



## **Seminaire sur le multilinguisme et la terminologie**

18 Avril 2013

Auditorium Colbert

Institut national d'histoire de l'art (INHA)

<http://www.inha.fr>

2 rue Vivienne

75002 Paris

### **Description :**

Le multilinguisme est essentiel, surtout dans un contexte européen. Les institutions culturelles sont face à des milliers et des milliers de contenus numériques. De nombreuses initiatives européennes se sont développées afin d'assurer et de promouvoir l'accès à ce patrimoine culturel numérique qui suggère implicitement un accès multilingue aux collections. Les institutions culturelles utilisent dans une grande majorité des cas leurs propres vocabulaires contrôlés, thésaurus, classifications, ontologies ou tout autre type de terminologie. Si les projets et les initiatives européennes sont un moyen de réunir ces différentes terminologies, les technologies du Web sémantique (avec un format interopérable comme SKOS (Simplified Knowledge Organisation System)) sont un moyen de favoriser le multilinguisme et un premier pas vers le Linked Data, le principe des données liées.

Le but de ce séminaire est de présenter des projets nationaux et européens où la question de la terminologie et le multilinguisme est traitée. Certains projets créent des terminologies multilingues sur des sujets très spécifiques, certains autres se concentrent sur la mise en œuvre d'outils pour créer, gérer et réutiliser ces terminologies. Le séminaire présentera l'une des réalisations du lot de travail (WP3) du projet européen Linked Heritage avec la plate-forme de gestion de terminologies (Terminology Management Platform : TMP) et réunira d'autres initiatives européennes et nationales qui créent et développent des ressources terminologiques multilingues ou des outils pour leur gestion ou réutilisation.

**Inscription :** <http://culture-terminology.eventbrite.com>

### **Programme**

9h-9h30 : Inscriptions des participants

9h30 – 9h45 : Accueil - Claire Lamboley (Chef du service de la coordination des politiques culturelles et de l'innovation, MCC)

Terminologie et multilinguisme : un défi à l'échelle européenne

9h45 – 10h : Introduction: le multilinguisme dans les projets européens - Rossella Caffo (MIBAC)

10h – 10h20 : Alignement multilingue et terminologique chez Europeana – Vivien Petras (Université Humboldt de Berlin)

10h20 – 10h40 : Recommandations et bonnes pratiques pour la terminologie (projets ATHENA et Linked Heritage) - Marie-Véronique Leroi (DREST, MCC) et Eva Coudyzer (KMKG)

*10h40 – 11h10 : Pause Café*

#### Des outils pour la création, l'harmonisation et la gestion de terminologies multilingues

11h10- 11h30 : Terminology Management Platform (TMP) (Linked Heritage) -Marie-Véronique Leroi (DREST, MCC) et Florent André (Université de Savoie)

11h30-11h50 : SIERA : un projet de partage de connaissance multilingue – Christophe Roche (Universidade Nova de Lisboa - Faculdade de Ciências Sociais e Humanas)

11h50-12h15 : Questions & discussion

*12h15- 13h30 : Temps libre pour le déjeuner*

14h-14h20 : GINCO, un outil ouvert pour la conception de vocabulaires – Katell Briatte (DGPat, MCC)

14h20 – 14h40 : La publication et l'exploitation des structures de connaissance avec OpenSKOS – Mark Linderman (Picturae Database publishing)

#### Thésaurus, classifications et autres ressources terminologiques spécialisées et multilingues

14h40 –15h: Thésaurus multilingue sur la photographie (Europeana Photography) – Nasha Van Steen (KMKG)

15h – 15h20 : Rôle, outils d'édition, exploitation et exposition du Thésaurus multilingue MIMO– Rodolphe Bailly (Cité de la musique)

15h20 – 15h40 : Partage Plus – Des vocabulaires contrôlés pour l'enrichissement de métadonnées internationales permettant l'Art Nouveau dans Europeana – Christiane Pagel (Fotomarburg)

*15h40 –16h10 : Pause Café*

#### Linked Data : Réutilisations et applications

16h10-16h30 : Semanticpédia : Enjeux culturels et linguistiques autour du Web de données – Thibaut Grouas (DGLFLF, MCC)

16h30 –16h50 : L'utilisation de VIAF dans data.bnf.fr: créer des liens pour importer des informations – Romain Wenz (Bnf)

16h50 – 17h10 : Graphes sociaux (Eclap) – Paulo Nesi (Université de Florence)

17h10 -17h30 : Questions, discussion ouverte et conclusion

Avec le soutien de :



## **Terminologie et multilinguisme : un défi à l'échelle européenne**

### Le multilinguisme dans les projets européens

Europeana – en tant que portail d'accès des biens culturels multilingues de l'Europe - reflète la riche diversité culturelle de l'Europe et incarne leur expression dans une forme tangible. La gestion du multilinguisme et des terminologies est l'un des défis majeurs pour améliorer l'accès à Europeana et au patrimoine culturel numérique.

Depuis 2006, plusieurs projets européens coordonnés par l'ICCU ont traité ou traitent de ces questions importantes.

MINERVA Plus a mené une enquête d'envergure afin d'avoir un aperçu de la situation concernant l'usage de la langue sur les sites internet culturels et de savoir si ces sites ont utilisé des outils de recherche tels que les vocabulaires contrôlés ou des thésaurus et si des outils multilingues étaient disponibles.

Le service multilingue MICHAEL, grâce auquel les utilisateurs sont en mesure de trouver et d'explorer les collections numériques européennes du patrimoine culturel, fait usage de terminologies multilingues partagées.

ATHENA a développé et élargi l'intégration de tous les types de ressources utilisées par les musées et les autres secteurs du patrimoine, se focalisant sur les terminologies et les outils sémantiques.

Linked Heritage et ATHENA Plus travaillent sur l'élaboration de guides et le développement d'une plate-forme de gestion des terminologies, un registre qui servira d'archive de terminologies, consulté, utilisé et lié pour le secteur du patrimoine culturel dans un environnement sémantique.

Afin de mieux exploiter les résultats au niveau européen, il est très important de partager les résultats de ces projets et d'activer la coopération avec d'autres institutions et projets actifs dans ce domaine: MIMO, Partage, les collections d'Europeana 14-18 etc.

**Rossella Caffo** est la directrice de l'Institut Central pour le Catalogue collectif des bibliothèques italiennes (ICCU) à Rome, qui est un institut du ministère italien du Patrimoine et des Activités Culturelles. L'ICCU est en charge de la délivrance, la traduction et la personnalisation des règles et des règlements pour le catalogage des documents de la bibliothèque de tous types, des manuscrits au multimédia, et fournit aux bibliothèques des informations et des guides relatifs aux normes de la bibliothèque.

Elle coordonne les grands projets numériques du patrimoine culturel au niveau national, par exemple, Internet Culturale, qui a développé un portail d'accès intégré aux ressources numériques des bibliothèques italiennes, archives et autres institutions culturelles, et CulturaItalia, le portail national sur le patrimoine culturel, le principal contributeur italien à Europeana. Au niveau européen, elle a été la coordonnatrice de projets tels que MINERVA, MICHAEL, ATHENA DC-net, INDICATE, et elle est la coordonnatrice de DCH-RP, LINKED HERITAGE et ATHENAPlus.

Elle représente aussi l'Italie dans le groupe d'experts des États membres sur la numérisation et la conservation numérique du contenu culturel numérique, un groupe mis en place par la Commission européenne en 2007.

### L'alignement de terminologie multilingue à Europeana

Cette présentation donnera un bref aperçu des défis posés par les terminologies multilingues à la bibliothèque numérique Europeana et des différentes approches pour l'alignement des terminologies qui ont été développées au cours des projets Europeana. La présentation donnera également un aperçu des travaux sur les terminologies multilingues dans le tout nouveau portail Europeana (www.europeana.eu), basé sur le modèle de données Europeana qui s'appuie lui-même sur le principe des données liées (Linked Data).

**Vivien Petras** est titulaire d'un doctorat en gestion de l'information et des systèmes de UC Berkeley, Etats-Unis et est maintenant professeur de recherche d'information à l'École Library and Information Science, Université Humboldt de Berlin. Elle est la coordinatrice du projet FP7-ICT DM2E (Manuscrits numériques à Europeana). Elle dirige également les tâches sur l'accès multilingue dans le Réseau Thématique Europeana v2.0 et participe au projet FP7-ICT PROMESSE (laboratoire de recherche participative pour l'évaluation des systèmes d'information multimédia et multilingues). Ses intérêts de recherche se situent dans les domaines de l'organisation de l'information pour assurer l'interopérabilité et la recherche d'information (multilingue).

### Recommandations et bonnes pratiques pour la terminologie

Marie-Véronique Leroi et Eva Coudyzer présenteront le travail réalisé au sein des projets ATHENA et Linked Heritage. ATHENA a fourni un ensemble de recommandations et de lignes directrices pour la gestion de la terminologie et du multilinguisme. ATHENA a été principalement consacré aux musées européens qui pour la plupart utilisent leurs propres vocabulaires pour la description de leurs collections. L'objectif des travaux sur la terminologie faits dans le cadre d'ATHENA était la sensibilisation des institutions culturelles au principe des données liées et les technologies du Web sémantique. Comme Europeana le recommande, la priorité a été donnée au format SKOS (Simplified Knowledge Organisation System). Ce travail s'est poursuivi dans le cadre du projet Linked Heritage où le travail théorique d'Athena a été mis en pratique avec le développement de la plate-forme de gestion des terminologies (TMP).

**Marie-Véronique Leroi** a une formation en traitement automatique des langues (TAL) et elle est chef de projet au Secrétariat Général du Ministère français de la Culture et de la Communication, dans le département de la recherche, de l'enseignement supérieur et de la technologie (DREST). Marie-Véronique Leroi est responsable de l'instance nationale française du projet MICHAEL-Patrimoine Numérique et dirigeait le lot de travail 4 d'Athena dédié aux terminologies et co-dirige le lot de travail 3 de Linked Heritage avec KMKG.

**Eva Coudyzer** a obtenu une maîtrise en histoire de l'art et archéologie à l'Université Libre de Bruxelles (VUB). Depuis 2009, elle a été employée par les Musées royaux d'Art et d'Histoire (KMKG) en tant que responsable du développement de thésaurus. Elle a travaillé sur les métadonnées, terminologies multilingues et les descriptions des objets. Elle est également coordonnatrice du projet de recherche terminologique multilingue (français, néerlandais et anglais) pour le développement et l'intégration des thésaurus scientifiques sémantiquement enrichis (MULTITA), financé par la Politique scientifique fédérale belge.

### **Des outils pour la création, l'harmonisation et la gestion de terminologies multilingues**

#### Terminology Management Platform (TMP – Plateforme de gestion de terminologies)

La TMP (Terminology Management Platform) est une plate-forme de gestion de terminologies développée dans le cadre du projet Linked Heritage. La TMP a été développée selon les processus et étapes fonctionnelles définies dans le cadre du projet ATHENA. La TMP est conçue comme une boîte à outils permettant d'enregistrer un thésaurus, une classification, une liste de termes ou tout autre type de terminologie dans un entrepôt dédié, de transformer cette terminologie dans le format SKOS, de l'enrichir et d'aligner des concepts avec ceux d'une autre

terminologie. Dans le même esprit des données liées (Linked Data), il s'agit ici de lier les concepts.

### SIERA: Un projet de partage de connaissances multilingues

La croissance rapide des marchés transfrontaliers et des préoccupations mondiales crée une énorme demande pour faciliter le partage des connaissances entre les sociétés. Numériser l'information et permettre que cette information soit accessible et compréhensible par d'autres communautés est d'une importance fondamentale. La diversité des langues, des cultures et des normes sont les principaux obstacles à l'échange de connaissances et de consommation.

L'objectif général du projet est de renforcer la coopération scientifique plus étroite entre l'UE et les scientifiques palestiniens dans le domaine des technologies multilingues et multiculturelles de partage des connaissances. Il comprend:

- Une exploration du support arabe dans le portail de partage de connaissances multilingue;
- Une extension de thésaurus multilingues / ontologies pour les domaines de la culture et de l'écologie (localisation arabe du thesaurus Michael, SKOSification du thésaurus Bethléem)
- Établir un cadre de mise en correspondance entre Wordnet et l'ontologie arabe

Cet exposé présentera les principes de localisation basés sur une approche axée sur l'ontologie et leur application en SKOS.

**Christophe Roche** est professeur titulaire à l'Université de Savoie, en France, et chercheur associé au Centre de Recherche Linguistique de l'Université Nouvelle de Lisbonne, au Portugal. Il est actuellement en charge du Groupe de recherche Condillac sur «Ontologie et terminologie». Ses principaux domaines d'intérêt sont l'ingénierie des connaissances et de l'ontologie, de la terminologie et de la linguistique, et leurs applications: recherche d'information sémantique (multilingue), cartographie des connaissances, applications orientées pour l'ontologie et la terminologie. Christophe Roche a mis en place et géré plusieurs projets européens: FP4, COST, EUREKA, INTERREG II, Interreg III, Tempus, FP6, FP7, etc. Il est actuellement impliqué dans de nouveaux projets européens: INTERREG IV et le 7e PC. Pour l'Université Nouvelle de Lisbonne, il est en charge du lot de travail 2 «mise en réseau et configuration de la recherche» du projet européen FP7 SIERA dont l'un des objectifs principaux est d'intégrer le multilinguisme dans les systèmes de partage des connaissances. Enfin, Christophe Roche est un expert international de l'ISO (International Organization for Standardization – Organisation internationale pour la standardisation) Comité technique TC 37 de l'ISO 704 et ISO 1087-1 Normes terminologiques (Principes, méthodes et vocabulaire).

### GINCO, un outil ouvert pour la conception de vocabulaires

L'application GINCO sera présentée pendant cette intervention ainsi que les objectifs du projet, les choix de réalisation, l'état d'avancement (couverture fonctionnelle et perspectives). Une démonstration de l'outil sera aussi faite.

**Katell Briatte** est chef de projet au ministère de la culture et coordonne le programme HADOC (Harmonisation de la production des données culturelles). Elle s'intéresse en particulier aux problématiques de modélisation des données culturelles, d'interopérabilité et d'articulation entre les besoins de production et de diffusion des données, en particulier dans le contexte de l'accès ouvert et des données liées. Dans le cadre du programme, elle pilote la réalisation de l'application GINCO.

## La publication et l'exploitation des structures de connaissances avec OpenSKOS

De nombreux vocabulaires dans le domaines des Humanités et de la culture numériques peuvent être, et sont de plus en plus souvent, convertis en SKOS. La plate-forme de services Internet OpenSKOS fournit des moyens faciles de publier, télécharger, mettre à jour, moissonner, interroger et distribuer les données de vocabulaire SKOS. Cela a des avantages pour les constructeurs et consommateurs de vocabulaire, et les constructeurs d'outils qui exploitent les vocabulaires. Dans cette présentation, nous présentons le système OpenSKOS et un certain nombre de ses applications.

**Mark Lindeman** est Directeur Informatique de Picturae, une société néerlandaise axée sur les solutions pour le patrimoine culturel (numérisation et logiciels). Mark a travaillé pendant plus de 15 ans dans la construction de solutions logicielles pour les Archives, Musées et Bibliothèques. Mark a été impliqué dans plusieurs solutions Open Source, comme par exemple OpenSKOS.

## Thésaurus, classifications et autres ressources terminologiques spécialisées et multilingues

### Thésaurus multilingue sur la photographie (Europeana Photography)

Nacha Van Steen présentera un bref aperçu de l'évolution du vocabulaire dans le projet *Europeana Photography* (sources, structure, travail, problèmes rencontrés), ainsi que d'un bref coup d'œil sur l'avenir de l'utilisation et de la gestion des vocabulaires au sein du projet *Europeana Photography*.

**Nacha Van Steen** a étudié la science de l'art à l'Université de Gand et a commencé à travailler pour les Musées royaux d'Art et d'Histoire (KMKG) à Bruxelles en 2010 en tant qu'analyste des collections, assistant la numérisation des collections de KMKG dans une base de données centrale. Depuis 2012, elle a été nommée sur deux projets européens, *Europeana Fashion* et *Europeana Photography*, et sur plusieurs projets internes concernant la publication et l'échange de données. Elle a ainsi pu acquérir une expertise sur les normes, les thésaurus multilingues et l'interopérabilité des données, et dirige maintenant l'évolution des vocabulaires multilingues pour *Europeana Fashion* et *Europeana Photography*.

### Rôle, outils d'édition, exploitation et exposition du Thésaurus multilingue MIMO.

**Rodolphe Bailly**, responsable du Système d'information documentaire et de la numérisation à la Cité de la musique, a 18 ans d'expérience dans le développement Web, la numérisation et la mise en ligne de patrimoine numérique. Il a conçu et dirigé le développement du portail de la Médiathèque de la Cité de la musique en 2003. Il dirige l'équipe chargée de la diffusion et du développement informatique de [www.citedelamusiquelive.tv](http://www.citedelamusiquelive.tv). Il a dirigé le développement et la numérisation de plusieurs projets nationaux et européens de mise en ligne de patrimoine numérique : [www.spectaclesdumonde.fr](http://www.spectaclesdumonde.fr), portail des orchestres, Inventaire des instruments de musique dans les musées Français, archives Pleyel, MIMO. Il est membre de groupes de travail internationaux d'élaboration de modèles de données dans le domaine culturel : EDM, LIDO, SCHEMABIBEX.

### Partage Plus –Des vocabulaires contrôlés pour l'enrichissement de métadonnées internationales permettant l'Art Nouveau dans Europeana

<http://www.partage-plus.eu>

L'ambition des artistes à travers l'Europe et l'Amérique du Nord, visant à développer une

nouvelle compréhension de l'expression artistique adaptée pour répondre aux exigences de leur époque, caractérise la production esthétique, connu sous le nom d'Art Nouveau (*Jugendstil*). Dépassant les frontières de la tradition académique, les artistes ont découvert de nouveaux champs pour leurs activités, englobant l'architecture, les arts décoratifs et le design graphique afin d'influencer tous les aspects de l'environnement visuel. Le spectre de types d'objets inclut l'environnement construit, les peintures, sculptures, dessins, gravures, affiches, matériaux publicitaires, céramique, verrerie, métaux, bijoux, meubles, textiles, costumes, etc. Des expositions internationales et de nombreux magazines ont organisé le transfert d'idées et amélioré la diffusion. Ainsi, l'Art Nouveau est un cadre idéal pour un projet collaboratif de publication sur le patrimoine culturel européen, offrant de bonnes conditions pour le développement d'un vocabulaire multilingue de base couvrant le contenu des arts visuels.

Réunissant le contenu de nombreuses collections européennes recueillies par les 21 partenaires du projet, collaborant avec d'autres musées de 17 pays, "Partage Plus" fournira une part essentielle de 75.000 objets dans Europeana en Mars 2014. La tâche de la MINUAR dans le lot de travail 3 est d'intégrer le contenu du projet dans Europeana, en facilitant les conditions de recherche optimisées. La structure de métadonnées et la définition des champs obligatoires de catalogage de base ont été conçues comme un élément de base intermédiaire reliant les lots de travail 2 et 3 au début du projet. La création d'une terminologie contrôlée pour l'enrichissement des métadonnées se rapporte à cette structure compilant ensuite les parties intégrantes du vocabulaire du projet destinées à couvrir au mieux le spectre des contenus des partenaires du projet. La présentation portera sur l'approche et le développement de la terminologie.

Après des études aux Beaux-Arts, Histoire de l'art et des médias à Kassel et Braunschweig, **Christiane Pagel** a travaillé au département des Estampes et des Dessins du musée Herzog Anton Ulrich-Musée Brunswick<sup>1</sup>, musée d'art des Landes Niedersachsen (HAUM) , en mettant l'accent sur la documentation et la gestion des collections. De 2007 à 2011, elle été impliquée dans un projet de collaboration de l'Herzog Anton Ulrich-Museum avec l'Herzog August Bibliothek Wolfenbüttel<sup>2</sup> (HAB). Financé par la Fondation allemande pour la recherche (DFG), les premières copies modernes des deux institutions historiquement connectées ont été reliées virtuellement pour l'accès à la recherche en ligne dans la base de données "Virtuelles Kupferstichkabinett" (Salle d'impression virtuelle).

Actuellement, elle travaille comme coordonnatrice du projet européen Partage Plus - numérisation et activation de l'Art Nouveau pour Europeana " pour le projet allemand partenaire UNIMAR, Deutsches Dokumentationszentrum für Kunstgeschichte - Bildarchiv Foto Marburg, Philipps-Universität Marburg. En tant que leader du lot de travail 3 (enrichissement des métadonnées) UNIMAR est responsable de l'élaboration d'un vocabulaire contrôlé couvrant le spectre d'objets Art Nouveau. En outre UNIMAR fournit du contenu à travers une campagne photographique documentant l'architecture Art Nouveau allemande et coordonne la participation au projet de cinq musées allemands en matière de numérisation et de catalogage des objets Art Nouveau.

### **Linked Data : Réutilisations et applications**

Semanticpédia : Enjeux culturels et linguistiques autour du web de données

Le Ministère de la culture et de la communication, l'Inria et Wikimédia France ont souhaité unir leurs efforts en créant le programme Sémanticpédia. Il s'agit d'une plate-forme destinée à réaliser

<sup>1</sup> <http://www.3landesmuseen.de/Herzog-Anton-Ulrich-Museum.304.0.html>

<sup>2</sup> <http://www.hab.de>

des programmes de recherche et développement appliqués à des corpus ou des projets collaboratifs culturels, utilisant des données extraites des projets de Wikimedia. Le premier projet développé dans le cadre de cette plate-forme a été « Dbpédia en français » : il vise à proposer des données structurées à partir des informations contenues dans l'encyclopédie Wikipédia en français. Il a déjà trouvé une application concrète sur le site portail de l'histoire des arts : HDA-Lab.

**Thibault Grouas** est responsable de la mission des langues et du numérique au sein de la Délégation générale à la langue française et aux langues de France (DGLFLF) du ministère de la Culture et de la Communication. à ce titre, il est chargé de développer la dimension numérique de la politique en faveur du français, des langues de France et du plurilinguisme et de mettre les technologies au service de cette politique.

#### L'utilisation de VIAF dans data.bnf.fr: créer des liens pour importer des informations

Cette présentation expliquera l'intérêt que représente l'utilisation de VIAF dans data.bnf.fr. Pour générer des pages comme celle de Boris Vian, [http://data.bnf.fr/13091689/boris\\_vian/](http://data.bnf.fr/13091689/boris_vian/), nous utilisons la base de VIAF, qui fournit un ensemble de liens dans différentes langues (<http://viaf.org/viaf/9983585/>). On intègre le lien vers cette base, mais aussi vers DBpedia, retrouvé grâce aux liens directs.

Cette mécanique permet aussi de retrouver les divers noms qu'un auteur a pu avoir dans les langues représentées dans VIAF, et de fournir un lien exploitable par les machines quelle que soit la langue des utilisateurs.

Par la suite cela devra permettre de consulter l'interface dans différentes langues.

**Romain Wenz**, conservateur des bibliothèques, archiviste-paléographe, a rejoint en 2009 le département de l'Information Bibliographique et Numérique de la BnF comme expert métadonnées. Il travaille à l'élaboration de l'outil "data.bnf.fr" qui vise notamment à fournir des pages synthétiques sur les auteurs et les oeuvres."