de (groupes de) termes. Vous avez ensuite traduit techniquement ce mapping à l'étape A7 : Produisez techniquement votre thésaurus. Maintenant nous vous proposons d'affiner et d'améliorer ce mapping de termes grâce aux possibilités que SKOS vous offre. Nous vous recommandons de suivre la procédure suivante en commençant par les termes de langue maternelle :

- Rendez explicites les relations sémantiques entre les labels.
 - Documentez grâce à la propriété Scope Note tout changement de terme dans votre thésaurus.
- Après avoir ainsi affiné votre mapping de termes, vous pouvez désormais définir et traduire techniquement le mapping entre les termes équivalents qui ont été identifiés à l'étape A6 : Trouvez des termes équivalents dans d'autres langues. Pour cela nous vous recommandons :
- De fournir pour chaque concept un label équivalent dans chaque langue concernée par votre terminologie.
 - D'utiliser le même système d'identifiants de langues (tags) pour définir la langue de chaque label ainsi créé.

Objectifs

L'objectif maintenant est de préciser et d'améliorer votre mapping de termes puisque le mapping des modèles de terminologies a été effectué auparavant, et de permettre le multilinguisme en exprimant les équivalences sémantiques de termes en différentes langues. Plus vous avez de termes (équivalents) reliés les uns aux autres et plus votre thésaurus sera exploitable (en différentes langues).

Associez à chaque concept un label équivalent pour toutes les langues concernées par la terminologie.

B6 MAPPEZ VOS TERMES (MULTILINGUES)

Une attention toute particulière doit être accordée aux labels multilingues exprimant les concepts. Ces labels multilingues doivent être définis de manière correcte selon les différentes langues de la terminologie pour qu'ensuite les équivalences puissent être traitées informatiquement à partir de la représentation SKOS des concepts.

Utilisez le même système d'identifiants de langues (tags) pour définir la langue de chaque label. Il existe plusieurs systèmes normalisés et équivalents : par exemple les trois tags « en », « en-GB » ou « en-Latn » sont différents systèmes d'identifiants de langue qui se réfèrent tous trois à une même langue, en l'occurrence l'anglais de Grande Bretagne en alphabet latin. Dans le cas de terminologies impliquant des langues de différents alphabets, le système de tags « langue-alpha-bet » (par exemple « en-Latn ») peut être utile afin de fournir plus de précisions. Nous recommandons d'utiliser un seul et même système de tags pour déclarer tous les attributs de langues de la terminologie. Si aucun système d'identifiants de langue en particulier ne vous est imposé, nous recommandons d'utiliser ceux définis dans la norme ISO 639-1 où les tags de langues sont encodés selon deux lettres dans le cas plus simple.

Exemple

Par exemple, si votre terminologie sur l'Architecture a un concept « dwelling » pour lequel « dwelling » est le label préféré et « houses » est le label alternatif, et si votre terminologie est bilingue français – anglais, vous aurez alors à fournir en français des équivalents à ces labels de concepts.

```
ex:dwelling rdf:type skos:Concept skos:prefLabel
« dwelling »@en; skos:prefLabel « habitation »@fr;
skos:altLabel « houses »@en; skos:altLabel « maisons »@fr;
```

B6 MAPPEZ VOS TERMES (MULTILINGUES)

Vous pouvez vous référer à une terminologie existante pour enrichir la vôtre, ou la soumettre dans sa version SKOSifiée à l'expertise d'une personne du domaine et d'une autre origine linguistique pour obtenir les termes exacts pour exprimer vos concepts.

Méthodes et outils

27 <http://europeanalabs.eu/wiki/WP2LanguageResources>

28 <http://incubator.apache.org/opennlp/index.html>

A l'instar du mapping de concepts, il n'existe aucun outil dédié pour vous aider et vous accompagner dans le mapping de termes, bien que vous puissiez trouver quelques outils spécifiques développés par des professionnels du Traitement Automatique des Langues (TAL, ou en anglais NLP - Natural Language Processing)²⁷ permettant l'alignement multilingue. Europeana a établi un répertoire d'outils et de ressources lexicales développés et utilisés dans le champ du TAL, et il existe également un incubateur destiné au développement en open source de ces technologies²⁸. Ces outils vous aident à extraire automatiquement des termes similaires sur la base de leurs lemmes. Cela peut être utile pour une première extraction automatique à affiner ensuite par un expert. Bien sûr vous pouvez utiliser les éditeurs de SKOS (ThManager ou Protégé avec le plugin SKOSed...) pour procéder au mapping de termes à réaliser à la suite du travail conceptuel préalable.

B7 VÉRIFIEZ QUE VOS CONCEPTS SONT DOCUMENTÉS

Actions	<p>Cette tâche de documentation consiste à donner de l'information sur les changements intervenus dans le temps, en se servant de la séparation entre les concepts et les labels. Nous vous conseillons :</p> <ul style="list-style-type: none">• De fournir de la documentation pour chaque changement impactant un concept et ses labels• De fournir autant que possible de la documentation sur les concepts à l'aide de la propriété Scope Notes
Objectifs	<p>Fournir de la documentation pour chaque changement impactant un concept et ses labels</p> <p>Le modèle de données de SKOS propose un grand nombre de propriétés de documentation permettant d'affiner le sens d'un concept ou de garder la trace de tous les changements de sens ou de labels. Pour des raisons d'administration et de maintenance de la terminologie, chaque changement doit être enregistré dans la terminologie SKOSifiée à l'aide des notes de modifications (skos:changeNote) ou des notes éditoriales (skos:editorialNote).</p> <p>Fournir autant que possible de la documentation sur les concepts à l'aide de la propriété Scope Notes</p> <p>Comme mentionné plus haut, la documentation sur les concepts aide à préciser le sens d'un concept. L'utilisation des Scope Notes (skos:scopeNote) peut véritablement aider à mieux comprendre les concepts grâce à l'information contextuelle ainsi fournie. Des exemples peuvent aussi être donnés en complément en se servant de la propriété skos:example. La documentation des concepts est particulièrement demandée quand leurs labels ont des formes homographiques et/ou homonymiques soit dans une seule et même langue, soit dans des langues différentes. Des Scope Notes et des exemples peuvent lever des ambiguïtés sémantiques et ainsi aider l'utilisateur à bien comprendre de quoi il retourne.</p>

B7 VÉRIFIEZ QUE VOS CONCEPTS SONT DOCUMENTÉS

	<p>Pour construire votre documentation interne à la terminologie, vous pouvez plus ou moins préciser les propriétés de notes qui sont proposées dans le format SKOS :</p> <ul style="list-style-type: none">• 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) <p>La propriété skos:note peut être utilisée pour exprimer une documentation générale sur un concept. Et tous les autres types de notes sont des spécialisations de cette propriété générale. Les propriétés skos:changeNote et editorialNote sont utiles principalement pour des questions d'administration et de maintenance. Et celles de skos:definition, skos:example, et skos:historyNote sont utiles quand vous souhaitez fournir de l'information sur le sens d'un concept afin de le rendre plus intelligible. Enfin, comme pour les labels, les propriétés de documentation peuvent être utilisées dans différentes langues en se servant des tags de langues associés à l'élément de code : xml:langattribute.</p>
Exemple	<p>L'utilisation des notes peut aider à garder une trace de l'historique d'un concept ou à donner des détails. Par exemple, si vous avez un concept « art gothique » dans votre terminologie sur l'architecture, vous pouvez utiliser une Scope Note introduite par la propriété skos:definition dans laquelle vous pouvez déclarer que l'art gothique est apparu durant la seconde moitié de la période médiévale. Vous pouvez aussi rendre explicite une connaissance qui est pour vous implicite en définissant ainsi vos concepts.</p>
Méthodes et outils	<p>Vous pouvez réaliser cette tâche de documentation sur vos concepts et vos termes à l'aide des éditeurs SKOS déjà mentionnés.</p>

Actions

La SKOSification est une procédure de conversion des éléments de votre thésaurus en un format spécifique. Cela signifie que la conversion est régie par des règles, et que le résultat d'une telle procédure doit avoir une syntaxe correcte eu égard à la « grammaire » du format SKOS. C'est pourquoi vous devez vérifier à la fin de la procédure si la version SKOSifiée de votre thésaurus est correcte ou non. Cette étape consiste en la validation des mappings de concepts et de labels, dans le respect du formalisme SKOS. Pour cela nous vous conseillons d'utiliser le service Web : Pool Party.

Objectifs

29 <http://www.w3.org/2004/02/skos/validation>

30 <http://demo.semantic-web.at:8080/SkosServices/index>

Le W3C met à disposition en ligne un outil de validation, mais il n'est pas configuré pour tenir compte de la version la plus récente du modèle SKOS²⁹. Pool Party, un système de gestion de thésaurus, offre un service en ligne³⁰ pour convertir et vérifier la cohérence de votre thésaurus SKOSifié.

D'un point de vue technique, nous vous recommandons d'utiliser Pool Party pour vérifier la cohérence de votre terminologie convertie en SKOS. Le service de validation en ligne de Pool Party est libre et gratuit, et peut s'appliquer à des fichiers SKOS, que ceux-ci soient déjà en ligne ou qu'ils se trouvent dans un répertoire local.

Pour vérifier la cohérence de votre terminologie SKOSifiée, Pool Party traite tous les points que nous avons présentés précédemment dans nos directives :

- URIs valides : l'outil vérifie s'il n'y a pas dans l'URI de caractère non autorisé. Toutefois si une URI est utilisée deux fois pour identifier deux concepts différents, il n'y aura aucune alerte ou aucun message d'avertissement.
- Tags de langues manquants : l'outil vérifie si tous les labels et les notes ont bien un tag de langue.
- Labels manquants : l'outil vérifie que chaque concept a au moins un « Preferred label ».

- Concepts égarés : tout concept isolé et non lié à d'autres concepts est signalé comme concept égaré (« loose concept »).
- Classes OWL séparées : certains éléments du modèle SKOS sont compatibles avec les éléments OWL, du coup l'outil vérifie le cas échéant la cohérence des éléments OWL qui se trouveraient dans la terminologie SKOSifiée.
- Usage correct des labels : les règles d'utilisation des labels sont vérifiées par l'outil pour éviter que soit utilisé un même label comme label préféré et label alternatif (ou comme label caché aussi), pour éviter aussi que soient utilisés deux labels préférés dans la même langue, ...
- Usage correct des propriétés de mapping : l'outil vérifie la consistencecohérence des relations de mapping.
- Usage correct des relations sémantiques : l'outil vérifie qu'il n'y a pas de mélange de relations hiérarchiques avec des relations associatives.

Du point de vue du contenu, seuls les administrateurs et les utilisateurs de la terminologie peuvent valider in fine la migration du thésaurus en format SKOS au moins en ce qui concerne la procédure initiale de transformation. En effet ces personnes seront capables de confirmer ou de modifier la conception globale de la terminologie et ses relations sémantiques en fonction des tests d'efficacité de l'indexation pour accéder aux ressources.

Exemple

Ci-dessous se trouve un résultat de contrôle fourni par le service de validation SKOS de Pool Party. Toutes les sections principales en vert sont vérifiées : dans ce cas la terminologie est bien SKOSifiée.



Méthodes et outils

Les éditeurs comme Protégé (+ SKOSed) procèdent à un premier contrôle grossier de la terminologie SKOSifiée avant de l'ouvrir et d'autoriser son édition. Cependant ce contrôle est très primaire. Si vous voulez être sûr que votre terminologie est bien SKOSifiée vous aurez besoin d'outils tels que Pool Party. Vous pouvez trouver de l'information sur l'outil Pool Party et l'utiliser en ligne à l'adresse : <http://poolparty.punkt.at/>

5. Lier sa terminologie à un réseau

Après avoir rendu interopérable votre terminologie en la SKOSifiant, nous vous conseillons maintenant de la lier à un réseau de ressources.

En effet plus votre terminologie est liée à d'autres, et plus les termes qu'elle contient seront efficacement trouvés et exploités par un moteur de recherche sémantique. Pour vous aider à lier votre terminologie à d'autres, nous vous proposons de suivre une procédure composée de 4 étapes :

- C1 : Définissez les métadonnées de votre terminologie
- C2 : Identifiez les autres ressources avec lesquelles mapper votre terminologie
- C3 : Mappez votre terminologie avec les autres ressources
- C4 : Validez l'interopérabilité

Une fois encore, même si nos recommandations sont présentées sous la forme d'une séquence d'actions linéaire, il est préférable pour vous d'en suivre les étapes de façon itérative si besoin.

DÉFINISSEZ LES MÉTADONNÉES DE VOTRE TERMINOLOGIE

Actions

Avant de lier techniquement votre terminologie à un réseau de ressources, nous vous recommandons de lui associer une documentation générale en définissant des métadonnées qui la décrivent dans sa globalité. Il n'existe pas de schéma de métadonnées spécifique que vous pourriez utiliser, mais nous pensons qu'une fiche en « Dublin Core extended » devrait vous fournir un bon point de départ. Voici les informations que devraient fournir les métadonnées sur votre terminologie :

- Nom de la terminologie
- Propriétaire
- Domaines de description
- Langues
- Contributeurs
- Date de création
- Dates de modifications
- Type de terminologie (ex : thésaurus)
- Licence
- Statut (ex : brouillon, publiée)
- Normes (ex : RDF/SKOS, RDF/OWL)
- Type de structure (ex : liste de termes à plat, arborescence stricte, mélange d'arborescence et de groupes transversaux)

Objectifs

La première étape ici consiste à décrire la terminologie dans sa globalité afin de l'identifier comme un élément unique et précis à connecter ensuite à un réseau de ressources. Toutes les métadonnées sur votre terminologie sont prévues pour intégrer une base de données spécifique et associée à un répertoire de ressources. Parmi tous les outils dédiés à la gestion de terminologies, certains permettent l'édition de métadonnées. Par exemple ThManager vous permet, voire exige, que vous définissiez certaines informations sur la terminologie avant de pouvoir l'éditer. Cet outil génère un fichier de métadonnées en relation avec une base de données et un répertoire.

DÉFINISSEZ LES MÉTADONNÉES DE VOTRE TERMINOLOGIE

A propos des droits

Puisque vous envisagez de lier votre propre thésaurus à un réseau d'autres terminologies, vous devez vous confronter au problème des droits. Pour vous permettre de mapper ou de dupliquer des termes externes, les autres terminologies source doivent être libres d'utilisation. Et réciproquement, si vous souhaitez un mapping bilatéral, vos propres termes doivent être également libres d'utilisation. Nous vous recommandons de placer votre thésaurus sous licence Creative Commons de type CC-by-share alike, ou CCo (domaine public universel).

Tags

Pour déclarer techniquement vos métadonnées, nous vous recommandons :

- De les rassembler toutes dans des tags de type `<rdf:Description>`.
- D'utiliser des tags DC (Dublin Core) pour définir le titre (`<dc:title>`), l'identifiant (`<dc:identifier>`), l'auteur (`<dc:creator>`), les contributeurs (`<dc:contributor>`), le format (`<dc:format>`), les langues (`<dc:language>`), la description (`<dc:description>`) et le statut (`<dc:status>`).
- D'utiliser des tags de type « DCterms » pour préciser les normes utilisées par la terminologie comme SKOS et RDF (`<dcterms:conformsTo>`), en indiquant les dates de création et de finalisation (`<dcterms:created>` et `<dcterms:issued>`), et en déclarant les éléments de droits (`<dcterms:RightOwner>` et `<dcterms:license>`).

Exemple

Vous pouvez voir comment le Thésaurus Athena est documenté en regardant l'en-tête de cette page : http://www.athenaeurope.org/athenawiki/index.php/RDF/XML_version
ThManager³¹ est un outil d'édition de thésaurus SKOSifiés qui permet d'enregistrer plusieurs thésaurus dans une base de données interne.

³¹ <http://thmanager.sourceforge.net>

C1

DEFINISSEZ LES METADONNES DE VOTRE TERMINOLOGIE

Voici une capture d'écran du registre de ThManager :



Méthodes et outils

Le Dublin Core est le format le plus élémentaire pour associer des métadonnées à une ressource. De plus il permet le moissonnage des métadonnées qu'il met en oeuvre par des protocoles d'échange comme OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting)³², ce qui garantit à votre terminologie d'être visible en tant qu'élément d'un registre existant.

Le projet Linked Heritage, qui a été présenté rapidement en introduction, fournira un registre de terminologies où les institutions pourront déclarer leurs terminologies et leurs besoins.

32 OAI-PMH: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

C2 IDENTIFIEZ LES AUTRES RESSOURCES AVEC LESQUELLES MAPPER VOTRE TERMINOLOGIE

Actions

Avant de lier votre terminologie avec d'autres, vous avez besoin d'identifier celles qui peuvent en effet être intéressantes dans le cadre d'un mapping. Pour identifier des terminologies pertinentes, nous vous recommandons :

- De consulter des répertoires de terminologies à l'aide de deux critères :
 - o Domaines de description que vous pouvez lier aux vôtres.
 - o Langues qui vous intéressent dans le cadre d'un mapping de termes équivalents.
- De vérifier les droits d'utilisation de chacune des terminologies identifiées.
- D'identifier, dans la terminologie avec laquelle mapper la vôtre, les (groupes ou listes de) termes qui vous intéressent.

Objectifs

Lier votre thésaurus à un réseau de ressources implique que vous vouliez profiter d'une meilleure exploitation sémantique de toutes les relations que vous êtes en train de créer entre les termes internes à votre thésaurus et ceux externes proposés dans d'autres terminologies. Ce qui implique notamment que vous envisagez de profiter également du multilinguisme en mappant des termes qui sont équivalents d'une langue à l'autre. Si bien que nous pouvons dire qu'il y a au moins deux principales et complémentaires raisons clefs à ce que vous cherchiez d'autres terminologies :

1. Pour trouver des termes appartenant potentiellement aux mêmes domaines de description que les vôtres.
2. Pour trouver des termes potentiellement équivalents aux vôtres dans d'autres langues.

DÉFINISSEZ LES MÉTADONNÉES DE VOTRE TERMINOLOGIE

Lier votre thésaurus à un réseau de ressources implique que vous vouliez profiter d'une meilleure exploitation sémantique de toutes les relations que vous êtes en train de créer entre les termes internes à votre thésaurus et ceux externes proposés dans d'autres terminologies. Ce qui implique notamment que vous envisagez de profiter également du multilinguisme en mappant des termes qui sont équivalents d'une langue à l'autre. Si bien que nous pouvons dire qu'il y a au moins deux principales et complémentaires raisons clefs à ce que vous cherchiez d'autres terminologies :

1. Pour trouver des termes appartenant potentiellement aux mêmes domaines de description que les vôtres.
2. Pour trouver des termes potentiellement équivalents aux vôtres dans d'autres langues.

Afin d'atteindre ces objectifs, un répertoire-dépôt de terminologies apparaît comme une solution très utile. C'est pourquoi nous en avons ouvert un disponible en ligne à l'adresse : http://www.athenaeurope.org/athenawiki/index.php/Inventory_of_resources

Il est le fruit de l'inventaire de ressources terminologiques que nous avons réalisé au cours du projet Athena. Ce répertoire-dépôt est dédié aux terminologies libres d'utilisation en provenance des musées européens. « Europeana datacloud »³³ ou DBPedia³⁴ sont d'autres répertoires-dépôts qui peuvent aussi vous être utiles.

33 <http://eculture.cs.vu.nl/europeana/www/datacloud.html>

34 <http://wiki.dbpedia.org/OnlineAccess>

Exemple

Votre terminologie contient une liste de noms de lieux, et vous voudriez mapper votre liste avec une autre terminologie de référence. Vous pouvez jeter un œil à quelques répertoires-dépôts de terminologies pour voir avec quelle ressource vous pourriez mapper vos concepts. Vous pouvez regarder sur le « Europeana datacloud » quelles sont les terminologies déjà mappées et utilisées par les moteurs de recherche sémantiques :

Lier sa terminologie à un réseau

C2 IDENTIFIEZ LES AUTRES RESSOURCES AVEC LESQUELLES MAPPER VOTRE TERMINOLOGIE

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

35 TGN-Getty: <http://www.getty.edu/research/tools/vocabularies/tgn/>
36 Geonames: <http://www.geonames.org/>

Le TGN (Thesaurus of Geographic Names)³⁵ du Getty et Geonames³⁶ sont les principales ressources pour les noms de lieux. Vous pouvez donc commencer de mapper vos noms de lieux avec ceux du TGN et de Geonames.

Méthodes et outils

Vous pouvez consulter notre inventaire de ressources libres d'utilisation à l'adresse :

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

Vous pouvez aussi consulter le « Europeana datacloud » : <http://eculture.cs.vu.nl/europeana/www/datacloud.html>

Ou le LOD datacloud de DBPedia: <http://wiki.dbpedia.org/About>

Le mapping de vos concepts avec une ressource comme DBpedia vous garantit un enrichissement de votre terminologie puisque il s'agit de la version RDF des articles disponibles sur Wikipedia.

MAPPEZ VOTRE TERMINOLOGIE AVEC LES AUTRES RESSOURCES

Actions

Tout d'abord mappez les concepts de votre thésaurus avec des concepts externes en reproduisant ici ce que vous avez fait à l'étape B3 : Définissez précisément les labels exprimant les concepts. Puis mappez les termes de votre thésaurus avec des termes externes en reproduisant maintenant ce que vous avez fait à l'étape B6 : Mappez vos termes (multilingues). Si cela est possible et utile, dupliquez les termes qui vous intéressent dans votre propre thésaurus.

Objectifs

Les recommandations formulées dans la partie B : Rendre sa terminologie interopérable proposaient un mapping interne des concepts et des termes de votre thésaurus. Maintenant, dans cette partie C, nous vous invitons à en faire de même avec les concepts et les termes qui appartiennent à d'autres terminologies.

Si on compare alors avec ce que nous vous avons recommandé aux étapes B3 : Définissez précisément les labels exprimant les concepts et B6 : Mappez vos termes (multilingues), seules quelques différences surviennent.

En ce qui concerne le mapping de concepts, la différence principale est que désormais vous avez besoin d'obtenir un identifiant pour chaque terminologie concernée par votre mapping. Nous considérons que la racine de votre système de nommage des URLs devrait être utilisée comme identifiant pour votre terminologie.

Si ensuite pendant votre recherche de termes à mapper avec les vôtres, vous avez remarqué des groupes de termes que vous voudriez avoir dans votre terminologie plutôt que de procéder à un mapping avec eux, vous pouvez enrichir votre terminologie en les intégrant. L'intégration de termes peut être intéressante si vous envisagez de donner accès en ligne à votre terminologie pour que des utilisateurs puissent la consulter.

MAPPEZ VOTRE TERMINOLOGIE AVEC LES AUTRES RESSOURCES

logie pour que des utilisateurs puissent la consulter. Vérifiez alors que vous avez le droit (c'est-à-dire si les licences des terminologies sources vous autorisent à dupliquer des termes sans conditions). Ensuite exprimez-les dans votre format habituel.

Pour réaliser techniquement le mapping de termes et de concepts entre plusieurs terminologies, vous pouvez utiliser le Format Athena qui s'appuie sur SKOS.

Exemple

Chaque terminologie est conçue dans un but spécifique. Comme ce fut présenté dans la section A des recommandations, c'est pour répondre à vos besoins propres que votre terminologie a été conçue. Dès lors, vous devrez probablement porter votre attention sur la structure de votre terminologie afin de la mapper correctement avec les concepts que vous avez définis à l'aide de SKOS.

Par exemple si votre terminologie porte sur l'architecture et que vous envisagez de la mapper avec un thésaurus sur l'environnement, vous devez porter votre attention à la structure de ces deux terminologies et définir le degré de correspondance. Si vous mappez le concept « patrimoine culturel » de votre terminologie avec le même concept du thésaurus GEMET, vous pourrez enrichir l'information induite par vos concepts grâce au lien de mapping et à son inférence.

Vous pouvez voir aussi comment le Thésaurus Athena est documenté en regardant l'en-tête de cette page : http://www.athenaeurope.org/athenawiki/index.php/RDF/XML_version

ThManager est un outil d'édition de thésaurus SKOSifiés qui permet d'enregistrer plusieurs thésaurus dans une base de données interne.

MAPPEZ VOTRE TERMINOLOGIE AVEC LES AUTRES RESSOURCES

Voici une capture d'écran du registre de ThManager qui représente le concept de « patrimoine culturel » dans le thésaurus GEMET. Le mapping de concepts pourrait ici être d'un réel bénéfice pour vous puisque le concept est déjà mappé avec d'autres ressources terminologiques telles que Agrovoc³⁷ ou EuroVoc³⁸, ainsi qu'avec l'article de Wikipedia lui correspondant.

37 Agrovoc: <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>

38 EuroVoc: <http://eurovoc.europa.eu/>



MAPPEZ VOTRE TERMINOLOGIE AVEC LES AUTRES RESSOURCES

Méthodes et outils

Le Format Athena est présenté en détails à l'adresse : <http://www.athenaeurope.org/athenawiki/index.php/References>

L'URI est un élément crucial quand vous envisagez de mapper votre terminologie avec d'autres et d'en faire un élément du LOD (Linked Open Data).

Comme pour le mapping de concepts et de termes, présenté dans la section B (étapes B5 et B6), il n'existe aucun outil libre et open source pour vous accompagner dans la procédure de mapping, mais vous pouvez jeter un œil à certains outils propriétaires si besoin.

Veillez vous reporter à la section Benchmark du Wiki Athena : <http://www.athenaeurope.org/athenawiki/index.php/Benchmark#Tools>

C4 VALIDEZ L'INTEROPÉRABILITÉ

Actions

Ici la procédure de validation ressemble à celle présentée à l'étape B8 : Validez votre SKOSification. Procédez exactement comme à l'étape B8. Puis :

- Vérifiez et validez votre terminologie SKOSifiée une fois que le mapping est réalisé.
- Vérifiez l'ingestion de vos collections dans Europeana et la qualité de leur rendu une fois transmises.
- Vérifiez l'interopérabilité de votre terminologie à l'aide d'un moteur de recherche sémantique.

Objectifs

L'objectif de cette étape est de valider la procédure complète vous ayant permis de concevoir votre terminologie, de la rendre interopérable et de la lier à d'autres vocabulaires.

Cette étape permettra également le multilinguisme grâce à l'interopérabilité, indépendamment des langues utilisées dans votre terminologie.

Exemple

39 Europeana Thought lab:
<http://www.europeana.eu/portal/thought-lab.html>

Le « semantic SearchLab »³⁹ développé dans le cadre d'Europeana est un moteur sémantique qui n'est pas finalisé pour l'heure, mais dans sa version d'exploitation à venir il pourra vous permettre de vérifier certaines requêtes utilisant votre terminologie et les inférences que le mapping a générées.

Si votre terminologie est bien structurée, SKOSifiée et liée à d'autres ressources, une requête du type « Léonard de Vinci » peut retourner les résultats suivants :



C4 VALIDEZ L'INTEROPÉRABILITÉ

Les résultats sont présentés en grappes selon leurs statuts (auteur, figure représentée, ...), et les objets indexés par les descripteurs « Léonard de Vinci » ou « Leonardo Da Vinci » sont proposés comme résultats indépendamment de la langue utilisée lors de la requête.

Méthodes et outils

40 Pearltrees: <http://www.pearltrees.com>

Un outil libre comme Pearltrees⁴⁰ peut vous aider à avoir un aperçu synoptique de votre terminologie grâce à une interface graphique.

La meilleure façon de valider l'interopérabilité est de vérifier la syntaxe et la cohérence de votre terminologie, et de les tester ensuite en procédant à des requêtes plus ou moins simples.

6. Conclusion

Toutes les recommandations que nous avons formulées dans ce livret dépendent de l'état de l'art du moment.

A l'avenir de nouvelles ressources, normes et directives et de nouveaux standards et outils apparaîtront pour vous aider à gérer votre terminologie dans le contexte du LOD.

Le projet Linked Heritage est l'héritier d'Athena qui était un projet «eContent plus» visant à fournir à Europeana des objets et des collections en provenance des musées européens. Le lot de travail WP4 d'Athena était dédié aux questions de terminologie et de multilinguisme. Le lot de travail WP3 de Linked Heritage tient compte des résultats issus de l'activité du WP4 d'Athena.

Grâce à l'expérimentation menée par le WP4 d'Athena, quelques enjeux et difficultés propres à la gestion de terminologies ont été soulevés. Pour commencer à y répondre, les membres du groupe de travail du WP4 ont fait l'effort de rédiger des spécifications techniques correspondant à un outil qui se veut idéal pour la gestion de terminologies et qui soit spécifiquement dédié à des utilisateurs non experts. A la suite de cet effort le WP3 de Linked Heritage a pour objectif de développer un prototype d'une plateforme logicielle intégrée de gestion de terminologies.

Les principaux résultats attendus de la part du WP3 de Linked Heritage sont un état de l'art des terminologies en usage dans les institutions européennes, la définition des besoins fonctionnels concernant la gestion de terminologies, le développement d'un répertoire-dépôt de terminologies et celui d'une plateforme logicielle de gestion de terminologies.

Le WP3 de Linked Heritage rassemblera les résultats du WP4 d'Athena et ceux d'Europeana pour atteindre ses objectifs et fournir une solution pérenne de gestion de terminologies.

7. Annexes

7.1 Acronymes

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 International Authority File

7.2 Références

Ici vous pouvez trouver un aide-mémoire structuré de toutes les références qui ont été mentionnées au long de ce livret. Elles sont organisées en 5 catégories :

- Répertoires-dépôts
- Vocabulaires
- Modèles de données et formats
- Recommandations / directives
- Outils

7.2.1 Répertoires des dépôts

L'inventaire de ressources du D4.1 d'Athena est disponible :

- En version pdf à l'adresse : <http://www.athenaeurope.org/getFile.php?id=398>
- En version wiki à jour à l'adresse : 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 Vocabulaires

Agrovoc : <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>
Eurovoc : <http://eurovoc.europa.eu/>
GEMET est disponible à l'adresse :
<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 – la terminologie MIMO est visible en ligne à l'adresse : 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 Modèles de données et 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 référence : <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>
SKOS wiki : <http://www.w3.org/2001/sw/wiki/SKOS/Datasets>

7.2.4 Recommandations / directives

Les directives du D4.2 pour la SKOSification sont disponibles :

- En version pdf à l'adresse : <http://www.athenaeurope.org/getFile.php?id=684>
- En version wiki à jour à l'adresse : <http://www.athenaeurope.org/athenawiki/index.php/Guidelines>

« Persistent identifiers: Recommendations for institutions » : <http://www.athenaeurope.org/getFile.php?id=779>

Normes :

- BS8723: Structured Vocabularies for Information Retrieval
- ISO 25964-1 Thesauri and interoperability

with other vocabularies: Thesaurus for information retrieval

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>

7.2.5 Outils

Les résultats de notre état de l'art sur les outils sont disponibles à l'adresse : <http://www.athenaeurope.org/athenawiki/index.php/Benchmark#Tools>

Parmi l'ensemble des outils existants :

- Annocultor : <http://annocultor.eu/>
- ITM3 : <http://www.mondeca.com/Products/ITM>
- Pearltrees : <http://www.pearltrees.com>
- Pool Party : <http://poolparty.punkt.at/>
- Protege + 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 Tableau de correspondance

Le tableau de correspondance ou « mapping sheet » doit vous aider à voir rapidement à quels éléments de SKOS la structure de votre terminologie se réfère.

Ma terminologie (ex: Architecture)	Thésaurus	Format Athena	Explication	Ma terminologie (ex: Architecture)	Thésaurus	Format Athena	Explication
Microthésaurus ex: Architecture	Microthésaurus ex: Architecture	skos:ConceptScheme (classe) skos:hasTopConcept (propriété) ex: Architecture	Si votre terminologie a une entrée Architecture dans son microthésaurus, vous pouvez la décrire comme une structure de concepts correspondant au modèle SKOS.	Terme de niveau N-1 ex: Palais	Terme spécifique (TS) ex: Palais	skos:Narrower ex: Palais	Le terme spécifique de votre terminologie peut être défini ainsi : skos:Narrower concept
				Terme de niveau N+1 ex: Architecture	Terme générique (TG)ex: Architecture	skos:Broader ex: Architecture	Le terme générique de votre terminologie peut être défini ainsi : skos:Broader concept
Groupe de termes ex: Bâtiments	Tableau de thésaurus ex: Bâtiments	skos:Collections (classe) ex : Bâtiments	Si votre terminologie a des groupes de termes thématique ou d'autres groupes de termes spécifiques, les classes de collection SKOS vous permettent de reproduire ces groupes de termes.	Terme de niveau N ex: Bâtiments	Terme associé (TA) ex: Bâtiments	skos:Related ex: Bâtiments	Le terme associé de votre terminologie peut être défini ainsi : skos:Related concept
				Notes	Notes Note d'application Définition Date d'entrée ...	Notes skos:ScopeNote skos:Definition skos:HistoryNote ...	Le modèle SKOS vous offre un large choix de notes qui peuvent être facilement transposées depuis votre terminologie si c'est un thésaurus.
Terme de niveau N ex: Monument	Descripteur, vedette ex: Monumento Non-descripteur Employé pour (EP) ex: Construction	Concept - Forme lexicale préférentielle skos:Preflabel ex: Monumento@it - Forme lexicale alternative skos:Altlabel ex: Construction -Forme lexicale cachée skos:HiddenLabel ex: Hutte	Les descripteurs/termes de votre terminologie peuvent être définis comme préférentiels, alternatifs ou cachés.				



Réseau européen

proposant de nouvelles
expériences du
patrimoine culturel numérique européen

www.michael-culture.eu

Une organisation à but non
lucratif, fondée en avril 2007
sous droit belge.

- Soutient le réseau Minerva des professionnels européens œuvrant pour le patrimoine culturel numérique.

Il fédère les professionnels et experts à travers le soutien et le développement de la transmission et de l'échange d'informations, via des publications, des conférences et des événements. Ses membres viennent de toute l'Europe, regroupant des agences publiques, des institutions culturelles et des organisations privées.

- Participe à des initiatives européennes majeures

Michael Culture Association vise à améliorer la numérisation et la promotion du patrimoine culturel Européen. Il participe à la construction actuelle d'Europeana, la Bibliothèque Numérique Européenne. L'association est membre du comité exécutif de la Fondation EDL et est également partenaire de projets européens majeurs.

- Promeut et développe le portail Michael

Le catalogue européen des collections culturelles numérisées en ligne offre un accès à plus de 10 000 collections de haute qualité culturelle provenant d'archives, de musées, de bibliothèques, etc. rassemblant ainsi des initiatives nationales de toute l'Europe.



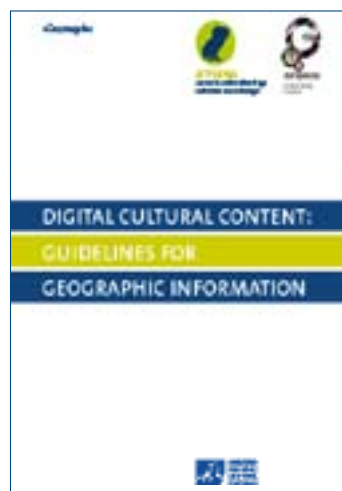
www.michael-culture.eu

contact
info@michael-culture.eu

Avec le soutien du Programme Culture de la
Commission Européenne



Dans la même série



La « terminologie » recouvre différents types de vocabulaires : lexique, thesaurus, vocabulaire contrôlé, ontologie...

Le type de vocabulaire est hautement connecté à son emploi (indexation, recherche d'information, traduction...). SKOSifier votre terminologie est une étape cruciale afin de la partager et de la connecter aux autres ressources terminologiques et ainsi améliorer le multilinguisme au niveau européen.

www.athenaeurope.org/athenawiki



<http://www.linkedheritage.org> | info@linkedheritage.org