

	<p style="text-align: right;">Date : 15/07/2007</p> <p><b>Les vocabulaires contrôlés de la Bibliothèque du Congrès, le Fichier d'autorité international virtuel et leur application au Web sémantique</b></p> <p><b>Barbara B. Tillett &amp; Corey Harper</b></p> <p style="text-align: right;"><i>Traduit de l'anglais par Yves-Gaël Lanchec, Bibliothèque nationale de France</i></p>
<b>Meeting:</b>	<b>147 IFLA-CDNL Alliance for Bibliographic Standards (ICABS)</b>
<b>Simultaneous Interpretation:</b>	<b>Yes</b>
<p style="text-align: center;"><b>WORLD LIBRARY AND INFORMATION CONGRESS: 73RD IFLA GENERAL CONFERENCE AND COUNCIL</b>  19-23 August 2007, Durban, South Africa  <a href="http://www.ifla.org/IV/ifla73/index.htm">http://www.ifla.org/IV/ifla73/index.htm</a></p>	

Library of Congress controlled vocabularies, the Virtual International Authority File, and their application to the Semantic Web] / Barbara Tillett et Corey Harper

-----

**Résumé:**

Notre exposé est en grande partie tiré d'un article écrit avec M. Corey Harper pour [la revue] **"Cataloging & Classification Quarterly"**

"Les vocabulaires contrôlés de la Bibliothèque du Congrès et leur application au Web sémantique," / par Corey A. Harper et Barbara B. Tillett, v. 43, n° 3/4, 2006.

Il fait le point sur les vocabulaires contrôlés de la Bibliothèque du Congrès (Vedettes-matière de la Bibliothèque du Congrès, Classification de la Bibliothèque du Congrès, et Fichiers d'autorité NACO/ Bibliothèque du Congrès) et présente le projet VIAF (Fichier d'autorité international virtuel). Il y a tout lieu de penser que les vocabulaires contrôlés pourront rapidement être les éléments constitutifs du Web sémantique. Au niveau international, il sera possible de lier les données fiables des notices d'autorité de chaque pays, pour le bénéfice des utilisateurs au niveau mondial.

En 2001, Tim Berners-Lee faisait paraître un article sur le Web sémantique<sup>1</sup> dans la revue :**"Scientific American"**, qui préfigurait l'idée d'un Web de données universelles associées entre elles.

---

<sup>1</sup> Berners-Lee. T., Hendler, J., & Lasilla, O. "The Semantic Web" [Electronic version]. Scientific American, 284, no.5, 2001, p. 34-43.  
<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

Il mentionnait le fait que des groupes indépendants travaillaient sur des domaines similaires et avaient besoin d'un langage commun pour faire converger leurs efforts.

**Les bibliothèques et les développeurs du Web sémantique partagent les mêmes buts : nommer des concepts, nommer des entités, et rapprocher les différentes formes de ces noms.**

Depuis plusieurs décennies, les bibliothèques développent des outils produisant de très importantes sources de données associées entre elles. Aujourd'hui, nous devons juste les traduire dans de nouveaux outils facilitant la réalisation de l'infrastructure du Web sémantique. Nous disposons désormais d'une masse de connaissances suffisante pour parvenir à un accord qui nous aide à aller de l'avant pour construire les outils du futur ; dans ce domaine les bibliothèques ont une place importante à prendre.

Une étape importante va consister à traduire les vocabulaires contrôlés dans les normes du Web sémantique, tels que OWL (Web Ontology Language), afin qu'ils puissent servir dans de nouveaux contextes . Mentionnons aussi la technologie du Web sémantique appelée SKOS Core (Simple Knowledge Organization System) pour le codage du contenu des thésaurus. Dès lors, pour les données d'autorité, tant SKOS que OWL sont un apport important pour la conversion des données d'autorité actuelles. Tout aussi utile pour le Web sémantique, la conversion de nos données bibliographiques selon la norme RDF (Resource Description Framework), et ce sont les travaux sur le Dublin Core et le RDA (Resource Description and Access) qui nous conduisent dans cette voie.<sup>2</sup>

La qualité et la fiabilité des vocabulaires contrôlés utilisés en bibliothèques sont reconnues et avérées, ce qui est profitable pour les actions futures à mener.

La plus grande part des efforts fournis jusqu'à présent pour les bibliothèques numériques ont visé à accroître leur visibilité sur le Web - grâce à la numérisation du contenu de collections choisies. Les données d'autorité des bibliothèques présentent un potentiel encore supérieur en fournissant aux utilisateurs la possibilité de trouver une réponse à leurs recherches sur le Web grâce au balisage de chemins qui les mèneront à l'information pertinente.

L'expression Web 2.0 est communément utilisée pour décrire les systèmes participatifs proposés à l'utilisateur de l'Internet - applications qui permettent de personnaliser les éléments collectés, ajouter ses propres commentaires, ses "folksonomies" ou mots-clés - ce que nous appelons vedettes-matières ou points d'accès - et lui permettent de retrouver plus aisément l'information dont il a besoin et de l'utiliser ensuite à sa guise dans son espace personnel.

Pour donner l'exemple de vocabulaires contrôlés produits par la profession, citons la Classification Décimale Dewey (CDD), la Classification de la Bibliothèque du Congrès (LCC). Nous disposons également des termes contenus dans les listes d'autorité-matières : la LCSH (Library of Congress Subject Headings) ; et deux Thésaurus pour les documents graphiques : le TGM I pour les mots-sujets, le TGM I I pour la Typologie et les caractéristiques matérielles. Citons encore le GSAFD : Guide pour l'accès-sujet aux oeuvres romanesques et au théâtre, etc ; et le Thésaurus Ethnographique (Etnographic Thesaurus). Actuellement, les outils de recherche disponibles ne sont pas en mesure de tirer profit de la structure syndétique (des systèmes de renvois croisés) présents dans les vocabulaires contrôlés.

---

<sup>2</sup> Voir l'annonce des résultats du colloque du 30 avril-1er mai 2007 Londres "Data Model Meeting" : <http://www.bl.uk/services/bibliographic/meeting.html>

Pourtant ces derniers permettraient une augmentation significative de l'usage du Web sémantique ainsi que des services et outils du Web2.0 - en apportant la structure et les liens nécessaires pour les sujets, les noms de personnes, les choses, etc.

La technologie du Web sémantique apporterait un autre avantage aux vocabulaires contrôlés. En effet, représenter la structure de tels vocabulaires grâce aux cadres communs offerts par OWL ou SKOS rendrait leur interopérabilité plus plausible. Représenter et modéliser les vocabulaires de cette manière facilite l'identification, l'exploitation et les relations d'équivalences entre concepts issus de vocabulaires contrôlés hétérogènes. Ceci permettrait par exemple, d'augmenter une part des possibilités de regroupement par grappes qui se font jour, et préfigurent l'architecture des catalogues de la prochaine génération.

OCLC a déjà commencé l'expérimentation de certaines utilisations des vocabulaires contrôlés avec des logiciels existant, comme Microsoft Word, par exemple. Grâce au Projet Terminologies d'OCLC, les utilisateurs peuvent avoir accès sous Word à des vocabulaires contrôlés, et faire usage de mots-matières parallèlement à la création de leurs documents de travail - sans avoir besoin de quitter (le logiciel) Word<sup>3</sup>. On peut tout à fait imaginer les catalogueurs disposer d'un système similaire et accéder rapidement aux termes recherchés issus de vocabulaires contrôlés, ce qui économiserait du temps et des coûts de catalogage. Le Projet " Terminologies " pose déjà les fondations des avancées futures en intégrant les outils disponibles sous Microsoft Office à ceux présents dans le logiciel de catalogage "Connexion" d'OCLC.

Nous devons aussi envisager que les autorités noms de personnes existantes pourront être accessibles dans un environnement Web plus convivial, et dès lors, utilisables dans les applications qui en découlent. Espérons aussi que le VIAF (Fichier d'Autorité International Virtuel) pourra aussi à terme évoluer dans un tel environnement.

Le VIAF est un projet dont on rêve depuis fort longtemps. Une version est actuellement testée par la Deutsche Nationalbibliothek, la Library of Congress, la Bibliothèque nationale de France et OCLC. Les étapes de tests sont terminées, mais il reste désormais à travailler sur le développement des applications utilisateurs, et les moyens de rendre les données plus facilement accessibles sur le Web. Je me bornerai à rappeler ici l'existence de nombreuses publications de l'IFLA, et de rapports déjà publiés sur le projet VIAF, et ne reviendrai donc pas sur tous les détails déjà connus. Retenez seulement l'idée à la base d'un ensemble partagé au niveau mondial, reliant entre eux les principaux réservoirs d'autorité dans le contexte d'un système totalement gratuit. Il établirait des liens entre les notices d'autorité présentes pour une même entité et permettrait d'être utilisé pour diverses autres applications.

Ce fichier partagé permettrait alors à une bibliothèque déterminée de réduire ses coûts de catalogage, serait également un outil permettant à l'utilisateur de choisir la langue ou l'écriture pour les noms de personnes (et à terme les noms de collectivités et les noms géographiques). Un système unique tel que celui-ci serait bien plus utile que d'avoir à consulter séparément les différents fichiers d'autorité du monde entier . Il pourrait de plus être utilisé par les systèmes Web pour ajouter de la précision lors d'une requête, élément faisant cruellement défaut dans les moteurs de recherche tels que

---

<sup>3</sup> Vizine-Goetz, Diane. "Terminology services: making knowledge organization schemes more accessible to people and computers." OCLC Newsletter, no. 266, 2004. Available from : <http://www.oclc.org/news/publications/newsletters/oclc/2004/266/>

Google. Cette façon de concevoir les notices d'autorité signifie que toutes les variantes de forme sont regroupées dans une notice d'autorité unique, et par extension, les notices d'autorité associées pourront être cherchées et affichées par les utilisateurs. Dans le passé, les bibliothèques ne retenaient à l'affichage qu'une seule forme du nom, un choix par défaut qui subsisterait si l'utilisateur ne précisait pas son choix pour telle langue ou telle écriture.

Nous espérons à terme pouvoir étendre le projet VIAF aux caractères non-latins.

En outre, si nous étendons le VIAF aux communautés du Web sémantique, la puissance de notre système d'autorité pourrait servir pour bien d'autres applications. Avec pour conséquences d'ouvrir plus de voies vers les oeuvres de tel auteur ou écrites sur celui-ci, mais aussi la connection possible à d'autres éléments d'information : un blog, une entrée wikipedia, des objets de cet auteur ou sur lui présents dans la bibliothèque de votre quartier, des éléments biographiques disponibles dans des dictionnaires ou des encyclopédies, des articles de journaux écrits par lui ou sur lui, les choses à vendre via le e-commerce, et plus encore. Ces interconnexions seront utiles pour évaluer la fiabilité et l'authenticité des résultats obtenus sur le Web, et suggérer à l'utilisateur final la part de subjectivité dans les points de vue émis et susceptibles d'influer sur le contenu des documents recherchés. Mais comment parvenir à ce résultat à partir des notices d'autorité actuelles ? Nos notices d'autorité doivent avant tout être disponibles sous forme informatique dans les formats conçus pour le Web sémantique, notamment OWL et SKOS. Une autre étape est d'assurer l'identification de chaque terme en le dotant d'un identifiant URI (Uniform Resource Identifier) ou de faire le choix de la conversion des notices d'autorité d'XML en SKOS. Une autre option encore, consisterait à faire à la volée la conversion à partir du format d'origine. Ainsi par exemple, en utilisant le VIAF stocké en MARCXML, il serait possible de convertir les données aussi bien en RDF qu'en SKOS selon chaque type d'application. Il pourrait augmenter aussi le potentiel des systèmes proposant des folksonomies.

Les informations contenues dans les fichiers d'autorité pourraient alors être réutilisées dans de nombreuses autres applications de manière ad-hoc et ciblée. L'identification des informations portées par les autorités au travers d'URI, permet la réutilisation de ces URI afin de lier d'autres descripteurs de personnes aux notices d'autorité, qui à leur tour pointent sur leurs oeuvres. Des normes émergent destinées à coder les relations entre personnes, par exemple le projet Friend of A Friend (FOAF)<sup>4</sup>, qui apportera une importante plus-value à cette information. Le FOAF fournit les moyens de détailler les relations entre différentes personnes mais aussi entre les personnes et les différents objets de leurs créations. La capacité à intégrer des URI aux notices VIAF dans les descriptions FOAF contribue à accroître le rôle des professionnels des bibliothèques pour préciser la fiabilité et la crédibilité de telle ou telle information. L'identification d'un contributeur à un blog ou à une entrée Wikipedia en tant qu'expert réputé sur un sujet donné serait plus facile.

Dans un article paru en 2007, nous notions : "Le Web sémantique vise pour partie à faciliter la découverte de ressources en créant des outils pour aider les chercheurs à affiner et développer leurs recherches, ainsi qu'une aide à la navigation dans les résultats de la recherche."<sup>5</sup>

---

<sup>4</sup> Voir la page d'accueil de The Friend of a Friend project. <http://www.foaf-project.org/>

<sup>5</sup> Harper Corey A. & Tillett, Barbara B., "Library of Congress controlled vocabularies and their application to the Semantic Web," *Cataloging & Classification Quarterly*, v.43, no. 3/4, 2006.

Un important colloque s'est tenu à Londres du 30 avril au 1er mai 2007, sous l'égide des différents responsables des RDA. Il rassemblait des représentants du SKOS, du Web sémantique, du Dublin Core, de l'IEEE/LOM et des RDA (Resource Description and Access).

Initialement l'objet de ce colloque était de passer en revue différents modèles de données (dont les FRBR et les FRAD) et d'étudier en détail l'utilité de la norme des contenus RDA par quelques groupes d'utilisateurs de métadonnées. Les conclusions de ce colloque furent diffusées aussitôt, et des recommandations suivirent afin d'obtenir des subsides pour le développement d'un Profil d'Application RDA en mesure de :

- Elaborer un vocabulaire RDA d'éléments,
- Elaborer un profil d'application RDA DC (Dublin Core) basé sur les FRBR et FRAD et
- Faire connaître des vocabulaires de valeurs RDA qui utiliseraient les RDF/ RDFS/SKOS.

La déclaration d'ouverture soulignait les profits à tirer de ces actions :

- les professionnels des bibliothèques se dotent d'un standard de métadonnées compatible avec l'architecture du Web et totalement interopérable avec d'autres chantiers du Web sémantique.
- les acteurs du DCMI se dotent ainsi d'un profil d'application pour les bibliothèques s'appuyant complètement sur le DCAM et les FRBR (modèle de référence à suivre pour le futur)
- les acteurs du Web sémantique disposent d'un réservoir important et bien conçu d'éléments de métadonnées réutilisables ;
- les RDA suscitent un intérêt élargi.

L'apport de l'expérience des bibliothèques, des normes et des vocabulaires contrôlés au Web sémantique bénéficiera aux bibliothèques et aux communautés de métadonnées, et les utilisateurs en seront les premiers bénéficiaires.