



Ecole de Bibliothécaires Documentalistes

Vers les bibliothèques 3.0

Synthèse préalable à la mise en place d'une veille pour la direction de la bibliothèque de Fels de l'Institut Catholique de Paris (ICP) sur les nouvelles normes de catalogage en bibliothèque

La mise en ligne des catalogues a donné aux bibliothèques une meilleure visibilité sur le web. Cependant, l'accès des internautes au contenu des catalogues, via des moteurs de recherche généralistes, n'est toujours pas possible. Pour s'adapter à ce type de recherche et franchir ainsi l'obstacle du web invisible, tout en améliorant la navigation à l'intérieur du catalogue, certaines bibliothèques cherchent à appliquer à la structuration de leurs données les apports du web sémantique. Les nouveaux modèles et normes de catalogage, FRBR (Functional Requirements for Bibliographic Records), RDA (Resource Description and Access), RDF (Resource Description Framework, cadre pour décrire des ressources), apparaissent désormais comme des outils d'avenir. Quels changements ces nouvelles normes vont-elles apporter dans les bibliothèques ? Le présent document se veut un état de l'art sur les travaux et expérimentations en normalisation documentaire à ce jour, permettant d'en cerner les enjeux et les acteurs.

Le web sémantique : quand les modèles conceptuels rendent les machines intelligentes

Une nouvelle étape dans l'évolution du web est en cours de préparation : le web sémantique ou web 3.0. Le projet du web sémantique, lancé par Tim Berners Lee¹, le fondateur du World Wide Web et président du W3C (World Wide Web Consortium), désigne un nouveau paradigme de structuration des données. Celui-ci vise à permettre aux ordinateurs, et plus particulièrement aux moteurs de recherche, de « comprendre » le contenu des documents.

Un nouveau paradigme pour le web favorisant la compréhension de contenus par les machines.

Actuellement, les moteurs de recherches ne comprennent pas le sens des mots-clés ou des questions posées par les internautes : basés sur des algorithmes statistiques, ils analysent la récurrence des mots-clés de la requête dans les pages web indexées par des robots (web crawler). Ils ne sont, de ce fait, capables de répondre qu'à deux types de questions : "Quelles sont les pages contenant le terme X ?" et "Quelles sont les pages les plus populaires au sujet de Y ?". L'ordinateur ne peut pas comprendre la nature et la relation entre deux entités nommées. Par exemple, il peut identifier, dans la page Wikipédia consacrée à Jean-Paul Sartre, un lien vers la page « Paris », mais il ne peut pas comprendre que Jean-Paul Sartre est un personnage historique, que Paris est un lieu géographique et que la relation entre les deux est de type : « A est né dans B »².

Le web sémantique fournit les outils permettant de décrire une entité, sa nature et ses relations, d'une manière qui soit compréhensible par la machine. De nouvelles possibilités de recherche d'information s'offrent alors. Par exemple, un moteur de recherche sémantique peut répondre à une question posée en langage naturel comme « Où est né Jean-Paul Sartre ? ». Il pourra également distinguer un contenu « sur » Jean-Paul Sartre d'un contenu écrit « par » Jean-Paul Sartre. Enfin, grâce à ce type de distinctions relationnelles, il sera en mesure de fournir automatiquement la liste des œuvres de Jean-Paul Sartre, ou encore la liste des auteurs influencés par Jean-Paul Sartre.³

Adapter la description de contenu à la machine grâce au RDF, aux triplets et aux URI

Le web sémantique s'appuie sur les contenus du web et des bases de données ainsi que sur les normes et standards existants : protocole HTTP, les URI (Uniform Resource Identifier, standard d'identification des ressources physiques ou abstraites), le langage XML (eXtensible Mark-up Language, « langage par balises

extensible ». Il y ajoute des outils spécifiques de description de contenu (RDF, OWL⁴, ontologies)⁵. La description de contenu, qui permet la compréhension de celui-ci par la machine, obéit au standard RDF, qui est un modèle abstrait de données fondé sur des triplets (Sujet, Prédicat, Objet)⁶. Ces triplets d'information associent deux entités, le sujet et l'objet (par exemple : Jean-Paul Sartre et Paris) par une relation typée appelée prédicat (ex : est né à, est l'auteur de, est le père de, etc.). Par exemple, un triplet dont Victor Hugo est l'objet pourrait être : (Quatre-vingt-treize, auteur, Victor-Hugo)⁷.

La définition d'une entité se présente sous la forme d'un nœud relationnel, chaque entité possédant plusieurs relations avec d'autres entités, et ainsi de suite. Chaque entité et type de relation est présent sous la forme d'URI⁸. Par exemple, dans l'ontologie Relationship⁹, la relation « child of » correspond à l'URI suivante : <http://purl.org/vocab/relationship/childOf>. Cette structuration sous forme de triplets relationnels permet, lors de la navigation sur un site ou moteur intégrant la technologie sémantique, de pouvoir orienter la recherche de manière plus souple. L'objet peut devenir sujet d'une autre recherche : on peut ainsi glisser de Jean-Paul Sartre à Simone de Beauvoir ou à Paris.¹⁰

Le web sémantique toujours en débat, malgré de premières applications très prometteuses

Les sceptiques dénoncent le caractère utopique du web sémantique¹¹. L'écrivain canadien Cory Doctorow, dans un article du 26 août 2001 intitulé « Les 7 illusions de la méta-utopie »,¹² affirme qu'il y a plusieurs façons de décrire un contenu et que le choix n'est pas neutre en ce domaine. Les créateurs de métadonnées peuvent agir en fonction de leurs intérêts (économique, politique, générer des spams, etc.)

Il existe néanmoins déjà un certain nombre d'applications sémantiques¹³, telles que :

- Des agrégateurs d'information comme [Freebase](#), ou [Twine](#), ou encore [TriplIt](#)
- Des moteurs de recherche en langage naturel comme [Powerset](#), [TrueKnowledge](#), [Wolfram Alpha](#) et [Hakia](#), qui, à terme, seraient capables de proposer des réponses directes à des questions posées en langage naturel.

Concernant les bibliothèques, selon Emmanuelle Bermès de la BnF¹⁴, RDF permettrait de sortir les catalogues du web profond, caché aux moteurs de recherche, et d'intégrer ainsi les bibliothèques au web de données.

FRBR et RDA : les chaînons manquants entre les bibliothèques et le web de données

FRBR, un modèle conceptuel « entités – relations »

Les FRBR ont été rédigées en 1998 sous l'égide de l'IFLA et traduites en français en 2001 par une équipe de la BnF sous le titre : « Spécifications fonctionnelles des notices bibliographiques ».¹⁵

L'édition 2010 du Métier de bibliothécaire¹⁶ définit les FRBR comme un modèle conceptuel, qui regroupe les données bibliographiques en :

- entités, entendues comme un ensemble d'éléments fondamentaux et pertinents pour les utilisateurs des données bibliographiques .
- attributs de ces entités, c'est-à-dire les caractéristiques importantes permettant de décrire chaque instance d'une entité.
- relations entre ces différentes entités.

Les entités sont elles-mêmes composées de trois groupes distincts détaillés dans les FRBR :

Le groupe 1 comporte 4 grandes divisions : œuvre, expression, manifestation, document (ou item).

Le groupe 2 comporte des personnes et des collectivités.

Le groupe 3 comporte des concepts, objets, événements, lieux.

Ce modèle FRBR_{ER} (ER pour « entités-relations »), doit être complété par FRAD (Functional Requirements for Authority Data) pour les autorités et par FRSAD (Functional Requirements for Subjects Authority Data) pour les vedettes matières.

Il existe aujourd'hui une autre version des FRBR, les FRBR_{OO}, FRBR « orienté objet » qui permettent une modélisation commune entre bibliothèques et musées.¹⁷

RDA, un nouveau code de catalogage

RDA (Resource Description and Access) se veut selon ses promoteurs anglo-américains du JSC (Joint Steering Committee) le standard de catalogage pour le 21^e siècle.¹⁸ Il repose sur les modèles FRBR, FRAD et FRSAD. Les notices bibliographiques y sont remplacées par un réseau de relations entre des « notices » correspondant aux entités du modèle FRBR.¹⁹ L'ISBD n'est alors plus qu'une option d'affichage.²⁰

Le JSC en assure le développement, en collaboration avec des représentants de plusieurs bibliothèques anglo-américaines dont la Library of Congress et le Comité canadien du catalogage.

Par ailleurs, un groupe de travail, le DCMI RDA Task Group (DCMI Dublin Core Metadata Initiative²¹) a été formé dans le but de définir les composants du RDA dans un vocabulaire RDF afin de le rapprocher de Dublin Core, standard de méta-données.²²

Une étape intermédiaire pourrait permettre d'utiliser RDA en maintenant la structure traditionnelle des catalogues basée dans les pays anglo-américains sur AACR2 (Anglo-American Cataloguing Rules 2nd edition).

La France en phase de réflexion.

En France, un groupe d'experts (GE) travaille spécifiquement sur RDA avec l'AFNOR et sous la tutelle de l'ABES.

Il est présidé par Françoise Leresche de la BnF et a comme identifiant AFNOR CG46/CN357/GE6 (CG pour Commission générale, CN pour Commission de normalisation).

Il dispose d'une newsletter commune à tout CG46, dans le cadre plus général de la révision des normes pour l'information-documentation. Plus spécifiquement, un site web devrait prochainement être créé pour diffuser de l'information sur les travaux du GE6.²³

Travaux sur la norme et premières expérimentations en bibliothèque

RDA : entre scepticisme et enthousiasme

D'après deux sondages évoqués lors de la conférence du groupe international de catalogage et d'indexation en RDA (CIG : Cataloguing & Indexing Group, chargé du test de RDA) à Exeter, 20% des bibliothèques au Royaume Uni et 50% en Europe envisageraient d'adopter RDA. Des difficultés subsistent pour mettre en relation MARC21 et RDA.²⁴

Plusieurs pays européens, réunis lors du séminaire EURIG (European RDA Interest Group) du 8 août 2010, envisagent une traduction de RDA. L'Allemagne serait la plus avancée en Europe dans ce travail.²⁵ En France, d'après les actualités de l'ABES, le groupe d'experts de l'AFNOR collabore avec l'ASTED (Association pour l'Avancement des Sciences et des Techniques de la Documentation) pour cette traduction.²⁶ Finlande, Danemark, Norvège, Italie, Portugal sont également très intéressés par RDA et attentifs aux tests en cours dans les bibliothèques engagées dans le projet.

Alan Danskin du JSC les incite vivement à signaler les difficultés d'adaptation nationales à RDA.²⁷

Les catalogueurs français défendent un « profil français » d'adoption de RDA. Ils regrettent la référence limitée à l'ISBD et privilégient le maintien des normes françaises et internationales sur certains points comme le traitement des noms géographiques.²⁸

Exemples de mise en place de RDF et FRBR : Dbpedia et Trove

Dbpedia, projet initié par des chercheurs de la Freie Universität de Berlin, utilise Wikipedia comme une base de données interrogeable par un moteur d'extraction de données, Scala, qui est un logiciel libre.²⁹ Il s'agit à la fois d'extraire des données structurées de Wikipedia pour les rendre disponibles sur le web et de lier les données du Web à celles de Wikipedia. Des liens sont créés entre les données et les résultats des requêtes proviennent de plusieurs sources (dans Wikipedia et sur le web). La structuration des données s'appuie sur RDF.³⁰

Le blog Bibliomancienne³¹ mentionne une innovation intéressante, le portail Trove de la bibliothèque nationale d'Australie, apparu en novembre 2009. Trove se présente comme une interface de recherche qui récupère les données de bibliothèques, musées et archives, structurés selon le modèle FRBR. L'affichage des résultats est organisé comme un tableau de bord selon 9 catégories prédéfinies.

D'autres exemples sont à signaler dans le domaine musical, comme Scherzo, catalogue de test du FRBR de la bibliothèque de l'université d'Indiana.

Les évolutions engendrées par l'apparition des catalogues 3.0

Le bibliothécaire de demain, un acteur du web sémantique

Actuellement les normes de catalogage ne permettent pas aux catalogueurs de récupérer la structure interne des documents pour renseigner les métadonnées. L'emploi des nouvelles normes de catalogage ouvrirait des possibilités de récupération automatique des métadonnées à partir des styles des documents XML par exemple.³² Un même format serait utilisé pour entrer aussi bien des métadonnées descriptives (contenant) que des métadonnées thématiques (contenu)³³.

Les nouvelles normes pourraient changer le métier de bibliothécaire³⁴. En effet le catalogueur sortirait de son isolement et pourrait se rapprocher d'autres corps de métiers comme les concepteurs de ressources sur le web.³⁵ Il deviendrait ainsi acteur du Web sémantique.³⁶ Une importante partie de son travail se situerait en amont de la relation avec l'utilisateur.

Mais cette évolution implique de nouvelles formations. Le bibliothécaire devra se former à la bibliothèque 3.0 et à un catalogage totalement renouvelé pour permettre un affichage modulaire. Pour le moment, les formations sont des stratégies personnelles d'apprentissage (learning strategy), sous forme de conférences ou de tutoriels³⁷. Cependant, ces formations personnelles peuvent passer au stade professionnel dans le cas de certaines bibliothèques qui en facilitent l'accès.

Evolutions des relations entre usagers et bibliothécaires : une médiation numérique accrue

Bruno Menon³⁸ distingue trois modèles de systèmes d'information liés au web sémantique, qui coexistent le plus souvent dans l'espace documentaire. Dans le premier, dit « classique », l'intermédiation bibliothécaire-usager est forte. Elle disparaît dans le second, dit « moderne » pour laisser l'utilisateur seul devant une interface unique. Dans le troisième système, qui correspond à la tendance actuelle (web 3.0), la médiation devient invisible aux yeux de l'utilisateur.

Evolutions des outils de catalogage : une offre démultipliée

La modification de la structure des catalogues a un impact sur la conception des logiciels. Ceux-ci deviendraient communs à plusieurs métiers et ne constitueraient donc plus des niches commerciales.³⁹ Le bibliothécaire pourrait alors avoir des choix multiples au niveau du prix et des fonctionnalités.

Pour aller plus loin:

Calendrier des événements concernant l'expérimentation de RDA (d'après ABES⁴⁰ et Gillian Cantello⁴¹)

Dates-clés	Événements autour de RDA, du web sémantique et de la norme
Juin 2009	Contenu de la norme RDA
Juin juillet 2009	Essai Bêta de RDA
4e trimestre 2009	Première version
08/08/2010	Séminaire <i>RDA in Europe - making it happen</i> (Copenhague, Danemark)
24/08/2010	Constitution de 2 groupes de travail à l'AFNOR
31/08/2010	Fin de la période de test gratuit du RDA Toolkit

Synthèse documentaire

Septembre 2010 à mars 2011	Mise à l'essai et préparation pour la mise en application, depuis Juillet 2010 pour les bibliothèques américaines.
13-15 septembre 2010	Conférence du CIG à Exeter (Royaume Uni)
16/09/2010	1ère réunion du groupe de travail <i>ad hoc</i> de l'AFNOR
15 octobre 2010 28 octobre 2010 19 novembre 2010 2 décembre 2010 Fin 2011	AFNOR : chapitres relatifs à la description bibliographique choix des accès notices d'autorité pour les <i>œuvres</i> et les <i>expressions</i> notices d'autorité pour les personnes, familles, collectivités et lieux). Décision finale en France
Avril – Juin 2011	Synthèse des tests RDA en Amérique du Nord et prise de décision sur l'adoption de RDA

Goasdoué, François et Léger, Alain (dir.), *Web sémantique : de la terminologie aux services web*, Hermes Science Publication, 2009

« Tim Berners-Lee on the next Web » (03/2009) [en ligne], disponible sur : http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html (consulté le 28/10/2010)

¹ W3C (30/04/2010), « Tim Berners Lee », [en ligne], disponible sur : <http://www.w3.org/People/Berners-Lee/> (consulté le 28/10/2010)

² Luly (14/06/2010), « Le web de données », [en ligne], disponible sur : <http://bibliotheques.wordpress.com/2010/06/14/le-web-des-donnees/> (consulté le 28/10/2010)

³ Id.

⁴ Herman, Ivan (15/10/2007), «Web Ontology Language (OWL)», [en ligne], disponible sur : <http://www.w3.org/2004/OWL/> (consulté le 28/10/2010).

⁵ Caribou (24/10/2006), « L'autre sémantique. Le Web Sémantique et les systèmes de recherche d'information. 1 : Définition et principes », [en ligne], disponible sur : <http://www.webmaster-hub.com/publication/L-autre-semantique-Le-Web.html> (consulté le 28/10/2010)

⁶ Id.

⁷ Institut National de Recherche en Informatique et en Automatique, «Métadonnées : mutations et perspectives», in *Séminaire INRIA, 29 septembre - 3 octobre 2008*, ADBS éd., 2008

⁸ Luly (14/06/2010), « Le web de données », [en ligne], disponible sur : <http://bibliotheques.wordpress.com/2010/06/14/le-web-des-donnees/> (consulté le 28/10/2010)

⁹ Davis, Ian et Vitiello, Eric (11/02/2004), « RELATIONSHIP: A vocabulary for describing relationships between people », [en ligne], disponible sur : <http://vocab.org/relationship/.html> (consulté le 28/10/2010)

¹⁰ Luly (14/06/2010), « Le web de données », op. cit.

¹¹ Caribou (24/10/2006), « L'autre sémantique. Le Web Sémantique et les systèmes de recherche d'information. 3 : Le Web Sémantique : une utopie condamnée ou une révolution en marche ? », [en ligne], disponible sur : <http://www.webmaster-hub.com/publication/L-autre-semantique-Le-Web,164.html> (consulté le 28/10/2010)

¹² Doctorow, Cory (26/08/2001), « Metacrap: Putting the torch to seven straw-men of the meta-utopia », [en ligne], disponible sur : <http://www.well.com/~doctorow/metacrap.htm> (consulté le 28/10/2010)

¹³ Guillaud, Hubert (21/01/2008), « Web sémantique : y aura-t-il une application qui tue ? », [en ligne], disponible sur : <http://www.internetactu.net/2008/01/21/web-semantique-y-aura-t-il-une-application-qui-tue/> (consulté le 28/10/2010)

¹⁴ Bermès, Emmanuelle (12/03/2010), « Le Web de données : perspectives pour les métiers de l'information documentation », [en ligne], disponible sur : <http://www.adbs.fr/le-web-de-donnees-perspectives-pour-les-metiers-de-l-information-documentation-79361.htm?RH=1266334869518> (consulté le 28/10/2010)

¹⁵ Groupe de travail IFLA sur les spécifications fonctionnelles des notices bibliographiques , *Spécifications fonctionnelles des notices bibliographiques : rapport final*, Bibliothèque nationale de France, 2001 (disponible sur : http://www.bnf.fr/documents/frbr_rapport_final.pdf)

¹⁶ Association des Bibliothécaires de France, *Le métier de bibliothécaire*, Ed. du Cercle de la Librairie, 2010

¹⁷ « BnF, modèles FRBR,FRAD et FRSAD », [en ligne], disponible sur : http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html (consulté le 30/10/2010)

¹⁸ « Rdabrochure-fre-fr.pdf », [en ligne], disponible sur : <http://www.rda-jsc.org/rda.html> (consulté le 24/10/2010)

¹⁹ « BnF – Lettre d'information – Actualités du catalogue : produits et services bibliographiques », [en ligne], disponible sur : <http://multimedia.bnf.fr/lettres/produits/produits21.htm> (consulté le 30/10/2010)

- ²⁰ Leresche, Françoise (20 et 21/05/2008), « Atelier6_Fleresche.pdf », [en ligne], disponible sur : http://www.abes.fr/abes/documents/reseau/journees_reseau/Atelier6_FLeresche.pdf (consulté le 30/10/2010)
- ²¹ Bellec, Joëlle (26/05/2010), « OAI_Dublincore.pdf », [en ligne], disponible sur : http://www.bnf.fr/documents/oai_dublincore.pdf (consulté le 30/10/2010)
- ²² « Declaring FRBR entities and relationships in RDF », [en ligne], disponible sur : <http://www.ifla.org/files/cataloguing/frbrg/namespace-report.pdf> (consulté le 30/10/2010)
- ²³ Op. cit. note 19
- ²⁴ Chapman, Anne (07/10/2010) « CILIP CIG Conference 2010 », [en ligne], disponible sur : <http://blogs.ukoln.ac.uk/cultural-heritage/2010/10/07/cilip-cig-conference-2010/> (consulté le 29/10/2010)
- ²⁵ Leresche, Françoise et Bourdon, Françoise (08/08/2010) « French Libraries Moving to RDA ? Key Issues » [en ligne], disponible sur : <http://www.slainte.org.uk/eurig/documents.htm> (consulté le 26/10/2010)
- ²⁶ ABES, « L'actualité de la normalisation », [en ligne], disponible sur : <http://www.abes.fr/abes/page.612.normes-et-documentation.html> [consulté le 24/10/2010]
- ²⁷ EURIG, (08/08/2010), « RDA in Europe: Making it happen. Summary of presentations by European countries on plans for moving to RDA » [en ligne], disponible sur : <http://www.slainte.org.uk/eurig/documents.htm> (consulté le 26/10/2010)
- ²⁸ BNF (11/08/2010), « Resource Description and Access" (RDA) : en France » [en ligne], disponible sur : http://www.bnf.fr/fr/professionnels/rda/s.rda_en_france.html?first_Art=non (consulté le 24/10/2010)
- ²⁹ Becker, Christian (2008), « Dbpedia : un hub de liens dans le web des données », [en ligne], disponible sur : <http://www.slideshare.net/christianbecker/dbpedia-un-hub-de-liens-dans-le-web-des-donnees> (consulté le 27/10/2010)
- ³⁰ Poupeau, Gautier (27/06/2008), « Le web de données et les bibliothèques », [en ligne], disponible sur : http://www.aula.asso.fr/index.php?option=com_docman&task=cat_view&gid=18&Itemid=44&mosmsg=Vous+tentez+d%27acc%27a%27cc%27E9der+%27E0+un+domaine+non+autoris%27E9.+%2728www.google.fr%2729 (consulté le 27/10/2010)
- ³¹ Martel, Marie D. (20/10/2010), « Bibliothèque numérique sociale : Australia Trove », [en ligne], disponible sur : <http://bibliomancienne.wordpress.com/2010/10/20/bibliotheque-numerique-sociale-australia-trove/> (consulté le 26.10.2010)
- ³² Le Boeuf, Patrick, *Functional requirements for bibliographic records (FRBR): hype or cure-all ?*, The Haworth information press, 2005
- ³³ id.
- ³⁴ Saw, Grace et Todd, Heather (19-23/08/2007), « Bibliothèque 3.0 : où en sont nos compétences ? », [en ligne], disponible sur : http://archive.ifla.org/IV/ifla73/papers/151-Saw_Todd-trans-fr.pdf (consulté le 01/11/2010)
- ³⁵ Op. cit., note 34
- ³⁶ Op. cit., note 32. slide27
- ³⁷ Op. cit., note 32. slide71
- ³⁸ Valderrama, Asuncion, *Documentaliste - Sciences de l'information*, vol. 48 n°3, 2009
- ³⁹ Op. cit., note 34
- ⁴⁰ Op. cit., note 25
- ⁴¹ Cantello, Gillian (29/05/2009), « RDA : pourquoi, quoi, qui et quand », [en ligne], disponible sur : <http://tsig.wikispaces.com/file/view/CantelloCLARDAgc-FRrev.pdf> (consulté le 01/11/2010)