

2^{ème} réunion BEINFRAT

Institut national d'Histoire de l'art, Paris 16 décembre 2019

Présents : Sarah Amsler (SARI), Els Angenon (KMKG-MRAH), Erik Buelinck (KIK-IRPA), Marie-Christine Claes (KIK-IRPA), Dominique Filippi (INHA), Nathalie Guénette (RCIP), Thomas Hänslì (SARI), Evelyne Sinigaglia (Maison de l'Archéologie et de l'Ethnologie René Ginouvès René, CNRS).

Excusée : Eva Coudyzer (IRPA).

Erik Buelinckx rappelle la kick-off meeting, puis la réunion prévue en Allemagne qui a été annulée, d'où la demande à Dominique Filippi de nous accueillir à Paris, pour voir comment les différents partenaires ont avancé. Remerciements pour l'accueil.

1. Présentation des intervenants/lien avec le AAT-FR

Dominique Filippi : l'idée initiale était d'organiser un réseau au niveau national à l'initiative de l'INHA, pour financer la traduction de l'AAT (pour le coût salarial, une subvention partielle était prévue, et les établissements devaient compléter). A l'issue du projet, un rapport devait être établi et l'INHA devait contacter des investisseurs. [Evelyne : Frantiq est prêt à collaborer]

Marie-Christine Claes, cellule recherche en histoire de l'art de l'Institut royal du Patrimoine artistique, et responsable des thésaurus des banques de données ADLIB (logiciel utilisé par l'IRPA) .

Evelyne Sinigaglia : directeur adjoint de la Maison de l'Archéologie et de l'Ethnologie René Ginouvès et directeur adjoint de Frantiq pour le thésaurus Pactols multilingue, qui est en pleine restructuration. Également membre de Dariah (*Digital Research Infrastructure for the Arts and Humanities*) qui a développé le « back-bone thesaurus ».

Nathalie Guénette : remplace Madeleine Lafaille. Elle a travaillé avec Madeleine depuis octobre 2017 au RCIP. Auparavant, elle travaillait au Musée canadien d'histoire (collections d'ethnographie, d'histoire, d'art autochtone) et y a développé une expertise en catalographie, thésaurus, terminologie.

Erik Buelinckx : responsable de l'infodhèque de l'IRPA : photothèque, bibliothèque et dossiers d'intervention (sur les œuvres d'art) et dossiers européens, car intérêt pour la mise en commun de données.

Le thésaurus du Getty est en cours depuis les années 1990'. Il y a eu un manque d'intérêt ces dernières années, puis un regain. Pour la Belgique, il y avait une possibilité de projet. En juin 2018, Erik a travaillé avec Eva pour rédiger le projet BEINFRAT.

Sarah Amsler : travaille au projet *Swiss Art Research Infrastructure* (SARI), Universität Zurich, plateforme pour la traduction de l'AAT.

Thomas Hänsli : Directeur de la *Swiss Art Research Infrastructure* (SARI), Universität Zurich. Est en contact avec Patricia Harping, directrice des bases de données du Getty (AAT, ULAN, TGN). La Suisse étant multilingue, elle a manifesté un intérêt pour des plateformes qui peuvent gérer le multilinguisme. Thomas et elle sont donc heureux de partager leur projet.

Els Angenon : responsable de la gestion des collections. Les MRAH utilisent les thésaurus TGN et AAT du Getty. Elle travaillait avec Eva, mais depuis son départ gère le thésaurus et s'occupe aussi de la traduction de l'AAT.

2. Informations administratives par Erik Buelinckx

Le projet permet de financer les voyages pour les rencontres (au départ, il y a eu quelques difficultés pour rembourser les frais du Canada, mais c'est arrangé). L'idée était d'avoir un budget pour voyager. Il faut justifier les déplacements et frais d'hôtel (une nuit pour les réunions en Europe), mais pour les repas, un *per diem* de 100 euros est octroyé sans devoir les justifier. Les activités ne coûtant pas beaucoup, Erik va demander de pouvoir dépenser davantage par réunion. Le budget prévu pour les USA ne sera pas dépensé. Quatre nuitées à Los Angeles sont prévues. Avec l'argent économisé, Erik envisage d'organiser une réunion supplémentaire. Les instructions de Belspo (Politique scientifique fédérale belge) peuvent se comprendre de diverses manières, il faudra donc préciser ce qui est possible ou pas. Erik avait prévu un montant d'environ 1000 euros pour acheter un outil de gestion de projet (Basecamp), mais on utilise maintenant un outil gratuit (SLACK), et donc il faut voir si on peut donner un peu d'argent pour l'outil suisse. La conférence finale se tiendra aux Musées royaux d'Art et d'Histoire, normalement fin 2020 (TBC). Il faut voir si on a économisé suffisamment d'argent pour inviter un key-note speaker.

Erik va envoyer les documents pour remboursement des tickets de trains ou avion, hôtels et *per diem*.

3. Powerpoint d'Eva Coudyzer sur la traduction de l'AAT en néerlandais et en allemand

Il y a eu des hauts et des bas selon les institutions partenaires. La Flandre avait de l'argent pour payer les traducteurs. Les Pays-Bas ont fourni des moyens en nature (temps de personnel). Au début, il y avait une licence à payer au Getty. Pour les Pays-Bas, c'était le RKD (*Rijksdienst voor Kunsthistorische Informatie*) et un ministère qui participaient mais pour la Flandre, les partenaires étaient des universités, ce qui n'était pas logique. Erik a proposé que l'IRPA participe pour avoir un institut national (en Flandre, aucun organisme régional n'étant intéressé), et donc l'IRPA a participé.

Depuis quelques années, le AAT a été publié en « linked open data », on ne se tracasse donc plus des licences. Tout est sur le site web : <http://website.aat-ned.nl/home>

Le système a changé : auparavant, c'était basé sur des *termes*, maintenant c'est *concept-based*. Cependant, l'AAT n'est pas entièrement conceptualisé.

Le site de l'AAT-Ned est développé par Adlib.

- Termes : il y en a 150737 en EN, 63198 en NL, 11824 en FR.

[rappel : au début des années 1990', le RCIP a fourni 3000 termes, mais c'était une opération unique].

Les Brésiliens ont fourni 247 termes en Portugais. Ils ont cherché des budgets mais n'en ont pas trouvé.

- Concepts : 45235 en EN, 37298 en NL, 8951 en FR.

Réunions à Anvers à partir de 2012. Les provinces ont été supprimées depuis en Flandre. Les institutions provinciales sont passées soit sous l'administration de la Région flamande, soit sous l'administration de villes importantes où elles ont leur siège, d'où un peu de flottement dans les ex-institutions provinciales. Il y a quelques réunions par an, et chacun consacre au projet soit du temps libre, soit du temps pris sur le travail dans son institution. Il faut en outre avoir la possibilité d'aller aux réunions. A l'Université d'Anvers, c'est surtout la bibliothèque qui est active. Le travail de l'AAT-terme, sa définition en EN et NL, ainsi que trois sources. Erik a négocié pour que l'IRPA devienne une source reconnue. Quand c'est fait, il faut envoyer le tout, et les termes sont renvoyés avec une décision : OK, refusé, à retravailler. Parfois des termes locaux sont acceptés. On ne sait pas toujours pourquoi un terme est refusé, c'est un peu une « boîte noire ».

Il ne faut pas oublier que l'IRPA a son thésaurus, plus grand que celui du Getty, et n'utilise pas (encore) celui du Getty. Mais le numéro d'identifiant du Getty est encodé pour pouvoir donner à terme un biais d'accès aux anglophones et aux autres langues qui utilisent l'AAT, le chinois par exemple. Les métadonnées ne seront pas en anglais, mais au moins, ils auront les images.

Els : nous avons aussi notre propre thésaurus *nom d'objet et géographie*, mais depuis un an, nous avons une possibilité de lier avec l'AAT et le TGN (à côté des thésaurus locaux). L'identifiant est repris dans le module thésaurus du système, mais la liaison des concepts qui existaient dans le système avant l'introduction de l'AAT et le TGN, doit être fait de façon manuel.

Erik : xTree est développé par DigiCult en Allemagne, c'est une plateforme pour la gestion des thésaurus. On ne sait pas faire des imports/exports de thésaurus de soi-même, il faut toujours passer par DigiCult. On est donc totalement dépendants d'eux. Le RKD avait reçu de l'argent pour utiliser cette plateforme, mais la rédaction a décidé de ne plus payer pour xTree parce que ce n'était pas utile pour les adaptations/changements dans le AAT (seulement traduction). Maintenant, les termes à traduire sont fournis dans un export Excel via Google drive.

AAT : écran « proposer un terme ». Les nouveaux termes sont souvent proposés par des spécialistes. Les anglophones utilisent plutôt le pluriel, tandis qu'aux Pays-Bas, en Belgique et en France, on utilise plutôt le singulier. Dans les fiches de concept, on trouve le singulier et le pluriel, le masculin et le féminin.

Dominique : y-a-t-il toujours des volontaires ? Erik : Oui, mais en même temps on cherche de l'argent pour avancer autrement.

Monica Hagedorn-Saupe¹ est directrice de l'Institut pour la recherche muséale à Berlin.

¹ <https://zugang-gestalten.org/programm/biographie-prof-monika-hagedorn-saupe/>

Dominique : qui héberge l'AAT-DE ? Thomas : c'est une association privée avec quelques institutions au Nord de l'Allemagne. Dominique : pour l'AAT-Fr, il faudra créer une instance qui assure la gouvernance, qui porte le projet, qui l'affiche. Avoir différents sites web n'est pas la meilleure solution, car l'information serait diluée. Il faut ménager les susceptibilités. MC : il faut éviter un site avec un suffixe national. Erik : il y a peut-être de l'argent à tirer de la francophonie, et en même temps cela internationaliserait davantage le projet. Evelyne : cela pourrait intéresser des pays africains. MC : et le Vietnam ? Dominique : certains pays n'adhèrent à la francophonie que pour avoir une visibilité, comme la Roumanie, mais ne sont pas très actifs.

Erik : la Flandre a fait un audit, on ne sait pas comment cela va évoluer pour l'AAT-Ned. La province du Limbourg (*Provincie Limburg*) a importé l'AAT en 2016, mais la Flandre va implémenter un nouveau système utilisable par toutes les provinces.

Sarah : les bibliothèques l'utilisent-ils pour l'indexation ? J'ai seulement trouvé l'ANET² qui le fasse (ce portail est celui des bibliothèques de l'Université d'Anvers, un peu élargi à d'autres institutions. Il est probable que la plupart des institutions ont leur propre thésaurus, d'où la question : jusqu'où doit-on aller au niveau national ?).

Dominique : il faut parfois être prudent. Pour les Chinois, la statue de *Bouddha* n'est pas un objet, mais c'est « le Bouddha ». MC : A l'IRPA, on a « lieux mythiques » dans le domaine géographie du thésaurus, où se trouvent l'Olympe, le Styx, l'El Dorado, mais pas Paradis, Purgatoire et Enfer qui sont dans le thésaurus iconographie chrétienne.

Erik : l'AAT sera donc souvent un méthode « d'enrichissement », mais ne remplacera dans beaucoup de cas pas les thésaurus des institutions.

4. Présentation du site SARI

a) Tom Hänsli (démonstration en ligne)

La Suisse a les mêmes problèmes que ceux évoqués par Erik pour la Belgique. Multilinguisme (FR, DE, IT). Il y a eu déjà un *dataset* de traductions livré à Getty dans le passé.

Il nous faut donc une plateforme « open source ». Le problème est la gestion du format des données. Les Allemands fournissent des données en Excel mais les Suisses ne veulent pas le faire. Le processus de transmission est trop long. Il faut un processus de workflow complet (de traduction au validation), en format RDF. Il faut donner des API qui permettent de retirer les données dans différents formats. L'outil a été développé par Metaphacts.

[roadmap par Thomas]

En janvier, on va ajouter les sources et de nouveaux concepts, et faire un update du Getty.

² <https://www.uantwerpen.be/nl/projecten/ Janet/> (réseau d'une vingtaine de bibliothèques scientifiques de la région d'Anvers et du Limbourg).

b) Présentation du site du SARI par Sarah Amsler

La dernière fois [kick-off meeting], nous avons examiné les formulaires. Depuis, nous avons surtout travaillé sur les sources, sur un interface public et sur un outil pour la traduction. Une personne non identifiée devrait pouvoir accéder à l'outil.

<https://aat.swissartresearch.net/resource/Default:Start>

Login comme « traducteur »

Sarah se logge d'abord en tant que traducteur. On a le choix entre recherche par terme et recherche hiérarchisée. Recherche sur le terme « *champagne bottles* » dans le AAT-hiérarchies → fiche concept avec hiérarchie et autres termes. Il existe des filtres pour mieux naviguer. Le traducteur voit immédiatement les concepts à traduire. Les règles Getty sont acceptés, e.a. l'ontologie Getty est respecté.

L'outil a été conçu par Metaphacts. C'est une plateforme open source, mais il existe des options payantes.

Les termes à traduire sont présentés de manière hiérarchique, c'est plus facile pour le traducteur qui voit dans quel contexte ils sont utilisés.

L'ontologie de base est conservée, mais par exemple on peut choisir le genre (masculin, féminin), le singulier ou le pluriel, ou la langue.

On peut ajouter des particularités linguistiques, comme par exemple le français canadien (fr-ca) ou le français suisse.

Chez Getty, c'est Rob Sanderson qui est le point de contact pour le lien vers SKOS³ et LOD ou le standard CIDOC-CRM.

Il existe un historique du workflow. On peut faire une requête pour chercher les termes qui ne sont pas encore attribués, pour ensuite les attribuer à un traducteur. Un traducteur peut aussi regarder quels sont les définitions non traduites, et demander à les faire.

On a séparé les concepts en « containers » par langue. C'est pratique parce que tout le monde ne travaille pas au même rythme dans les différentes langues.

Evelyne : il existe déjà deux vocabulaires du ministères qui sont alignés avec le Getty. Est-il possible de les voir en Json ?

³ Simple Knowledge Organization System (TMP – Terminology management platform). *TMP - SKOS* : http://cidoc2014.de/images/sampleddata/cidoc/papers/F-1_Coudyzer_paper.pdf et https://fr.wikipedia.org/wiki/Simple_Knowledge_Organization_System

Les sources

Sarah : un des points sur lesquels on a travaillé, ce sont les sources. Il y a une *scroll-list* : on doit taper au moins trois caractères, et l'on obtient les sources du Getty et nos sources propres. Nous avons la possibilité de créer une nouvelle source (en respectant le « Chicago style source citations»), puis on l'envoie aux administrateurs pour qu'ils la valident. Il est important de bien gérer le processus de création des sources, mais nous n'avons pas envie d'attendre qu'ils valident une source avant de la réutiliser.

(MC doit envoyer à Thomas et Sarah la liste des termes dont on a déjà la traduction en français).

Nathalie nous informe l'existence du logiciel bibliographique Zotero.⁴

[IRPA : il faut réfléchir à mieux structurer les sources. Actuellement on encode « source, page » dans le même champ, d'où il grande multiplicité d'entrées (c'est un champ libre). Il y a bien un autre champ, « numéro », mais on n'y encode que les identifiants de glossaires en ligne (AAT, dictionnaire néerlandais Van Dale, etc.). Il faudrait un champ distinct pour la page et une liste d'autorité.]

Types de sources :

- *Local* : approuvée localement
- *Proposed* : vient d'être créée. L'administrateur doit la vérifier.

Les sources doivent être visibles pour tous pour éviter les doublons. On peut sauter d'une source aux termes qui les ont utilisées.

Login comme « admin »

Sarah poursuit la démonstration en tant qu'administrateur. L'administrateur voit tous les termes encore à traduire et ceux qui ont été traduits déjà. L'administrateur assigne des traductions aux traducteurs.

Il y a un administrateur par institution. Dans « *content review* » il faut donner la tâche de contrôler à quelqu'un avec un profil adéquat pour évaluer le contenu (un spécialiste). Donc l'administrateur doit valider (si les trois sources sont présentes, si c'est fait en « Chicago Style » (= contrôle formel), puis cela va chez un spécialiste du domaine, pour valider le contenu).

On peut aussi laisser un commentaire.

Nouveaux termes/lots de termes à importer et exporter

Des lots peuvent être importés et exportés. Le fonctionnalité d'introduire de nouveaux termes est prévu pour un autre trajet. C'est finalement Getty qui va décider si un terme est accepté ou non.

Dominique : ne vaut-il pas mieux avoir l'accord pour le concept avant de faire tout le travail, s'il risque d'être refusé par le Getty ?

Thomas : il y a des termes qui peuvent vivre dans l'ontologie même s'ils sont refusés par le Getty.

⁴ <https://carrefour.uquebec.ca/zotero>

Erik : il y a des institutions qui veulent utiliser le Getty mais qui ont besoin de termes en plus, donc c'est important qu'ils puissent proposer des termes.

Livraison à Getty

Thomas : les traductions peuvent être livrées à Getty via un SPARQL endpoint (RDF repository).

Phase de tests

Dominique : il est possible de tester la plateforme, mais ensuite il faut désigner des administrateurs et des éditeurs. On ne prévoit pas un « bac à sable ».

Thomas : on fait un backup. Il serait utile d'intégrer d'autres thésaurus.

Sarah : on y travaille depuis mars, mais il y a encore des tests, des mauvaises données ajoutées.

Thomas : il faut établir un processus de feed-back très rapide. On espère que fin janvier cela ne sera plus seulement un grand projet, mais l'outil terminé. Vous pouvez toujours nous envoyer des propositions. La plateforme sera testée par une organisation italienne pour la traduction en italien.

En tout cas, il faut une équipe pour tester la plateforme et le workflow, Sarah et Thomas sont administrateurs/éditeurs et ne peuvent pas tout tester. Il y a un budget pour des autres organisations pour tester/travailler avec la plateforme.

Le plateforme est en « open source », donc un autre développeur peut continuer à développer s'il est intéressé.

5. Présentation de Nomenclature par Nathalie Guénette

Traduction de Nomenclature

1992 : Système de classement des collections historiques du service canadien des parcs.

1997 : Voir et nommer nos objets de collections (dictionnaire visuel)

2009 : site *Dictionnaire descriptif et visuel d'objets de Parcs Canada*

2015 : le RCIP propose d'harmoniser le Dictionnaire visuel et Nomenclature

À partir de 2015 : Rosemary Campbell est retraitée de Parcs, mais travaille encore 82 jours par an. Elle s'assure que les correspondances fonctionnent bien. Elle reprend les termes équivalents. Il y a actuellement 3000 termes anglais équivalents aux termes français. 6500 sont revus par un terminologue. Validation avec Madeleine Lafaille pour éliminer les doublons (fichier Excel). [Exemple : article de jardin → tonnelle]. On est obligé d'ajouter beaucoup de qualificatifs pour éviter les doublons.

Validation et transition vers *Poolparty*⁵. Erik : c'est assez cher. Nathalie : *Poolparty*, c'est seulement le back. Erik : Êtes-vous contents de ce logiciel ? Nathalie : Oui, mais on a été aidés pour faire le modèle

⁵ https://en.wikipedia.org/wiki/PoolParty_Semantic_Suite

des sources en connexion avec Zotero directement dans *Poolparty*. En bleu = SKOS, en vert = Nomenclature. En quatre ans, on y est arrivé.

14917 termes ont été envoyés au Getty. En juillet 2019, 5843 termes étaient disponibles dans le Getty grâce à une correspondance automatisée.

Exemple de hiérarchie : <https://www.nomenclature.info/nom/743>

On peut faire un requête SPARQL pour trouver les termes de Nomenclature [Thomas va envoyer 4900 concepts].

Variantes linguistiques (canadien) : le nombre n'a pas encore été évalué. Quand tout sera implémenté, cela sera plus stable. On voudrait ajouter l'espagnol. 2018 était l'année Unesco de protection des langues autochtones.

Sources

Culture.fr

FR.DBPedia

Integrated Authority File

Parfois des termes sont difficiles à traduire, et même à identifier, parce qu'on n'a pas ce genre d'objet partout (par exemple *chaise de bar de boissons gazeuses = Soda fountain chair*), donc on invente des traductions. Evelyne : on a le même problème dans PACTOLS pour les traductions français-arabe.

Voir *top terms* dans Nomenclature : souvent basé sur la fonction d'un objet.

Dominique : c'est quoi le prix de *Poolparty* ?

Nathalie : quelques millier de dollars canadiens (20.000 par an, contrat de 3 ans).

Erik : Eva a regardé, c'est très cher.

Nathalie : cela permet de gérer le workflow. En 2006, il n'y avait pas beaucoup de logiciels de gestion de thésaurus qui offraient ces possibilités.

6. Entretien de suivi

Erik : quand on veut demander des ressources supplémentaire on doit pouvoir dire combien de temps ça prend de traduire et livrer un terme. Un collègue de Sarah a estimé que ça prend un demi-journée par concept (intégration → validation).

La plateforme de traduction créée par SARI peut être testé quand le développement est complètement finalisé. Maintenant on est dans la phase de échanger et préparer.

À la fin du projet on peut chercher des fonds financiers supplémentaires pour mettre en marche la plateforme.

Est-ce que chaque participant de BEINFRAT fait une liste de tout les termes qui existent en français ?
Après la réunion à Los Angeles on aura une nouvelle entretien pour voir comment on continue.
BNF/INHA n'a pas de ressources pour tester.

International Thesaurus Group Meeting (ITWG), Getty, 6-7 février 2020

Eric présentera le projet.

Thomas et Sarah présenteront le SARI.

Thomas : ils sont en train de tout changer (Robert Sanderson) : *Linked Art* = format standard global pour l'art (<https://linked.art/>). On va essayer de contacter Rob Sanderson.

Communication

Tout le monde est d'accord de continuer à utiliser SLACK comme moyen de communication entre partenaires.

7. Visite guidée de la bibliothèque par Dominique