

Vers le web sémantique ?

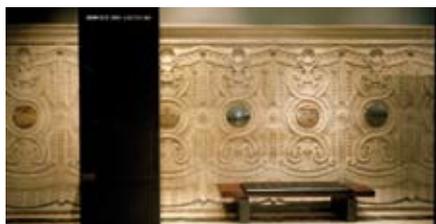
Les bibliothèques face aux normes du web

C'est la normalisation qui a fait le succès du web : c'est elle qui garantit son existence comme espace global et interopérable, indépendant des systèmes matériels et logiciels. Le W3C* est l'organisme chargé de développer ces normes.

Les bibliothèques, elles aussi, disposent d'une tradition normative riche et ancienne : avant même l'explosion du web, elles ont cherché le moyen de mettre leurs catalogues à disposition, en ligne, et à échanger des données informatiques pour rationaliser l'effort de catalogage. Toutefois, cette tradition normative est questionnée par les difficultés d'intégration de ces formats antérieurs au web au nouvel environnement des bibliothèques et à l'enjeu que représente leur visibilité en ligne.

Les standards du web sémantique, un ensemble de normes développées par le W3C autour du modèle RDF*, changent la donne en faisant miroiter la promesse de réconcilier ces deux univers...

Un peu d'histoire



<http://www.bnf.fr>

Dès la fin des années 1990, les fondateurs du web s'interrogeaient sur la nécessité de l'étendre pour que puisse être développé, en plus d'un réseau de documents, un réseau qui permettrait de relier les données

elles-mêmes. Le web sémantique est né : reposant sur un modèle logique de représentation des données structurées sous forme de triplets (sujet-prédicat-objet), il a vocation à permettre la création d'un espace global d'information, constitué de ressources connectées par des relations explicites, et exploitable par des machines. Une mission qui semble répondre exactement aux besoins des bibliothèques : échanger en ligne des données structurées et normalisées.

Dès lors, pourquoi les bibliothèques ne se sont-elles pas intéressées plus tôt aux technologies du web sémantique ? Sans doute en raison du paradoxe de l'œuf et de la poule... Pour que les bibliothèques fassent l'effort d'adopter de nouveaux standards, de nouveaux formats, il fallait que ceux-ci aient fait leurs preuves. Or, les normes du web sémantique ne pouvaient faire leurs preuves sans que soient disponibles d'importantes masses de données... par exemple celles des bibliothèques.

L'initiative *Linking Open Data*, en 2007, permet d'enrayer ce cercle vicieux. Encourageant la publication ouverte d'ensembles importants de données en RDF*, elle a permis de démontrer qu'il était possible d'appliquer les technologies du web sémantique de façon simple et directe, en suivant quelques principes de base.

La publication de DBPedia, une version RDF* issue d'une extraction des contenus de Wikipedia, a ouvert la voie : avec ces données disponibles, il devenait possible de démontrer l'intérêt du modèle, et d'agréger de nouvelles communautés, en leur offrant un point d'ancrage encyclopédique et universel.

L'important, c'est le lien

Car ce qui compte, sur le web de données, comme sur le web d'ailleurs, c'est le lien. Une information publiée sur le web n'a de valeur que si elle est reliée à d'autres, autorisant ainsi la découverte intuitive, par navigation d'une information à l'autre, en suivant les liens.

Rien de très exotique pour les bibliothèques, qui ont déjà intégré dans leurs modèles de données cette notion de liens : d'abord de façon pragmatique, en reliant les notices bibliographiques avec des notices d'autorité qui décrivent les personnes, les concepts, les lieux... Puis de façon plus formalisée en développant des modèles adaptés : le modèle FRBR*, suivi de FRAD* pour les autorités personnes, collectivités et familles, puis de FRSAD* pour les sujets. Le nouveau code de catalogage anglo-saxon RDA* formalise cette évolution en la rendant applicable dans les processus de production.

Avec ces évolutions, les modèles de données de bibliothèques, basés sur la description d'entités reliées entre elles par des relations, semblent naturellement destinés à être exprimés suivant les standards du web sémantique.

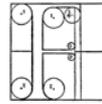
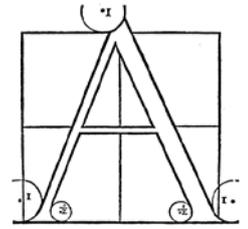
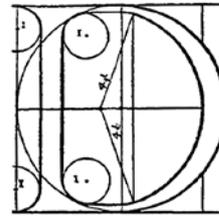
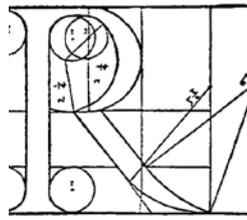
Qu'est-ce qu'on attend ?

En quelques années, de nombreux projets ont éclos : VIAF* pour les autorités et Open Library pour le livre, les bibliothèques nationales de Suède et de Hongrie qui ont rendu l'intégralité de leur catalogue accessible dans le web de données, suivies par des expérimentations du côté de la Deutsche Nationalbibliothek et de la British Library... En France également, des projets ont vu le jour (le site IdRef de l'ABES), d'autres se préparent (le projet data.bnf.fr de la Bibliothèque nationale de France).

Ces expérimentations en ordre dispersé révèlent une véritable dynamique d'adoption, qui nous autorise à dire que le web sémantique et le web de données sont devenus des concepts indispensables à maîtriser pour le bibliothécaire aujourd'hui.

Et pourtant, il existe encore un grand besoin de normalisation. Il ne s'agit pas d'élaborer de nouvelles normes : on laissera ce champ ouvert pour le W3C, qui poursuit le chantier web sémantique en y ajoutant progressivement les briques nécessaires pour l'adoption industrielle. Il s'agit bien plutôt de savoir comment appliquer ces normes, les décliner en fonction des besoins de notre communauté, de ses exigences, et de l'existant : toutes les données rétrospectives des catalogues actuels, qu'il faudra faire évoluer.

Opera del modo de fare le littere
maiuscole antique
Francesco Torniello da Novara (1490-1589)
Typographe milanais, écrivain et frère franciscain



Francesco Torniello da Novara a publié, en 1517, son ouvrage intitulé *Opera del modo de fare le littere maiuscole antique* (Gotardo da Ponte, Milan) présentant une police de caractères accompagnée de spécifications mathématiques permettant de les élaborer « au moyen de compas pour les courbes et avec une règle » (*col circino in curvo e libella recta*).

Le groupe Library Linked Data (LLD) fondé par le W3C en mai 2010, s'est donné pour mission d'améliorer l'interopérabilité globale des données de bibliothèque en étudiant, justement, la possible convergence des normes existantes dans notre domaine avec celles qui sont élaborées par le W3C. En se basant sur le recueil de plus de 40 études de cas, il dresse le tableau du voyage commencé par les bibliothèques vers le web de données : quels sont les objectifs, les moyens employés pour les atteindre, et surtout, quels sont les obstacles à franchir ?

Du côté de l'IFLA* et de RDA, des initiatives ont commencé pour exprimer les modèles bibliothéconomiques avec les formats du web sémantique. Le groupe LLD du W3C encourage cet effort en confrontant ces travaux à la communauté des experts du web sémantique.

Par-dessus tout, l'enjeu qui se cache derrière ces efforts convergents de normalisation, c'est celui de la création d'une véritable communauté d'intérêt et de pratique, autour de l'adoption des standards du web sémantique en bibliothèque. De nombreuses questions se posent, et nous apprenons encore en marchant : il est donc vital de disposer d'un réservoir de compétences, d'expertise, et de retour d'expérience sur ces premières applications. Les organismes de normalisation peuvent mettre au service de cette cause leurs experts et une infrastructure rodée.

Demain...

Quels sera le bénéfice de cet effort ? Sans jouer les diseuses de bonne aventure, on peut supposer que la convergence des normes de bibliothèque vers le web sémantique devrait permettre une implantation beaucoup plus naturelle des données de bibliothèque dans le web de demain.

Les échanges seront facilités, de même que la réutilisation des données en dehors du domaine des bibliothèques, et la construction d'applications innovantes pour lesquelles les données de la bibliothèque ne seront qu'un des maillons de la chaîne.

Dans un premier temps, cette évolution s'articulera avec nos systèmes et nos formats actuels. Mais si les normes du web sémantique tiennent leurs promesses, il faudra progressivement remettre en cause la logique même qui gouverne nos applications de production et d'accès, et réinventer complètement l'informatique documentaire. Ce qui ne se fera pas en un jour. Ni sans de bonnes normes bien solides !

Emmanuelle Bermès

BNF <http://www.bnf.fr>
 Direction des services et des réseaux
 Département Information bibliographique et numérique
 Service Prospective et services documentaires
 Emmanuelle Bermès emmanuelle.bermes@bnf.fr

Pour aller plus loin

Le web de données : perspectives pour les métiers de l'information-documentation
 Vidéo du 5 à 7 de l'ADBS du 18 février 2010, en ligne <http://www.adbs.fr/le-web-de-donnees-perspectives-pour-les-metiers-de-l-information-documentation-79361.htm>
 Pages « Web sémantique, Web de données » sur le site de la BNF http://www.bnf.fr/fr/professionnels/web_semantique_donnees/s.web_semantique_intro.html
 Le site du groupe W3C « Library Linked Data » <http://www.w3.org/2005/Incubator/ld/>



Le 27 mai **La nouvelle jeunesse des données d'autorité**
 La journée annuelle AFNOR-BNF sera, en 2011, consacrée aux données d'autorité dans la perspective du web sémantique. Son thème, Référentiels et données d'autorité à l'heure du web sémantique, confirme la montée en puissance des données d'autorité dans le nouveau contexte des métadonnées bibliographiques.
www.afnor.fr www.bnf.fr

* Sigles utilisés

- FRBR : Functional Requirements for Bibliographic Records
- FRAD : Functional Requirements for Authority data
- FRSAD : Functional Requirements for Subject authority data
- IFLA : International Federation of Library Associations
- RDA : Resource Description and Access
- RDF : Resource Description Framework
- VIAF : Virtual International Authority File
- W3C : World Wide Web Consortium

