

## Programme « AGORA »

### Rapport final – volet « Synthèse de recherche »<sup>1</sup>

CONTRAT DE RECHERCHE : AG/LL/169

ACRONYME DU PROJET: MULTITA

TITRE: **MULTILINGUAL TERMINOLOGICAL RESEARCH (FRENCH, DUTCH AND ENGLISH) FOR THE DEVELOPMENT AND INTEGRATION OF SEMANTICALLY ENRICHED SCIENTIFIC THESAURI (MULTITA)**

EQUIPE:

COORDINATEUR : **EVA COUDYZER (MRAH)**

PROMOTEUR(S): **ERIK BUELINCKX (RICH)**

CHERCHEUR(S): **VERONIQUE LHEUREUX, EVA COUDYZER, ERIK BUELINCKX**

DATE: 30/01/2015

---

<sup>1</sup> Voir art. 5.5.2 du contrat de base.

## Liens vers les sites Internet où les travaux de l'équipe dans le cadre du projet sont mentionnés:

- Site web MRAH: [KMKG-MULTITA](#)
- Site web RICH: [KIK-MULTITA](#)
- Linked Heritage, Seminar Multilingual Terminologies, April 13 2013 (Paris, France), presentation " Recommendations and Guidelines for Terminologies": [MULTITA-Linked Heritage](#)
- AthenaPlus, Access to Cultural Heritage Networks for Europeana (Rome, Italy), Signed Memorandum of Understanding between MULTITA and AthenaPlus: [MULTITA-AthenaPlus](#)
- FRANTIQ, Fédération et Ressources sur l'Antiquité (Nanterre, France), PACTOLS-thesaurus:
  - [FRANTIQ\\_01](#)
  - [FRANTIQ\\_02](#)
  - [FRANTIQ\\_03](#)
  - [FRANTIQ\\_04](#)
- Mémoires des archéologues et des sites archéologiques (MASA-consortium), Réunion de travail « Ontologies et redocumentarisation » , October 13 2014 (Nanterre, France) : [MULTITA-MASA](#)
- RKD, Rijksbureau voor Kunsthistorische Documentatie (Den Haag, Nederland), Art and Architecture Thesaurus Nederlandstalig: [MULTITA-AAT](#)
- Getty Institute,(Los Angeles, USA), Art and Architecture Thesaurus. Mention of MULTITA-project in presentation RICH: [MULTITA-Getty Institute](#)
- l'UMR8546 CNRS/ENS Archéologie et Philologie d'Orient et d'Occident – AOROC, Presentation project MULTITA (meeting November 13 2013): [MULTITA-AOROC](#)
- Fédération Wallonie-Bruxelles, AICIM-thesaurus: [MULTITA-AICIM](#)
- University of Savoie, TOTH-workshop, December 5 2014 (Brussels, Belgium), presentation and paper project: [MULTITA-TOTH](#)
- VIAA, Vlaams Instituut voor Archivering, Unified Thesaurus Haalbaarheidsstudie uitgevoerd door iMinds (Media Innovation Center), Eindrapport, augustus 2013-januari 2014: [MULTITA-VIAA](#)

\*\*\*\*\*

## Synthèse

### MULTITA Multilingual Terminological Research for the Development of Semantically Enriched Thesauri AG/LL/169

Véronique l'Heureux (IRPA), Eva Coudyzer (KMKG)

Musées Royaux d'Art et d'Histoire

Bruxelles, janvier 2015

### Résumé

MULTITA est un projet sur les thésaurus et le multilinguisme, établi sur la nécessité d'intégrer des thésaurus normalisés, enrichis et multilingues dans les établissements scientifiques fédéraux (ESF). Beaucoup d'institutions culturelles en Belgique utilisent un ou plusieurs thésaurus qui sont développés par l'organisation même. Il s'agit de listes alphabétiques de mots simples, ou bien de thésaurus structurés de manière sémantique. Souvent, ces thésaurus locaux sont développés spécifiquement pour une ou plusieurs collections. Ils varient donc entre eux, suivant les règles internes de l'organisation ou la nature des collections. Ceci veut dire qu'il est difficile de lier les différents thésaurus (ou les différentes collections) des ESF en ligne.

Les activités prioritaires du projet MULTITA étaient l'alignement et l'enrichissement de ces thésaurus locaux, ainsi que le développement de quelques micro-thésaurus (dans un domaine spécifique), normalisés et multilingues. Ces thésaurus sont des exemples de thésaurus scientifiques enrichis, construits suivant les règles et normes internationales. Le projet nous a aussi permis d'étudier les difficultés et les défis qui surgissent pendant le développement et la traduction des thésaurus.

**Key words:** thésaurus, multilinguisme, thesaurus management tool, SKOS

### 1. Introduction

MULTITA est un projet de recherche subventionné par la Politique Scientifique, un département pour la recherche scientifique du Gouvernement Fédéral de Belgique. Le projet a débuté en juin 2012 et s'est terminé en décembre 2014. MULTITA est coordonné et promotionné par deux établissements scientifiques fédéraux : les Musées d'Art et d'Histoire (MRAH) et l'Institut royal du Patrimoine artistique (IRPA).

Plusieurs organisations belges et internationales ont travaillé ensemble dans ce projet : les Musées Royaux des Beaux-Arts (Bruxelles), le Musée du Costume et de la Dentelle (Bruxelles), le Bureau d'État pour la Documentation relative à l'Histoire de l'Art (RKD, La Haye) et la Fédération et Ressources sur l'Antiquité (FRANTIQ, Paris). Le projet a été soutenu par un comité de suivi avec comme membres Jef Malliet (Province de Limbourg, Erfgoedplus), Els Angenon (Digital Collections MRAH) et Bert Lemmens (PACKED).

L'idée de Multita est née après que les Musées Royaux d'Art et d'Histoire ont débuté avec la numérisation de leurs collections il y a quelques années. Un système de gestion des collections (CMS) a été acheté et les premiers encodages ont alors commencés. À partir de ce moment, il était évident que le musée avait besoin de thésaurus pour explorer les collections dans le CMS ou en ligne. Pendant la recherche de thésaurus relatifs à l'histoire de l'art ou d'archéologie, il est devenu clair qu'il n'existait pas beaucoup de thésaurus en français, néerlandais et anglais et qu'il y avait des grandes variations de qualité.

### 2. Les données

Les partenaires culturels du projet MULTITA utilisent des vocabulaires contrôlés principalement pour gérer les collections. Cependant, les thésaurus varient selon la structure, le domaine, le nombre de concepts et le niveau de multilinguisme. Certaines organisations ont fait des efforts

pour créer des thésaurus structurés, au moins au niveau hiérarchique, suivant les normes internationales comme ISO (International Standard Organization). d'Autres utilisent des listes alphabétiques de mots simples sans relations conceptuelles. Souvent ces listes sont unilingues. Les vocabulaires de travail dans MULTITA sont:

- Les thésaurus des Musées d'Art et d'Histoire (sauf les instruments de musique): matières, techniques, noms de l'objet et références géographiques.
- Les thésaurus de l'Institut royal du Patrimoine artistique: matières, techniques, noms d'objets.
- PACTOLS-thésaurus de FRANTIQ: anthroponymes, chronologie, lieux, œuvres, peuples et sujets.
- AICIM-thésaurus de la Fédération Wallonie-Bruxelles: matières, techniques et dénominations.
- Thésaurus du Musée royal des Beaux-Arts: noms d'objet, matières et techniques.
- Art and Architecture Thesaurus Néerlandais (AAT): concepts abstraits, activités, attributs physiques, matières, objets, acteurs et organismes, styles et périodes.

Le "Art and Architecture thesaurus" a été considéré comme thésaurus de référence, tandis que les autres thésaurus pouvaient être importés et manipulés dans le projet.

### 3. Résultats

#### 3.1 Alignement de concepts des différents thésaurus

Nous avons comparé et aligné les concepts pertinents entre les thésaurus. Le « Art et Architecture Thesaurus », développé par le Getty Research Institute, a été considéré comme thésaurus de référence puisque ce thésaurus est internationalement considéré comme un norme de thésaurus sur l'histoire de l'art. En outre ces concepts sont traduits dans plusieurs langues, dont le néerlandais, chinois et espagnol. La partie néerlandaise est gérée par le RKD, partenaire des Pays-Bas dans le projet MULTITA.

*Exemple d'alignement:* le thésaurus nom de l'objet du MRAH a été aligné avec le AAT. Si des concepts équivalents existaient dans les deux thésaurus, la structure, scope note et traduction des concepts ont été adoptées. Les URI's des concepts de l'AAT – les « identifiants uniques » attribués à chaque concept en ligne – ont été adoptés dans un fichier Excel pour garantir un réemploi ultérieur (par ex. lier les collections). Les scope notes de l'AAT étaient – lorsqu'elles étaient pertinentes – adoptées dans le CMS des MRAH.

#### 3.2 Enrichissement des thésaurus locaux

Souvent de l'information importante manquaient dans les thésaurus des partenaires, comme par exemple les scope notes, traductions, synonymes et relations (hiérarchiques, associatives ou équivalentes). Une partie du travail dans MULTITA consistait à ajouter cette information manquante dans les thésaurus.

*Exemple d'enrichissement:* les thésaurus AICIM et les thésaurus PACTOLS ont été traduits en néerlandais pendant la phase d'alignement. Les traductions pour les thésaurus PACTOLS pouvaient être encodées directement dans le CMS de FRANTIQ. Ça veut dire que les traductions étaient immédiatement visibles sur le site web du PACTOLS. Les thésaurus des MRAH ont été enrichis avec les scope notes de l'AAT. Quand des concepts des thésaurus belges ne se trouvaient pas dans le AAT, ce qui était le cas par exemple pour beaucoup de concepts archéologiques, ceux-ci étaient encodés directement dans le AAT néerlandais comme « concepts candidats » dans le service web du RKD.

### 3.3 Échange d'expertise

Durant le projet Multita, on a cherché des moyens d'élargir le réseau et d'obtenir plus de connaissances et une plus grande compréhension dans le domaine des thésaurus sémantiques. *Exemple d'échange d'expertise* : le projet a conclu un contrat de collaboration avec AthenaPlus, un projet européen qui crée et rend accessible en ligne du contenu culturel européen. Ce contrat a permis à MULTITA d'assister à des réunions et de s'informer par des mailings. De plus, il était possible dans le cadre d'AthenaPlus, d'utiliser le système de gestion de thésaurus xTree, un outil indispensable dans le développement des thésaurus sémantiques. xTree est un service web qui permet de créer des thésaurus, lier des thésaurus, manipuler des concepts et structures et exporter les données dans un format SKOS (un format d'échange pour des thésaurus, nécessaire pour les publier et les lier en ligne).

### 3.4 Réemploi d'URI's

Nous avons réemployé des URI's (*Uniform Resource Identifier*) de l'AAT pour les importer dans nos propres thésaurus. De cette manière, un lien entre des concepts exacts ou inexacts (voir supra) pouvaient être garanti.

*Exemple de réemploi des URI's*: dans les thésaurus de l'IRPA et des MRAH (nom de l'objet), les URI's de l'AAT étaient adoptés pour chaque concept dans les thésaurus. En ce qui concerne l'IRPA, l'URI pouvait directement être encodé dans le CMS. Pour les MRAH les URI's ont été conservés dans un fichier Excel. Les URI's de l'AAT ont aussi été repris dans le champ « URI » dans xTree. Cela signifie que les concepts de ces micro-thésaurus ont reçu automatiquement l'URI de l'AAT.

### 3.5 Développement de micro-thésaurus

Dans le projet MULTITA nous avons aussi créé des micro-thésaurus structurés, normalisés et multilingues dans quelques domaines spécifiques.

Le développement de quelques micro-thésaurus pour l'art et l'archéologie en anglais, néerlandais et français était une des priorités du projet. Les partenaires avaient proposé leur préférence pour certains domaines spécifiques (des micro-thésaurus). En créant ces nouveaux thésaurus, on a voulu proposer aux partenaires un modèle de thésaurus scientifique multilingue, suivant les règles et normes appropriées. Il est évident que ce n'était pas possible de développer un thésaurus général, comme le AAT. C'est pour cette raison qu'on a choisi de nous diriger vers deux domaines d'intérêt :

- a. *Techniques, matières, acteurs et objets de textile* : textile est un sujet compliqué qui demande une connaissance profonde sur le sujet. Les thésaurus textile des partenaires étaient très variés, par exemple dans la manière dont les concepts ont été catégorisés. On a trouvé beaucoup d'erreurs et d'inconsistances dans les thésaurus.
- b. *Styles et périodes*: aucun partenaire ne disposait d'un thésaurus étendu de styles et périodes artistiques ou culturels. Il était difficile dans le secteur culturel de trouver des thésaurus qualitatifs sur des styles et périodes dans l'art ou l'archéologie.

Les concepts/termes, les relations et les scope notes des thésaurus des partenaires ont été inclus dans les nouveaux thésaurus. Dans le thésaurus textile nous avons aussi inclus les termes provenant du Musée de Textile de Tilburg (Pays-Bas). Comme mentionné avant, l'AAT était considéré comme une référence, bien qu'on y remarquait des fautes et des inconsistances. Celles-ci ont été corrigées dans notre micro-thésaurus et on les a fait passer à l'AAT.

Le système de gestion de thésaurus allemand xTree était à ce moment le meilleur système pour développer des thésaurus. Le système propose des fonctionnalités comme :

- Développement de nouveaux thésaurus ou importation de thésaurus existants.

- *Preferred et non-preferred terms* pour chaque concept, disponible dans toutes les langues ISO.
- *Labelled concepts*: choix entre concepts, *guide terms* (ou *node label*), hiérarchies et facettes.
- Des identifiants uniques pour chaque concept et la possibilité d'ajouter des URI's externes.
- Structure hiérarchique et relations génériques (type de), partitives (part de) ou instantielles (instance de).
- Liaison de concepts se trouvant dans différents thésaurus.
- *Notes*: notices éditoriales (*editorial*), notices historiques (*historical*), scope notes, définitions (*définitions*).

Le développement de ces deux micro-thésaurus est achevé. Les thésaurus sont offerts comme références scientifiques au secteur culturel, ou pour les importer dans les CMS locaux. Les thésaurus sont disponibles en néerlandais, français et anglais, ainsi qu'en format SKOS.

#### 4. Concernant les normes et schémas

Plusieurs normes et schémas ont été étudiés dans le projet MULTITA, dont les plus importants **SKOS** et **ISO 25964-2**.

*Simple Knowledge Organization System* ou SKOS est une norme d'échange développée par le World Wide Web Consortium (W3C) pour représenter des thésaurus en ligne, plus précisément l'information sémantique inhérente à ces thésaurus. SKOS utilise le *Research Description Framework* (RDF), ce qui permet aux applications en ligne d'échanger de l'information. Le principe central de SKOS/RDF est l'emploi d'URIs au lieu d'hypertexte (HTML) pour publier de l'information en ligne.

SKOS propose des *classes* et des *properties* qui sont définies dans un SKOS manuel.

*SKOS:concept*, par exemple, a été développé pour représenter des termes dans un thésaurus.

*Property:prefLabel* est utilisé pour représenter des *preferred terms* dans un thésaurus. Les micro-thésaurus de MULTITA ont été créés en SKOS dans xTree.

À côté de SKOS on a aussi consulté la norme ISO 25964-2. Cette publication donne des précisions sur le développement et la gestion de thésaurus, ainsi que sur l'interopérabilité entre différents thésaurus.

Quelques constatations après avoir étudié ces normes:

- La norme ISO 25964-2 n'est pas toujours claire et peut être interprétée de manière différente. Un exemple est la différence entre les *micro-thésaurus* et les *facettes*. Un micro-thésaurus est décrit comme un « designated subset of a thesaurus that is capable of functioning as a complete thesaurus ». Une *facette* est un « grouping of concepts of the same inherent category ». Suivant ces deux définitions, le concept « architecture » peut être à la fois une *facette* et un *micro-thésaurus*.
- Malgré le fait que ISO fait une différence sémantique entre une *facette* et un micro-thésaurus, SKOS ne prévoit rien pour représenter ces deux structures individuellement.
- Des *termes guides* ou *node labels* peuvent être traduits en SKOS par *skos:Collection*. Par contre, en réalité *skos:Collection* est souvent utilisé pour des facettes ou micro-thésaurus.
- L'*Art and Architecture Thesaurus* structure le thésaurus au moyen de *facettes*, *hiérarchies* et *termes guides*. Par contre, les *hiérarchies* ne figurent pas dans le schéma proposé par SKOS.

- Dans xTree il existe la possibilité de sélectionner le type de concept ou structure : *catégorie OU concept OU élément structurel*. Pour ce qui concerne l'*élément structurel*, on a ensuite le choix entre *hierarchy labels*, *nonindexing concepts*, *node label* et *facet label*. Il n'est pas toujours clair de découvrir la définition exacte de toutes ces étiquettes. La terminologie est vague.
- La norme ISO prévoit trois types de relations hiérarchiques : *générique* (un livre est « un type de » support d'information), *instance* (le Louvre est « un exemple de » musée) et *partitif* (Bruxelles est « une partie de » Belgique). Dans SKOS ces relations ne peuvent pas être représentées : toutes les relations sont de type simple générique *broader* ou *narrower*. Une partie de l'information gérée dans un thésaurus peut être perdue en SKOS.

Il n'est donc pas toujours simple de suivre ou d'implémenter des normes. Souvent on prend encore des décisions subjectives. Cela peut être problématique dans le cas où des fichiers SKOS sont importés ou manipulés dans des applications, notamment quand on a interprété le model SKOS différemment.

## 5. Problèmes et défis du multilinguisme

Pendant le développement de ces thésaurus multilingues et le travail de traduction dans nos propres thésaurus, certains problèmes liés à la problématique de multilinguisme ont été constatés. Il est important de se méfier de ces difficultés pendant la création de thésaurus multilingues ou pendant le mapping de concepts multilingues entre différents thésaurus. Quelques exemples pour illustrer :

Ex. *Entablature* peut être traduit *hoofdgestel* (IRPA) et *entablement* (AAT) en néerlandais.

- Un terme peut avoir plusieurs traductions différentes. Ces termes sont des équivalents corrects. Il est important de stimuler l'encodage de synonymes dans le thésaurus, comme termes alternatifs (*alternative terms*). Les utilisateurs auront plusieurs possibilités pour faire des recherches dans la collection.

Ex. *Pichet* en français ne peut pas être traduit en néerlandais ou en anglais.

- Il y a un manque d'équivalence entre ces langues : le terme français n'a pas de synonyme en anglais ou en néerlandais. On peut traduire par paraphrase : *small jug*, *kleine kruik*.
- *Factoring* ou la combinaison de plusieurs concepts est aussi possible : *gros bétail* (FR) peut être traduit en anglais avec *cattle* (EN) + *horses* + *goats* etc.

Ex. *Cup*, *trophy*, *bowl*, *champagne glass* se traduisent par *coupe* en français.

- Cette situation est déconseillée parce-que elle crée une ambivalence dans le thésaurus. Dans un thésaurus il faut éviter l'emploi d'un terme pour désigner plusieurs concepts. La créativité peut aider : *trophy* peut être traduit en *trophée*, *champagne glass* devient *coupe à champagne* etc.

Les problèmes et discordances entre langues sont d'une part liés aux caractéristiques intrinsèques des langues, mais peuvent d'autre part avoir un rapport avec la nature ou la culture en général. Une langue est toujours le reflet de la vision d'une culture sur le monde. Un objet peut être utilisé intensivement dans une culture et avoir par conséquent un concept pour représenter cet objet. Mais ce même objet peut-être complètement absent dans une autre culture et par conséquent n'aura pas de terme dans sa langue.

Ex. *Daubière* en français: c'est un type de marmite utilisé pour faire le *daube*, une spécialité d'origine provençale. L'objet et le pot sont utilisés en France et le concept est donc inconnu dans les autres cultures. Du coup il existe une non-équivalence pour ce concept en anglais et en néerlandais. Dans ce cas, il est préférable de simplement reprendre le mot français *daubière*.

Dans la même optique il faut remarquer que les concepts culturels n'ont pas la précision et la clarté de concepts des sciences (exactes). Un concept mathématique, par exemple, a une définition claire et limitée. La définition est logique et ne peut être interprétée que d'une manière. Un concept représente un terme. Des concepts culturels par contre sont souvent difficiles à définir et les limites sont un peu vagues. Nous rencontrons déjà ce principe chez des gens qui vivent dans le même milieu culturel et encore plus quand les nationalités, les cultures et les langues sont différentes.

Par exemple *Trapèze*. Le terme *trapèze* fait référence à une figure géométrique qui répond aux critères suivants :

- Un trapèze est qualifié de trapèze rectangle dès qu'il possède un angle droit.
- Deux angles adjacents à une même base sont égaux.
- Les côtés non parallèles sont de même longueur.
- Les deux bases du trapèze ont la même médiatrice, et celle-ci est un axe de symétrie du trapèze.
- etc.

Une figure géométrique ne peut pas être « plus ou moins trapézoïdal ». Ou bien la figure correspond à tous les critères et est ainsi reconnue comme trapèze ; ou alors les critères ne sont pas remplis et la figure n'est pas un trapèze. Il n'y a pas de zone floue parce-que le concept a des limites sémantiques bien définies.

Par exemple *Chaise*. Le terme "chaise" fait référence à un certain type de meuble conçu pour s'asseoir. Une chaise peut avoir des caractéristiques variées, mais répond généralement aux critères suivants:

- A une fonction de siège
- A généralement quatre pieds
- A généralement un dos
- Peut avoir des bras

Une chaise ne doit pas forcément correspondre à tous ces critères pour être une chaise. Il n'existe pas de limites spécifiques pour ce concept. En outre, ils peuvent être différents culturellement. Ce qui est considéré comme une chaise dans une culture, peut ne pas l'être dans une autre culture où l'objet est considéré plutôt comme un sofa.

Beaucoup de ces exemples précédents relatifs au multilinguisme et traductions ont été décrits par Aitchison et Gilchrist (Londres, 1987) :

- Équivalence exacte: *mosaïque* (FR), *mosaic* (EN), *mozaïek* (NL) ; *amulette* (FR), *amulet* (EN), *amulet* (NL)
- Équivalence inexacte: *chapeau* (FR), ***top hat*** (EN), *hoed* (NL) ; *crucifix* (FR), ***cross*** (EN), *kruisbeeld* (NL)
- Équivalence simple-à-multiple : *bol* (FR), ***small bowl*** (EN), ***kommetje/kleine kom*** (NL) ; *javeline* (FR), *javelin* (EN), ***kleine werpspeer*** (NL)



- Équivalence partielle: *carriage* (EN), *voiture* (FR), *rijtuig* (NL) ; *model* (EN), *mannequin* (FR), *paspop* (NL)
- Non-équivalence : *statuette* (FR), *statuette* (EN), *beeldje (van bepaalde hoogte)* ; *figurine* (FR), *figurine* (EN), *(menselijk) beeldje*.

## 6. Case study: *Kandelaar* (NL), *Chandelier* (FR) et *Candlestick*, *Candelabrum* (EN)

Les quatre termes dans le titre désignent un objet qui sert à porter des bougies. Par contre, il existe deux termes en anglais et un terme en français et néerlandais. *Kandelaar* et *chandelier* peuvent référer à un étendard qui porte soit une ou plusieurs bougies. Le dictionnaire Van Dale définit le *kandelaar* (NL) clairement comme : « Étendard sur lequel sont posés une OU plusieurs bougies ». En anglais il n'existe pas un terme qui correspond à un type général de porteur de bougies, malgré le nombre de bougies. L'Anglais propose deux concepts : un *candlestick* pour une bougie ; et un *candelabrum* pour deux ou plus de bougies.

Le problème est de nature conceptuelle. Le fait que dans la langue française ou néerlandaise il n'existe pas de terme pour désigner un objet ne portant qu'une bougie, démontre que les utilisateurs du néerlandais ou de français n'enregistrent pas de différence : ils identifient un porteur de bougies comme un objet ou un concept. Puisqu'il n'y a pas de différence conceptuelle, il n'y a pas de différence terminologique. Un mot/terme suffit pour décrire le concept. En anglais par contre, les deux types sont considérés comme des objets différents, du coup ils ont besoin de deux termes.

Quand un thésaurus multilingue est développé ou quand des traductions sont ajoutées à un thésaurus existant, cette donnée peut poser de soucis. Regardons l'AAT, un thésaurus conçu aux États-Unis, donc fortement ancré dans la culture anglo-saxonne. Le AAT propose deux concepts pour deux termes : *candlestick* et *candelabrum*. Les scope notes décrivent clairement que le *candlestick* porte une bougie et le *candelabrum* en porte deux ou plus. Si on traduit ces concepts vers le néerlandais il n'est pas possible dans le AAT d'utiliser le terme *kandelaar* pour désigner les deux concepts. Le AAT néerlandais a résolu le problème de telle manière :

- *Candlestick* = *Kandelaar*
- *Candelabrum* = *Kandelaber*

Cependant, cette solution n'est pas totalement correcte. D'abord le AAT choisit d'ignorer une partie de la définition de *kandelaar* : le AAT fait seulement référence à l'objet portant une bougie. Il est cependant possible que les utilisateurs finaux ne connaissent pas la définition du mot dans le AAT et utilisent *kandelaar* pour chercher tous les objets portant des bougies. Le *kandelaar* en néerlandais est considéré dans Van Dale comme un hypéronyme de *kandelaber*. Cela veut dire dans un contexte de thésaurus : *kandelaar* NT *kandelaber* ou *kandelaber* BT *kandelaar*. Un *kandelaar* est donc un «étendard pour une ou deux bougies ». Dans le AAT les deux concepts sont considérés comme *narrower terms* du concept général *candleholder*. Le terme *kandelaber* est assez formel et peut avoir une définition désuète. Il est peu probable que les utilisateurs finaux vont utiliser ce terme pour faire une recherche. Pour conclure, il est donc important de prêter attention à deux choses quand on traduit un thésaurus :

- La disparité entre un concept vs un terme et les différences entre langues : certaines langues ne disposent pas de termes pour référer à un concept (il n'y a pas de terme en néerlandais pour désigner un *candlestick*, i.e. porteur de bougies à un bras). Ou, en sens inverse, ce qui est considéré comme un concept dans une langue ou culture, est

considéré dans une autre langue/culture comme deux concepts différents (en anglais on voit la différence entre un *candlestick* et un *candelabrum*).

- Il est important de penser à l'utilisateur final quand on traduit ou quand on crée un thésaurus. Il ne suffit pas de résoudre les problèmes linguistiques avec des interventions artificielles : l'utilisateur va toujours employer sa langue comme il le fait dans la vie de tous les jours.

## 7. Conclusion

Le projet MULTITA nous a permis d'améliorer et d'enrichir les thésaurus locaux des partenaires, ainsi que de créer deux nouveaux micro-thésaurus sur le textile et sur les styles et périodes artistiques et archéologiques. Le projet comprenait une phase préliminaire qui consistait en une analyse des thésaurus des partenaires et une recherche vers un système de gestion de thésaurus etc. ; une phase de diffusion de contenu, où les thésaurus des partenaires ont été envoyés au coordinateur en format CSV ou SKOS, puis importés dans un Excel ou dans le système de gestion de thésaurus xTree (automatiquement ou manuellement). Après suivait une phase de recherche terminologique, qui convergeait principalement avec la phase d'implémentation : dans cette phase les concepts provenant de différents thésaurus ont été comparés, traduits et enrichis avec des relations sémantiques. Des nouvelles scope notes ont été écrites et échangées avec le AAT. Dans la phase d'agrégation de contenu les deux nouveaux thésaurus ont été développés dans xTree sur base du contenu (concepts) des thésaurus des partenaires. Dans le projet MULTITA, les problèmes et défis de multilinguisme ont aussi été étudiés.

Un *memorandum of understanding* a été signé avec le projet européen AthenaPlus et des collaborations étroites ont été établies avec FRANTIQ-CNRS (Paris), RKD (La Haye) et les partenaires belges. Un total de quatre-vingt réunions ont eu lieu, dont cinq avec tous les partenaires nationaux et internationaux. Le comité de suivi s'est rassemblé trois fois. Le 12 décembre 2012 un symposium a eu lieu dans le Musée du Cinquantenaire pour les partenaires de MULTITA et le 5 décembre 2014 les résultats du projet MULTITA ont été présentés durant la journée d'étude scientifique TOTh qui se déroulait à Bruxelles sur le sujet de « Terminologie et Multilinguisme ».

Les résultats pratiques du projet sont:

- Traduction en néerlandais du thésaurus PACTOLS (> 5000 concepts)
- Traduction en néerlandais du thésaurus AICIM (> 5000 concepts)
- Remise de > 200 termes candidats dans le AAT, en français, néerlandais et anglais
- Enrichissement des thésaurus de l'IRPA : ajout de scope notes (d'AAT ou propre), traductions manquantes en néerlandais et français, améliorations structurelles, encodage dans le CMS des URI's AAT.
- Enrichissement des thésaurus des MRAH : ajout de scope notes (d'AAT ou propre), traductions manquantes en anglais, néerlandais et français, améliorations structurelles, encodage dans un Excel des URI's AAT.
- Développement d'un thésaurus textile (objets, matières, techniques, acteurs) en anglais, français et néerlandais, exportable en SKOS
- Développement d'un thésaurus styles/périodes en anglais, français et néerlandais, exportable en SKOS

Il n'était pas possible de résoudre tous les problèmes des institutions partenaires concernant leur thésaurus dans le projet MULTITA, mais ce projet a définitivement apporté de nouvelles

visions sur le développement des thésaurus dans le secteur culturel. De grandes parties de thésaurus ont été améliorées, enrichies et préparées pour de futures extensions ou pour réemploi de ces thésaurus dans d'autres applications.

## 8. Sources

### Livres et articles:

AITCHISON et al. 1987

Jean Aitchison, Alan Gilchrist, *Thesaurus Construction*, London 1987<sup>2</sup>.

AITCHISON et al. 2005

Jean Aitchison, Alan Gilchrist, David Bawden, *Thesaurus construction and use : a practical manual*, London 2005.

CAGNOT et al. 2011

Stéphane Cagnot, Johann Holland, Marie-Véronique Leroi, Michael Culture Aisbl, *Your Terminology as Part of the Semantic Web. Recommendations and Guidelines for Design and Management*, Linked Heritage 2011.

DE BOER et al. s.d.

Victor de Boer, Jan Wielemaker, Judith van Gent, Marijke Oosterbroeck, Michiel Hildebrand, Antoine Isaac, Jacco van Ossenbruggen, Guus Schreiber, *Amsterdam Museum Linked Open Data*, EuropeanaConnect research project, s.d.

HUNTER 2001

Jane Hunter, *MetaNet: A metadata term thesaurus ton enable semantic interoperability between metadata domains*, Journal of Digital Information, 1(8), 2001.

MAY et al. 2011

Keith May, Ceri Binding, Doug Tudhope, Stuart Jeffrey, *Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources*, United Kingdom 2011.

MILLER 2003

Miller U., *Thesaurus and new information environment*. In: M. Drake and M. N. Maack (Eds.); *Encyclopaedia of library and information science*, 2nd ed. Boca Raon: Taylor & Francis Group.

MORSHED et al. 2011

Ahsan Morshed, Caterina Caracciolo, Gudrun Johanssen, Johannes Keizer (2011), *Thesaurus Alignment for Linked Data Publishing*, Proc. International Conference on Dublin Core and Metadata Applications 2011.

SHIRI 2012

Ali Shiri, *Powering Search. The Role of Thesauri in New Information Environments*, New Jersey 2012.

SOERGEL 2003

Dagobert Soergel, *Functions of a Thesaurus/Classification/Ontological Knowledge Base*, University of Maryland 2003.

VAN DALE, *Elektronisch Woordenboek*, versie 6.10/14.10, Van Dale Uitgevers 2014.

### Normes thésaurus:

ANSI/NISO Z39.19-2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, 2005.

ISO 25964-1, *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*, Geneva 2011.

ISO 25964-2, *Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies*, Geneva 2013.

## 9. Sites web pertinents :

AICIM-thesauri: [AICIM thesauri](#)

Art and Architecture Thesaurus: [AAT Getty](#)

Art and Architecture Thesaurus Nederlandstalig: [AAT Ned](#)

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

British Museum Object Names Thesaurus: [British Museum](#)

Carmentis: [www.carmentis.be](http://www.carmentis.be)

Erfgoedplus: [Erfgoedplus](#)

Inventarisatie Terminologiebronnen, DEN (Digitaal Erfgoed Nederland) 2010, [Inventarisatie terminology sources DEN](#)

Glossarium CEST: [www.projectcest.be](http://www.projectcest.be)

Mondeca – Linked Open Vocabularies: [MONDECA](#)

Multilingual websites and multilingual thesauri, Minerva survey 2004-2005, [Minerva survey](#)

PACKED: [PACKED](#)

PACTOLS: [PACTOLS](#)

TOTh-workshop: [TOTh](#)

## 10. Personnes de contact:

[e.coudyzer@kmsg-mrah.be](mailto:e.coudyzer@kmsg-mrah.be)

[erik.buelinckx@kikirpa.be](mailto:erik.buelinckx@kikirpa.be)