

Le virage Linked Open Data en bibliothèque : étude des pratiques, mise en œuvre, compétences des professionnels

Jasmin HUGI
Bibliothèque cantonale et universitaire, Lausanne

Nicolas PRONGUE
<https://orcid.org/0000-0003-0938-6029>
Haute école de gestion de Genève

Résumé

Les Linked Open Data (LOD) mettent peu à peu en évidence des enjeux considérables pour les professionnels de l'information et les bibliothèques. Cet article fait un tour d'horizon de la question en abordant plusieurs aspects. Après avoir introduit les concepts-clés de cette thématique, il s'attache à déterminer l'utilité des LOD en bibliothèque en présentant quelques applications innovantes. Prérequis pour la création de telles applications, la conversion des données en LOD est décrite sous la forme d'un procédé généralisable. L'article change ensuite de perspective et se penche sur le professionnel de l'information, en identifiant les compétences les plus pertinentes à acquérir pour faire face à ces évolutions. Enfin, il décrit la réalisation concrète d'une formation à distance sur les LOD, accessible à tous sur le web.

1. Introduction

La thématique émergente des Linked Open Data (LOD) représente peut-être l'aube d'une révolution en bibliothèque. Avec cet article, nous avons souhaité brosser un aperçu des enjeux auxquels seront probablement bientôt confrontés les professionnels de l'information.

Tout d'abord, les concepts principaux entourant les LOD sont introduits (les termes et abréviations plus spécifiques sont définis dans un glossaire en fin d'article). Ensuite, ce texte s'attache à décrire l'utilité des LOD : pourquoi ces technologies sont-elles pertinentes et opportunes pour les bibliothèques ? Dans une troisième partie, la question est abordée d'un point de vue plus pragmatique, par la présentation d'un procédé générique de transformation de données en LOD. Ce modèle peut être réutilisé pour tout type de projet similaire. Une quatrième partie se penche ensuite sur les adaptations nécessaires des professionnels de l'information pour faire face à ces évolutions et se les approprier : quelles sont les compétences principales qu'ils devront acquérir ? Enfin, la dernière partie permet à tout un chacun d'approfondir individuellement ses connaissances sur les LOD : une formation sur la thématique y est présentée, librement accessible sur le web.

Ces divers aspects ont été traités dans trois travaux pratiques ou académiques que nous avons réalisés dans le cadre du Master en Information documentaire de la Haute école de gestion de Genève.

2. Définitions

Le terme Linked Open Data résulte de la fusion de deux concepts distincts : Linked Data et Open Data. Nous proposons ci-dessous une définition de ces deux concepts.

Linked Data est un terme créé en 2006 par Tim Berners-Lee dans le contexte du web sémantique. L'objectif du web sémantique est de briser les silos de données sur le web et de les rendre interopérables (Byrne, Goddard, 2010). Le web *classique* non sémantique se base sur un réseau hypertexte dans lequel des documents, les pages web, sont liés entre eux. Le web sémantique quant à lui veut créer un nouveau réseau, mais au sein duquel les données sont reliées entre elles, en plus des pages web. Pour obtenir des données liées ou des Linked Data, quatre principes doivent être respectés (Berners-Lee, 2009) :

1. Utiliser des adresses URI pour identifier les choses ;
2. Utiliser des adresses URI HTTP pour que l'on puisse consulter ces identifications ;
3. Fournir des informations utiles sous forme de standards (RDF, SPARQL) lors d'une recherche d'adresse URI ;
4. Inclure des liens vers d'autres adresses URI qui permettent de découvrir d'autres informations.

En d'autres termes, il faut donc attribuer des identifiants aux ressources sous forme d'adresses URI HTTP, les décrire en utilisant des standards du web sémantique et créer des liens vers des données externes pour créer un réseau. Si plusieurs acteurs du web utilisent les mêmes standards et les mêmes modèles, il devient plus facile d'échanger, de réutiliser et de lier les données. Mais ce scénario n'est possible qu'à condition que les données soient disponibles sous licence ouverte. C'est pour cela que les Linked Data à elles seules ne suffisent pas dans

l'optique du web sémantique ; elles doivent également être des Open Data, ou données ouvertes en français.

Ce second concept est défini ainsi par l'Open Knowledge Foundation (2012) :

« Une donnée ouverte est une donnée qui peut être librement utilisée, réutilisée et redistribuée par quiconque - sujette seulement, au plus, à une exigence d'attribution et de partage à l'identique. »

Par déduction, cette définition stipule donc qu'une donnée ouverte doit pouvoir être réutilisée à des fins commerciales également.

De manière générale, pour faciliter la réutilisation des données, trois aspects doivent être pris en compte (Chignard 2012 : p. 13-14) :

1. Aspect technique : un format le plus ouvert possible
2. Aspect juridique : une licence ouverte
3. Aspect économique : des redevances nulles ou limitées

L'Open Data s'inscrit dans une mouvance plus générale d'ouverture des connaissances, incluant entre autres l'Open Source (pour les logiciels), l'Open Research Data ou encore l'Open Access (pour les ressources d'information). Le concept inclut par ailleurs l'Open Government Data, qui concerne les données produites par des institutions de droit public.

L'interopérabilité technique et la capacité sémantique des Linked Data, couplées à la libre utilisation juridique et à la gratuité de l'Open Data, forment la puissance du concept de Linked Open Data. L'aspect ouvert des données permet ainsi au web sémantique de se populariser et facilite la création de données liées en donnant accès à des jeux de données externes de manière libre.

3. Quel intérêt pour les bibliothèques ?

Les LOD représentent un thème évoqué de plus en plus fréquemment en bibliothèque, mais dont la substance est parfois difficile à saisir. L'aspect très technique qui caractérise le web sémantique, ainsi que le jargon qui l'accompagne, contribuent largement à en faire une thématique mal comprise parmi les professionnels de l'information.

Afin de concrétiser ce sujet, nous avons réalisé une étude (Hügi, Prongué, 2014), proposant une analyse de diverses applications déjà existantes et basées sur les LOD en bibliothèque. D'après les résultats de ce travail, quatre grandes catégories de plus-values ont été identifiées. Elles sont présentées ci-dessous, illustrées par des exemples réels.

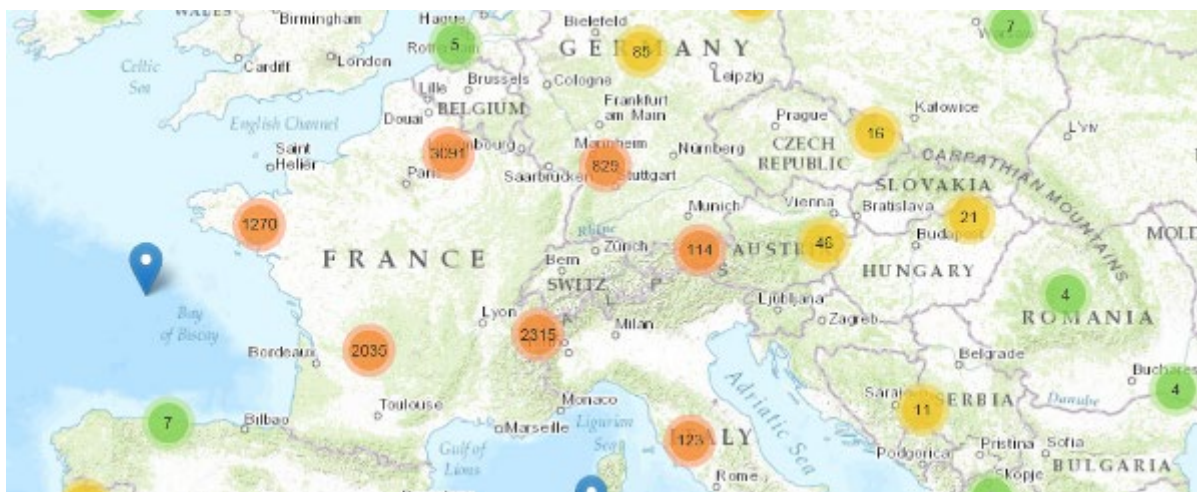
3.1. Une meilleure interopérabilité des données

Les données des bibliothèques sont actuellement enregistrées sous forme de notices en format MARC. Il s'agit d'un standard métier établi dans le monde entier depuis plus de 50 ans. Néanmoins, du fait de son utilisation en bibliothèque uniquement, ce format pose certains problèmes d'interopérabilité. Les autres acteurs de l'information utilisent actuellement des standards basés sur XML ou sur d'autres formats plus modernes.

MARC s'appuie fortement sur les chaînes de caractères : trop peu de champs sont contrôlés ou utilisent des identifiants, ce qui limite le traitement automatique des données par des machines. De plus, ce format est difficilement extensible et adaptable aux évolutions. En effet,

L'utilisation d'un standard répandu sur le web et d'une licence ouverte permet également aux données des bibliothèques d'être réutilisées par des acteurs externes au domaine. Au final, cette meilleure interopérabilité peut déboucher concrètement sur des gains de temps et d'argent pour les bibliothèques.

Les LOD se caractérisent par des liens avec d'autres jeux de données. Grâce à ces informations externes supplémentaires, des requêtes plus précises sont rendues possibles. L'enrichissement par des données géographiques telles que latitudes et longitudes permettent ainsi la mise en place de fonctionnalités de recherche au moyen d'une carte (illustration 1).



Sur le même principe, des liens avec des thésaurus similaires dans d'autres langues peuvent constituer la base de fonctionnalités de recherches multilingues. Le portail *The European Library* (<http://www.theeuropeanlibrary.org/tel4/>) a implémenté une fonction de ce genre proposant des termes de recherche en trois langues à l'utilisateur. Le portail finlandais Kulttuurisampo propose quant à lui de nombreuses fonctionnalités expérimentales, parmi lesquelles la recherche de relations entre deux personnes comme les peintres Rubens et Miró (illustration 2).



Illustration 2: Recherche de relation entre deux personnes

L'amélioration des possibilités de recherche peut élargir le rôle actuel du catalogue d'un dispositif de localisation vers un véritable outil de recherche.

3.3. Une meilleure sérendipité

Au-delà de l'amélioration des fonctionnalités de recherche, les liens vers des jeux de données externes augmentent la sérendipité de l'utilisateur. Plusieurs applications en proposent, par exemple la Deutsche digitale Bibliothek (<https://www.deutsche-digitale-bibliothek.de/>). Ces liens permettent de rediriger l'utilisateur vers des ressources complémentaires, comme des notices d'autorités d'autres bibliothèques ou des articles Wikipédia. VIAF (<http://viaf.org/>), un agrégateur de ressources de bibliothèques sur les personnes et les organisations, permet de faciliter la création de ce genre de relations.

Une institution culturelle ne possède pas que des données bibliographiques. En intégrant tous types de données au sein d'une application unique basée sur RDF, on offre la possibilité à l'utilisateur de trouver des informations qu'il ne cherchait pas forcément. Ainsi, le Centre Pompidou à Paris a converti des données de bibliothèque et de musée en un seul format, et les a reliées grâce à RDF. De cette manière, une seule recherche donne accès à des livres, des œuvres d'art numérisées, des dossiers pédagogiques ainsi que d'autres types de documents (illustration 3).

MAX ERNST

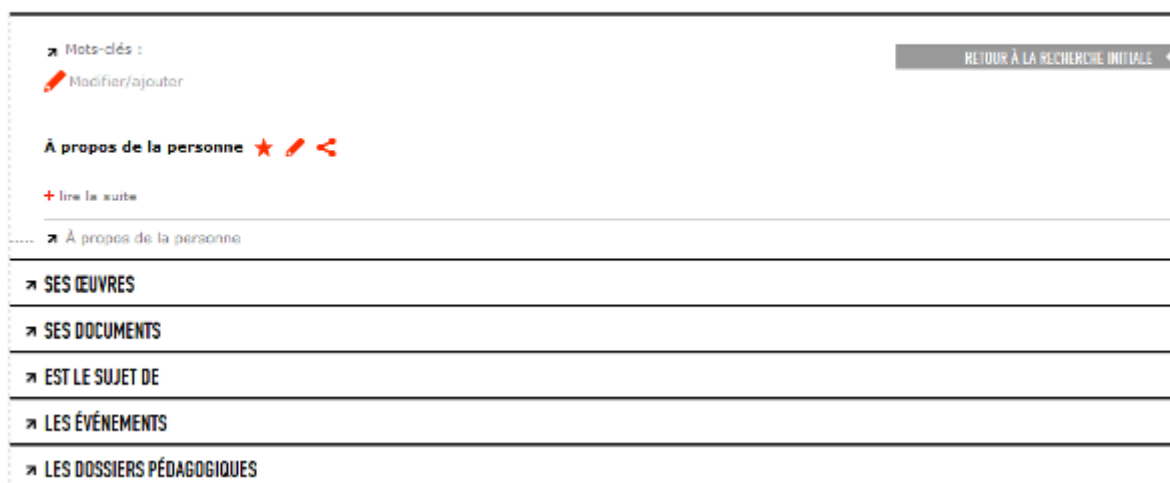


Illustration 3: Une seule interface pour les ressources variées du Centre Pompidou, ici sur Max Ernst

La variété des données liées ouvre des possibilités intéressantes de représentation de l'information. On s'oriente alors vers une recherche par navigation plutôt que par requêtes comme cela se fait dans les catalogues traditionnels. Dans ce sens, l'application data.bnf.fr propose la recherche géographique, mais également des représentations chronologiques des données.

Les méthodes de recherche évoluent également vers de nouvelles fonctionnalités d'auto-complétion, notamment grâce aux classes sémantiques attribuées aux données. L'application bêta d'Europeana (<http://eculture.cs.vu.nl/europeana/>) offre ainsi des propositions de types d'entités au cours de la saisie de la requête (illustration 4).

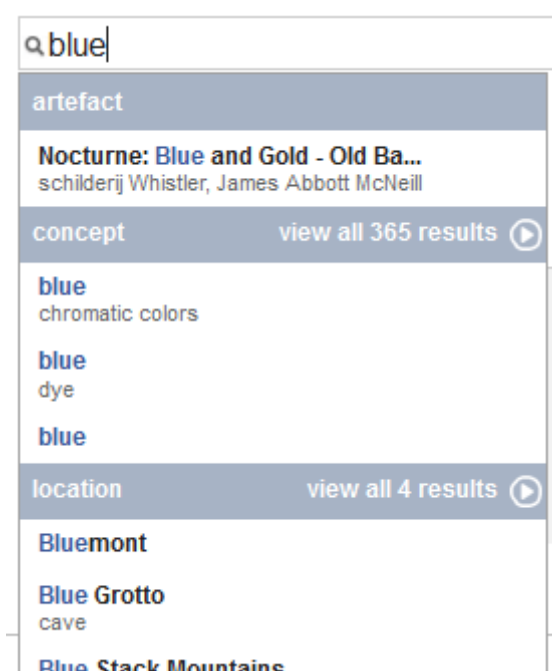


Illustration 4: Recherche par type d'entité

3.4. Une meilleure visibilité des données sur le web

Se basant sur des standards du web, les LOD proposent une interopérabilité appréciée par les moteurs de recherche, qui ont plus de facilité à indexer les contenus (Ostrowski & Pohl, 2012 : p. 272). A l'inverse, les données MARC paraissent en silos sur le web, car elles ne sont accessibles qu'au travers de portails, et ne communiquent pas avec le reste du web. L'utilisation d'identifiants uniques et pérennes pour les LOD permet de créer ces relations manquantes entre les ressources. Grâce à cela, des liens tiers pointant vers le catalogue améliorent la visibilité des données des bibliothèques. Par exemple, près de 80 % des visiteurs du service data.bnf.fr proviennent des moteurs de recherche, alors même que le catalogue traditionnel n'y est absolument pas visible (Simon & Peyrard, 2014).

4. Mise en place d'un projet Linked Open Data en bibliothèque

Les diverses plus-values présentées au chapitre précédent se basent sur les LOD, des données structurées selon le modèle RDF. Un projet LOD en bibliothèque commence donc forcément par la conversion des données préexistantes en RDF. Ce processus est présenté ci-dessous.

4.1. Les étapes d'un projet LOD

En collaboration avec RERO, le Réseau des bibliothèques de Suisse occidentale, nous avons entamé un projet de transformation des métadonnées de bibliothèque en LOD (Prongué, 2014). Sur cette base en a été déduit un processus générique composé de huit étapes. Ce processus s'inspire des bonnes pratiques issues de projets similaires, notamment au sein du réseau de bibliothèques suédois Libris (Malmsten, 2009), de la Bibliothèque nationale espagnole (Vila-Suero & Gómez-Pérez, 2013 : p. 580) ou du projet européen LOD2 (Auer, 2014 : p. 3-4).

Etape 1 : revue de la littérature

Cette première étape consiste à analyser les applications déjà existantes et les pratiques des autres bibliothèques face au web sémantique. Il s'agit également de se tenir au courant des évolutions techniques du domaine ainsi que des nouveaux standards et modèles pour les données bibliographiques. Ceci est particulièrement important dans une époque charnière comme celle-ci du point de vue des métadonnées en bibliothèque.

Etape 2 : analyse des données

Afin de déterminer précisément les données qui seront publiées en LOD, il faut les étudier. L'analyse peut se porter sur des jeux de données entiers (lesquels faut-il publier?) ou au niveau plus fin des champs de données au sein d'une base de ressources. Divers critères peuvent avoir leur importance, tels que la pertinence des données, leur qualité, leur quantité, leur normalisation, etc.

Etape 3 : modélisation

L'étape de la modélisation est cruciale, car en fonction du modèle de données retenu, des fonctionnalités différentes pourront être implémentées au sein d'une application sémantique.

Les données traditionnelles sont structurées en notices bibliographiques et en notices d'exemplaires. C'est un modèle simple, à deux niveaux, qui est dit *plat*. L'initiative BIBFRAME,

visant à remplacer MARC, a développé un modèle innovant basé sur trois niveaux : œuvre, instance et item. Enfin, l'IFLA promeut depuis plusieurs années le modèle FRBR, structuré en quatre niveaux : œuvre, expression, manifestation, item. Chaque solution possède ses avantages et ses inconvénients. A titre d'exemple, une trop grande complexité pourrait représenter un obstacle pour la création de liens avec d'autres types de ressources, mais pourrait également permettre de construire une application plus performante.

Selon le modèle choisi, il s'agit alors de l'appliquer à ses propres données en identifiant des équivalences : quelles données correspondent à quelles entités du modèle ? La solution peut se représenter sous la forme d'un schéma de type entités-relations.

Pour être reliée, chaque ressource devra ensuite recevoir un identifiant de type URI. Une réflexion sur la forme que ces URIs doivent adopter, variant selon les types de données, est nécessaire.

Etape 4 : mapping

Une fois qu'un modèle complet et cohérent a été conçu, il faut sélectionner des vocabulaires du web sémantique pour décrire au mieux les données à publier. Il est possible de créer son propre vocabulaire, mais il en existe déjà un très grand nombre sur le web, dont certains sont devenus des standards de facto. Créer son propre vocabulaire pourrait nuire à l'interopérabilité des données, car ce nouveau langage serait utilisé uniquement dans une institution (Dunsire et al., 2012 : p. 11). Le choix des vocabulaires peut dépendre de divers critères, tels que leur pertinence, leur niveau d'utilisation en général, leur niveau d'utilisation par les bibliothèques, leur précision, etc. Les plus utilisés sur le web sont Dublin Core (DC) pour décrire des ressources bibliographiques, Friend Of A Friend (FOAF) pour des personnes et leurs relations, et Simple Knowledge Organization System (SKOS) pour des systèmes de connaissance tels que des thésaurus (Jentzsch et al., 2011). Des éléments de plusieurs vocabulaires différents peuvent être utilisés pour décrire un même jeu de données.

Quand ce cadre théorique et conceptuel est posé, des règles de conversion pour chaque élément de donnée (chaque champ) peuvent être rédigées.

Etape 5 : liens externes

Partie intégrante du processus, la génération de liens externes intervient avant ou après l'étape de la transformation. Ces liens sont créés avec des référentiels du web sémantique : il s'agit de jeux de données de référence disponibles en LOD, et généralement spécialisés dans un domaine. Ainsi, pour les descriptions de personnes et d'organisations, on créera des liens vers le référentiel VIAF, ou directement vers des fichiers d'autorités de bibliothèques. Les lieux seront par exemple reliés aux données géographiques du référentiel GeoNames, les langues à Lexvo.org, des entités diverses à DBpedia, etc.

La génération de liens externes peut se faire directement d'après les données de base, au moyen de codes ou de descripteurs contrôlés préexistants (par exemple les codes de langue MARC). Une autre méthode, plus complexe, consiste à effectuer des alignements : il s'agit de comparer les chaînes de caractères présentes dans chaque jeu de données afin d'établir des équivalences.

Etape 6 : transformation

La transformation consiste à formuler en langage informatique des règles de transformation jusqu'ici décrites en langue humaine, et à les appliquer sur les données pour les convertir.

Étape 7 : contrôle qualité

A ce stade, les premières données RDF ont été créées, mais sans doute avec des erreurs. Contrôler la qualité de l'output est donc essentiel dans le processus. Cette tâche se réalise au moyen d'échantillons représentatifs des données. Pour plus de précision, il est envisageable d'avoir un échantillon spécifique à chaque règle de transformation.

Étape 8 : publication

Lorsqu'une qualité satisfaisante a été atteinte, les données peuvent être publiées sur le web. Pour être de vraies LOD, elles doivent respecter les standards du web sémantique et posséder une licence ouverte.

La publication peut se faire de diverses manières, la plus simple étant de déposer les données en fichiers téléchargeables (nommés dumps dans le jargon) sur un serveur. Un dump doit être mis à jour périodiquement. Il est aussi possible de proposer des données RDF grâce à des interfaces pour clients informatiques (API REST ou SPARQL endpoint) ou en les intégrant dans des pages web pour les moteurs de recherche (grâce à RDFa, aux microdata ou aux microformats).

4.2. Un processus itératif

Les huit étapes mentionnées plus haut mènent à la publication de LOD, mais pas de manière linéaire (illustration 5). Il est en effet recommandé d'adopter une approche itérative pour parvenir à ses fins, comme l'affirme Martin Malmsten (2009 : chap. 1), spécialiste travaillant à la Bibliothèque royale de Suède : « a "data first" approach is better than "perfect metadata first" ». Le principe d'une telle méthode consiste à réaliser, dans une première phase, toutes les étapes du processus assez rapidement avec des éléments basiques. Ensuite, on peut perfectionner et complexifier progressivement le modèle, le mapping, les liens externes et la transformation en repassant par certaines phases du processus, autant de fois que nécessaire.

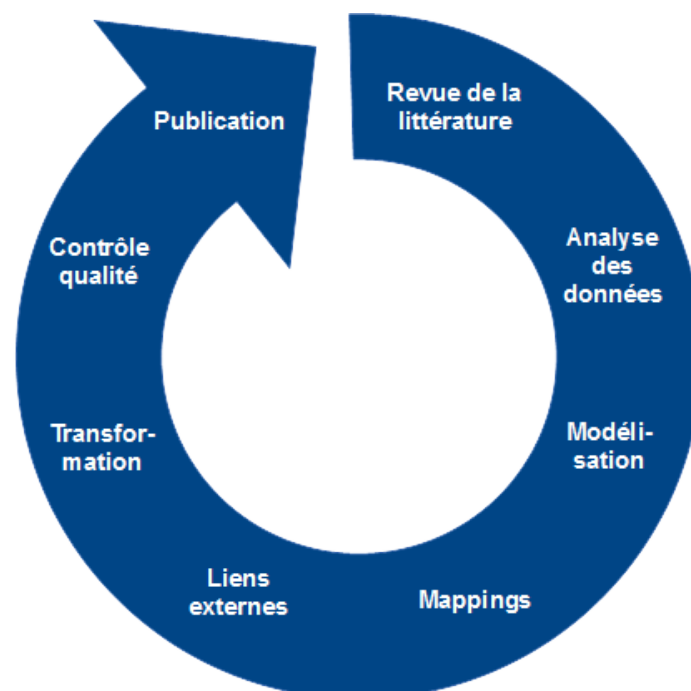


Illustration 5: Processus itératif

Après la l'étape de la publication, le processus continue. Si des dumps ou un SPARQL endpoint sont mis à disposition, les données RDF doivent être actualisées selon une périodicité régulière. Par ailleurs, le modèle et le mapping doivent être adaptés continuellement en fonction des évolutions du domaine.

4.3. Après la publication

La publication est un acte de simple mise à disposition des données afin qu'elles puissent être réutilisées librement par des tiers. Pour tirer un réel bénéfice des LOD dans une institution, il faut concevoir sur cette base une application destinée aux utilisateurs finaux. Cela peut générer des plus-values telles que celles mentionnées au chapitre 2.

Le site web data.bnf.fr illustre bien ce potentiel. Convivial, novateur, fluide et évolutif, il a déjà trouvé sa raison d'être au sein des services de la Bibliothèque nationale de France, grâce à la visibilité qu'il apporte à l'institution (Wenz, 2012 : p. 24).

Cependant, il existe aujourd'hui encore peu d'applications telles que celle-ci, basées sur les LOD et réellement innovantes. Leur conception demeure un processus complexe et peu connu, qui relève plutôt de l'expérimentation. Les technologies du web sémantique sont encore émergentes, mais possèdent un fort potentiel de développement, qui apportera sans doute bientôt de nombreuses innovations en bibliothèque.

5. Développer les compétences des professionnels de l'information

Dans les chapitres précédents, les avantages des LOD pour les bibliothèques et leurs usagers ont été présentés et les étapes d'un projet de transformation de données bibliographiques en LOD ont été expliquées. Il est évident qu'un tel projet n'est pas anodin à réaliser. De plus, la transformation de données en LOD ne représente que la moitié du chemin, car par la suite, il faut encore réussir à les exploiter. Se pose alors la question suivante : quelles sont les compétences nécessaires pour mener à bien un tel projet ? Pour y répondre, nous avons mené une étude (Hügi, Prongué, 2014) qui vise à identifier les compétences en matière de LOD pour les professionnels de l'information.

Cette étude a pour but d'informer les professionnels de l'information qui souhaitent se spécialiser dans ce domaine sur les nouvelles compétences à acquérir, apporter une aide aux managers devant engager un professionnel de l'information pour travailler sur cette thématique ou organiser des formations continues pour ses employés, ainsi que guider les responsables de plans d'études dans l'établissement de curriculums sur les LOD pour les professionnels de l'information.

5.1. Méthodologie

Afin de répondre à la question ci-dessus, nous avons décidé d'évaluer des offres d'emploi. Cette méthode permet d'avoir un aperçu des préférences en matière de compétences demandées de la part des employeurs vis-à-vis des nouveaux employés, ainsi que de leurs attentes et futurs besoins (Park et al., 2009 : p. 844). En outre, des experts ont été consultés afin de compléter les informations trouvées dans les offres d'emploi. Pour les évaluer, une analyse de contenu est préconisée par la littérature scientifique (Stanton et al., 2011 ; Marion et al., 2005 ; Orme, 2008).

Pour ce faire, une liste de plates-formes spécialisées s'adressant aux professionnels de l'information a été constituée. Seules les plates-formes couvrant un pays entier ont été retenues. Les pays suivants ont notamment été pris en considération : Suisse, Allemagne, Autriche, France, Belgique, USA. Dans les archives de chaque liste de diffusion, deux recherches par mot-clé ont été effectuées. La première avec le terme *Linked* et la seconde avec *RDF*. Seules les offres d'emploi exigeant un diplôme en science de l'information, en plus des compétences liées aux LOD, ont été retenues, pour éviter l'identification de compétences purement informatique. Ainsi, quinze offres d'emploi pertinentes ont été sélectionnées pour l'analyse de leur contenu.

En plus des offres d'emploi, nous avons décidé de nous adresser à des experts du domaine travaillant ou ayant travaillé sur un projet LOD en bibliothèque. Nous leur avons demandé quelles compétences ils exigeraient d'un professionnel de l'information pour collaborer à un projet LOD si c'était à eux d'effectuer le recrutement. Grâce aux réponses des experts, onze descriptions de compétences exigées se sont ajoutées aux quinze offres d'emploi identifiées, ce qui a permis de faire une analyse de contenu sur 26 cas en tout.

5.2. Résultats

Huit compétences clés ont pu être identifiées grâce à l'analyse de contenu de quinze offres d'emploi et onze avis d'experts sur les compétences des professionnels de l'information travaillant sur un projet LOD. Dans l'illustration 6, ces catégories sont énumérées selon le nombre de cas dans lesquels elles apparaissent. La signification de chaque catégorie est détaillée ci-après.

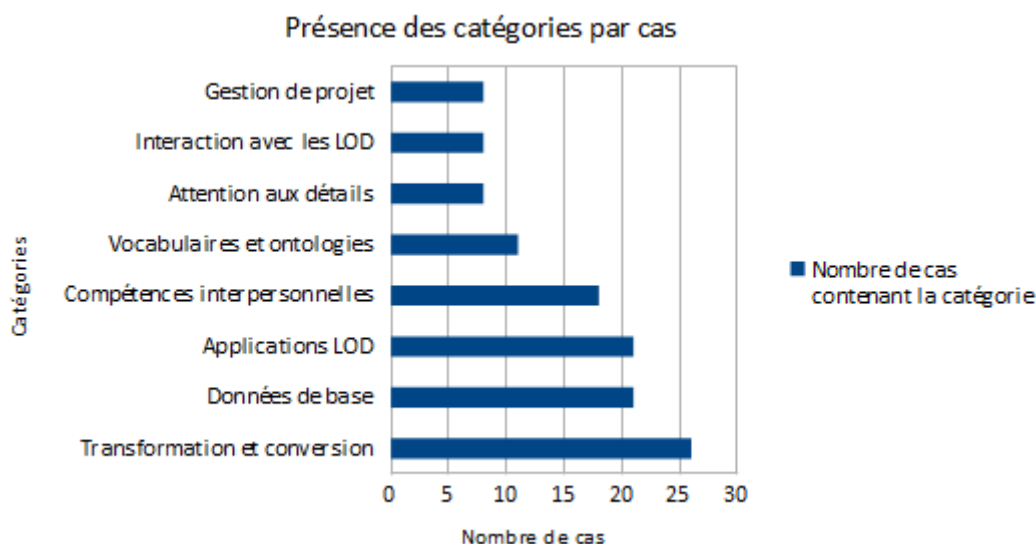


Illustration 6 : Huit compétences clés

Transformation et conversion

La catégorie *Transformation et conversion* comprend les compétences essentielles et nécessaires pour transformer et convertir des données en RDF afin d'en faire des Linked (Open) Data. Il s'agit donc de compétences liées à RDF et à la structure des données

correspondante. Cette catégorie englobe également des compétences en lien avec les identifiants persistants, l'attribution d'URIs, la modélisation des schémas de données, etc. La maîtrise de XML est également attribuée à cette catégorie, car ce métalangage sert souvent à la transformation de données en RDF. Cette catégorie figure au premier rang, ce qui n'est guère étonnant puisqu'elle correspond au critère utilisé pour faire la sélection des offres d'emploi.

Données de base

La catégorie *Données de base* regroupe toutes les compétences requises pour un projet LOD ayant lieu en bibliothèque. Puisque des projets LOD peuvent avoir lieu dans d'autres contextes également, cette catégorie représente l'aspect spécifique pour les bibliothèques. Afin de pouvoir transformer et convertir des données en RDF, il est important de bien connaître les données de base. En bibliothèque, ceci implique une bonne connaissance des règles de saisie (telles que les AACR2), des formats d'enregistrement (tels que MARC) et des systèmes de saisie (tels que les SIGB). De manière plus générale, il s'agit de compétences de catalogage et d'indexation ainsi que de connaissances des standards de métadonnées bibliographiques.

Applications LOD

Cette catégorie décrit les compétences informatiques nécessaires pour la conception, le développement et la mise en route d'une application basée sur les LOD. De plus en plus de projets en bibliothèque ne se contentent pas uniquement de transformer des données bibliographiques en RDF. Leur but est de créer une réelle plus-value pour les utilisateurs à travers des applications en ligne qui se basent sur les données converties en RDF. Pour ceci, un savoir-faire par rapport à la conception, la rédaction de spécifications, l'évaluation et la mise en œuvre de tests est requis. En outre, des connaissances utiles pour la mise en ligne et la création d'interfaces ainsi que du savoir-faire en lien avec le développement web (tels que HTML, CSS, Javascript) sont exigés. Le développement d'applications nécessite quant à lui des compétences en programmation dans des langages tels que PERL, Java ou Ruby.

Compétences interpersonnelles

Cette catégorie regroupe les aptitudes de communication, de collaboration et de réseautage. La proactivité ainsi que la maîtrise de l'anglais sont également incluses dans cette catégorie.

Vocabulaires et ontologies

Cette catégorie représente des compétences en lien avec les schémas de métadonnées sous forme de vocabulaires et d'ontologies. L'utilisation et la réutilisation de vocabulaires et d'ontologies existants contribuent grandement à la création de Linked Data. Il est donc important de connaître les plus spécifiques à son propre domaine. En outre, la capacité à choisir des éléments de schémas de métadonnées pertinents pour représenter ses propres données est également considérée sous cette catégorie.

Attention aux détails

La catégorie *Attention aux détails* décrit des caractéristiques personnelles englobant des qualités telles que la rigueur, la précision et un esprit logique. Le sens de l'organisation est également inclus dans cette catégorie.

Interaction avec les LOD

Cette catégorie consiste à de la recherche dans des ensembles de données LOD grâce à des interfaces spécifiques (SPARQL Endpoint) en utilisant le langage de requête SPARQL. En outre, la visualisation de données disponibles en tant que LOD et la connaissance d'outils permettant cette tâche représentent également une compétence de cette catégorie.

Gestion de projet

La catégorie *Gestion de projet* comprend les compétences telles que la planification, l'organisation, l'implémentation et l'établissement d'un budget.

L'analyse d'offres d'emploi montre la combinaison de compétences jugées idéales pour une fonction spécifique. Cette combinaison idéale exige très souvent une double compétence en science de l'information et en informatique. Ainsi, parmi les quinze offres d'emploi analysées, six demandaient cette double compétence pour un expert LOD. Néanmoins, un projet LOD se fait rarement de manière isolée et englobe d'habitude plusieurs employés avec plusieurs profils. Il n'est donc pas forcément nécessaire de réunir toutes les compétences demandées en un seul profil. En outre, la réalité se montre souvent très différente. Dans beaucoup de bibliothèques, le budget manque pour engager du nouveau personnel et les collaborateurs se forment sur place selon les besoins pour participer à un projet LOD. Les résultats de cette étude sont à interpréter dans leur contexte ; un profil idéal y est décrit, mais cela n'empêche pas le fait que, dans la réalité, beaucoup de ces compétences sont acquises sur le terrain.

Par ailleurs, le travail avec les LOD représente une fonction en plein développement. Il sera donc important de réévaluer périodiquement les compétences requises pour ce type d'emploi. Il se peut que les tâches décrites ici soient dans un futur plus ou moins proche directement intégrées dans les SIGB avec lesquels les professionnels de l'information saisissent les métadonnées. Les critères d'embauche pourront alors se développer dans une autre direction.

Le travail avec des LOD en bibliothèque nécessite visiblement des compétences très spécifiques qui peuvent, au premier abord, sembler intimidantes. Néanmoins, les professionnels de l'information possèdent déjà plusieurs de ces compétences. Ainsi, la catégorie touchant aux ontologies et aux métadonnées qui sont utilisées pour la description des ressources a toujours fait partie des connaissances de base des professionnels de l'information. Concernant les standards de catalogage, il est inutile d'évoquer que cela appartient à la bibliothéconomie. En outre, les futurs professionnels de l'information en Suisse acquièrent des compétences de développement web et de XML.

Ce qui manque réellement en ce moment, ce sont les compétences strictement liées au web sémantique (principes, standards et technologies). Selon nous, l'enseignement de ces compétences devrait être inclus dans le curriculum des professionnels de l'information. Une introduction obligatoire à la thématique ainsi que des cours à choix devraient être dispensés au niveau bachelor, alors que les étudiants de niveau master devraient acquérir une compréhension globale de la thématique afin de pouvoir de prendre des décisions stratégiques. Par ailleurs, des cours de formation continue sont à mettre en place pour que les professionnels sur le terrain puissent, eux aussi, se former aux LOD. En effet, le web sémantique est bien plus qu'une tendance et le monde des bibliothèques ne peut se permettre d'attendre.

6. Un exemple de formation en ligne

Comme le montrent les chapitres précédents, les LOD ont dépassé le stade de simple *buzz* et sont devenues des pratiques établies. Parmi toutes les évolutions qui touchent les bibliothèques en ce moment, à l'instar du nouveau code de catalogage RDA ou de l'initiative BIBFRAME, les LOD ne sont pas la seule à attirer l'attention des bibliothécaires. La transformation des données bibliographiques en LOD est-elle plus importante que le passage à RDA ? Dans quel projet une bibliothèque doit-elle investir en priorité ses précieuses ressources ? Face à ces évolutions, se positionner et effectuer des choix en toute connaissance de cause est nécessaire, même si les décisions prises consistent à attendre. Le choix d'une orientation stratégique touchant à cette thématique requiert en tous les cas une compréhension minimale du fonctionnement et des enjeux des LOD, un concept qui demeure plutôt abstrait et dont la plus-value pour l'utilisateur n'est pas évidente au premier abord.

En ce moment, il existe très peu de formations en ligne qui ne s'adressent pas à des informaticiens. C'est dans ce contexte-là qu'une formation courte en ligne a été développée (voir Hügi, 2014) sous mandat de la Bibliothèque cantonale et universitaire de Lausanne (BCUL). Le but était de transposer une formation en présentiel déjà existante (décrite dans Hügi, Schneider, 2014) dans un environnement numérique, et de permettre aux collaborateurs de la BCUL de se sensibiliser à la thématique des LOD en bibliothèque.

6.1. Méthodologie

La méthodologie ADDIE a été choisie pour la réalisation de la formation en ligne. Elle consiste en cinq phases (Herridge Group, 2004 ; Boling et al., 2011 ; Molenda, 2003) : l'analyse des besoins, la conception, le développement, l'implémentation et l'évaluation. Dans un premier temps, les besoins et les attentes des futurs apprenants et du prescripteur ont été recueillis, permettant de définir l'étendue et les objectifs de la formation, ainsi que le prérequis nécessaire pour pouvoir la suivre. Dans une deuxième phase, ces informations ont servi à concevoir le cours en ligne et à définir les modes d'accompagnement, d'interaction et d'évaluation, ainsi qu'à choisir la plate-forme en ligne. Par la suite, les supports de cours existants ont été transposés dans l'environnement numérique et complétés par de nouveaux contenus, des exercices et des tests. Dans cette étape, la structure générale et la navigation ont été créées, et une première évaluation de la formation a été effectuée. En ce qui concerne les phases de l'implémentation et de l'évaluation, qui ne sont pas couvertes par ce projet, des recommandations et un script ont été rédigés pour permettre de terminer ces étapes avec succès.

6.2. Résultats

La formation sur les LOD dans les bibliothèques est maintenant disponible sur la plate-forme Moodle de l'Université de Lausanne. Elle est constituée de quatre modules pas très techniques, qui proposent une sensibilisation et une introduction à la thématique.

Voici plus en détails le contenu couvert par les modules :

- **Module 1: L'évolution du web et son impact sur les bibliothèques**
Dans ce premier module, la naissance et l'évolution du web sont retracées. Cette évolution a mené à une déclinaison du web en plusieurs versions: le web 1.0, 2.0, 3.0 et 4.0. Les caractéristiques de chacune de ces versions sont expliquées et l'on démontre comment cette évolution a impacté le monde des bibliothèques.

- **Module 2: Linked Data, Open Data et Linked Open Data: définition des concepts**
Dans le deuxième module, le terme Linked Open Data (données ouvertes et liées) et les concepts annexes sont définis et expliqués. Les principes de base sont présentés et des exercices donnent un premier aperçu des possibilités qu'offrent les LOD.
- **Module 3: Présentation d'exemples de bibliothèques appliquant les LOD**
Dans ce troisième module, des exemples concrets de bibliothèques utilisant les LOD sont donnés. Une présentation des applications de la Bibliothèque nationale de France et du Centre Pompidou est fournie, accompagnée par des articles concernant ces deux projets.
- **Module 4: Les enjeux et les avantages des LOD pour les bibliothèques**
Dans le quatrième module, les avantages des LOD ainsi que les enjeux de ces dernières en bibliothèque sont présentés. La plus-value pour les usagers est démontrée à travers des exemples et le potentiel des LOD à moyen et long terme est illustré.

Les contenus sont médiatisés grâce à des présentations, des articles ainsi que des vidéos. La formation est disponible en ligne sur la plate-forme Moodle de l'Université de Lausanne et ouverte à tous. Les personnes disposant d'un identifiant d'une haute école ou d'une université suisse peuvent en plus accéder à des tests de connaissances.

La formation est accessible via le lien suivant :

<http://moodle2.unil.ch/course/view.php?id=2960>

Le cours est protégé par le mot de passe suivant : biblod

Une Licence Creative Commons « Attribution – Pas d'utilisation commerciale – Partage dans les mêmes conditions » a été attribuée à tous les contenus originaux, donc à tous les textes sur la plate-forme ainsi qu'aux présentations. Par conséquent, si quelqu'un souhaite un jour actualiser les contenus ou corriger des propos devenus obsolètes, il a le droit de le faire.

Cette formation en ligne sur les LOD en bibliothèque est la première en son genre à s'adresser aux professionnels de l'information et à être en français. Cependant, le travail ne s'arrête pas là ; il existe bien plus de contenus à enseigner, comme la logique des triplets, la structure de RDF, la modélisation des données, les différents vocabulaires, le fonctionnement d'une ontologie, etc.

Cette formation a le potentiel d'être complétée et peut faire partie d'un ensemble plus grand. Nous invitons les professionnels de l'information à la suivre, voire à la réutiliser dans un autre contexte.

7. Conclusion

Les LOD représentent une nouvelle tendance dont l'impact sur les métadonnées peut être énorme. Dans le monde entier, la communauté des bibliothèques se penche intensivement sur la question, tentant d'approprier ces technologies émergentes et d'adapter ou de renouveler les standards existant, pour certains, depuis plus de 40 ans.

Cet article a tenté de faire le point sur les problématiques majeures des LOD en bibliothèque, en introduisant tout d'abord la thématique et son contexte. Ensuite, l'importance parfois méconnue des LOD pour la communauté des bibliothèques est illustrée par divers avantages,

tels que des gains d'interopérabilité, de visibilité et de fonctionnalités. Une fois convaincues de l'intérêt de ces technologies, les institutions doivent alors les implémenter sur leurs propres données, en fonction de leurs propres besoins. Cet article donne un exemple type de processus à suivre pour la mise en place d'un tel projet, en huit phases clés.

Qui dit nouvelles technologies dit également nouvelles compétences. La quatrième partie de ce texte passe en revue les principaux domaines de qualification dans lesquels les professionnels de l'information devront se former pour mettre en place des initiatives LOD dans leur milieu.

Par ailleurs, nous invitons le lecteur à poursuivre sa lecture au moyen d'une formation en ligne sur le sujet. Celle-ci, présentée en fin d'article, apporte concrètement une pierre à l'édifice des LOD en bibliothèque.

Cet édifice n'en est actuellement qu'à l'étape de ses fondations. Les technologies évoluent encore et la courbe d'apprentissage pour les professionnels est raide. Pour pallier ce problème, des efforts de sensibilisation à la thématique sont nécessaires. L'intégration des LOD dans le cursus académique et au sein de la formation continue en science de l'information est une piste à considérer, tout comme l'augmentation de l'offre de formations en ligne en libre accès.

Du côté de la mise à disposition de services à valeur ajoutée, nous encourageons les acteurs à faire preuve d'audace et à se lancer dans le développement d'applications innovantes, sans forcément connaître à l'avance la forme exacte du résultat ni la plus-value pour l'utilisateur. Il s'agit alors, par l'expérimentation, de remettre en question les fondements mêmes de la recherche d'information tels qu'on les connaît actuellement. L'utilisation de méthodes de créativité ou la recherche exploratoire peuvent être de bons moyens pour découvrir des opportunités de plus-value encore ignorées que les LOD peuvent offrir aux bibliothèques.

GLOSSAIRE

AACR	Abr. de Anglo-American Cataloguing Rules. Règles de catalogage qui donnent des directives sur la formulation des données pour la description et l'exploration des ressources. AACR2 est la version no 2.
BIBFRAME	Abr. de Bibliographic Framework. Initiative internationale, dirigée par la Library of Congress, dont le but final est le remplacement du format MARC et la création d'un nouveau modèle et d'une ontologie pour la description des données bibliographiques.
Dublin Core	Schéma de métadonnées créé pour décrire des ressources numériques.
EAD	Abr. de Encoding Archival Description. Standard qui sert à l'encodage des instruments de recherche qui se trouvent en ligne et à la description de collections et de matériaux archivistiques.
FRBR	Abr. de Functional Requirements for Bibliographic Records (spécifications fonctionnelles des notices bibliographiques). Modélisation conceptuelle des informations contenues dans les notices bibliographiques, utilisant le modèle entity-relationship.
HTTP-URI	Voir sous URI
LOD	Abr. de Linked Open Data. Données structurées en RDF qui sont disponibles sous une licence ouverte.
MARC	Abr. de MACHine Readable Cataloging. Format de catalogage qui sert à la représentation et à la communication d'informations bibliographiques et connexes. Le format se décline en plusieurs versions, parmi lesquelles MARC21, INTERMARC, UNIMARC, USMARC, etc. MARC/XML est un format basé sur le langage XML développé afin de faciliter les échanges de données.
RDA	Abr. de Resource Description & Access. Nouvelles règles de catalogage censées remplacer les AACR.
RDF	Abr. de Resource Description Framework. Recommandation du W3C proposant des spécifications pour la description conceptuelle et la modélisation d'informations issues de ressources web. RDF utilise une variété de notations, de syntaxes et de formats de sérialisation. RDF est basé sur l'idée d'identifier des choses avec des HTTP URIs et de représenter les informations sous forme de triplets.
SIGB	Abr. de Système intégré de gestion de bibliothèque. Logiciel qui permet la gestion informatique des différentes tâches et activités d'une bibliothèque.

SPARQL	Abr. de SPARQL Protocol and RDF Query Language. Langage de requête qui sert à interroger des données RDF. Un SPARQL endpoint est une interface qui accepte des requêtes SPARQL d'autres logiciels et qui renvoie des résultats dans la forme prévue par SPARQL. Un formulaire SPARQL est une interface qui permet à des êtres humains de faire des requêtes SPARQL.
URI	Abr. de Uniform Resource Identifier. Standard qui permet d'identifier des ressources web et comprend des URL pour la localisation et des URN pour l'appellation.
XML	Abr. de eXtensible Markup Language. Métalangage et spécification du W3C pour créer des documents textuels structurés.

BIBLIOGRAPHIE

7.1. Les trois études à la base de cet article

HÜGI, Jasmin, 2014. *Développement d'une formation e-learning sur les Linked Open Data dans les bibliothèques*. Genève : Haute école de Gestion. Disponible prochainement à l'adresse : <http://doc.rero.ch/>

HÜGI, Jasmin et PRONGUÉ, Nicolas, 2014. *Les bibliothèques face aux Linked Open Data* [en ligne]. Genève. Haute école de gestion de Genève. [Consulté le 28 octobre 2014]. Disponible à l'adresse : http://doc.rero.ch/record/209598/files/M7-2014_memoire_HUGI-PRONGUE.pdf

PRONGUÉ, Nicolas, 2014. *Modélisation et transformation des métadonnées de RERO en Linked Open Data*. Genève : Haute école de Gestion. Disponible prochainement à l'adresse : <http://doc.rero.ch/>

7.2. Autres références

AUER, Sören (éd.), 2014. Introduction to LOD2. In : AUER, Sören (éd.), *Linked Open Data: creating knowledge out of interlinked data: results of the LOD2 project* [en ligne]. Heidelberg : Springer International Publishing. Lecture notes in computer science, 8661. p. 1-17. [Consulté le 8 décembre 2014]. ISBN 978-3-319-09846-3. Disponible à l'adresse : http://link.springer.com/content/pdf/10.1007%2F978-3-319-09846-3_1.pdf

BAKER, Thomas, BERMÈS, Emmanuelle, COYLE, Karen et DUNSIRE, Gordon, 2011. *Library Linked Data Incubator Group Final Report* [en ligne]. S.I. W3C. [Consulté le 19 juin 2014]. Disponible à l'adresse : <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>

BERMÈS, Emmanuelle, 2013. *Le web sémantique en bibliothèque* [en ligne]. Paris : Ed. du Cercle de la librairie. Collection bibliothèques. ISBN 978-2-7654-1417-9. Disponible à l'adresse : <http://www.electrelaboutique.com/ProduitECL.aspx?ean=9782765414179>

BERNERS-LEE, Tim, 2009. *Linked Data*. In : *Design Issues* [en ligne]. 18 juin 2009. [Consulté le 4 juin 2014]. Disponible à l'adresse : <http://www.w3.org/DesignIssues/LinkedData.html>

BOLING, Elizabeth, EASTERLING, Wylie V., HARDRE, Patricia L., HOWARD, Craig D. et ROMAN, Tiffany Anne, 2011. ADDIE: Perspectives in Transition. In : Educational Technology. 2011. Vol. 51, n° 5, p. 34 38.

BYRNE, Gillian et GODDARD, Lisa, 2010. The Strongest Link: Libraries and Linked Data. In: D-Lib Magazine [en ligne]. novembre 2010. Vol. 16, n° 11/12. [Consulté le 19 juin 2014]. DOI 10.1045/november2010-byrne. Disponible à l'adresse : <http://www.dlib.org/dlib/november10/byrne/11byrne.html>

CHIGNARD, Simon, 2012. Comprendre l'ouverture des données publiques. S.I. : FYP Editions. Collection entreprendre. ISBN 9782916571706

DUNSIRE, Gordon, HARPER, Corey, HILLMANN, Diane et PHIPPS, Jon, 2012. Linked Data vocabulary management: infrastructure support, data Integration, and interoperability. *Information standards quarterly*. 2012. Vol. 24, n° 2/3, pp. 4-13. DOI <http://dx.doi.org/10.3789/isqv24n2-3.2012.02>

GANDON, Fabien, CORBY, Olivier et FARON-ZUCKER, Catherine, 2012. Le web sémantique: Comment lier les données et les schémas sur le web ? S.I. : Dunod. ISBN 9782100581405.

HERRIDGE GROUP, 2004. The Use of Traditional Instructional Systems Design Models for eLearning [en ligne]. 2004. S.I. : s.n. [Consulté le 17 juillet 2014]. Disponible à l'adresse : <http://www.herridgroup.com/pdfs/The%20use%20of%20Traditional%20ISD%20for%20eLearning.pdf>

HÜGI, Jasmin et SCHNEIDER, René, 2014. Linked Open Data Literacy for Librarians. In : European Conference on Information Literacy. Dubrovnik : s.n. paraître 2014

JENTZSCH, Anja, CYGANIAK, Richard et BIZER, Christian, 2011. State of the LOD cloud. *LOD Cloud* [en ligne]. 19 septembre 2011. [Consulté le 24 juin 2014]. Disponible à l'adresse : <http://lod-cloud.net/state/>

MALMSTEN, Martin, 2009. Exposing library data as Linked Data. *Satellite meetings IFLA 2009: Emerging trends in technology, : libraries between web 2.0, semantic web and search technology* [en ligne]. Florence. 19 août 2009. [Consulté le 25 juillet 2014]. Disponible à l'adresse : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.181.860&rep=rep1&type=pdf>

MARION, Linda, KENNAN, Mary Anne, WILLARD, Patricia et WILSON, Concepción S., 2005. A Tale of two markets: employer expectations of information professionals in Australia and the United States of America. In : 71th IFLA General Conference and Council: « Libraries - A voyage of discovery » [en ligne]. Oslo : s.n. 2005. [Consulté le 4 septembre 2013]. Disponible à l'adresse : <http://arizona.openrepository.com/arizona/handle/10150/105082>

MOLEND, Michael, 2003. In Search of the Elusive ADDIE Model. In : Performance Improvement [en ligne]. 2003. Vol. 42, n° 5. [Consulté le 26 mars 2014]. Disponible à l'adresse : <http://search.proquest.com/eric/docview/62177054/584334380C63442EPQ/11?accountid=15920>

OPEN KNOWLEDGE FOUNDATION, 2012. Le manuel de l'opendata [en ligne]. Release 1.0.0. Open Knowledge Foundation. [Consulté le 13 juin 2014]. Disponible à l'adresse : <http://opendatahandbook.org/fr/>

ORME, Verity, 2008. <IT>You will be</IT> ...: a study of job advertisements to determine employers' requirements for LIS professionals in the UK in 2007. In : *Library Review*. 5 septembre 2008. Vol. 57, n° 8, p. 619-633. DOI 10.1108/00242530810899595

OSTROWSKI, Felix et POHL, Adrian, 2012. Zur Entwicklung eines Linked-Open-Data-Dienstes für Bibliotheksdaten. In : *Vernetztes Wissen: Daten, Menschen, Systeme: 5.-7. November 2012: Proceedingsband* [en ligne]. Jülich : Forschungszentrum Jülich, Zentralbibliothek, Verl. 2012. pp. 271-291. [Consulté le 3 juillet 2013]. Schriften des Forschungszentrums Jülich, Reihe Bibliothek ; Bd. 21. ISBN 978-3-89336-821-1. Disponible à l'adresse : <http://hdl.handle.net/2128/4699>

PARK, Jung-ran, LU, Caimei et MARION, Linda, 2009. Cataloging professionals in the digital environment: A content analysis of job descriptions. In : *Journal of the American Society for Information Science and Technology*. 2009. Vol. 60, n° 4, p. 844-857. DOI 10.1002/asi.21007

SIMON, Agnès et PEYRARD, Sébastien, 2014. Visibilité et web de données: réflexions autour du projet data.bnf.fr. In : *SemWeb.Pro 2014* [en ligne]. Paris. 5 novembre 2014. [Consulté le 8 décembre 2014]. Disponible à l'adresse : <http://semweb.pro/file/3396/raw/11-asimon.pdf>

STANTON, Jeffrey M., KIM, Youngseek, OAKLEAF, Megan, LANKES, R. David, GANDEL, Paul, COGBURN, Derrick et LIDDY, Elizabeth D., 2011. Education for eScience Professionals: Job Analysis, Curriculum Guidance, and Program Considerations. In : *Journal of Education for Library and Information Science*. 1 avril 2011. Vol. 52, n° 2, p. 79

VILA-SUERO, Daniel et GÓMEZ-PÉREZ, Asunción, 2013. datos.bne.es and MARiMbA: an insight into Library Linked Data. In : *Library hi tech*. 2013. Vol. 31, n° 4, p. 575-601. DOI 10.1108/LHT-03-2013-0031

W3C, 2013. Semantic Web. In : W3C [en ligne]. 2013. [Consulté le 14 juillet 2014]. Disponible à l'adresse : <http://www.w3.org/standards/semanticweb/>

WENZ, Romain, 2012. Linked Open Library Data in practice: lessons learned and opportunities for data.bnf.fr. *SWIB 2012* [en ligne]. Cologne. 27 novembre 2012. [Consulté le 13 octobre 2014]. Disponible à l'adresse : http://swib.org/swib12/slides/Wenz_SWIB12_108.ppt