

Journée d'étude « Les données en bibliothèques, les enjeux des linked open data », Lausanne 1er octobre 2013

Pierre Boillat
Conservatoire et Jardin botaniques de la Ville de Genève

Mots-clés

LOD, Linked Open Data, RDF, Web sémantique

L'importance des données ne va faire que s'accroître ces prochaines années. Les professionnels de l'information documentaire de Suisse romande ne s'y sont pas trompés en s'inscrivant en nombre à la journée d'étude du 1er octobre 2013 à Lausanne sur le web des données libres (en anglais les linked open data) en lien avec les bibliothèques. Septante-deux inscrits ont conduit les organisateurs, la filière Information documentaire de la Haute école de gestion de Genève, à déplacer le lieu de la rencontre afin de recevoir tous les participants dans de bonnes conditions.

La journée s'est organisée autour de six exposés présentés par des intervenants du monde des bibliothèques provenant de plusieurs pays européens et ayant, pour la plupart, une solide expérience de la pratique du web des données libres. En guise d'introduction, Madame Emmanuelle Bermès, du Centre Pompidou (Paris), a dressé un tableau synthétique de la problématique du web sémantique en abordant notamment le Resource description framework (RDF)¹. Elle a insisté sur l'importance du partage des données, qui doit primer sur la possession de données en elle-même (le succès du libre accès – l'open access – en est un brillant exemple). Cet échange passe par le renouvellement des normes et standards. MARC et ses dérivés nationaux ou internationaux ne peuvent plus convenir à une société qui échange globalement et fait fi des barrières des métiers. L'utilisateur du web n'a que faire de nos catalogues ou portails documentaires bridés par leurs formats limités, il veut l'information directement. De nouveaux standards au-delà de l'environnement des bibliothèques (tels que MARC) émergeront et les bibliothèques devront se battre pour rester dans la course. Ce combat ne réussira que si les bibliothèques s'unissent ; si une association faîtière telle que l'IFLA peut apparaître peu réactive au premier abord face aux métamorphoses continues des pratiques du web, elle dispose néanmoins de compétences humaines véritables par sa section dédiée au web sémantique (*cf. infra*). Les initiatives suivantes ont été citées : W3C Library Linked Data Incubator Group² (qui a cessé ses activités en avril 2012), Dublin Core Metadata Initiative³, Resource Description and Access (RDA)⁴, Bibliographic Framework Transition Initiative (BIBFRAME)⁵, Linked Open Data in Libraries, Archives & Museums (LODLAM)⁶, Schema Bib Extend Community Group (W3C Community and Business Groups)⁷ ou encore le Semantic Web Special Interest Group de l'IFLA⁸.

Ces concepts se sont concrétisés avec la communication de Monsieur Stefan Gradmann, professeur au Département de l'art de la Katholieke Universiteit Leuven (Louvain) et directeur de la bibliothèque universitaire de cette université. M. Gradmann nous a donné à voir une application en bibliothèque du web des données libres à travers la bibliothèque numérique européenne Europeana et son Europeana data model. Europeana data model est un modèle de représentation du patrimoine culturel dans Europeana. Il montre bien les relations entre les données internes à Europeana et celles récupérées dans d'autres bases de données ou catalogues de référence tels que le Virtual International Authority File (VIAF)⁹ pour les autorités, le Getty Thesaurus of Geographic Names¹⁰ pour les noms de lieu, la Library of Congress pour les noms de langues, etc. Europeana data model est un modèle valable autant pour les bibliothèques, que pour les archives ou les musées. Il pourrait devenir un standard même si pour l'heure aucun logiciel commercial ne l'utilise. Par ailleurs, et d'une manière générale, la masse des ressources en ligne constitue un vrai défi pour les chercheurs qui doivent développer de nouvelles stratégies d'utilisation de l'information à l'aide d'outils encore à créer. Les bibliothèques ont ici une opportunité à saisir pour se mettre en avant. Ce développement s'accompagne aussi d'une nouvelle terminologie que le bibliothécaire doit

apprivoiser. Dans ce contexte, Europeana, qui est devenu une source majeure des humanités numériques et qui propose une nouvelle manière de rechercher l'information, permet aux bibliothécaires de se positionner au même niveau que les chercheurs (ou presque).

Un autre exemple nous a été présenté par Monsieur Romain Wenz, conservateur à la Bibliothèque nationale de France (BnF), avec l'outil « data.bnf.fr ». Lancé en 2011, « data.bnf.fr » réutilise les données contenues dans les différents catalogues de la BnF (à ce jour 5,5 millions de documents sont concernés, soit 40% des notices de catalogues) et les présente aux utilisateurs, à l'aide des outils du web sémantique, d'une manière synthétique et originale (voir les graphiques par frises chronologiques pour les différentes éditions d'une œuvre) alors qu'elles étaient jusque-là dispersées dans lesdits catalogues. « data.bnf.fr » est indexé par les moteurs de recherche. Ainsi, l'internaute accède à l'information sans passer par les catalogues. Ces données sont complètement libres de droit. Cette approche novatrice permet de mettre en exergue des auteurs peu connus ou des contributions peu visibles des grands auteurs. Le succès est au rendez-vous avec plus de 70'000 visiteurs uniques par mois, dont 80% passent par l'entremise des moteurs de recherche.

Le CERN est un grand producteur de données et Monsieur Jens Vigen, chef du Service d'information scientifique dudit CERN, a présenté les défis d'une institution qui rencontre déjà des problèmes de lecture de données vieilles d'à peine dix ans ! Aux projets d'open access pour les publications que le CERN a mis en oeuvre avec succès depuis déjà un certain temps, s'ajoutent ceux concernant les données scientifiques. Monsieur Vigen s'est réjoui que l'accès aux données soit une priorité reconnue par les instances politiques. Il a cité le rapport remis en octobre 2010 par le High Level Expert Group on Scientific Data à la Commission européenne, document qui présente une vision de la problématique des données scientifiques à l'horizon 2030¹¹.

Le Réseau des bibliothèques de Suisse occidentale (RERO) est un acteur régional majeur dans le monde des bibliothèques. Monsieur Miguel Moreira, chef de projet à RERO, a présenté les implications actuelles et potentielles du web des données libres sur le réseau romand. Le web sémantique est une priorité pour RERO comme en témoigne son inscription au plan stratégique 2013-2017 sous le chiffre 1.5¹². La bibliothèque numérique RERO DOC apparaît comme le plus bel exemple d'une valorisation des données par RERO. L'adhésion en 2010 de RERO à VIAF a constitué une étape forte vers le web de données libres. Les projets s'articulent autour du passage au RDF et au choix « des vocabulaires et métadonnées à adopter pour les relations ».

La dernière présentation, un duo mené par Madame Jasmin Hügi et Monsieur Nicolas Prongué, tous deux étudiants en Master en science de l'information à la Haute école de gestion de Genève, s'est divisée en deux parties. D'une part, ils ont décrit les cinq exemples suivants d'applications existantes dans le web des données libres en bibliothèque : lobid.org (Linking Open Bibliographic Data)¹³, British National Bibliography¹⁴, VIAF¹⁵, Centre Pompidou¹⁶ et « data.bnf.fr ». Ils ont relevé qu'en Suisse, rien n'a encore dépassé le stade du projet. Néanmoins, des recherches sont menées par la Bibliothèque nationale suisse, le CERN, RERO et Swissbib. D'autre part, ils ont effectué une étude des compétences requises pour les professionnels de l'information documentaire souhaitant œuvrer dans le web des données libres en bibliothèque. Ils ont analysé huit offres d'emploi et ont pris contact avec dix experts dans le domaine. Leurs résultats leur permettent d'esquisser trois axes de compétence des

candidats : transformation de données bibliographiques en linked open data, publication des linked open data et accès en ligne et, finalement, collaboration. Ils émettent aussi deux constats généraux : le premier sur la nécessaire métamorphose du travail des catalogueurs tel qu'il se pratique aujourd'hui et, le deuxième, sur l'importance de la veille et du suivi des évolutions métier.

Cette journée a ouvert une fenêtre sur l'évolution de nos catalogues de bibliothèque ou plutôt a permis de démontrer que les innombrables données qu'ils contiennent ont un avenir. Des générations de bibliothécaires se sont succédées pour les alimenter et les développer. Il est heureux de savoir que les outils du web des données libres accroissent encore les possibles de leur utilisation. Peut-être devrons-nous, bibliothécaires, faire le deuil du contrôle du contenu de « nos » notices ; les relations entre ressources estomperont les limites des réservoirs des uns et des autres. Et ces applications permettront d'offrir aux utilisateurs des accès à l'information en phase avec leurs demandes, en adéquation avec les besoins de la société. Toute nécessité indispensable à l'inscription pérenne des bibliothèques dans notre temps.

NOTES

- (1) <http://www.w3.org/RDF/> (Page consultée le 01.12.2013).
- (2) <http://www.w3.org/2005/Incubator/> (Page consultée le 01.12.2013).
- (3) <http://dublincore.org/> (Page consultée le 01.12.2013).
- (4) <http://www.rdata toolkit.org/> (Page consultée le 01.12.2013).
- (5) <http://www.loc.gov/bibframe/> (Page consultée le 01.12.2013).
- (6) <http://lodlam.net/> (Page consultée le 01.12.2013).
- (7) <http://www.w3.org/community/schemabibex/> (Page consultée le 01.12.2013).
- (8) <http://www.ifla.org/swsig> (Page consultée le 01.12.2013).
- (9) http://www.bnf.fr/fr/professionnels/donnees_autorites/a.viaf.html (Page consultée le 01.12.2013).
- (10) <http://www.getty.edu/research/tools/vocabularies/tgn/> (Page consultée le 01.12.2013).
- (11) « Riding the wave : how Europe can gain from the rising tide of scientific data » ; http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707 (Page consultée le 01.12.2013).
- (12) http://www.rero.ch/pdfview.php?section=infos&filename=plan_strategique_rero_2013_2017.pdf (Page consultée le 01.12.2013).
- (13) <http://lobid.org> (Page consultée le 01.12.2013).
- (14) <http://bnb.bl.uk/> (Page consultée le 01.12.2013).
- (15) <http://viaf.org/> (Page consultée le 01.12.2013).
- (16) <http://www.centre pompidou.fr/> (Page consultée le 01.12.2013).