Communauté UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Sciences de l'Information et de la Communication**

Arrêté ministériel : 25 mai 2016

Présentée par

Hiba MELHEM

Thèse dirigée par **Laurence BALICCO**

préparée au sein du **Laboratoire GRESEC (Groupe de recherche sur les enjeux de la communication)**
dans **l'École Doctorale Langues, Littératures et Sciences Humaines**

Usages et applications du web sémantique en bibliothèques numériques

Volume 1 : Mémoire

Thèse soutenue publiquement le **27 octobre 2017**,
devant le jury composé de :

Madame Viviane COUZINET
Professeure en Sciences de l'Information et de la Communication
Université Paul Sabatier, Toulouse3, Rapporteur

Monsieur Madjid IHADJADENE
Professeur en Sciences de l'Information et de la Communication
Université Paris 8, Président et Rapporteur

Madame Viviane CLAVIER
Maître de conférences HDR en Sciences de l'Information et de la Communication
Université Grenoble Alpes, Examinateur

Madame Laurence BALICCO
Professeure en Sciences de l'Information et de la Communication
Université Grenoble Alpes, Directrice de thèse

Remerciements

Ce travail de thèse n'aurait pas vu le jour sans le soutien et l'aide de plusieurs personnes. J'aimerais donc remercier ceux qui m'ont accompagné tout en long de ce travail, ceux qui ont été à coté de moi pendant les périodes de doute, ceux qui n'ont pas arrêté à m'encourager et ceux qui ont cru en moi et qui m'ont permis d'arriver au bout de cette thèse.

Je souhaite remercier en premier lieu ma directrice de thèse, Madame Laurence Balicco, d'avoir accepté de diriger cette thèse. Je lui suis reconnaissante pour le temps qu'elle m'a accordé, et pour ses qualités pédagogiques et scientifiques.

J'exprime toute ma gratitude aux membres du jury d'avoir bien voulu évaluer mon travail. Madame Viviane Couzinet, Madame Viviane Clavier et Monsieur Madjid Ihadjadène.

Je remercie Madame Isabelle Pailliart de m'avoir accepté au sein du GRESEC, et Razika pour son soutien et son aide professionnelle et amicale. Je remercie également mes collègues doctorants pour leur soutien moral et leur encouragement.

Mes remerciements vont également à tous mes collègues de l'IUT2 Grenoble. Merci à Dominique, Isabelle, Patrick, et Talal pour leur soutien, leur encouragement et les échanges que nous avons eu autour de mon sujet de thèse.

Je remercie toutes les personnes qui ont participé au questionnaire et aux entretiens. Merci de m'avoir accordé votre temps et de m'avoir fait confiance.

J'adresse aussi mes remerciements à tous mes amis qui m'ont encouragée et soutenue. Merci à Darine, Maguy, Weloré, Aude, Julien, Thomas, Audrey, Violaine, Cassandre, ainsi qu'à tous mes amis du théâtre.

Je souhaite remercier spécialement Manutea pour son soutien et sa patience tout au long de cette thèse. Merci d'avoir relu mon travail afin d'y apporter tes corrections et tes remarques. Merci beaucoup.

Enfin, je remercie ma famille, malgré la distance, vous êtes pour moi une source d'énergie et d'amour.

إلى أمي إخوتي وأخواتي، أنتم لي كل الحب والفرح.....

إلى أبي، برحيلك يصعب علي أحياناً ان أستعيد هذا السلام الهادئ بداخلي.....

Résumé

Ce travail de recherche se situe dans le champ interdisciplinaire des sciences de l'information et de la communication (SIC) et a pour but d'explorer la question de l'usage du web sémantique en bibliothèques numériques. Le web oblige les bibliothèques à repenser leurs organisations, leurs activités, leurs pratiques et leurs services, afin de se repositionner en tant qu'instituts de références pour la diffusion des savoirs. Dans cette thèse, nous souhaitons comprendre les contextes d'usage du web sémantique en bibliothèques numériques françaises. Il s'agit de s'interroger sur les apports du web sémantique au sein de ces bibliothèques, ainsi que sur les défis et les obstacles qui accompagnent sa mise en place. Ensuite, nous nous intéressons aux pratiques documentaires et à leurs évolutions suite à l'introduction du web sémantique en bibliothèques numériques. La problématique s'attache au rôle que peuvent jouer les professionnels de l'information dans la mise en place du web sémantique en bibliothèques numériques. Après avoir sélectionné 98 bibliothèques numériques suite à une analyse de trois recensements, une enquête s'appuyant sur un questionnaire vise à recueillir des données sur l'usage du web sémantique dans ces bibliothèques. Ensuite, une deuxième enquête réalisée au moyen d'entretiens permet de mettre en évidence les représentations qu'ont les professionnels de l'information du web sémantique et de son usage en bibliothèque, ainsi que de l'évolution de leurs pratiques professionnelles. Les résultats montrent que la représentation des connaissances dans le cadre du web sémantique nécessite une intervention humaine permettant de fournir le cadre conceptuel pour déterminer les liens entre les données. Enfin, les professionnels de l'information peuvent devenir des acteurs du web sémantique, dans le sens où leurs rôles ne se limitent pas à l'utilisation du web sémantique mais aussi au développement de ses standards pour assurer une meilleure organisation des connaissances.

Mots clés : bibliothèques numériques, web sémantique, web de données, usage, pratiques professionnelles, pratiques documentaires, professionnels de l'information

Abstract

This research work deals with the interdisciplinary field of the information and communication sciences (CIS) and aims to explore the use of the semantic web in digital libraries. The web requires libraries to rethink their organizations, activities, practices and services in order to reposition themselves as reference institutes for the dissemination of knowledge. In this thesis, we wish to understand the contexts of use of the semantic web in French digital libraries. It questions the contributions of the semantic web within these libraries, as well as on the challenges and the obstacles that accompany its implementation. We are also interested in documentary practices and their evolutions following the introduction of the semantic web in digital libraries. The problem is related to the role that information professionals can play in the implementation of the semantic web in digital libraries. After selecting 98 digital libraries following an analysis of three censuses, a questionnaire survey aims to collect data on the use of the semantic web in these libraries. Then, a second interview-based survey consists of highlighting the representations that the information professionals have of the semantic web and its use in the library, as well as on the evolution of their professional practices. The results show that the representation of knowledge within the semantic web requires human intervention to provide the conceptual framework to determine the links between the data. Finally, information professionals can become actors of the semantic web, in the sense that their roles are not limited to the use of the semantic web but also to the development of its standards to ensure better organization of knowledge.

Keywords: digital libraries, semantic web, linked data, use, professional practices, documentary practices, information professionals

Table des matières

Introduction générale **1**

Partie I : État de l'art **9**

Introduction **11**

1 Le web sémantique: un objet de recherche en sciences de l'information et de la communication (SIC) **13**

1.1 Le web sémantique : un terme à clarifier …… 14

1.1.1 Historique du terme …….. 14

1.1.2 Clarification du terme ……... 16

1.1.3 Evolution du terme ……... 16

1.1.4 Web sémantique, web de données, web 3.0, quel usage dans les bases de données et les moteurs de recherche ? …….. 19

1.2 Le web sémantique: une nouvelle approche pour la représentation des connaissances ……. 23

1.2.1 Les modèles de représentation des connaissances …… 25

1.2.2 L'architecture du web sémantique …… 28

1.2.3 Donnée, information, connaissance, quelle représentation ? …… 34

1.2.4 La notion de document revisitée dans le contexte du web sémantique …… 37

Conclusion du chapitre ……. 39

2 Les bibliothèques numériques: des dispositifs info-communicationnels ? **41**

2.1 La bibliothèque numérique, une notion à définir …..... 42

2.1.1 Bibliothèque numérique, électronique et virtuelle quelle terminologie ? …. 43

2.1.2 Bibliothèque numérique : analyse de définitions …..... 45

2.1.2.1 Définitions proposées par la communauté de pratique …..... 46

2.1.2.2 Définitions proposées par la communauté de recherche …....….. 49

2.2 La bibliothèque numérique : nécessité d'ancrage conceptuel en SIC …..... 53

2.2.1 La bibliothèque numérique : un dispositif info-communicationnel …...….. 53

2.2.1.1 La notion de dispositif …… 53

2.2.1.2 Dispositif info-communicationnel …….. 54

2.2.1.3 Positionnement par rapport à la notion de dispositif …….. 55

2.3 La médiation documentaire en bibliothèques numériques …… 56

2.3.1 Le traitement documentaire …….. 57
2.3.2 Les métadonnées en bibliothèques numériques……... 58
2.4 L'accès aux connaissances en bibliothèques numériques: quelle médiation? … 60
2.4.1 Médiation culturelle, médiation de connaissances, médiation de savoirs, quelle terminologie employer? ……. 60
2.4.2 Médiation des savoirs en bibliothèques numériques …… 62
Conclusion du chapitre ……64
3 Les usages et les applications du web sémantique en bibliothèques numériques 65
3.1 Usages et pratiques en bibliothèque numérique : le professionnel en question ..66
3.1.1 Usage et pratique, quelle utilisation en SIC ? …..... 66
3.1.2 Usage et pratique dans un contexte professionnel …..... 68
3.1.3 Usager-professionnel : quelle pratique ? …..... 68
3.1.3.1 Le professionnel de l'information …..... 69
3.1.3.2 Les pratiques et les activités informationnelles dans un contexte professionnel …..... 70
3.1.4 Evolution des pratiques professionnelles de l'information …..... 71
3.1.5 Les pratiques professionnelles en bibliothèques numériques …..... 72
3.2 Les applications du web sémantique en bibliothèques numériques: repenser les pratiques professionnelles .....…….. 76
3.2.1 Les catalogues en bibliothèques numériques: vers une approche orientée usager du web …..... 77
3.2.1.1 Les formats de description des ressources numériques : structuration et échange des métadonnées …....78
3.2.1.2 La représentation des connaissances au sein du web sémantique …..... 80
3.2.1.3 La recherche d'information au sein du web sémantique …..... 83
3.2.2 Le web sémantique en bibliothèque numérique, quel(s) usage(s) ? …..... 84
3.2.3 Comparaison entre les bibliothèques numériques et le web sémantique ….. 87
3.2.4 Le rôle de professionnel de l'information dans le contexte du web sémantique …..... 88
3.2.5 Les défis du web sémantique …..... 89
Conclusion du chapitre …..... 92
Conclusion 93

Partie II : Méthodologie de recherche et analyse des résultats 95

Introduction 97

4 Enquête de terrain : méthodes et déroulements 98

4.1 La sélection des bibliothèques numériques …… 99
4.1.1 Objectif de recherche …… 99
4.1.2 Les critères de sélection des bibliothèques numériques …….. 99
4.1.3 Les bibliothèques qui n'entrent pas dans le cadre de notre recherche……..100
4.1.4 Les outils de recensement ……... 101
4.2 Enquête quantitative – le questionnaire : les usages du web sémantique en bibliothèques numériques en France …… 104
4.2.1 Le questionnaire …….. 105
4.2.2 L'objectif de l'enquête ……. 105
4.2.3 La sélection …….. 105
4.2.4 Les moyens matériels ……... 105
4.2.5 Descriptif du questionnaire …….. 106
4.2.6 Le pré-test : mise à l'œuvre du projet de questionnaire ……... 108
4.2.7 La rédaction et la réalisation matérielle de l'enquête …….. 108
4.2.8 L'analyse des données …….. 109
4.3 Enquête qualitative – Entretiens semi-directifs : les pratiques documentaires dans le contexte du web sémantique …… 110
4.3.1 Objectif de recherche et hypothèses …….. 110
4.3.2 Méthode : entretiens semi-directifs ……111
4.3.3 Les participants …….. 112
4.3.4 L'échantillon de l'enquête …….. 113
4.3.5 La grille d'entretien …… 114
4.3.6 Le déroulement de l'entretien ……. 117
4.3.7 Le mode de saisie …….. 117
4.3.8 L'analyse des données …… 117
Conclusion du chapitre …….. 119

5 **Présentation et analyse des résultats 121**

5.1 Analyse de la sélection ……. 122
5.1.1 Classement des bibliothèques numériques ……. 122
5.1.2 Caractéristiques des bibliothèques numériques …….. 125
5.1.3 Les limites de l'analyse de sélection ……... 127

5.2 Présentation et analyse des résultats tirés du questionnaire …….. 128
5.2.1 Le profil des bibliothèques ……. 128
5.2.2 Profil des professionnels de l'information ……. 129
5.2.3 Connaissances des professionnels de l'information sur le web sémantique 130
5.2.4 Usage du web sémantique en bibliothèques numériques …….. 133
5.2.5 Le web sémantique et les pratiques documentaires …… 138
5.2.6 Les apports du web sémantiques en bibliothèques ……. 140
5.2.7 Les obstacles à l'utilisation du web sémantique en bibliothèques …….. 141
5.2.8 Le web sémantique et le futur des bibliothèques …… 143
5.2.9 Formation des professionnels de l'information à l'usage du web Sémantique ……..145
5.2.10 Limites du questionnaire …… 145
5.3 Présentation et analyse des résultats des entretiens ……. 146
5.3.1 Informations sur les enquêtés ……..146
5.3.2 Définition d'une bibliothèque numérique ……148
5.3.3 Les enjeux du web sémantique en bibliothèques numériques ……151
5.3.4 Les obstacles à l'utilisation du web sémantique en bibliothèques numériques ……. 157
5.3.5 Le web sémantique et les pratiques des professionnels de l'information …159
5.3.6 Evolution du métier des professionnels de l'information ……165
5.3.7 Limites des entretiens ……..167
Conclusion du chapitre ……168
6 **Les bibliothèques numériques : des nouveaux acteurs du web 172**
6.1 Valoriser les pratiques documentaires ……...173
6.1.1 Le web sémantique : une technique pour la documentation ……173
6.1.2 Les langages documentaires comme modèles conceptuels ……177
6.1.3 Evolution des pratiques professionnelles ……181
6.2 Valoriser les collections numériques ……182
6.2.1 Améliorer la visibilité des bibliothèques numériques sur le web ……182
6.2.2 Favoriser l'interopérabilité technique et sémantique ……...185
6.3 Favoriser l'ouverture des données ……187
6.3.1 Ouverture et partage des données avec d'autres communautés ……..188
6.3.2 Le linked open data ……190
6.4 Les défis et les obstacles à l'usage du web sémantiques en bibliothèques numériques ……192

Conclusion du chapitre ……195
Conclusion générale **197**
Bibliographie **207**

Tables des figures

1.1 Le web sémantique vu par Tim Bernes Lee en 1994, [Source : W3C] ……...…… 15

1.2 Web 3.0, [Source : Deutshe Telekom laboratories] ……..…….. 18

1.3 Nombre de documents pour chaque terme recherché dans la base de données : Web of science ……. 21

1.4 Nombre de documents pour chaque terme recherché dans la base de données: ScienceDirect ……... 21

1.5 Nombre de documents pour chaque terme recherché dans la base de données : Sage journals …… 21

1.6 Nombre de documents pour chaque terme recherché dans la base de données : Cairn …. 21

1.7 Résultats de recherche sur Google …….....……. 22

1.8 Résultats de recherche sur Google scholar …….…… 22

1.9 Répartition des résultats dans les trois bases de données : web of science, ScienceDirect et Sage …… 22

1.10 Exemple schématique d'un réseau de la mémoire représentant une hiérarchie à trois niveaux (d'après Collins et Quillian, 1969) [Nevers, 2000] …….. 26

1.11 Exemple schématique d'un réseau sémantique proposée par Collins et Lotfus (1975) [Nevers, 2000] ……. 26

1.12 Layer cake, [Source :w3c] …….. 29

1.13 Exemple d'un triplet RDF …… 31

1.14 Un exemple schématique de RDFS [Source W3C] …….…... 32

1.15 Représentation cyclique des notions : donnée, information et connaissance dans le contexte du web sémantique …….....……... 36

2.1 Répartition des termes (Digital library, Electonic library, Virtual library) dans les bases de données (Cairn , Sage Journals, Scholarvox, Science Direct) ….....…….. 45

2.2 La médiation documentaire numérique …….. 59

2.3 Positionnement des notions : connaissances, savoirs et culture …….. 62

2.4 La médiation documentaire et la médiation des savoirs en bibliothèques numériques …... 63

3.1 Représentation en SKOS d'une notice issue d'un thesaurus ….....……. 82

5.1 Répartition de nombre de bibliothèques sélectionnées selon leur type ……. 125

5.2 Les catégories des bibliothèques numériques ……... 128

5.3 Taux de réponses au questionnaire selon le type de bibliothèque ……....…….. 129

5.4 Répartition de nombre de professionnels de l'information selon leur profil …… 129

5.5 Compréhension de l'expression "web sémantique" par les enquêtés …… 130

5.6 Compréhension de l'expression "web de données" par les enquêtés ……. 130

5.7 Compréhension de l'expression "Ressource Description Framework" par les enquêtés .. 131

5.8 Compréhension de l'expression "Web Ontology Language" par les enquêtés ……. 131

5.9 Compréhension de l'expression "Simple Knowledge Organization System (SKOS)" … 131

5.10 Connaissances de quelques notions liées au web sémantique par les professionnels de l'information …….. 132

5.11 Nombre de bibliothèques utilisant ou non le web sémantique …….. 133

5.12 Répartition des standards du web sémantique utilisés en bibliothèques ……... 137

5.13 Les apports du web sémantique en bibliothèques numériques ……...141

5.14 Les obstacles à l'utilisation du web sémantique en bibliothèques numériques……..….142

5.15 Répartition de nombre d'enquêtés souhaitant suivre ou non une formation sur le web sémantique …… 145

6.1 Le système d'annotations de « 5 stars linked open data » ……. 190

Liste des tableaux

2.1 Analyse des définitions de la notion de bibliothèque numérique (communauté de pratique)……. 46

2.2 Synthèse des définitions issues de la communauté de pratique ….....…… 48

2.3 Analyse des définitions de la notion de bibliothèque numérique (communauté de recherche) ……. 50

2.4 Synthèse des définitions issues de la communauté de recherche ….....….. 52

3.1 Classement des compétences et des connaissances des professionnels de l'information : étude réalisée par [Choi et Rasmussen, 2006] ….....…….73

3.2 Connaissances spécifiques à la discipline (LIS) pour la nouvelle génération des professionnels [Nonthacumjane, 2011] ….....….. 74

3.3 Comparaison des fonctionnalités liées au web de données dans cinq bibliothèques nationales [Hallo etal, 2016] ….....…… 86

3.4 Comparaison entre la bibliothèque numérique et le web sémantique ….....……87

4.1 Grille d'analyse du questionnaire …… 109

4.2 Les bibliothèques numériques qui ont répondu au questionnaire …….. 112

4.3 Le nombre d'entretiens réparti sur les quatre types de bibliothèques ……. 114

4.4 Grille d'analyse des entretiens …… 118

5.1 Une comparaison entre les deux typologies de bibliothèques numériques …….123

5.2 Caractéristiques des quatre catégories des bibliothèques numériques …….. 127

5.3 Analyse thématique de la question «pourquoi la bibliothèque utilise le web sémantique »…… 134

5.4 Analyse thématique de la question «pourquoi la bibliothèque n'utilise pas le web sémantique» ……... 135

5.5 Les raisons pour lesquelles le web sémantique a changé ou non les pratiques documentaires des professionnels de l'information enquêtés …… 139

5.6 Analyse thématique de la question « pourquoi le web sémantique n'a pas changé vos pratiques documentaires » ……...139

5.7 Analyse thématique de la question sur l'importance du web sémantique pour les bibliothèques numériques …….. 143

5.8 Les profils des enquêtés …….. 146

5.9 Les missions des enquêtés au sein des bibliothèques …… 147

5.10 Les définitions de bibliothèque numérique selon les enquêtés …….. 148

5.11 Les caractéristiques d'une bibliothèque numérique selon les enquêtés …….. 151

5.12 Analyse thématique des réponses à la question concernant l'utilisation de l'expression « web sémantique » et « web de données » …… 152

5.13 Analyse thématique de la question sur les enjeux du web sémantique en bibliothèques numériques ……. 154

5.14 Analyse des réponses à la question sur les obstacles à l'utilisation du web sémantique en bibliothèques numériques …….. 157

5.15 Analyse thématique des réponses à la question sur les pratiques professionnelles mobilisées pour accompagner la mise en place du web sémantique en bibliothèques numériques …… 161

5.16 Analyse thématique des réponses liées à la représentation des données dans le cadre du web sémantique ……. 163

5.17 Analyse thématique des réponses à la question de l'évolution du métier des professionnels de l'information ……165

5.18 Les bibliothèques qui utilisent le web sémantique (Résultats du questionnaire) ……168

5.19 Les bibliothèques qui n'utilisent pas le web sémantique (Résultats du questionnaire) .. 169

5.20 L'importance du web sémantique pour le futur des bibliothèques (Résultats du questionnaire) ……. 169

5.21 L'usage du web sémantique en bibliothèques numériques (Résultats des entretiens) ….170

6.1 Les techniques de traitement des données utilisées dans le cadre du web sémantique et des bibliothèques numériques …… 174

6.2 Les pratiques, les compétences et les connaissances nécessaires pour accompagner l'usage du web sémantique en bibliothèques numériques …….. 181

Introduction générale

Depuis son apparition au début des années 1990, le web a profondément transformé la société contemporaine. Il est désormais omniprésent dans nos vies, que ce soit dans notre façon de communiquer, de travailler, de nous divertir, d'acheter des produits etc. Il est maintenant l'application la plus utilisée de l'Internet pour créer, partager et utiliser de l'information.

Victime de son succès, le web est devenu un gigantesque réservoir d'informations rendant parfois la recherche d'information laborieuse, d'autant plus lorsqu'il s'agit de retrouver des informations fiables et pertinentes. Face à ce problème, l'inventeur du web Tim Berners-Lee a eu l'idée d'ajouter une « sémantique » aux documents du web [Berners-Lee, 1996]. Cette idée s'est concrétisée en 2001 quand Tim Berners Lee a parlé du web sémantique pour désigner une évolution du web qui permettrait aux données disponibles (contenus et liens) d'être plus facilement utilisables et interprétables à la fois par les hommes et par les machines [Berners-Lee et al., 2001].

Par ailleurs, le web a apporté des modifications considérables aux bibliothèques puisque «*c'est toute la chaîne de création, diffusion, et médiation qui est changée*» affirment Anne-Marie Bertrand et Yves Alix [Bertrand et Alix, 2015, p. 84]. Les bibliothèques se retrouvent obligées de repenser leurs organisations, leurs activités, leurs compétences et leurs services afin de se repositionner en tant qu'instituts de références pour la diffusion des savoirs. Leur rôle dans l'organisation, la conservation et la mise à disposition des documents doit garder une place fondamentale tout en créant un équilibre entre les innovations techniques et les conditions sociales de création et de diffusion des documents.

De ce fait, aborder la question du web sémantique en bibliothèques est particulièrement important. Il s'agit, d'une part, de mieux comprendre son contexte d'usage en bibliothèques numériques françaises, et d'autre part, de s'interroger sur son influence sur les pratiques documentaires et l'accès aux savoirs.

Les thèmes de recherche que nous abordons dans cette thèse s'intéressent à l'usage et à la mise en place du web sémantique en bibliothèques numériques. Cela nous amène à nous interroger sur les définitions des bibliothèques numériques et du web sémantique. Nous nous intéressons plus précisément aux pratiques des professionnels de l'information et à leurs évolutions suite à l'introduction du web sémantique en bibliothèques.

Dans le cadre de notre travail, nous pensons que les notions de bibliothèque numérique et de web sémantique nécessitent un ancrage conceptuel en sciences de l'information et de la communication. Les bibliothèques numériques peuvent être pensées comme des dispositifs de traitement de l'information, en référence au terme employé par Yves Jeanneret lorsqu'il étudie les médias

informatisés. Autrement dit, il s'agit de considérer les bibliothèques numériques comme des dispositifs techniques ayant comme effet social de faire circuler les informations et rendre possible leurs échanges et la production des connaissances et des savoirs dans la société [Jeanneret, 2000, p.59]. Ainsi, pour Hervé le Crosnier les bibliothèques numériques ont un effet social au regard de l'accès aux connaissances [Le Crosnier, 2007]. En d'autres termes, les bibliothèques numériques sont à la fois des dispositifs techniques permettant le traitement des données via des langages informatiques et des logiciels ; des dispositifs sociaux permettant la diffusion des savoirs ; et des dispositifs documentaires qui selon Cécile Gardiès, rassemblent *« des objets, objets de savoir, organisés par des actions humaines dans un but de stockage, de conservation et d'accès »* [Gardiès, 2012, p.117]. De ce point de vue, les bibliothèques numériques peuvent être pensées comme des dispositifs d'accès à l'information, une notion utilisée par (Chaudiron, Ihadjadène, 2008, 2010 ; Simonnot, 2012) pour désigner les activités et les pratiques liées à la recherche de l'information via l'usage d'un outil technique.

Quant au web sémantique, il peut être considéré comme un domaine d'application pour la représentation des connaissances. Pour le collectif Roger T. Pédauque, il fait partie de l'ingénierie documentaire [Pédauque, 2006]. Ainsi pour Bruno Bachimont, il est la figure de l'ingénierie des contenus [Bachimont, 2007]. Enfin, selon Jean-Michel Salaun il fait partie de l'architecture de l'information [Salaun, 2012]. Le web sémantique considéré comme un domaine de recherche de la représentation des connaissances exige de mettre en question les pratiques documentaires en bibliothèques numériques. Par ailleurs, l'usage du web sémantique s'inscrit dans un mouvement visant une évolution des normes et des modèles de catalogage dans les bibliothèques. En France, la bibliothèque nationale de France et l'ABES[1] mènent des travaux d'analyse et de remise en question des standards et des normes utilisés en bibliothèques françaises.

Dans ce travail de recherche, nous souhaitons tout d'abord comprendre les contextes d'usage du web sémantique en bibliothèques numériques. Il s'agit de s'interroger sur les apports du web sémantique au sein des bibliothèques numériques, ainsi que sur les défis et les obstacles qui accompagnent sa mise en place. Ensuite, nous nous intéressons à la relation qui existe entre les techniques utilisées dans le cadre du web sémantique et les techniques documentaires permettant la représentation des connaissances dans le contexte des bibliothèques numériques.

Cette thèse s'inscrit dans le champ interdisciplinaire des sciences de l'information et de la communication (SIC) et a pour but d'explorer la question de la mise en place des applications du web sémantique en bibliothèques numériques. Pour Bernard Miège, *«le développement des*

1 Agence bibliographique de l'enseignement supérieur

(nouvelles) techniques de l'information et de la communication amenant à s'interroger sur les changements qu'elles entraînent, ou qu'elles accompagnent, au sein des structures de médiation » [Miège, 2005, p.60]. Considéré comme une nouvelle technique pour la représentation de l'information, le web sémantique amène donc à s'interroger sur les changements qu'il entraîne en bibliothèques numériques.

Nous nous interrogeons plus particulièrement les changements intervenus sur les pratiques des professionnels de l'information. Ainsi, notre étude s'inscrit dans une approche info-communicationnelle qui, selon Céline Paganelli, repose *«sur la prise en considération des activités informationnelles du point de vue, d'un côté, des pratiques et usages et, de l'autre, de la production et de la conception des dispositifs.»* [Paganelli, 2013, p.237]. Cette approche permet d'étudier les usages sous l'angle de la médiation. Une médiation documentaire qui réintroduit des pratiques professionnelles, et une médiation des savoirs qui s'inscrit dans une politique d'ouverture des données des bibliothèques sur le web.

Le développement d'une nouvelle technique de représentation des connaissances (le web sémantique) au sein des bibliothèques numériques entraîne une reconfiguration des médiations exercées autour des données et des usagers.

Dès lors, la problématique est formulée comme suit :

En quoi l'usage du web sémantique en bibliothèques numériques peut affecter les pratiques documentaires des professionnels de l'information et quel rôle joue-t-il dans l'évolution de leurs métiers ?

Les objectifs de cette recherche peuvent être formulés ainsi :

- Étudier les bibliothèques numériques en tant que dispositifs info-communicationnels (dimension technique, sociale, médiation).
- Étudier le web sémantique en tant que nouvelle approche pour la représentation des connaissances.
- Repenser la médiation documentaire dans un contexte où le web sémantique est utilisé en bibliothèques numériques.
- Comprendre les enjeux et les intérêts de relier les données des bibliothèques avec d'autres données sur le web.

L'hypothèse principale sur laquelle se base notre recherche est que l'utilisation du web sémantique en bibliothèques numériques nécessite une médiation documentaire. Une donnée ne peut pas être traitée seulement de façon automatique, c'est-à-dire uniquement par l'utilisation des langages normalisés du web sémantique (XML, RDF...). La mise en contexte des données nécessite une intelligence sociale (Cotte, 2013). Autrement dit, dans le cadre du web sémantique l'enrichissement des données par des métadonnées doit reposer sur un traitement documentaire dont l'objectif est d'associer les données à des contextes d'usages, précis et non triviaux, qu'un algorithme seul ne peut réaliser.

De cette hypothèse principale dérive trois hypothèses secondaires :

Hypothèse H1

Les professionnels de l'information peuvent jouer un rôle dans la mise en œuvre et l'évolution du web sémantique en introduisant leur expertise documentaire relative à la gestion des métadonnées.

Hypothèse H2

La représentation des connaissances ne se fait pas seulement en utilisant les langages informatiques et standardisés du web sémantique, il faut une intervention humaine permettant de fournir à ces langages les données et les modèles conceptuels nécessaires pour représenter les connaissances.

Hypothèse H3

Les bibliothèques numériques peuvent se servir du web sémantique pour proposer de nouveaux services plus adaptés aux usagers du web et pour rendre les bibliothèques plus visibles.

Le plan de notre manuscrit s'articule suivant deux parties.

La première partie (chapitres 1, 2 et 3) vise à situer notre recherche par rapport aux domaines des bibliothèques numériques et du web sémantique. Le premier chapitre est consacré à la présentation du web sémantique. Tout d'abord, un détour historique nous permet de comprendre l'évolution de ce terme. Ensuite, nous mettons l'accent sur certaines approches liées à la représentation de l'information que nous mobilisons tout au long de notre recherche. Cette approche nous conduit à discuter des notions de donnée, d'information, de connaissance ainsi que du document dans le cadre du web sémantique. Dans le second chapitre, nous discutons de la notion de bibliothèque numérique. Après une analyse des différentes définitions nous clarifions cette notion. Ces bibliothèques peuvent être pensées comme des dispositifs info-communicationnel nécessitant de prendre en considération la notion de médiation sous ces différentes formes qui sont la médiation documentaire et la médiation des savoirs. Enfin, dans le chapitre 3, nous analysons les usages et les

applications du web sémantique en bibliothèques numériques. Nous nous intéressons plus particulièrement aux pratiques des professionnels de l'information et à leurs évolutions en lien avec la technique utilisée en bibliothèques. Nous traitons la question des apports du web sémantique au sein des bibliothèques numériques et nous discutons du rôle des professionnels de l'information dans la mise en œuvre et l'évolution du web sémantique.

Dans la deuxième partie, nous présentons la méthodologie de recherche mise en œuvre pour répondre à la problématique de cette étude et nous discutons ensuite des résultats que nous avons obtenus. Dans le chapitre 4, nous présentons les méthodes de recherche que nous avons mobilisées dans notre enquête de terrain. Nous décrivons le déroulement de chaque méthode : la sélection des bibliothèques numériques, le questionnaire et les entretiens. Nous présentons, ensuite, les résultats de notre enquête (chapitre 5). Dans un premier temps, nous présentons notre sélection constituée de 98 bibliothèques numériques. Ensuite, nous discutons les résultats du questionnaire. Trente trois bibliothèques ont répondu à notre questionnaire portant sur l'usage et les applications du web sémantique en bibliothèques numériques françaises. Enfin, nous présentons les résultats des entretiens. Nous avons pu réaliser dix entretiens avec des professionnels de l'information. L'analyse de leur discours nous a permis d'appréhender leurs représentations, leurs expériences et leurs attentes concernant l'usage du web sémantique en bibliothèques numériques. Enfin dans le chapitre 6, nous discutons de nos résultats par rapport aux travaux que nous avons cité dans l'état de l'art. Ceux-ci nous permettent de montrer les spécificités de l'usage du web sémantique en bibliothèques numériques en France.

Nous terminons ce mémoire de thèse par une conclusion générale portant sur les apports et l'intérêt de ce travail de recherche. Ainsi, nous abordons les limites de l'étude effectuée et les perspectives de recherche.

Partie I: État de l'art

Introduction

La première partie de cette thèse a pour objectif de présenter l'état de l'art de notre sujet de recherche. Un positionnement de notre part a été nécessaire pour apporter un éclairage sur certains éléments de recherche.

Cette thèse couvre deux champs de recherche : le web sémantique et les bibliothèques numériques. Les travaux traitant de ces deux champs sont issus de différentes disciplines scientifiques qui sont principalement l'informatique, les sciences de l'information et de la communication (SIC) et l'intelligence artificielle. Afin de se positionner en SIC, il était donc nécessaire de limiter notre recherche aux travaux qui nous apportaient une meilleure compréhension de l'usage du web sémantique en bibliothèques numériques, et à ceux qui traitaient la question de la médiation en bibliothèques numériques.

Cette partie s'articule en trois chapitres suivant un cheminement logique. D'abord, dans le chapitre 1 nous introduisons le web sémantique comme étant une nouvelle approche de la représentation des connaissances. Une étude empirique nous a parus nécessaire pour comprendre l'utilisation et l'évolution des termes «web sémantique», «web de données», «web 3.0», «linked data» et «linked open data». Cette approche nous a conduits à nous interroger sur les notions de donnée, d'information, de connaissance et de document ainsi que sur leur représentation dans le cadre du web sémantique.

Ensuite, dans le chapitre 2 nous présentons les bibliothèques numériques. Une analyse des différentes définitions proposées par la communauté de recherche et la communauté de pratique en SIC a été faite afin d'apporter des clarifications concernant la notion de bibliothèque numérique. Pour déterminer le cadre conceptuel de cette notion, une approche info-communicationnelle a été mobilisée permettant d'envisager les bibliothèques numériques comme des dispositifs info-communicationnels. Une telle approche nécessite d'étudier ces dispositifs en prenant en considération la notion de médiation.

Enfin, dans le chapitre 3 nous mettons en relation les deux champs de recherche étudiés dans les chapitres 1 et 2 sous un angle de recherche précis, celui des usages et des applications du web sémantique en bibliothèque numérique. Nous accordons un intérêt particulier aux pratiques des professionnels de l'information et leur évolution en lien avec la technique utilisée pour représenter l'information. Nous traitons la question des apports du web sémantique au sein des bibliothèques

numériques, en insistant aussi sur le fait que les professionnels de l'information peuvent à leur tour jouer un rôle dans la mise en œuvre et l'évolution du web sémantique.

Chapitre 1

Le web sémantique : un objet de recherche en Sciences de l'information et de la communication (SIC)

Ce premier chapitre a pour objectif de fournir des éléments théoriques et empiriques permettant de clarifier et de contextualiser notre objet de recherche : le web sémantique. Web sémantique, web de données ou encore web 3.0, sont les termes utilisés pour désigner cette « évolution » du web. D'après le fondateur du web Tim Berners Lee, le web est en état d'évolution permanente [Berners Lee et al., 2001]. Une évolution est définie comme une « *suite de transformation dans un même sens »*[2], une sorte de développement ou encore une progression. Cette évolution donna lieu au web 1.0, dit web documentaire [Bachimont et al., 2011], puis au web 2.0 désignant le web collaboratif. Considéré comme une troisième génération du web, le web sémantique est le web de contenu [Salaun, 2012]. Le web sémantique, présenté comme une extension du web actuel, constitue une nouvelle approche pour la représentation du contenu numérique [Chaudiron, 2007].

Le principe du web sémantique, tel qu'il est défini par Tim Berners Lee, est d'identifier chaque donnée, de l'enrichir avec des métadonnées et de la relier avec d'autres données sur le web. De ce fait, le web sémantique en proposant une forme d'interopérabilité[3] basée sur les standards du web et sur des liens entre les ressources, permet à plusieurs systèmes de communiquer entre eux, échanger et partager l'accès à l'information.

Dans un premier temps, afin de clarifier la notion du web sémantique, nous commençons par un retour historique permettant de suivre l'évolution du cette notion. Une étude empirique a été faite pour mieux comprendre l'utilisation des différents termes (web sémantique, web de données, web 3.0 etc.) dans des bases de données scientifiques et sur les moteurs de recherche Google et Google scholar. Dans la deuxième section, nous exposons les fondements théoriques permettant d'inscrire le web sémantique dans le domaine de la représentation des connaissances. Par la suite, pour comprendre l'architecture du web sémantique, nous exposons les différents standards permettant son fonctionnement. Cela nous a ramenés à nous interroger sur les notions « donnée », « information » et «connaissance » et leur représentation au sein du web sémantique. Enfin, la notion de document est examinée afin de comprendre ses caractéristiques dans le contexte du web sémantique.

2 Rey-Debove, J. (Ed.). (2004). *Le Robert Brio: analyse des mots et régularités du lexique.* Paris: Robert

3 L'interopérabilité est la capacité que possède un produit ou un système dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs, sans restriction d'accès ou de mise en œuvre. Définition proposée par le groupe de travail intéropérabilité de l'AFUL 'Association Francophone des Utilisateurs de Logiciels Libres'. [En ligne] Disponible sur http://definition-interoperabilite.info/

1.1 Le web sémantique : un terme à clarifier

Parler du web sémantique c'est avant tout parler du web. C'est pourquoi, il nous paraît intéressant de présenter un bref historique de ce dernier afin de comprendre le contexte dans le lequel il est né, et suivre son développement. C'est au début des années 90 au CERN[4] que le chercheur Tim Bernes Lee, et afin de pouvoir partager des documents entre les chercheurs en utilisant les ordinateurs, a eu l'idée du web. L'objectif était de penser aux problèmes liés à l'accès à l'information au CERN. Il s'agissait d'introduire l'idée des systèmes d'information liés permettant de retrouver d'une manière efficace les documents [Berners Lee, 1989]. Pour répondre à ce besoin, le web s'appuie sur trois principes : la mise en réseau avec le protocole de communication client-serveur HTTP[5] permettant de transiter un message entre deux machines ; la définition d'un système d'identifiants URI[6] permettant de localiser une ressource d'une manière uniforme sur les différentes machines ; la structuration du contenu des documents en utilisant le langage de balisage HTML[7] et en y incluant des hyperliens. Le web est une application du réseau Internet parmi d'autres, c'est un système hypertexte[8] pour partager les documents que nous pouvons retrouver grâce à leurs adresses et auxquels nous pouvons accéder par des liens.

À ce jour, les principaux outils permettant d'exploiter les contenus du web sont les moteurs de recherche. Offrant une recherche rapide et efficace, ces moteurs prennent peu en compte la notion de qualité des documents [Berners Lee, 1996]. Face à ce problème une proposition a été faite par Tim Berners Lee reposant sur l'idée d'ajouter une sémantique aux documents du web, qui devrait permettre des outils beaucoup plus puissants [Ibid].

1.1.1 Historique du terme

Le terme « web sémantique » est apparu au grand public pour la première fois dans un article de Tim Bernes Lee, James Hendler, et Ora Lassila publié en 2001 [Berner-Lee et al., 2001]. Cependant c'est lors de la première conférence du consortium du web W3C à Genève en 1994 que Tim Berners Lee évoqua pour la première fois la nécessité d'une sémantique du web. (voir fig. 1.1 ci-dessous).

4 Organisation européenne pour la recherche nucléaire
5 Hypertext Transfer Protocol
6 Universal Ressource Identifier
7 Hypertext Markaup Language
8 Un système hypertexte se présente comme un dispositif informatisé permettant l'interconnexion de documents de divers type, ceci non sur la base d'un modèle hiérarchique ou relationnel, mais par des mécanismes associatifs sous le contrôle de l'utilisateur [Shneiderman, Kearsly, 1989, Nielsen, 1990, Berk, Devlin, 1991]

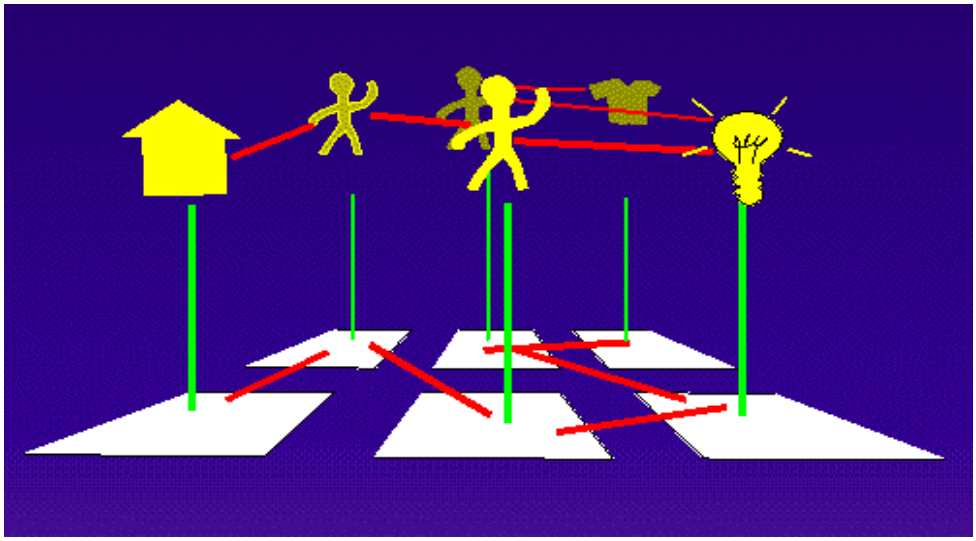


Figure 1.1: Le web sémantique vu par Tim Bernes Lee en 1994, [Source : W3C]

En 1998, Tim Bernes Lee publie un brouillon d'une feuille de route pour le web sémantique. Il indique que le web doit être conçu comme un espace d'informations qui permet non seulement la communication entre les humains (*human-human communication*) mais aussi la communication entre les machines et les humains. Un web où la machine serait en mesure de participer et d'aider les humains [Berners-Lee, 1998]. Dans cette feuille de route Tim Berners Lee met en place « un plan de travail » des différentes technologies à déployer pour la mise au point du web sémantique.

Dans un article publié en 2001, Tim Berners Lee précise que le web sémantique «*apportera une structure au contenu significatif des pages web* » en permettant aux agents[9] de logiciels d'effectuer facilement des tâches compliquées pour les usagers. Le web sémantique est donc présenté comme une extension du web actuel où l'information bien définie permettrait aux hommes et aux machines de travailler en coopération [Berners Lee et al., 2001].

1.1.2 Clarification du terme

L'expression « web sémantique » est composée de deux mots : web et sémantique. Le web est défini comme étant un système hypertexte public permettant de consulter des pages accessibles sur le réseau Internet. Le mot web est très important pour comprendre les enjeux du web sémantique. Selon le collectif Roger T. Pédauque l'expression web sémantique *« fait d'abord référence à la vision du web de demain comme un vaste espace d'échange de ressources entre êtres humains et*

9 « Dénomination générale de logiciels exécutant certaines tâches déterminées, en relation avec d'autres agents d'un système informatique ». [Source] Dictionnaire : Informatique, Internet et nouvelles technologies de l'information et de la communication, [Gualino, 2005]

machines [...] » [Pédauque, 2007, p.100]. Le web sémantique respecte les standards du web concernant la transmission des données sur le réseau. Cela veut dire que le web sémantique, comme le web, est caractérisé par son aspect « universel ».

Le terme « sémantique » quant à lui, utilisé comme adjectif, signifie « *qui concerne le sens* »[10]. Selon le dictionnaire culturel en langue française le mot sémantique utilisé comme adjectif signifie « *relatif au sens, à la signification, notamment dans le langage* »[11]. Et d'après la définition du centre national de ressources textuelles et lexicales CNTRL, l'adjectif sémantique signifie « *qui a rapport à la signification d'un mot ou d'une structure linguistique* ». Ces définitions nous permettent de dire que le mot employé dans l'expression « web sémantique » ne se réfère pas à la sémantique en tant que branche d'une science comme la linguistique ou la logique. De ce fait, le web sémantique peut être vu comme le web du sens, ce qui correspond bien à ce que Tim Bernes Lee a annoncé en 1994 au sujet d'une sémantique qui permettrait aux machines d'ajouter du sens au contenu du web [Berners-Lee, 1994].

Pour expliciter le sens des données au sein du web sémantique Bruno Bachimont et al. considèrent que *« le sens n'est pas autre chose que la paraphrase logique de ce qui a du sens, c'est-à-dire que le sens ne peut être capté que par une réécriture dans les termes de la logique ..»* [Bachimont et, 2011, p. 25]. Autrement dit, l'extraction du sens se fait à partir d'une représentation structurée des données, ceci grâce à l'utilisation des langages formalisés du web sémantique.

1.1.3 Evolution du terme

En 2001, le terme de web sémantique a été proposé par Tim Berners Lee pour désigner une évolution du web qui permettrait aux données disponibles (contenus, liens) d'être plus facilement utilisables et interprétables automatiquement, par des agents logiciels. L'article a décrit une évolution attendue du web existant vers un web sémantique [Berners-Lee et al., 2001].

L'utilisation du terme web sémantique a évolué. D'après Muriel Amar et Bruno Menon, *« l'appellation «web sémantique» a d'ailleurs été (en partie) désavouée par son initiateur, qui lui préfère le terme de «web de données». D'autres préfèrent parler de «données liées» ou « linked data » pour désigner le même ensemble de technologies (d'arts de faire) que certains ont aussi voulu appeler «web 3.0»* [Amar et Menon, 2011, p. 23].

De nouvelles expressions sont apparues : « linked data » traduit en français par l'expression « web de données », « web 3.0 » ou encore « web of data » et récemment « linked open data ». Ces différentes dénominations désignent-elles la même chose ? S'agit-il d'une évolution du terme ? Afin

10 REY-DEBOVE, J. (2004). *Le Robert Brio: analyse des mots et régularités du lexique* . Robert, Paris.
11 REY, A. (2005). *Dictionnaire culturel en langue française*. Le Robert, Paris.

de répondre à ces questions nous allons dans un premier temps interroger chaque terme à part.
Le terme « web 3.0 » est utilisé, d'une part, pour se référer à une évolution chronologique du web (web 1.0 ; web 2.0 et web 3.0), et d'autre part, il est utilisé comme synonyme de l'expression du web sémantique. Le Web 1.0 était le web des pionniers que l'on peut situer au début des années 90. Le web 1.0, selon Bruno Bachimont *« est le web dit documentaire, où les ressources sont publiées sans autre traitement que leur mise en forme et sans autre interactivité que l'activation des liens hypertextuels. »* [Bachimont et al, 2011, p. 24]. Le web 1.0 est donc caractérisé par une production de contenus qui se faisait essentiellement par les entreprises, des internautes passifs et des pages statiques.
En 1999, l'expression web 2.0 a été utilisé pour la première fois par Darcy DiNucci pour se référer à une nouvelle génération du web [DiNucci, 1999]. En 2004, le web 2.0 va être officiellement adopté lors d'une conférence «*brainstorming*» entre la société O'Reilly Média (du nom de son fondateur Tim O'Reilly) et la société MediaLive International. Au sein du web 2.0, nommé plus justement web social ou web collaboratif, les usagers entrent dans un mode actif. Au fur et à mesure de leur navigation, les utilisateurs peuvent ajouter du contenu à travers des liens hypertextes et d'autres tags, annotations ou commentaires. Le contenu est généré par les utilisateurs grâce à l'émergence des blogs, des wikis etc. L'utilisateur devient alors source d'informations et de création. Les données sont partagées, l'internaute fournit ses propres contenus à travers des réseaux sociaux (comme Facebook ou MySpace), des blogs, du contenu partagé.
L'expression «web 3.0 » a été utilisée pour désigner le web qui suit le web 2.0. La dénomination de web 3.0 va, quant à elle, être associée directement au web sémantique. Cependant, le web sémantique a été pensé en 1994 soit avant le web 2.0 (1999), mais il faut attendre quelques années pour que les outils et les standards du web sémantique soient développés. Nous pouvons dire que le web sémantique et le web 2.0 ont évolué parallèlement depuis la fin des années 90. Une étude publiée en 2006 par Wolfgang wahlster et Andreas Denge montre que le web 3.0 peut être vu comme une convergence entre le web 2.0 et le web sémantique [Wahlster, et al., 2006]. (voir fig. 1.2 ci-dessous).

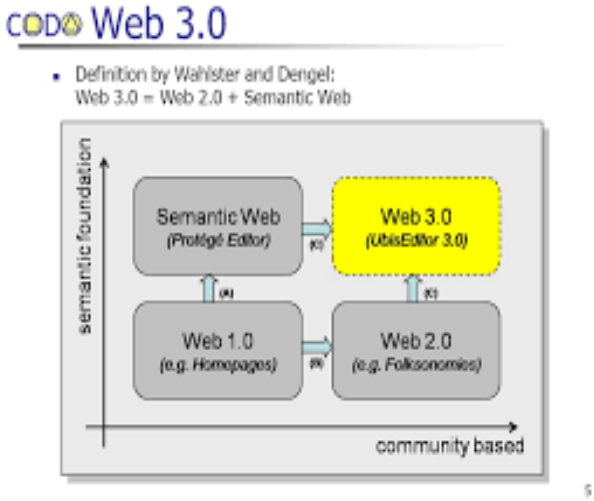


Figure 1.2 : Web 3.0, [Source : Deutshe Telekom laboratories

A cet égard, Manuel Zacklad parle du web socio-sémantique pour désigner la troisième vision du web. Selon Zacklad le web socio-sémantique est « *une approche qui combine les deux précédentes* [le web sémantique et le web 2.0] *en postulant l'existence d'une co-détermination des usages informationnels et communicationnels du web* » [Zacklad, 2005]. Le web 3.0 semble donc désigner une évolution du web permettant d'associer les outils et les applications du web sémantique à ceux du web 2.0.

L'expression « web of data » ou web de données a été utilisée pour la première fois par Tim Berners Lee pour désigner le web sémantique. Selon lui « *le web sémantique est un web de données, à un certains égards, comme une base de données globale* » [Berners Lee, 1998]. L'expression « web of data » est utilisée donc pour expliquer et clarifier l'expression «Semantic web ». Elle est considérée comme un synonyme de l'expression « Semantic web ».

En 2006, Tim Berners-Lee publie un article intitulé « linked data » traduit en français par l'expression web de données. Dans cet article Berners Lee indique qu'il ne s'agit pas seulement de mettre des données sur le web mais de les lier entre elles [Berners-Lee, 2006].

Tim Bernes Lee annonce en 2007 dans le magazine « la recherche » que :

> « *Le terme sémantique prête un peu à confusion car la sémantique s'intéresse au sens du langage pour en déduire des constructions logiques. Du coup, certains ont pensé qu'il s'agissait d'un Web qui permettrait par exemple d'effectuer des recherches sur*

Internet en posant des questions sous forme de phrases, en langage naturel. Or ce n'est pas son but. En fait, nous aurions dû l'appeler dès le départ "Web de données"[12].

A partir de 2007, l'usage de l'expression « linked data » ou « web de données » va prendre plus de place. Le terme « donnée » vient clarifier et préciser les objectifs visés par le « web sémantique ». Donc comme le précise Tim Berners Lee il ne s'agit pas d'extraire du sens à partir d'un traitement automatique des langues. Le web de données consiste à enrichir toute information avec des métadonnées, des annotations, etc. L'information ainsi structurée pourrait permettre aux données du web d'être connectées et reliées entre elles.

En 2010, un nouveau terme « linked open data » est utilisé par Tim bernes Lee pour désigner un web de données liées en licence ouverte, afin de favoriser la réutilisation et l'ouverture des données. Ce sujet sera traité dans le chapitre 6.

Dans la section suivante, nous présentons une étude empirique permettant au mieux la compréhension de l'évolution de ces termes. Pour cela, nous avons analysé l'utilisation de ces termes dans des bases de données scientifiques et sur les moteurs de recherche Google et Google scholar .

1.1.4 Web sémantique, web de données, web 3.0, quel usage dans les bases de données et les moteurs de recherche ?

Pour avoir une idée de l'utilisation de ces différentes expressions, nous avons analysé le corpus de quatre bases de données (ScienceDirect ; Sage journals online ; Web of science et Cairn) et des moteurs de recherche Google et Google scholar. Parmi ces bases de données, trois contiennent essentiellement des revues anglophones, tandis que la dernière est francophone. Le but était, d'une part, d'analyser la répartition d'usage des expressions (Semantic web, linked data, web 3.0, web of data, Linked open data, web de données et web sémantique) dans ces bases de données, et d'autre part, d'analyser leur répartition dans le temps et cela selon trois intervalles. Le premier se situe entre 1994 et 2000, le deuxième entre 2001 et 2005 et le troisième entre 2006 et 2016. La répartition en trois périodes s'appuie sur l'évolution historique du web sémantique vue par son fondateur Tim Berners Lee. La première période « 1994-2000 » est liée à la première feuille de route élaborée par Tim Berners Lee. La deuxième « 2001 -2005 » est liée au premier article publié par Tim Berners Lee sur le web sémantique. Et la dernière « 2006-2016 » est due à l'apparition de

12 « Le Web va changer de dimension », entretien avec Tim Berners-Lee par Marie-Laure Théodule. *La Recherche*, 2007, n° 413, p. 34-38

l'expression « Linked data » pour la première fois dans un article de Tim Berners Lee.

Les trois bases de données anglophones (web of science ; Sciencedirect et Sage journals) ont la particularité de contenir une large collection de revues scientifiques dans des domaines différents y compris en sciences de l'information et de la communication. La base de données Cairn qui a pour particularité de couvrir des revues scientifiques en français, contient elle aussi des articles dans le domaine de l'information et de la communication. En ce qui concerne le moteur de recherche Google, il est actuellement le moteur de recherche le plus utilisé par les usagers du web pour publier et diffuser de l'information. Selon Comscore Inc[13], pour le mois de février 2016, 16.8 trillion de requêtes ont été soumises vers les cinq principaux sites de recherche d'information en ligne : Google 64 %; Microsoft (Bing) 21.4 %; Yahoo 12,2 %; Ask network 1.6 % et AOL 0.9%.

En ce qui concerne le moteur de recherche Google scholar, qui permet d'effectuer des recherches sur des travaux universitaires, en plus d'être un outil de recherche convivial pour les étudiants [Cothran, 2011], une étude publié dans le magazine « Nature » a montré qu'il est l'outil le plus utilisé par les chercheurs pour faire des recherches sur des réseaux sociaux académiques [Van Noorden, 2014].

Par le biais des moteurs de recherche spécifiques à chaque base de données, nous avons effectué une recherche en utilisant les expressions exactes suivantes en texte intégral: semantic web, linked data, linked open data et web 3.0 dans les bases de données (web of science[14], ScienceDirect[15] et Sage journals[16]) ; nous avons ajouté à celles-ci les deux expressions web de données et web sémantique dans la base de données « Cairn »[17]. Cette recherche a été repartie sur les trois périodes définies précédemment. Le même protocole de recherche a été utilisé avec Google et Google scholar mais cette fois en utilisant les sept expressions utilisées dans les 4 bases, et sans répartition dans le temps, parce que ces moteurs de recherche ne permettent pas de faire une recherche par intervalle de temps. La recherche a été fait le 13 janvier 2017.

Les résultats sont affichées sous forme des diagrammes. (voir figures ci-dissous). Une description détaillée de ces résultats est donnée dans l'annexe I.

13 Étude menée par: « ComScore Releases February 2016 U.S. Descktop Search Engine Rankings », *Comscore Inc*, [enligne] <https://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings> (consulté le 16/05/2017).

14 Cette base signale la littérature scientifique mondiale dans tous les domaines. L'ensemble du contenu est constitué du dépouillement de plus de 10 000 périodiques et de plus de 110 000 actes de conférences. Les plus vieux articles remontent à 1900.

15 Permet l'accès aux 2500 revues de l'éditeur Elsevier. Collection pluridisciplinaire.

16 Permet l'accès au texte intégral de plus de 500 revues éditées par Sage Publications en sciences humaines et sociales.

17 Nous avons ajouté les deux expressions web sémantique et web de données parce que la base de données Cairn contient des articles en langue française.

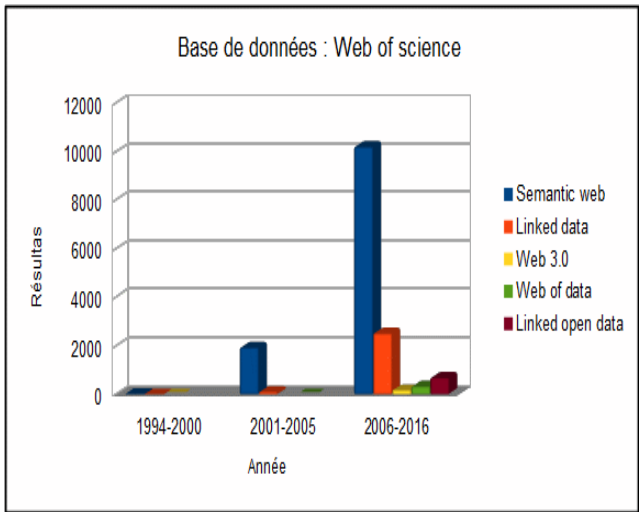


Figure 1.3 : Nombre de documents pour chaque terme recherché dans la base de données : Web of science

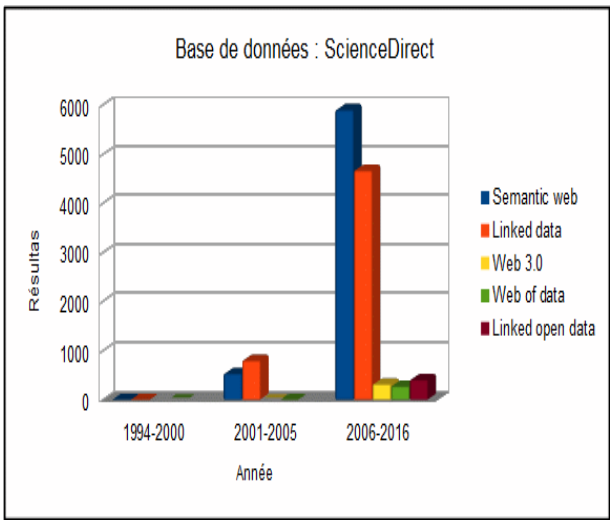


Figure 1.4 : Nombre de documents pour chaque terme recherché dans la base de données : ScienceDirect

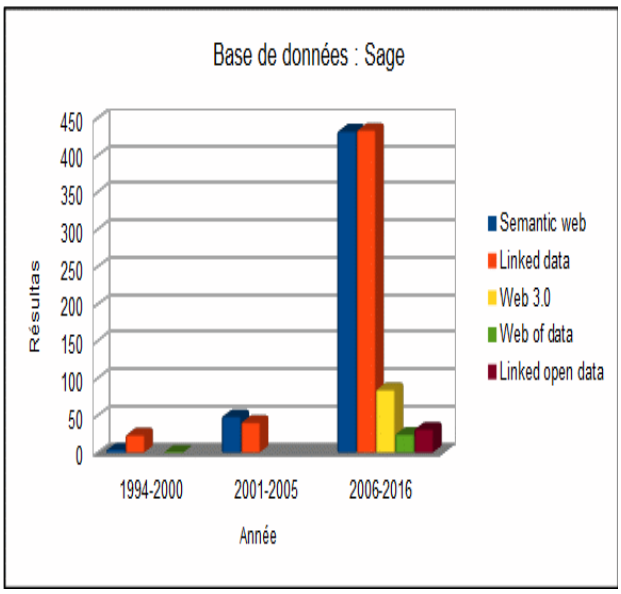


Figure 1.5: Nombre de documents pour chaque terme recherché dans la base de données : Sage journals

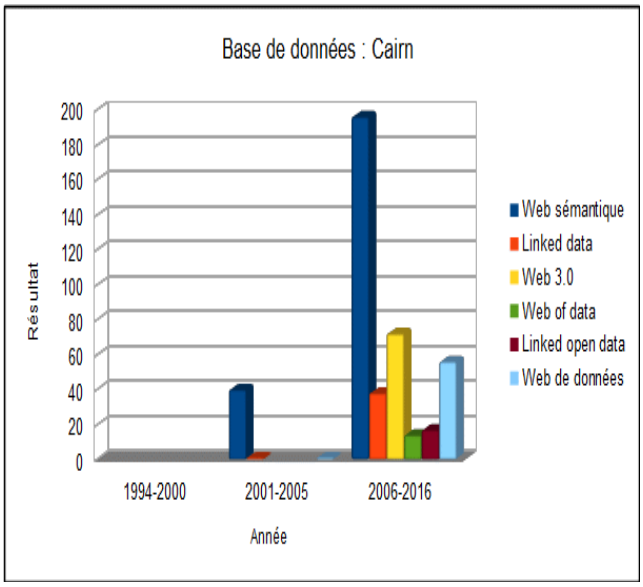


Figure 1.6 : Nombre de documents pour chaque terme recherché dans la base de données : Cairn

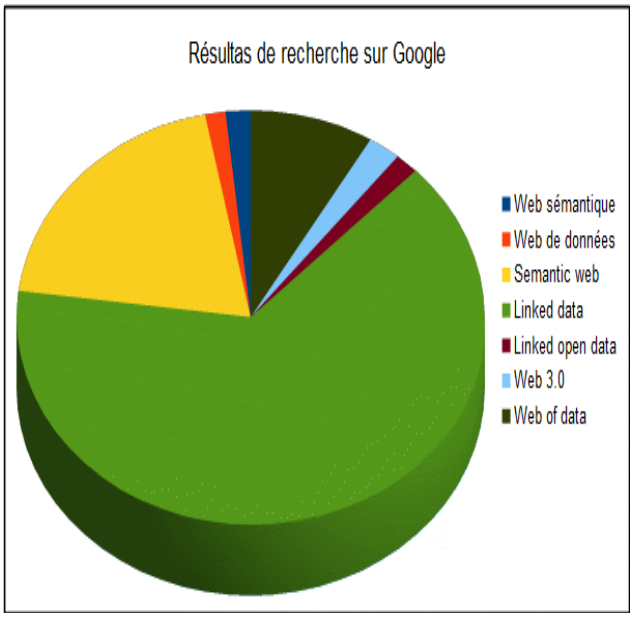


Figure 1.7: Résultats de recherche sur Google

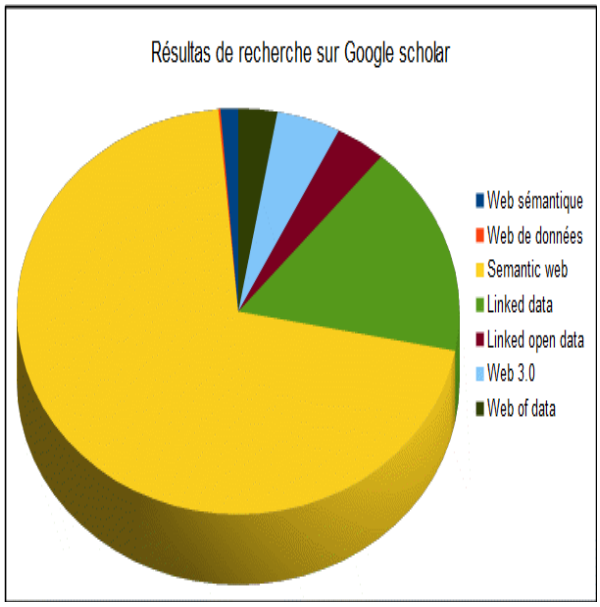


Figure 1.8: Résultats de recherche sur Google scholar

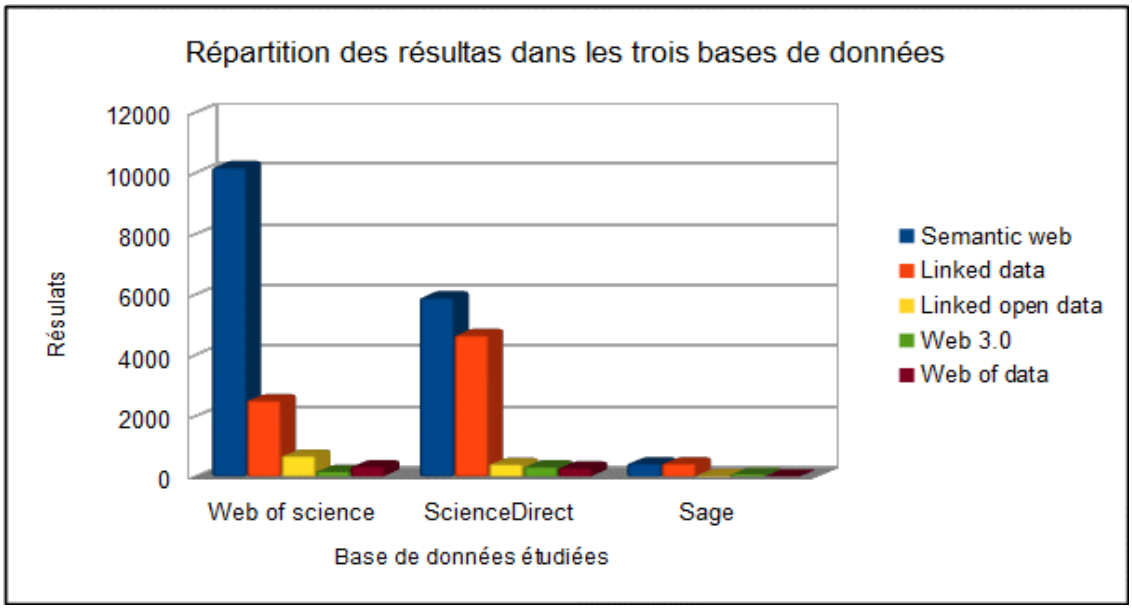


Figure 1.9 : Répartition des résultats dans les trois bases de données : web of science, ScienceDirect et Sage

Les résultats montrent que l'expression « semantic web » est la plus utilisée dans les trois bases de données anglophones et dans le moteur de recherche Google Scholar. De même, l'expression « web sémantique » est la plus employée dans la base de données Cairn. Cela signifie que l'expression « semantic web » est la plus utilisée dans les publications scientifiques et académiques (1994-

2016). D'autre part, l'augmentation de l'utilisation de ce terme pourrait être dû à son émergence comme une nouvelle technique pour la gestion et la diffusion de l'information sur le web, dans beaucoup de domaines : la recherche d'information , la représentation des connaissances, le développement des ontologies, l'intelligence artificielle, le data mining etc. [Ding, 2010].

Il est intéressant de remarquer que l'expression « linked data » apparaît en premier dans le moteur de recherche Google, et en deuxième dans les bases de données étudiées et dans Google Scholar. Plus précisément, dès l'année 2006 l'utilisation de l'expression « linked data » commence à s'accroître pour dépasser celle de l'expression « semantic web » dans la base de données « Sage journal » (voir figure 1.5). L'usage de l'expression « linked data » par les chercheurs résulte de son utilisation par Tim Berners Lee pour la première fois en 2006. Cette expression est utilisée pour clarifier les objectifs visés par cette nouvelle génération du web. Autrement dit, l'expression « linked data » est utilisée pour lever l'ambiguïté contenue dans l'expression «semantic web ». Finalement les résultats ont montré que l'expression « semantic web » est la plus utilisée dans les bases de données scientifiques et dans Google scholar, tandis que l'expression « linked data » est la plus utilisée dans le moteur de recherche Google.

Les expressions « Linked open data », « Web of data » et « Web 3.0 » apparaissent surtout dans la troisième période étudiée (2006-2016). L'expression « Linked open data », parue en troisième dans les bases de données (web of science, ScienceDirect et Sage), est utilisée par Tim Berners Lee en 2006. Il s'agit du web de données liées et ouvertes (Linked open data), notion que nous abordons dans le chapitre 6.

Une fois le web sémantique présenté, nous abordons dans la section suivante la question de la représentation des connaissances au sein du web sémantique. Nous commençons par un retour historique sur les modèles des représentations graphiques à l'origine de la structure du web sémantique.

1.2 Le web sémantique : une nouvelle approche pour la représentation des connaissances

La question de la représentation des connaissances en sciences de l'information et de la communication est abordée pour référer aux langages documentaires utilisés pour représenter l'information afin de faciliter son repérage et son accès. La représentation des connaissances constitue une branche de recherche dans une discipline en sciences de l'information et de la communication, celle de l'organisation des connaissances. Cette dernière porte sur la conception,

l'étude et la critique des méthodes d'organisation et de représentation des documents qu'une société juge qu'ils méritent d'être conservés [Tennis, 2008]. Tennis a étudié l'organisation des connaissances en s'intéressant à la relation entre le langage et l'action [Tennis, 2012]. Il s'interroge sur le langage que nous utilisons pour représenter les connaissances dans les systèmes d'informations et sur les questions éthiques et sociales liées à l'emploi de ce langage. Autrement dit, les usagers du système d'organisation des connaissances sont influencés par cette représentation. C'est pourquoi Selon Tennis, il faut faire attention à l'utilisation du langage car selon lui « *c'est une lourde responsabilité que d'être les gardiens de la mémoire du monde ».* [Tennis, 2012, p. 2]

Les premiers systèmes de classifications ont été construits par Melvil Dewey en 1876 pour faciliter le repérage des livres dans les bibliothèques. Ensuite au début du vingtième siècle, la classification décimale universelle a été établit par Paul Otlet et Henri Lafontaine pour classer les connaissances. Actuellement, des thésaurus et des ontologies se développent pour assurer l'organisation des connaissances. Issu d'une tradition bibliothéconomique d'organiser et de représenter les connaissances, le domaine de l'organisation des connaissances fut investi d'abord par les professionnels d'information, puis il s'est élargi pour accueillir des spécialistes d'autres disciplines (linguistes, informaticiens, terminologues et sociologues), cela pour faire face aux nouveaux supports par lesquelles les informations circulent [Hudon et El-Hadi, 2012]. A cet égard, Majdid Ihadjadène et Laurence Favier considèrent que la « représentation des connaissances » continue à être exploitée par les informaticiens, et que le « web sémantique » constitue un exemple de l'actualité de cette conception [Ihadjadène et Favier, 2008].

La représentation des connaissances est aussi une discipline de recherche qui peut être située entre l'intelligence artificielle et les sciences cognitives [Manes-Gallo et Paganelli, 2003]. Il s'agit d'imiter le comportement humain ou bien de le reproduire via un processus automatisable. Appelé l'hexagone cognitif [Bara, 1995] les sciences cognitives englobent six disciplines : psychologie , philosophie, linguistique, anthropologie, neurosciences et intelligence artificielle[18]. Selon Ronald J. Brachman et Hector Lavesque, la représentation des connaissances fait partie de l'intelligence artificielle qui se préoccupe de la façon dont un agent utilise ce qu'il sait pour décider quoi faire. La représentation des connaissances est donc le domaine de l'Intelligence Artificielle qui s'intéresse à la façon dont la connaissance peut être représentée symboliquement et manipulée de manière automatisée par des programmes de raisonnement. [Brachman et Levesque, 2004] .

18 Cacaly, Serge (dir.) 2001. Dictionnaire encyclopédique de l'information et de la documentation. Paris : Nathan. P510

La notion de « représentation des connaissances » peut être mieux comprise à partir des différents rôles qu'elle joue [Davis et al., 1993, p.17] :

- c'est une manière de raisonner sur le monde ;
- c'est un ensemble d'engagements ontologiques dans le sens où il faut penser comment on devrait représenter ce monde et par quels termes ;
- c'est une théorie de raisonnement intelligent exprimée en termes de trois composants : la conception, les inférences (les contraintes) et les inférences (les recommandes) ;
- c'est un moyen de calcul efficace ;
- c'est un milieu d'expression humaine.

1.2.1 Les modèles de représentation des connaissances

Plusieurs modèles ont été proposés par les différentes communautés de l'Intelligence Artificielle et de la science cognitive pour représenter des connaissances. Nous allons exposer brièvement les modèles de représentation des connaissances qui ont influencé le fonctionnement et la structure du web sémantique : les graphes conceptuels [Pradel, 2013]. Pour cela, nous allons d'abord introduire et discuter la notion de réseau sémantique, qui apparaît comme étant l'ancêtre des modèles de graphe conceptuel.

L'origine des réseaux sémantiques remonte aux graphes existentiels introduits par Charles Peirce en 1896. Par la suite, dans les années soixante, des chercheurs en sciences cognitives et linguistiques (Allan Collins, Ross Quillian, Elizabeth Lotfus) ont proposé un modèle pour stocker l'information sémantique dans la mémoire de l'ordinateur [Collins et Quillian, 1969, p.240]. Ce modèle sémantique de recherche d'information s'appuie sur un réseau ou un schéma de nœuds liés entre eux. Autrement dit, ils ont proposé un modèle pour l'organisation des connaissances sous forme de réseau de concepts. Chaque nœud est associé à des propriétés et à des relations avec d'autres nœuds. Ce modèle consiste à classer les concepts de façon hiérarchique dans une arborescence. (voir fig. 1.10 ci-dessous).

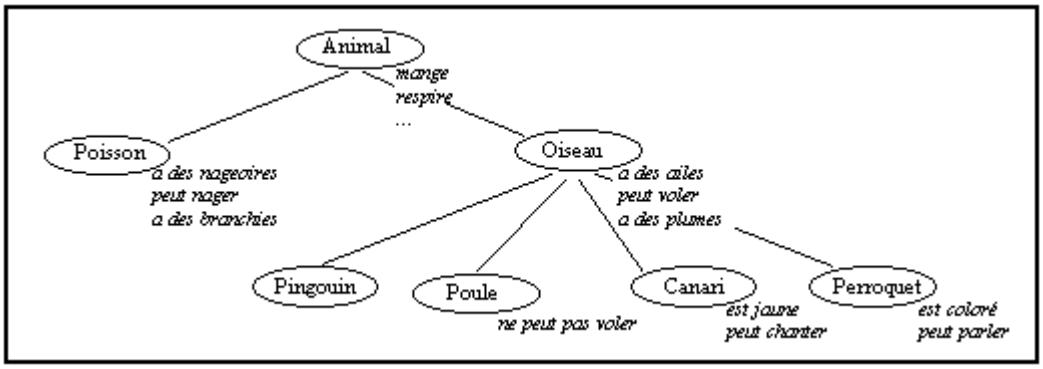


Figure 1.10: Exemple schématique d'un réseau de la mémoire représentant une hiérarchie à trois niveaux (d'après Collins et Quillian, 1969) [Nevers, 2000]

Plus tard, Elizabeth Lotfus rejoint Collins, ils ont modifié le modèle de Quillans en ajoutant que les relations entre les nœuds (les concepts) ne sont pas seulement de type hiérarchique mais elles peuvent être d'une nature sémantique [Collins et Lotfus, 1975]. (voir fig. 1.11 ci-dessous)

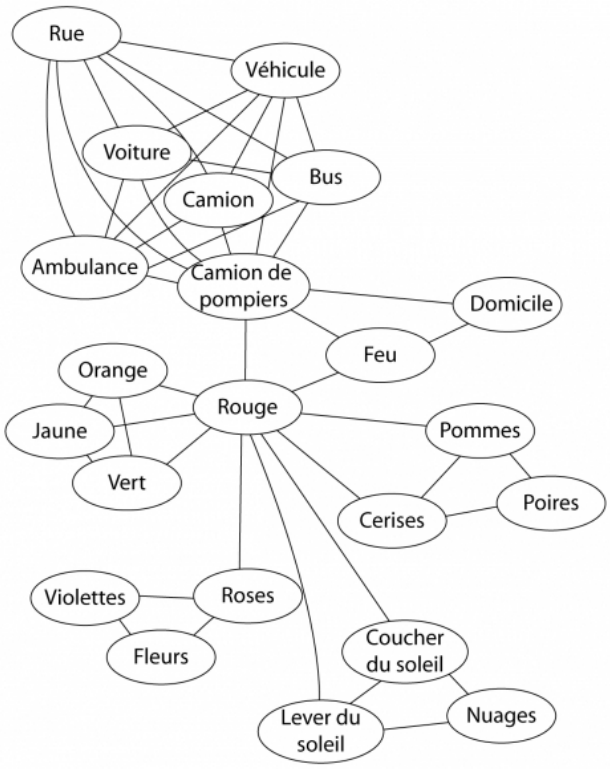


Figure 1.11: Exemple schématique d'un réseau sémantique proposé par Collins et Lotfus (1975) [Nevers, 2000]

Un autre modèle de représentation des connaissances a été proposé par John Sowa (1984). Ce modèle appelé « graphes conceptuels » s'appuie sur les réseaux sémantiques de l'intelligence artificielle et les graphes existentiels de Pierce [Sowa, 1984]. Il s'agit de représenter les connaissances sous forme graphique. Plus précisément, un graphe conceptuel est un graphe biparti étiqueté. Il comprend deux sortes de nœuds « les concepts » et les « relations ». Les nœuds sont liés par des arcs orientés. Un arc sert à lier un concept avec une relation. Autrement dit, la relation entre deux concepts est construite d'une manière explicite.

Ensuite, le concept de réseau a été étudié par François Rastier (linguiste spécialisé en sémantique). Selon Rastier un *« réseau est un graphe fini, orienté, étiqueté, généralement connexe et cyclique. Ses sommets appelés nœuds (nodes), représentent des concepts ; et ses arcs, appelés liens (links), des relations binaires entre ces concepts. »* [Rastier, 1991, p.121]. Un réseau sémantique est un graphe dont les nœuds représentent les concepts et les arcs les relations entre les concepts.

Les graphes conceptuels et les réseaux sémantiques semblent être proche du modèle de représentation du web sémantique par le biais des liens entre les concepts. D'une part, le web sémantique consiste à représenter les données, et d'autre part, il consiste à lier ces données entre elles. A cet égard, Tim Berners Lee indique que les graphes conceptuels sont facilement intégrés au web sémantique [Berners Lee et al., 2001].

En sciences de l'information et de la communication, plusieurs auteurs se sont intéressés à analyser le web sémantique en tant que nouvelle approche pour la représentation des connaissances. Selon Bruno Bachimont il y a deux objectifs ou contraintes pour représenter les connaissances *: « d'une part, être capable d'expliciter le sens d'un contenu, d'autre part, rendre exploitable cette explication par un traitement informatique. »* [Bachimont et al., 2011, p. 25]. Pour expliciter le sens des données au sein du web sémantique Bruno Bahimont considère *que « [...] le sens ne peut être capté que par une réécriture dans les termes de la logique [...] »* [Ibid, p.25]. Il rajoute que cette opération se fait grâce à *« des présentations de langages formalisés comme RDF, OWL, puisque ce sont ces langages qui auront la lourde tâche d'exprimer le sens et d'être exploités par les outils informatiques ».*[Bahimontet al., 2011, p.25].

Pour extraire du sens, il s'agit tout d'abord de représenter le langage naturel par des termes (des annotations, des métadonnées etc). Dans le cas du web sémantique, il faut que cette représentation soit formelle et standardisée. Ensuite, l'enjeu principal réside dans la capacité des outils informatiques d'exploiter ces métadonnées et ces annotations.

Selon Tim Berners Lee, pour que le web sémantique fonctionne, les ordinateurs doivent avoir accès

à des collections d'information structurées et à un ensemble de règles d'inférences utilisées pour conduire à un raisonnement automatisé [Berners-Lee et al., 2001, p.5]. Le défi du web sémantique est donc de fournir un langage qui exprime à la fois les données et les règles de raisonnement sur les données. Il s'agit des langages et des standards développés pour assurer le bon fonctionnement du web sémantique. Ces langages seront étudiés dans la section qui suit.

Dans la section suivante, nous présentons les standards permettant de définir l'architecture du web sémantique. L'objectif est de comprendre le fonctionnement du web sémantique dont l'intérêt est de décrire les données et de les lier entre elles afin d'offrir une représentation graphique des connaissances.

1.2.2 L'architecture du web sémantique

Parler de l'architecture du Web c'est se référer aux standards qui définissent l'infrastructure technologique dont il émerge. Il y a trois standards fondamentaux au cœur de l'architecture du web. Le premier est l'URI (Universal Resource Identifier). C'est un format d'identifiants uniques permettant de nommer les ressources sur le Web. Le deuxième est le protocole HTTP assurant une communication client-serveur. Le troisième, le HTML est le langage de balisage utilisé pour représenter, stocker, mettre en forme et publier les pages web. Étant compatible avec l'architecture du web, le web sémantique utilise les mêmes standards pour transférer les données entre les différents ordinateurs et applications au sein du réseau Internet.

D'après Jean Michel Salaun, pour comprendre ce qui se passe sur le web nous pouvons nous baser sur « les trois dimensions du document numérique » [Salaun, 2012]. La première dimension concerne le repérage du document, « le document doit être vu » ; la deuxième est celle de la médiation, « le document a une fonction social » ; enfin la troisième met en avant le contenu, « la signification du texte ». Le web 1.0 est le web du document, « Vu », tandis que le web 2.0 est celui des échanges « Su » et enfin le web 3.0 correspond au web de données « Lu ». Ainsi ce qui compte dans la troisième génération du web c'est le contenu, plus exactement c'est le sens des données. De ce fait Stéphane Chaudiron considère que la présentation de l'information est la solution pour créer du sens au sein du désordre informationnel et aider l'usager à s'orienter [Chaudiron, 2007]. Selon Chaudiron, il y a trois approches de la représentation du contenu : l'indexation contrôlée (langage documentaire contrôlé) ; l'indexation libre (web 2.0) ; l'indexation structurée (web sémantique). Le web sémantique consiste à identifier les données (en utilisant des URI), ensuite à les décrire d'une manière structurée (en utilisant le langage XML) et enfin à les lier en utilisant le triplet (RDF).

Nous allons maintenant présenter les principales couches du web sémantique tel qu'elles étaient décrites par Tim Bernes Lee. La figure 1.12 (Layer cake) regroupe les différents standards et technologies utilisés dans le contexte du web sémantique.

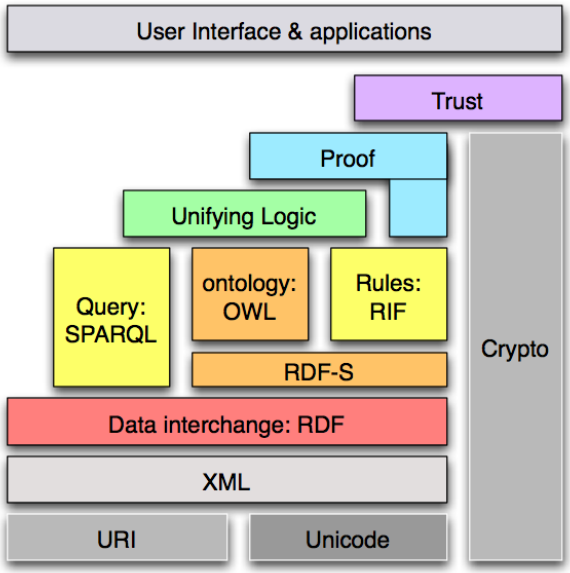


Figure 1.12: Layer cake - [Source :w3c]

<u>Première couche : Identification des ressources</u> :

Dans le web sémantique, la première couche consiste à identifier les ressources par des identifiants uniques, les URIs. Un objet d'information est "sur le Web" s'il possède un URI. Autrement dit, nous ne pouvons pas trouver un objet sur le web s'il n'a pas d'URI. Par le mot universel, Tim Berners Lee affirme que « *le web est sensé pouvoir contenir en principe chaque bit d'information accessible par les réseaux* » [Berners Lee, 1996]. Les URIs sont donc des identifiants universels dans le sens où il est possible d'identifier chaque objet sur le web que ce soit des objets réels ou des concepts abstraits. Autrement dit, les URIs peuvent se référer à des éléments tangibles (des personnes, des lieux etc.) ou à des éléments plus abstraits (services, types de relations, concepts etc).

<u>Deuxième couche : description du contenu en XML</u>

Pour décrire le contenu des ressources d'une manière structurée au sein du web sémantique il faut utiliser le langage de balisage XML. XML signifie en anglais 'Extensible Markup Language ' et en français 'langage de balisage extensible'. C'est une recommandation du W3C datant de 1998

inspirée de SGML[19]. Un document XML est représenté physiquement sous la forme d'un fichier texte structuré en éléments, à l'aide de balises éventuellement imbriquées.

Décrire des contenus en XML, permet d'une part, de structurer un document en vue de son traitement informatique, et d'autre part, de rendre l'information facilement réutilisable et diffusable sur le web. Cela facilite les échanges de données entre plusieurs systèmes d'informations. Autrement dit, XML favorise l'interopérabilité et les échanges entre des systèmes d'informations hétérogènes. C'est pourquoi XML constitue un pilier essentiel dans l'architecture du web sémantique. Le web sémantique vise à décrire les ressources d'une manière à les rendre exploitables par l'homme et par la machine, et à faciliter la diffusion, l'échange et la réutilisation des données sur le web.

Troisième couche : lier les données entre elles en utilisant RDF

La troisième couche dans le web sémantique consiste à lier les données entre elles et cela en utilisant RDF (Resource Description Framework). RDF est un langage de représentation de l'information au sein du web. C'est une recommandation du w3c publiée dans sa première version en 1999 et finalisée en 2004.

Selon le consortium du web w3c, RDF est un modèle de données simple (simple data model) qui permet de faire des liens entre des ressources identifiées par des URIs et décrites en XML. Selon Fabien Gandon, RDF :

> *« est la première brique des standards du web sémantique [..] elle fournit une structure de données standard et un modèle pour encoder des données et des métadonnées sur n'importe quel sujet sur le web ; ces sujets sont appelés des ressources et identifiés par des URI[…]. La plus petite structure de description en RDF est le triplet. Un triplet RDF décrit et relie une ressource en associant celle-ci à une propriété et à la valeur de cette propriété (Sujet, Prédicat, Objet*) » [Gandon et al., 2012, p.28]

RDF s'appuie sur le constat simple qu'une métadonnée est un couple (propriété, valeur), l'ensemble de ces couples constituant une description de la ressource à laquelle ils s'appliquent. La description RDF d'une ressource est donc un ensemble de triplets (sujet, prédicat, objet) où le sujet est la ressource à décrire, le prédicat une propriété applicable à ce sujet et l'objet une valeur de cette propriété. La puissance de RDF tient au fait que non seulement le sujet, mais aussi le prédicat lui-même, sont obligatoirement des ressources identifiées par des URI. L'objet, valeur de la propriété,

19 Standard Generalized Markup Language

peut être une ressource ou une donnée. RDF est donc un modèle ouvert, conçu pour permettre la réutilisation et l'intégration de vocabulaires définis de façon totalement indépendante.

Selon Fabien Gandon, RDF est *«un modèle sans précédent dans le domaine de la représentation des connaissances.* »[Gandon et al., 2012]. Le point fort de RDF réside dans sa capacité d'expliciter la nature de relation entre les deux ressources. Autrement dit, il s'agit non seulement de lier les données mais surtout d'interpréter le lien qui existe entre elles. (voir fig. 1.13 ci-dessous).

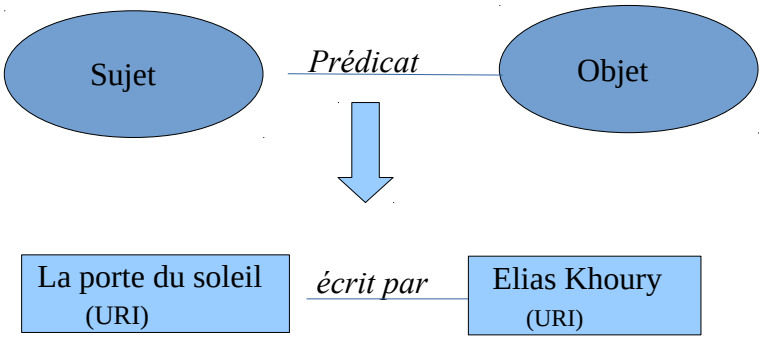


Figure 1.13 : Exemple d'un triplet RDF

<u>Quatrième couche : Représentation graphique des données RDFS et OWL</u>

RDFS (Resource Desciption Framework Schema) est une extension de l'RDF de base. RDFS est le langage de description des vocabulaires associé à RDF, il permet de spécifier des ontologies dites « légères » [Gandon et al., 2012]. Cela consiste à nommer des classes et des propriétés et à les organiser d'une manière hiérarchique [Ibid]. Les classes permettent de définir la nature des ressources et les propriétés correspondent aux prédicats dans les triplets et permettent d'exprimer les relations entre les ressources [Bermès et al., 2013]. Autrement dit, RDFS est un ensemble de triplets RDF dont les données sont liées entre elles pour former un schéma. (voir fig. 1.14 ci-dessous)

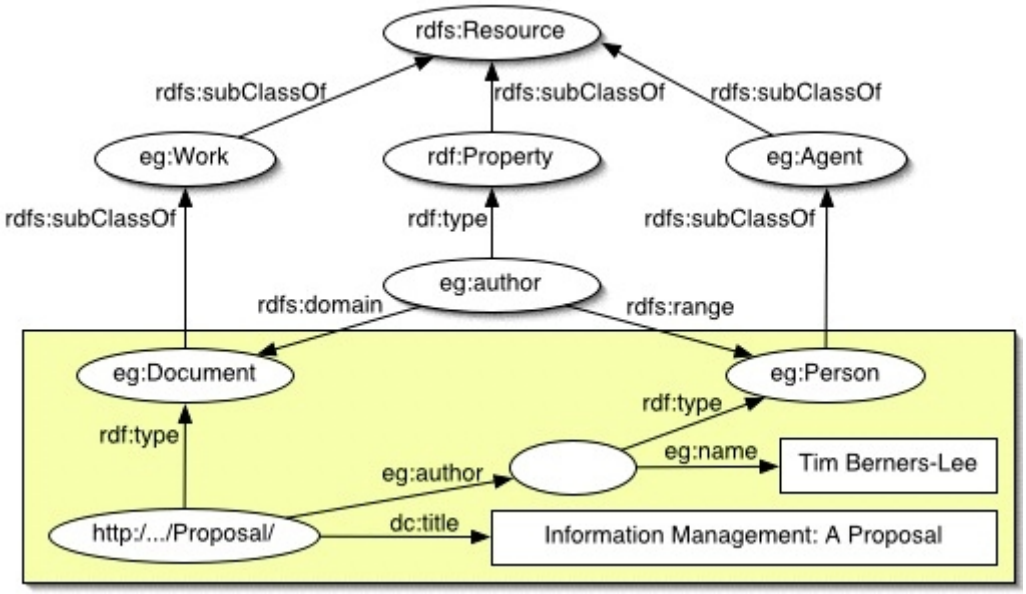


Figure 1.14: Un exemple schématique de RDFS [Source W3C]

RDFS permet de construire un langage primitif de description d'ontologies [Brunet et Xuan, 2010], tandis qu'une définition formelle et plus exhaustive des ressources relève des ontologies recommandées sous l'appellation OWL.

Avant de parler du langage OWL défini comme un langage d'ontologie web, nous abordons la notion d' « ontologie » pour comprendre son utilisation dans le contexte du web et dans le domaine des sciences de l'information et de la communication. En sciences de l'information et de la communication, une ontologie est définie comme une «*représentation des connaissances terminologiques relatives à un domaine, agréée par une communauté de personnes et sensée en faciliter le partage. Une ontologie a une taxonomie et un ensemble de règles d'inférence (raffinement, décomposition, prédiction, relativité, similarité).*»[20]

Les communautés de l'intelligence artificielle et la communauté de représentation des connaissances ont adopté le terme d'ontologie dans leur domaine pour désigner l'objet issu d'un processus de modélisation de connaissances permettant de représenter d'une façon formelle un domaine donné. La définition la plus consensuelle d'une ontologie est celle introduite par Gruber [Gruber, 1993] et étendue par Borst [Borst, 1997] : une ontologie est une spécification formelle explicite d'une conceptualisation partagée d'un domaine donné. Cette définition s'appuie sur deux dimensions : la conceptualisation qui permet de décrire d'une manière explicite les vocabulaires d'un domaine donné, et la spécification qui consiste à utiliser un langage formel pour cette description. Le but est de fournir « *une compréhension spécifique des définitions de termes et du vocabulaire utilisé dans un domaine* ». [Abascal et al., 2005, p.192].

20 Cacaly, Serge et al. (2004). Dictionnaire de l'information. Paris : Armans Colin

Pour Fabien Gandon, une ontologie «*inclut une organisation hiérarchique des concepts pertinents et des relations qui existent entre ces concepts*. » [Gandon, 2012 p. 84]. Il s'agit donc de représenter un domaine de connaissance par un vocabulaire spécifique. Ce vocabulaire est formé d'un ensemble de concepts et des relations qui les unissent. De ce fait, la composition d'une ontologie dépendrait du choix des concepts. Selon Bruno Bachimont, il y a deux difficultés pour élaborer une ontologie. La première concerne la définition des concepts d'un domaine tandis que la seconde concerne l'exploitation informatique de ces concepts [Bachimont, 2007]. Autrement dit, il s'agit tout d'abord de représenter un domaine de connaissances par des termes appelés « concepts ». Le terme concept reste en lui même un terme vague. Cela nous amène à nous interroger sur la notion du « concept » surtout son utilisation en domaine de sciences de l'information et de communication. Un concept est définie comme « *une idée générale désignée par un mot* »[21]. Le concept est le résultat d'une double opération intellectuelle : l'abstraction et la généralisation[22]. C'est un travail philosophique qui permet de transformer une notion vague en un concept précis et universel[23]. Cette définition met l'accent sur deux caractéristiques essentielles du concept, la précision et l'universalité. D'une part, un concept ne peut pas contenir en lui même d'ambiguïté, il doit être clair et précis. Et d'autre part, il doit être exprimé par un terme général dans le sens où son usage est commun entre les hommes.

Une étude approfondie réalisée par Birger Hjorland sur une « théorie de concept » considère que les systèmes d'organisation des connaissances (par exemple, systèmes de classification, thésaurus et ontologies) doivent être compris comme des systèmes fondamentalement de l'organisation de concepts et leurs relations sémantiques [Hjorland, 2009]. Par ailleurs, nous pensons qu'une organisation conceptuelle d'un domaine de connaissances est un travail assez complexe et rigoureux. Il nécessite une compréhension globale d'un domaine donné, une maîtrise des langages, une connaissance des usagers et une logique d'organisation. Pour cette raison, nous considérons que l'organisation des connaissances dans son sens large est une discipline qui englobe différentes expertises dans le domaine des sciences de l'information, de l'informatique, de la linguistique et des sciences sociales.

Après avoir présenté la notion d'ontologie, nous passons désormais à la définition du langage OWL. Le langage d'ontologie Web OWL est une recommandation du w3c permettant de définir des ontologies web. Basé sur la syntaxe de RDF/XML, OWL sert à décrire un modèle de données représentatif de l'ensemble des concepts dans un domaine donné. La particularité de ce modèle tient au fait qu'il est exploitable par les machines et les humains. Le langage OWL par opposition à RDF

21 Godin, Christian. (2004) Dictionnaire de philosophie. Paris : Fayard
22 Ibid
23 Ibid

Schema permet de définir un cadre formel et détaillé d'un domaine de connaissances en utilisant des contraintes appliquées aux classes et propriétés [Bermès et al., 2013]. D'une part, les ontologies devraient jouer un rôle majeur pour accéder à l'information. D'autre part, elles permettent de garantir l'interopérabilité et les échanges des données entre différentes applications.

Selon le consortium du web w3c, les ontologies Web possèdent les avantages suivants[24] :

- *Capacité d'être distribuées au travers de nombreux systèmes*
- *Mises à l'échelle pour les besoins du Web*
- *Compatibles avec les standards Web pour l'accessibilité et l'internationalisation*
- *Ouvertes et extensibles*

Le langage de requête : SPARQL

SPARQL[25] Protocol and RDF Query Language, est un protocole et un langage de requête pour les données RDF. C'est une recommandation conçue par le W3C datant de janvier 2008 et une des technologie du web sémantique. La puissance de ce langage d'interrogation réside dans sa capacité à interroger plusieurs bases de données. SPARQL prend également en charge l'agrégation, les sous-requêtes, la négation, la création de valeurs par des expressions, le test de valeur extensible et la contrainte des requêtes RDF. Selon Tim Berners-Lee, SPARQL facilite l'interrogation des informations issues des bases de données et d'autres sources répandues sur le Web. Selon W3C, SPARQL est conçu pour être utilisé à l'échelle du Web et donc faciliter les interrogations des sources de données distribuées, quel que soit leur format. Un point fort de ce langage d'interrogation est sa capacité à faire une recherche en se basant sur les liens décrits par les triplets RDF. Une recherche peut amener l'utilisateur à découvrir des informations auxquelles il n'avait pas pensé, cela grâce aux règles d'inférences.

Après avoir présenté l'architecture du web sémantique, nous abordons dans la section suivante la question de l'articulation de la donnée, l'information et la connaissance et leur place dans cette architecture.

1.2.3 Donnée, information, connaissance, quelle représentation ?

Les notions donnée, information et connaissance ont fait objet de recherches dans beaucoup de domaines : sciences de l'information et de la communication, mathématiques, physique, économie etc. En sciences de l'information et de la communication, elles constituent à la fois la base sur

24 https://www.w3.org/2004/01/sws-pressrelease.html.fr
25 https://www.w3.org/TR/sparql11-query/

laquelle s'appuie les recherches théoriques de cette science et la matière première des dispositifs par lesquels elles circulent.

Le but n'est pas d'analyser ces notions mais de voir comment elles s'articulent dans le contexte du web sémantique. Autrement dit, il s'agit de redessiner leur modèle de présentation qui est généralement pensé de manière linéaire ou sous fourme de pyramide comme suit:

Donnée → Information → Connaissance.

Le terme clé dans ce triptyque est l'information. Il s'agit d'un concept difficile à cerner puisqu'il englobe selon Francis Beau « *tout un ensemble de concepts apparentés qui lui sont subordonnés* » [Beau, 2017, p.119]. Considéré comme un concept interdisciplinaire [Capurro et Hjorland, 2003], il a été défini comme un concept technique en référence aux travaux de Shannon [Balicco, 2013], puis comme un concept social considérant l'information comme un processus de communication [Meyriat, 85, cité par Balicco], intelligible par l'esprit humain dans les échanges ou les interactions [Ibekwe-Sanjuan, 2012].

Pour Raphael Capurro et Birger Hjorland, la distinction la plus importante dans les différents concepts d'information se situe entre l'information en tant qu'objet (par exemple nombre de bits) et l'information comme concept subjectif, c'est-à-dire comme signe. Le sens est cependant déterminé dans les contextes sociaux et culturels [Capurro et Hjorland, 2003]. La donnée peut exister sous des formes variées, utilisables ou non, et elle n'a pas de sens en elle-même [Bellinger et al., 2004]. Autrement dit, la donnée devient une information une fois qu'elle est mise dans un contexte permettant sa communication et son interprétation.

Concernant le terme de connaissance et son lien avec l'information, Laurence Balicco indique que la connaissance existe quand « *l'information est utilisée d'une manière appropriée par son récepteur* » [Balicco, 2013, p.111]. De cette façon, elle vient modifier le savoir de la personne qui la reçoit [Meyriat, 1985]. Autrement dit, la connaissance est une information reçue qui modifie les savoirs des personnes qui l'utilisent.

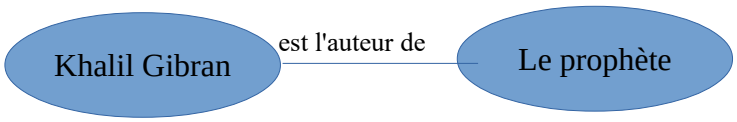
Dans le contexte du web sémantique, la présentation de ces trois notions peut se faire de manière cyclique et non pas seulement linéaire. D'une part, le web sémantique permet de décrire chaque donnée indépendamment de son contexte, et d'autre part, il favorise la réutilisation de ces données. Une donnée devient donc non seulement une donnée brute et indépendante mais aussi une donnée significative, et cela grâce aux métadonnées et aux langages utilisés pour décrire cette donnée. Une donnée identifiée et décrite a pour caractéristique d'être liée à d'autres données et en même temps d'être exploitables par les machines.

En résumé, une donnée dans le contexte du web sémantique possède les caractéristiques suivantes :

elle est identifiable, structurée, réutilisable, liée, exploitable par l'homme et la machine.

Concernant l'information, celle-ci peut être vue comme un ensemble de données liées, intelligibles et qui prend un sens. L'extraction du sens est dû au lien explicite existant entre les deux données liées. Autrement dit, une fois que la donnée est liée à une autre par un lien explicite, elle devient une information parce qu'elle est mise dans un contexte permettant sa compréhension par l'homme et par la machine.

Par exemple, un triplet RDF pourrait représenter une information :



Enfin, la connaissance peut, quant à elle, être vue comme l'ensemble des triplets RDF. Autrement dit, la connaissance serait représentée par une ontologie permettant de définir les concepts et les relations liés à un domaine donné. La connaissance est donc une représentation formelle de données et leur relations d'un domaine de connaissances donné. Elle est à la fois exploitable par l'homme et la machine.

Le schéma de la figure 1.15 illustre notre représentation de ces trois notions dans le contexte du web sémantique.

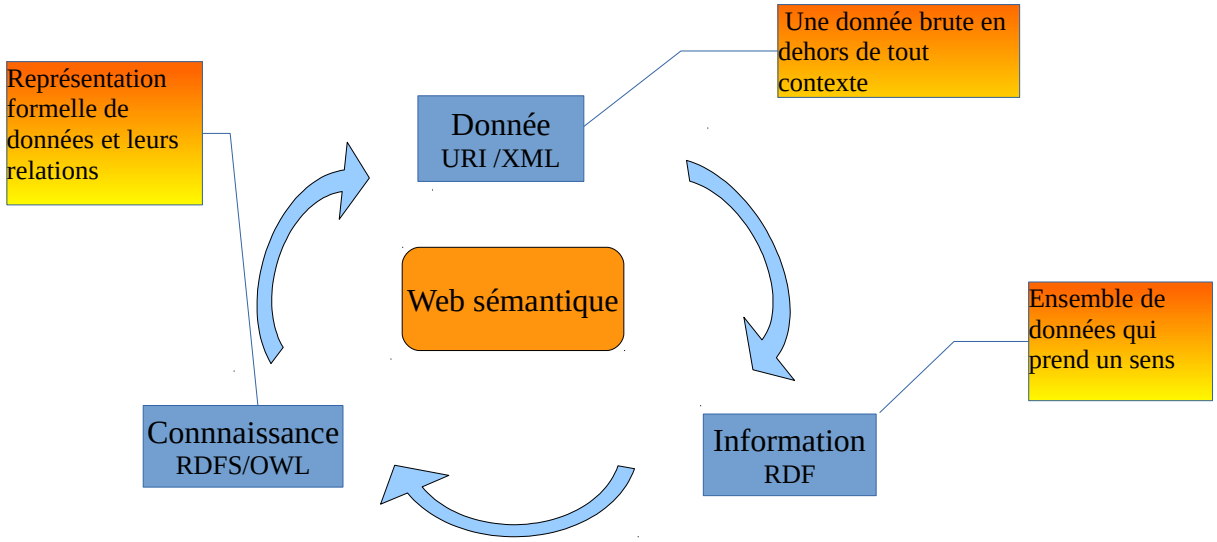


Figure 1.15 : Représentation cyclique des notions : donnée, information et connaissance dans le contexte du web sémantique

En résumé, cette représentation peut être vu comme une représentation cyclique du modèle de couche (Layer cake) de Tim Berners Lee. La première couche est formée par les données identifiées et décrites (URI / XML). La deuxième couche est constituée de données liées sous la forme de triplets RDF. La troisième couche, quant à elle, est l'ensemble de ces triplets (RDFS/OWL). Le fait que les données soient identifiées et décrites d'une manière indépendante de leur contexte, les rend facilement réutilisables et connectables avec d'autres données qui se trouvent sur le web. La réutilisation consiste à utiliser les données dans d'autres contextes, à les exploiter et à créer de nouvelles données ainsi que de nouveaux savoirs [Carmes et Noyer, 2013]. D'où l'importance d'une représentation cyclique signifiant que les données se transforment en informations, puis que ces dernières se transforment à leur tour en connaissances et qu'enfin celles-ci peuvent se retransformer en données grâce à la réutilisation et ainsi de suite.

Cela nous ramène à nous interroger sur la notion du document. Un sujet que nous traitons dans la section suivante afin de voir comment cette notion pourrait être envisagé dans un contexte du web sémantique.

1.2.4 La notion de document revisitée dans le contexte du web sémantique

Avec le web sémantique l'attention se porte sur la donnée elle-même. A cet égard, Jean-Michel Salaun indique que *«dans le web de données, l'unité de base n'est plus le document comme précédemment, mais la donnée. On ne relie plus des documents entre eux, mais des données entre elles. »* [Salaun, 2012, p.4]. Donc avec le web sémantique, les données (le contenu) se libèrent du document (le contenant). Avec le web sémantique la notion de document elle même doit être réinterrogée. Cela nous amène à la question suivante. Quelle place tient le document au sein du web sémantique ?

La notion de document porte en elle même une forte ambiguïté et fut l'objet de recherche de grands auteurs en sciences de l'information et de communication (Meyriat, Briet, Estivals etc). Pour Jean Meyriat (1981), le document est *« un objet qui supporte de l'information, qui sert à la communiquer, et qui est durable »* [Meyriat cité par Couzinet, 2001, p.143]. Cette définition nous permet de distinguer deux élément essentiels : l'espace du document et la communication. L'espace contient l'information (le contenu) et le support (le contenant). Nous constatons l'importance donnée aussi à l'usage du document et sa communication. C'est l'utilisateur, le récepteur du message, qui fait le document. De ce point de vue, Robert Estivals et Jean Meyriat (1981) indiquent que *«dès lors qu'il est employé et décodé par un récepteur, le document devient réel : il a trouvé son utilisateur »* [Meyriat cité par Couzinet, 2001, p. 490].

Susanne Briet va encore plus loin et considère que tout objet est un document ou peut le devenir, si et seulement si, il transmet une information. Elle définit le document comme *«tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène physique ou intellectuel »* [Briet, 1951]. Un document est donc un objet qui supporte de l'information, lui permettant d'être transmise à un récepteur ou à un utilisateur.

La notion de « document » ne cesse d'être interrogée surtout avec l'émergence du numérique. Un document numérique est un document qui a pour caractéristiques d'être sur un support électronique ainsi que d'être perceptible via la technologie numérique. Le document numérique a permis de séparer l'aspect présentation (les techniques de mise en page par exemple) de l'aspect information (structuration du texte, données). De ce fait, les données se sont libérées de leur support permettant ainsi leur utilisation en dehors de leur contexte et de leur mise en forme.

Un support physique délimite les dimensions d'un document, comme par exemple le nombre de pages d'un livre. Mais quelles sont les limites d'un document numérique circulant sur le web ? Étant donné que l'information présente sur le web est une information structurée de manière hypertextuelle, il devient possible de naviguer sur le web indéfiniment jusqu'à décider de s'arrêter. Par ailleurs, sur le web, l'utilisateur ne devient-il pas lui même créateur du document qu'il consulte ? Cela revient, comme le signale Viviane Couzinet [Couzinet, 2001], à parler de « document par attribution » selon le terme employé par Jean Meyriat pour décrire le fait que la réception du document fait le document. Comment alors envisager la nature d'un document qui circule sur la toile ? Selon Couzinet, un document présent sur le Web est un objet qui se caractérise davantage en terme de flux qu'en terme de fixité [Couzinet, 2001]. Autrement dit, le document sur le web n'est plus un objet fixe mais un objet qui se propage sur la toile.

Comme nous l'avons mentionné auparavant, le numérique a permis de séparer le contenant (le support) du contenu (l'information) au sein d'un document. Ainsi, avec le web, le document est devenu un flux d'information. Jean-Michel Salaun indique que sur le web un document devient « *un nœud ou une tête de réseau qui permet de passer d'un document à un autre* » [Salaun, 2012, p. 71]. Autrement dit, le web a facilité en quelque sorte la consultation immédiate d'un document et les documents en lien avec celui-ci ; navigation permise et facilitée par les liens hypertextes. Rapidement le web va devenir la plus grande plate-forme d'accès à l'information. Cependant le problème de l'hypertexte réside aussi dans cette ouverture vers d'autres informations. L'utilisateur perd facilement son chemin de départ à force de naviguer d'un lien à un autre sans parfois avoir un fil conducteur. Le document circulant sur le web se caractérise de plus en plus par sa nature de flux d'informations, obligeant l'utilisateur lui même à construire son document. Une mission qui n'est pas toujours facile à réaliser, vu la masse immense d'informations qui circulent sur le web.

Dans le contexte du web sémantique, il s'agit de relier les données entre elles et non pas les documents [Salaun, 2012]. Chaque donnée pourrait être liée à d'autres données sur le web et cela grâce à l'utilisation du langage RDF. Ainsi, en partant de l'hypothèse que l'utilisateur construit lui même le document, le web sémantique lui permet de passer moins de temps à naviguer parce qu'il lui sera proposé l'information demandée ainsi que l'information liée à celle-ci. De ce fait, le document peut être vu comme l'ensemble des données liées. Il n'est construit que lorsqu'un utilisateur le demande [Salaun, 2012].

En résumé, dans le contexte du web sémantique, le document prend la forme d'un ensemble de données liées qui permet à l'utilisateur une navigation facilitée et rapide sur le web.

Conclusion du chapitre

Ce chapitre a été consacré à la présentation du web sémantique en tant que nouvelle approche pour la représentation des connaissances. En s'appuyant sur les publications de Tim Berners Lee, nous avons pu suivre l'évolution de cette expression. Plus précisément, l'expression web sémantique a été employée par Tim Berners Lee publiquement en 2001, puis, en 2006, le chercheur a utilisé une autre expression « linked data », traduite en français par « web de données », en précisant que ce changement a pour but d'éviter la confusion que porte le terme « sémantique ». Malgré la tentative de Tim Berners Lee pour rediriger les chercheurs vers l'utilisation de l'expression « linked data », les résultats de notre recherche ont montré que l'expression « web sémantique » reste la plus employée par les chercheurs.

Dans notre travail, nous avons essayé de définir un cadre conceptuel de notre objet de recherche : le web sémantique. Pour se faire, nous avons mobilisé des approches théoriques liées à la représentation des connaissances, plus précisément les réseaux sémantiques et les réseaux conceptuels (Collins, Quillian, Lotfus, Sowa, Rastier). Ces réseaux sont basés sur la notion simple de graphe [Sabah, 1990], formé de nœuds (les concepts) et de relations entre ces concepts. De même, le web sémantique est basé sur un modèle de graphe où les données sont liées entre elles.

Le web sémantique peut être considéré comme une nouvelle approche de la représentation des connaissances. Il s'agit d'identifier et de relier les données dans le but de former un graphe conceptuel et formel appelé ontologie. Ce graphe est composé de données identifiées grâce à des URI et de liens explicites entre ces données. Ce qui compte dans le web sémantique n'est pas la donnée elle même mais les métadonnées utilisées pour décrire les données et ensuite les rendre exploitables par les machines.

Nous pouvons comparer ce travail d'enrichissement au travail documentaire réalisé par les documentalistes depuis des années (l'identification, le catalogage et l'indexation). Dans le contexte du web sémantique, afin d'enrichir les données il est nécessaire de les décrire en utilisant un langage normalisé compréhensible par l'homme et la machine. Autrement dit, ce travail d'enrichissement ne se fait pas seulement par la machine toute seule mais à l'aide d'un travail intellectuel humain. Ce travail consiste donc à exprimer le sens derrière chaque donnée avec des métadonnées exploitables par la machine. Ce sujet sera traité d'une manière plus détaillée dans le chapitre 3.

Pour finir, nous avons interrogé la notion de document . Le but étant de s'interroger sur la nature du document dans le contexte du web sémantique. Dans un premier temps, nous avons abordé les trois notions de base : donnée, information et connaissance. Un schéma cyclique a été réalisé pour montrer leur représentation au sein du web sémantique. Dans un deuxième temps, nous avons évoqué la notion de document vu par Jean Meyriat et Susanne Briet. Une caractérisation a été élaborée pour spécifier la nature du document dans le contexte du web sémantique. Le document peut être pensé comme un ensemble de données liées reconstruite d'une manière automatisée à la demande d'un utilisateur.

Dans le second chapitre, nous abordons la notion de bibliothèque numérique. Il s'agit, d'une part, de clarifier cette notion, et d'autre part, de s'interroger sur la bibliothèque numérique comme un dispositif info-communicationnel.

Chapitre 2

Les bibliothèques numériques: des dispositifs info-communicationnels ?

Ce chapitre a pour objectif d'étudier la notion de bibliothèque numérique. Le but étant à la fois de clarifier cette notion et de proposer un cadre conceptuel en sciences de l'information et de la communication permettant de l'encadrer. Selon Joachim Schopfel et Jacques Creusot, le concept de la bibliothèque numérique est apparu au milieu des années 1970 comme étant un concept fondamentalement informatique [Creusot et Schopfel, 2005]. En s'appuyant sur une analyse des revues scientifiques, ces auteurs ont constaté que 75 % des articles consacrés à la bibliothèque numérique ont été publiés dans des revues informatiques, contre 15 % en sciences sociales et seulement 5 % dans le domaine de la bibliothéconomie. D'une part, la domination technologique est dû à une négligence du facteur humain réduisant son rôle à un simple opérateur technique. D'autre part, la prise de conscience par les professionnels de l'information sur l'utilisation de cette technologie a pris 5 à 10 ans de retard [Ibid].

Cependant, l'idée d'une bibliothèque numérique n'est pas nouvelle. En effet, dans les années trente, Paul Otlet (1934), traditionnellement considéré comme le «père de la documentation », eut l'intuition d'une « bibliothèque numérique » lorsqu'il parla d'un «*télescope électrique, permettant de lire de chez soi des livres exposés dans la salle teleg des grandes bibliothèques, aux pages demandées d'avance*» [Otlet, 1934, p.238]. En 1945, Vannevar Bush a écrit un article intitulé « as we make think » dans lequel montre comment la machine Memex, pourrait être utilisée pour stocker, organiser et récupérer des informations. Dans les années 1960, l'ambition de T. Nelson était la création d'une bibliothèque virtuelle, capable d'accueillir une infinité de documents écrits, visuels, sonores produits par l'humanité, interconnectés entre eux et consultables par tous, dans laquelle on se déplacerait librement [Compiègne, 2011, p.21]. Mais c'est seulement au début des années 90, que les chercheurs ont commencé à utiliser le terme de « bibliothèque numérique » et à lui conférer des préoccupations techniques et des missions sociales et culturelles [Le Crosnier, 2007].

Ce chapitre s'articule en quatre sections. Dans la première, nous analysons les différentes définitions proposées tant par la communauté de pratique que par la communauté de recherche en SIC [Borgman, 1999]. Nous mobilisons ensuite dans la deuxième section le concept de dispositif

info-communicationnel afin de déterminer un cadre conceptuel de la bibliothèque numérique en SIC. Nous discutons ensuite dans les troisième et quatrième sections des travaux traitant la notion de la médiation au sein des bibliothèques numériques. Il s'agit, d'une part, de mettre l'accent sur le rôle des langages documentaires et des métadonnées dans l'organisation et la représentation des connaissances en bibliothèques numériques (médiation documentaire), et d'autre part, de souligner le rôle des bibliothèques numériques dans la diffusion des savoirs (médiation des savoirs).

2.1 La bibliothèque numérique, une notion à définir

De nombreuses définitions ont été proposées pour définir les bibliothèques numériques (Fox, 1993 ; Levy, Marshall, 1995 ; Bishop, Star, 1996 ; Waters, 1998 ; Arms, 2000 ; Lesk, 2004 etc) . Selon Christine Borgman l'analyse de ces différentes définitions permet de distinguer deux groupes qui s'intéressent aux bibliothèques numériques : la communauté de recherche (research community), et la communauté de pratique (practice community) [Borgman, 1999]. La communauté de pratique travaille sur les défis et les problèmes auxquels sont confrontés les professionnels dans leurs pratiques professionnelles en bibliothèques. La communauté de recherche quant à elle a pour intérêt les aspects théoriques et conceptuels des bibliothèques numériques. [Borgman, 1999].

Autrement dit, la façon de décrire une bibliothèque numérique diffère selon la communauté de la recherche qu'elle l'étudie. A ce propos, Gary Cleveland indique qu'une bibliothèque numérique :

- est une grande base de données d'un point de vue de la recherche d'information ;
- est une application particulière de méthodes hypertextes pour ceux qui s'intéressent à la technologie de l'hypertexte ;
- est une application du web pour ceux qui s'intéressent à la diffusion de l'information en réseau informatique ;
- est une nouvelle étape dans la poursuite de l'automatisation des bibliothèques pour la science des bibliothèques ;
- est un ensemble d'algorithmes informatiques pour les informaticiens et les développeurs de logiciels [Cleveland , 1998].

Andrew M. Cox indique que le concept de la communauté de pratique prend sa place dans les « théories de pratique » en science de l'information. C'est un concept essentiel dans le domaine de la gestion des connaissances et de l'éducation avec (Wenger, 1991, 1999, 2002) [Cox, 2012]. La communauté de pratique est définie comme étant un groupe de personnes qui partagent des intérêts communs et qui approfondissent leurs connaissances et leurs compétences en interagissant de façon

continue [Wenger et al., 2002]. La communauté de pratique en bibliothèques numériques serait donc les groupes de professionnels de l'information qui mobilisent leurs connaissances et leurs compétences afin de résoudre des problèmes liés à leurs pratiques professionnelles en bibliothèques.

Aborder aussi les définitions de la communauté de recherche est indispensable pour définir le cadre conceptuel et théorique du travail de recherche mené dans cette thèse. Par ailleurs, nous nous appuyons sur des études et des recherches anglophones et francophones en SIC pour clarifier et comprendre la notion de bibliothèque numérique tout en analysant ses aspects théoriques.

Après avoir expliqué ce que nous entendons par communauté de recherche et communauté de pratique, nous allons maintenant analyser les différentes définitions de la bibliothèque numérique proposées par ces communautés.

2.1.1 Bibliothèque numérique, électronique et virtuelle, quelle terminologie ?

Les termes « bibliothèque électronique », « bibliothèque numérique » et « bibliothèque virtuelle » ont été utilisés d'une façon interchangeable pour se référer à l'utilisation de la technologie numérique au sein des bibliothèques : pour acquérir, stocker, conserver et diffuser leurs contenus à des utilisateurs distants. Selon Christine Borgman, ces différents termes ont été utilisés dans la littérature en SIC sans être clairement définis [Borgman, 1999].

Certains auteurs ont cependant essayé d'éclaircir la différence entre les trois notions. C'est le cas de Roy Tennant qui considère que la bibliothèque électronique est un terme plus large parce qu'il englobe tout le matériel nécessaire à la conception d'une bibliothèque numérique. Selon Tennant la bibliothèque électronique est constituée des documents électroniques et des services. Les premiers incluent des documents numériques et des formats analogiques [Tennant, 1999].

Ceci nous amène à nous interroger sur la différence entre les qualificatifs « numérique » et « électronique ». A ce sujet, Dominique Cotte indique qu'il serait plus exacte de parler d'un objet « électronique » car il s'agit d'une technologie qui repose sur le passage ou non du courant électronique. Cependant, l'auteur précise que le vocable « numérique », formé à partir d'une base technologique, « *est porteur de représentations idéologiques puisqu'il voisine avec virtuel, immatériel, intuitif, transparent* ». [Cotte, 2004, p.33]. Selon Marcello Vitali-Rosati, le terme « numérique » est initialement utilisé pour caractériser le mode d'enregistrement des données en opposition à l'analogique. A partir des années 1990, et avec la naissance d'Internet et du web, le terme va ensuite prendre une valeur sociale et culturelle [Vitali-Rosati, 2014]. En SIC, de nombreux travaux de recherche ont été menés sur l'analyse des enjeux sociaux liés au numérique, notamment

la relation entre la technique et les usages des dispositifs de l'information et de la communication.

Le terme électronique est un terme plus large qui englobe plusieurs catégories : le numérique (signal discret) ; l'analogique (signal contenu) ; la radio fréquence etc. Le numérique consiste à stocker l'information grâce à une technique d'échantillonnage permettant de transformer un signal continu et analogique en un signal discret et digital formé des bits. Nous parlons donc de numérique pour nous référer à un objet électronique dont l'information est de nature numérique.

Le terme « bibliothèque virtuelle » est simplement définit par un « *ensemble organisé de liens vers des articles sur le réseau* »[26], son objectif est d'économiser le temps et l'effort des utilisateurs qui cherchent des informations sur le Web. Ces bibliothèques ne possèdent pas de bâtiments, elles existent dans le cyberespace et permettent l'accès à des documents numériques [Chisenga, 2003].

Aux États-unis, les bibliothèques numériques « digital libraries » désignent un « *domaine d'application au défi national* » [Borgman, 1999]. Des conférences et des projets ont été dédiés aux recherches portant sur le thème de bibliothèques numériques. De ce fait, l'utilisation du terme «digital libraries» peut être attribuée au projet de « l'initiative des bibliothèques numériques » « Digital Libraries Initiative »[27] en 1994 [Harter, 1996]. Ensuite, ce terme va être largement adopté par les chercheurs et les organisations et des différentes définitions vont être proposées pour le décrire [Ibid].

Aujourd'hui, l'emploi des termes de bibliothèque numérique, de bibliothèque électronique ou de bibliothèque virtuelle sont confondus dans l'usage courant. Le terme numérique évoque davantage le contenu, tandis que, le terme électronique met l'accent sur la technique de support. Quant au terme de bibliothèque virtuelle, il désigne l'accès aux documents numériques via le réseau Internet ou via des espaces de stockage virtuels comme une disque dur par exemple.

Dans cette thèse, nous adoptons le terme « bibliothèque numérique » ou bien « digital library » en anglais, parce que d'une part, nous pensons que c'est le terme le plus adapté pour désigner les contenus numériques d'une bibliothèque ; et d'autre part, c'est désormais le terme le plus utilisé dans la littérature en SIC pour désigner les bibliothèques qui diffusent leur contenu sur le web.

Pour examiner l'utilisation de ces différentes expressions, nous avons analysé le corpus de quatre bases de données (ScienceDirect ; Sage journals online ; Scholarvox et Cairn). Trois bases de données contiennent essentiellement des revues anglophones et une propose des revues

26 Définition proposée par Charles Stuart University citée dans l'article « Digital libraries and virtual libraries : definitions, concepts and goals. (2003)

27 Le programme "Digital Libraries Initiative" (phase I) est lancé en 1994. Il est subventionné par la National Science Foundation, la Darpa, la Nasa. Cette dynamique relance de nombreux travaux, développements et recherches dans le domaine de l'information numérique.

francophones (Cairn). Le but est d'analyser la répartition d'utilisation des expressions (Digital library, electronic library, virtual library) dans ces bases de données.

Nous avons effectué une recherche dans les moteurs de recherche spécifiques à chaque base de données en utilisant les expressions exactes[28] suivantes en texte intégral: digital library ; electonic library ; virtual library. Les résultats montrent, pour chaque base de données interrogée, le nombre de documents correspondant aux termes étudiés (voir fig. 2.1 ci-dessous).

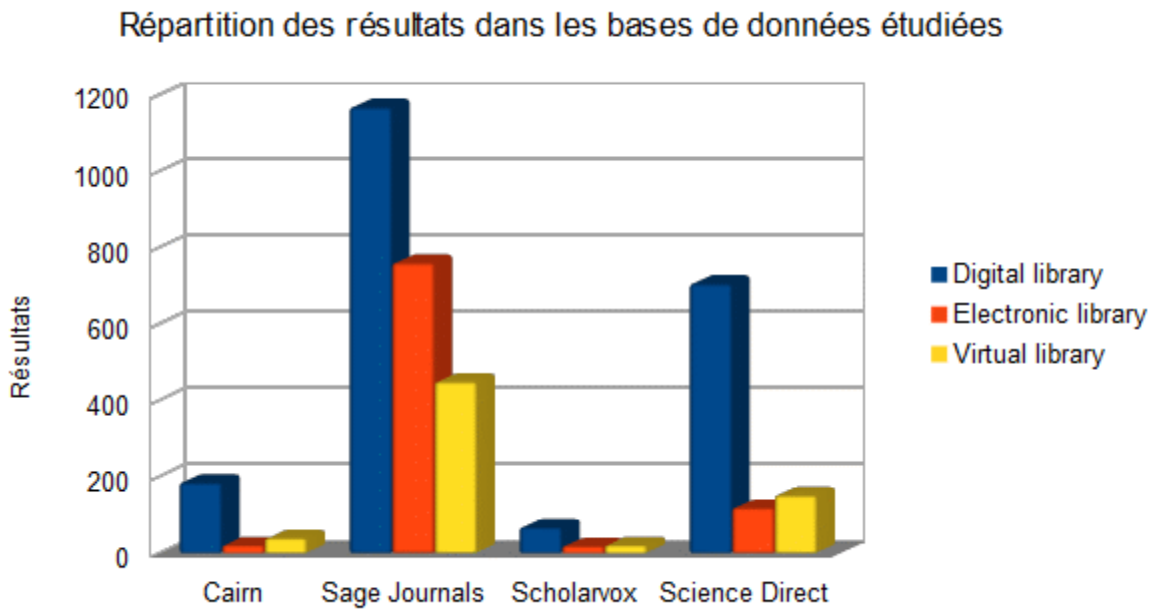


Figure 2.1 : Répartition des termes (Digital library, Electonic library, Virtual library) dans les bases de données (Cairn , Sage Journals, Scholarvox, Science Direct)

Les résultats indiquent que le terme « digital library » est le plus utilisé dans les quatre bases de données. Cela montre que le terme bibliothèque numérique, traduction française de « digital library » est le terme le plus employé par les chercheurs pour désigner les bibliothèques qui gèrent et diffusent des document numériques en ligne.

2.1.2 Bibliothèque numérique : analyse de définitions

A ce stade, le terme « bibliothèque numérique » reste flou et nécessite d'être clarifié. Pour cette raison, nous avons, dans un premier temps, analysé les définitions proposées par la communauté de

28 Nous avons traduit les trois termes en français (bibliothèque numérique, bibliothèque virtuelle et bibliothèque électronique) pour la recherche effectuée dans la base de données Cairn

pratique (practice community) et, dans un deuxième temps, celles proposées par la communauté de recherche (research community) en SIC.

2.1.2.1 Définitions proposées par la communauté de pratique

Nous avons repéré cinq organismes regroupant des professionnels qui s'intéressent aux bibliothèques numériques : l'association des bibliothèques de recherche (Association of Research libraries ARL) ; la fédération des bibliothèques numériques (Digital Library Federation DLF) ; la Fédération internationale des associations de bibliothécaires et d'institutions IFLA ; l'école nationale supérieure des sciences de l'information et des bibliothèques ENSSIB et le réseau d'excellence pour les bibliothèques numériques (DELOS Network of Excellence on Digital libraries). Le but est d'analyser les différentes définitions de bibliothèque numérique proposées par ces organismes.

Nous avons utilisé une méthode employée par Armand Brahaj, Matthias Razum et Julia Hoxha pour définir la notion de bibliothèque numérique. Cette méthode consiste à analyser les définitions issues des articles académiques, et elle s'appuie sur trois éléments caractérisant une définition : l'objet de la définition, la famille à laquelle appartient cette définition et ses attributs distinctifs [Brahaj et al, 2013]. (Voir tableau 2.1 ci-dessous).

Le tableau 2.1 présente l'analyse des définitions de la notion de bibliothèque numérique :

Une bibliothèque numérique n'est pas une entité unique ; elle nécessite la technologie pour relier les ressources de plusieurs bibliothèques et services d'information ; son objectif est d'assurer l'accès universel aux services numériques des bibliothèques ; les bibliothèques numériques ne sont pas limitées à des substituts de documents, elles s'étendent à des artefacts numériques qui ne peuvent pas être représentées ou distribués sous forme imprimée. (ARL, 1995)[29]	Famille	Technologie
	Attributs distinctifs	- relier les ressources - assurer un accès universel aux services numériques des bibliothèques - peut contenir des artefacts numériques
Les bibliothèques numériques sont des organisations qui fournissent des ressources, y compris le personnel spécialisé, pour sélectionner, structurer, offrir un accès intellectuel pour interpréter, distribuer, préserver l'intégrité de documents sous forme numérique, et assurer la maintenance des collections d'œuvres numériques, afin qu'ils soient facilement et économiquement utilisables par une communauté définie ou un ensemble de communauté. (DLF, 1998)[30]	Famille	Organisation
	Attributs distinctifs	- sélectionner des collections numériques - structurer des collections numériques - offrir un accès intellectuel aux collections numériques - distribuer des collections numériques

29 Definition and purposes of digital library (1995). Disponible en ligne <http://old.arl.org/resources/pubs/mmproceedings/126mmappen2>

30 A working definition of digital library (1998). Disponible en ligne <https://old.diglib.org/about/dldefinition.htm>

		- préserver des collections numériques - assurer la maintenance des collections
Une organisation, qui pourrait être virtuelle, qui collecte, gère et conserve à long terme un contenu numérique et offre à ses utilisateurs des fonctionnalités spécialisées, de qualité et selon des politiques codifiées. (DELOS, 2007)[31]	Famille	Organisation
	Attributs distinctifs	- virtuelle - collecte, gère et préserve des contenus numériques - fonctionnalités de qualité - des politiques codifiées
Une bibliothèque numérique est une collection en ligne d'objets numériques, de bonne qualité, créée ou rassemblée et gérée selon des principes internationalement acceptés pour le développement des collections et rendues accessibles de façon cohérente et permanente, soutenue par les services nécessaires pour permettre aux utilisateurs de retrouver et exploiter ces ressources. (IFLA, 2010)[32]	Famille	Collection en ligne d'objets numériques
	Attributs distinctifs	- collection de bonne qualité - collection créée ou rassemblée et gérée selon des principes internationalement acceptés
Les bibliothèques numériques proposent de véritables collections numériques, selon une politique documentaire déterminée. Elles sont alimentées soit par des opérations de numérisation (documents patrimoniaux ou non), soit par des documents nativement numériques. Les contenus sont organisés pour en faciliter la consultation (ENSSIB , 2012)[33]	Famille	Collections numériques
	Attributs distinctifs	- politique documentaire déterminée - documents numérisés - documents nativement numériques - contenus organisés

Tableau 2.1 - Analyse des définitions de la notion de bibliothèque numérique (communauté de pratique)

Une bibliothèque numérique selon la définition de l'association des bibliothèques de recherche (ARL) contient à la fois des documents numérisés et d'autres publiés directement sous forme numérique. Elle nécessite la technologie pour relier et mettre ses ressources et ses services en ligne dans le but d'y garantir un accès universel.

Pour la fédération des bibliothèques numériques (DLF), la bibliothèque numérique ne se résume pas

31 The DELOS digital library reference model : foundations for digital libraries (2007). Disponible en ligne <http://delosw.isti.cnr.it/files/pdf/ReferenceModel/DELOS_DLReferenceModel_096.pdf>

32 IFLA/UNESCO Manifesto for digital libraries (2010). Disponible en ligne <https://www.ifla.org/publications/iflaunesco-manifesto-for-digital-libraries>

33 Bibliothèques numériques (2012). Disponible en ligne <http://www.enssib.fr/le-dictionnaire/bibliotheques-numeriques>

à numériser des fonds et à les mettre en ligne. Elle exige une expertise de type documentaire, qui va mettre à disposition des outils et des langages documentaires permettant l'exploitation des données, cela pour faciliter l'accès à l'information et pour assurer sa conservation sur le long terme.

La définition de l'IFLA met l'accent sur trois points importants : la qualité de l'information ; les normes et les standards utilisés pour le traitement, le partage et la diffusion des collections numériques ; et les services permettant de faciliter l'accès à ces collections. Tandis que, la définition de l'ENSSIB met l'action sur le traitement documentaire exercé sur les collections numériques pour en faciliter la consultation.

L'analyse de ces cinq définitions nous a permis de distinguer deux grandes catégories auxquelles la communauté de pratique en SIC s'intéresse : l'organisation et la collection. Cela montre que cette communauté se préoccupe des problèmes liés à l'organisation des collections en bibliothèques numériques. Plus précisément, elle s'intéresse à la création, la sélection, la structuration, la distribution et la préservation des collections numériques . (voir tableau 2.2 ci-dessous).

Le tableau 2.2 présente la synthèse des définitions issues de la communauté de pratique :

Collections	**Organisation**
• Collection organisée • Collection de bonne qualité • Collection créée ou rassemblée et gérée selon des principes internationalement acceptés • Collection d'objets d'information • Collection de documents en format électronique • Collection numérisée • Artefact numérique	• Sélectionner des collections numériques • Structurer les collections numériques • Offrir un accès intellectuel aux collections numériques • Distribuer des collections numériques • Gérer les contenues numériques • Préserver des collections numériques • Assurer la maintenance des collections • Proposer des politiques documentaires

Tableau 2.2 - Synthèse des définitions issues de la communauté de pratique

Ces différentes définitions permettent de dégager les caractéristiques principales des bibliothèques numériques qui peuvent être résumées de la façon suivante. La bibliothèque numérique :

- gère des collections numériques (documents numérisés et numériques)
- permet l'accès universel aux documents
- nécessite la technologie pour lier et mettre les ressources et les services en ligne
- contient des informations de qualité

- exige une expertise de type documentaire
- met en disposition une information organisée pour en faciliter l'accès
- assure la conservation des informations sur le long terme.

Ces différentes définitions permettent donc de clarifier le terme de bibliothèque numérique d'un point de vue professionnel (la communauté de pratique). L'intérêt que nous portons à cette communauté provient du fait que nous considérons le professionnel de l'information comme étant un acteur central dans les études qui portent sur les usages et les pratiques professionnelles en bibliothèques numériques.

2.1.2.2 Définitions proposées par la communauté de recherche

Dans ce paragraphe, nous analysons le terme « bibliothèque numérique » du point de vue de la communauté de recherche. Des auteurs en sciences de l'information et de la communication anglophones (Lynch, 1995 ; Fox, 1995 ; Borgman, 1996, 1999, 2016 ; Lyman, 1996 ; Bishop, 1996; Cleveland, 1998 ; Waters, 1998 , Saracevic) et francophones (Papy, 2005, 20087, 2009, 2011 ;Saleh, Mkadmi 2008 ; Ihadjadene, ; Chaudiron, Le Crosnier, 2010 ; Salaun etc) s'intéressent dans leurs travaux aux bibliothèques numériques. Ces études portent , d'une part, sur les aspects théoriques et conceptuels de la notion de bibliothèque numérique, et d'autre part, sur les usages et les pratiques informationnelles, les langages documentaires, les métadonnées etc.

Nous exposons d'une manière chronologique les différentes définitions données par la communauté de recherche. Pour chaque définition nous distinguons sa famille et ses attributs distinctifs.(Voir tableau 2.3 ci-dessous)

Le tableau 2.3 présente l'analyse des définitions de la notion de bibliothèque numérique issues de la communauté de recherche.

Une bibliothèque numérique est un environnement technologique distribué qui réduit considérablement les obstacles à la création, la manipulation, la diffusion, l'intégration, le stockage et la réutilisation de l'information par les individus et les groupes (Fox 1993) *La bibliothèque de l'avenir sera basée sur des données électroniques ... contient à la fois du texte et des graphiques et sera largement disponible via les réseaux*	Famille	Environnement technologique distribué
	Attributs distinctifs	- réduire les obstacles à la création, la diffusion, l'intégration, le stockage et la réutilisation de l'information - contient à la fois du texte et

électroniques. Il est susceptible d'être décentralisé (Fox, 1995)		des graphes - disponible via les réseaux électroniques
Systèmes fournissant à une communauté d'utilisateurs un accès cohérent à un dépôt organisé d'informations et de connaissances ... enrichi par les capacités de la technologie numérique ... couvrant à la fois des matériaux imprimés et numériques ... tels que des données multimédia, géospatiales ou numériques ... [caractérisé par] la continuité [avec] les rôles et les missions traditionnels de bibliothèques ... (Lynch et Garcia-Molina, 1995)	Famille	Système
	Attributs distinctifs	- accès cohérent à l'information organisée - enrichi par les capacités de la technologie numérique - extension de la bibliothèque traditionnelle - multimédias
La définition de la bibliothèque numérique exigera une compréhension du rôle et de la nature des institutions publiques dans une société postindustrielle. *Un domaine de liberté d'expression ainsi qu'un marché de l'information . (Lyman , 1996, 1999)*	Famille	Marché de l'information
	Attributs distinctifs	- extension de la bibliothèque traditionnelle (rôle dans la société) - un domaine de liberté d'expression
Collections organisées d'informations numériques. Elles associent la structuration et la collecte de l'information, ce que les bibliothèques ont toujours fait, en permettant une représentation numérique de l'information que les ordinateurs ont rendue possible. (Lesk, 1997)	Famille	Collections organisées
	Attributs distinctifs	- informations numériques - extension de la bibliothèque traditionnelle - structuration de l'information - collecte de l'information - représentation numérique de l'information
La bibliothèque numérique un ensemble de ressources électroniques et de capacités techniques associées pour créer, repérer et utiliser l'information. *Les bibliothèques numériques sont construites - collectées et organisées - par [et pour] une communauté d'utilisateurs, et leurs capacités fonctionnelles prennent en charge les besoins en information de cette communauté* (Borgman, 2000)	Famille	Ressources électroniques
	Attributs distinctifs	- capacités techniques - créer, rechercher et utiliser l'information - répondre aux besoins de la communauté d'usagers
une bibliothèque numérique est une collection d'information organisée, accompagnée par des services, où les informations sont stockées en formats numériques et sont accessibles sur les réseaux (Arms, 2001)	Famille	Collection d'information organisée
	Attributs distinctifs	- services -informations stockées en formats numériques - accessibles sur les réseaux
«Systèmes sociotechniques - réseaux de technologie,	Famille	Système sociotechnique

d'information, de documents, de personnes et de pratiques». (Bishop et al., 2003)	Attributs distinctifs	Réseaux de : - technologie - information - documents - personnes - pratiques
« est une vrai bibliothèque car elle est une collection organisée, triée, présentée » (Papy, 2005)	Famille	Collection
	Attributs distinctifs	- organisée -triée - présentée - vrai bibliothèque
Une bibliothèque est différente d'un entrepôt par la place que prennent les métadonnées. Il ne s'agit pas de retrouver un document par son identifiant, mais de constituer une information sur le document, et de le relier par les métadonnées à la fois à l'organisation globale de la connaissance et aux autres documents disponibles. *Une bibliothèque numérique travaille sur le « contexte » : la contextualisation des documents numérisés devient un élément essentiel d'un projet alternatif au "quick and dirty" de l'indexation de Google. (Le crosnier, 2007, p.131)*	Famille	Organisation
	Attributs distinctifs	- gérer des métadonnées - relier l'information à l'organisation globale de la connaissance - contextualiser des documents numérisés
La bibliothèque numérique est une collection ciblée d'objets numériques, y compris le texte, la vidéo et l'audio, ainsi que des méthodes d'accès, de récupération, de sélection, d'organisation et d'entretien de la collection (Witten, 2009)	Famille	Collection ciblée d'objets numériques
	Attributs distinctifs	- texte, audio, vidéo - méthodes d'accès, de récupération , de sélection, d'organisation et d'entretien de la collection.
La bibliothèque numérique, en tant qu'innovation, se fonde sur plusieurs inventions, dont notamment les réseaux télématiques, la photographie numérique, les scanners, la reconnaissance optique de caractères. Ces inventions ont été combinées grâce à l'action de différents groupes professionnels pour aboutir à la création d'un artefact technique innovant, la bibliothèque numérique. (Bequet, 2013)	Famille	Innovation
	Attributs distinctifs	- réseaux télématiques - photographie numérique - artefact technique -différents groupes professionnels

Tableau 2.3 - Analyse des définitions de la notion de bibliothèque numérique (communauté de recherche)

L'analyse de ces définitions nous permet de distinguer quatre catégories auxquelles la communauté de recherche en SIC s'intéresse : la collection, le système, la technologie et l'organisation. (voir tableau 2.4 ci-dessous).

Le tableau 2.4 présente la synthèse des définitions issues de la communauté de recherche

Collection	Système	Technologie	Organisation
• collection organisée • ressources électroniques • métadonnées • multimédias	• système sociotechnique • pratiques • services • accessibilité • accès au contenu numérique • besoins des usagers	• réseaux • artefact technique numériques • photographie numérique	• Information liée à l'ensemble des connaissances • contextualisation des données • personnels • gestion de l'information numérique (création, collecte, gestion, diffusion, conservation)

Tableau 2.4 - Synthèse des définitions issues de la communauté de recherche

La communauté de recherche en SIC s'intéresse à la fois à des problématiques liées aux aspects techniques de la bibliothèque numérique (les définitions antérieures aux années 2000), et à celles liées à l'usage des bibliothèques numériques (besoins, usages et pratiques). Cela se concrétise par les deux approches orientée-système et orientée-usage que les auteurs en SIC mobilisent dans leurs recherches portant sur les systèmes et les dispositifs d'organisation, de recherche et d'accès à l'information.

Les définitions proposées par la communauté de recherche restent assez proches de celles proposées par la communauté de pratique en ce qui concerne les aspects liés à l'organisation des collections numériques. Cependant, la communauté de recherche s'intéresse aux usages et aux pratiques des utilisateurs des bibliothèques numériques, les professionnels y compris. La communauté de pratique devient donc un objet d'études pour la communauté de recherche qui s'intéresse aux pratiques des professionnels de l'information. Les deux communautés interagissent les unes avec les autres partageant un intérêt commun, celui de la bibliothèque numérique.

Après avoir analysé ces différentes définitions, notre objectif était dans un premier temps, de clarifier le terme de « bibliothèque numérique » et dans un deuxième temps, de proposer un cadre conceptuel de ce terme en sciences de l'information et de la communication. Dans la section suivante, nous mobilisons des concepts et des approches en sciences de l'information et de la communication sur lesquels se fonder pour déterminer un cadre conceptuel du terme bibliothèque

numérique.

2.2 La bibliothèque numérique : nécessité d'ancrage conceptuel en SIC

Yves Jeanneret, indique qu'il y a une différence essentielle de nature entre l'information au sens social du terme et l'information au sens mathématique. L'information numérique au sens social, est une information communiquée via des dispositifs afin d'être utilisée et interprétée par des groupes d'individus [Jeanneret, 2000, p. 45]. Yves Jeanneret évoque deux termes essentiels : communication et dispositif. Le premier terme, la communication, est un enjeu important car sans elle, l'information ne peut être échangée et perd ainsi son intérêt et sa potentielle, et le deuxième terme, le dispositif, correspond au moyen par lequel cette information est communiquée.

De son coté, Bruno Bachimont indique que les contenus numériques présentent une double face : une face technique et une face culturelle. La face technique concerne les outils par lesquels les contenus se manipulent. La face culturelle relève de la valeur de ces contenus en tant que fruit de la pensée humaine [Bachimont, 2007, p.70].

Les études ayant comme objet de recherche l'information numérique nécessitent de prendre en considération sa valeur technique ainsi que sa valeur sociale. Ainsi dans ce qui suit, nous prenons en compte les aspects techniques et sociaux du développement des bibliothèques numériques. Cela nous amène à mobiliser la notion de dispositif info-communicationnel qui prend en considération les interactions entre les utilisateurs et la technique.

2.2.1 La bibliothèque numérique : un dispositif info-communicationnel

Une bibliothèque numérique, considérée comme un lieu culturel, dispose d'un ensemble de moyens techniques et sociaux pour s'organiser et communiquer. Cet ensemble de moyens peut-il être considéré comme un dispositif info-communicationnel?

2.2.1.1 La notion de dispositif

Le dispositif se définit comme « *la manière dont sont disposés les pièces, les organes d'un appareil ; le mécanisme*»[34].

Cette notion est utilisée dans de nombreuses sciences. En sciences de l'ingénieur, considéré comme un système, le dispositif est utilisé pour son aspect technique [Couzinet, 2011]. En sciences

34 REY-DEBOVE, J. (2004). *Le Robert Brio: analyse des mots et régularités du lexique* . Robert, Paris.

sociales, l'usage du terme dispositif trouve son origine avec Michel Foucault. Selon Foucault (1977) c'est un assemblage intentionnel d'éléments hétérogènes (instruments, actions publiques, méthodes, lois, morales etc.) répartis spécifiquement en fonction d'une finalité attendue [Beuscart et Peerbaye , 2006]. Le dispositif tel qu'il est conçu par Foucault est « *une formation historique spécifique, issue du jeu de différents éléments hétérogènes »* [Ibid, p. 6]. Bruno Latour parle d'une association pour désigner les assemblages socio-techniques d'humains et de non-humains [Latour, 1996, 2006].

En sociologie, la notion de dispositif est utilisée, d'une part, pour désigner un assemblage ou une association de pratiques sociales permettant de mettre en œuvre une stratégie spécifique, d'autre part, pour désigner un ensemble d'instruments techniques et leur interaction avec le monde social.

2.2.1.2 Dispositif info-communicationnel

La prise en considération des usages sociaux de la technique a permis de mettre la notion de dispositif au centre des préoccupations des auteurs en SIC. Les dimensions sociales et techniques d'un dispositif se composent d'acteurs, de techniques et d'objets matériels, permettant la mise en commun d'informations à activer [Couzinet, 2009]. Dans ce contexte, Emmanuel Souchier indique que trois espaces d'analyse doivent être pris en compte : l'objet technique, le processus de communication et la façon dont les usagers s'approprient le dispositif [Souchier et al., 2003].

De la notion très générale du dispositif découle plusieurs notions plus restreintes appliquées en sciences de l'information et de la communication qui sont : le dispositif d'accès à l'information, le dispositif de traitement de l'information, le dispositif info-communicationnel, etc.

Yves Jeanneret a introduit la notion de dispositif de traitement de l'information pour désigner une technique ayant comme composants des appareils de traitement de l'information et *«ayant pour effet social de faire circuler des messages»,* de faciliter les échanges d'information et les productions des connaissances et des savoirs dans la société [Jeanneret, 2000].

Stéphane Chaudiron et Madjid Ihadjadène (2008, 2010), Brigitte Simonnot (2012) ont utilisé la notion de dispositif d'accès à l'information pour désigner les activités et les pratiques liées à la recherche d'information via l'usage d'un outil technique. Les dispositifs d'accès à l'information en ligne sont des lieux d'interaction entre trois acteurs : le producteur de l'information, celui qui assure son traitement et celui qui cherche à la retrouver. Ces dispositifs sont des lieux de médiations par excellence indique Brigitte Simonnot [Simonnot, 2012].

Ensuite, la notion de dispositif info-communicationnel a été abordé par Viviane Couzinet pour mettre en évidence l'interdépendance entre information et communication [Couzinet, 2009]. Cette position met en valeur le rôle communicationnel de l'information, abordé par Jean Meyriat qui

considère information et communication comme « des objets solidaires ». Selon Couzinet le dispositif info-communicaionnel doit être abordé *« comme un ensemble de liens sociaux unissant le producteur d'information, celui qui permet sa circulation grâce à des machines et à des tuyaux, celui qui intervient pour faciliter la diffusion et celui qui s'approprie l'information »* [Couzinet, 2009, p.22].

Nous pouvons constater que le dispositif info-communicationnel peut être vu sous trois dimensions: une dimension technique (la conception), une dimension sociale (les usages) et une dimension de médiation (les pratiques professionnelles). Dans cette thèse, nous nous intéressons à la dimension de médiation liée aux pratiques professionnelles qui interviennent pour faciliter l'accès à l'information, celle-ci est correspond à la médiation documentaire. Cette dimension sera abordée dans la section (2.3).

2.2.1.3 Positionnement par rapport à la notion de dispositif

Nous considérons que la notion de dispositif info-communicationnel est une notion large qui couvre à la fois les dimensions : technique, sociale et documentaire de l'information. C'est pourquoi nous considérons que l'approche info-communicationnelle est une approche permettant d'étudier la notion de bibliothèque numérique sous une dimension technique, sociale et documentaire. Autrement dit, les études qui portent sur les bibliothèques numériques sont souvent orientées soit système soit usager. Étant des lieux d'interaction entre technique et usage, les bibliothèques numériques pourraient être considérées comme des dispositifs info-communicationnels mettant en relation trois acteurs : les producteurs de l'information, les personnes qui gèrent l'information et les utilisateurs.

Les bibliothèques numériques peuvent être pensées comme des dispositifs info-communicationnels permettant d'établir une relation entre un utilisateur et une information. Cette relation est rendue possible par l'intermédiaire d'un professionnel, qui à son tour assure un traitement documentaire de l'information en se servant à la fois des langages documentaires et des outils techniques. Une bibliothèque numérique ne se résume pas à numériser des fonds et les mettre en ligne. Elle exige une expertise de type documentaire qui va mettre à disposition des outils et des langages documentaires permettant l'exploitation des données pour faciliter l'accès à l'information et pour assurer sa conservation sur le long terme.

Cela nous ramène à proposer une définition de la notion de bibliothèque numérique :

Une bibliothèque numérique peut être considérée comme un dispositif info-communicationnel permettant, grâce aux outils techniques et au traitement documentaire, l'accès à l'information

numérique.

Une bibliothèque numérique considérée comme un dispositif info-communicationnel est à la fois : un dispositif technique permettant la circulation de l'information et son traitement via des langages informatiques et des logiciels ; un dispositif social permettant la diffusion des savoirs ; et un dispositif documentaire qui, selon Cécile Gardiès, *« rassemble des objets, objets de savoir, organisés par des actions humaines dans un but de stockage, de conservation et d'accès »* [Gardiès, 2012, p.117].

Dans cette thèse, nous nous intéressons à la dimension documentaire de la notion de dispositif info-communicationnel. A ce propos, Viviane Couzinet indique que cette dimension implique un processus de médiation *« visant à faciliter l'accès au document et à son contenu dans un contexte particulier »* [Couzinet, 2011, p.117]. Dans la section suivante, nous mobilisons la notion de médiation afin d'analyser ses aspects en bibliothèques numériques.

2.3 La médiation documentaire en bibliothèques numériques

La médiation est définit par « *le fait de servir d'intermédiaire* »[35]. Ainsi, l' « intermédiaire » est « *qui, étant entre deux termes, forme une transition ou bien assure une communication* »[36].

La notion de médiation est devenue un champs de recherche important en sciences de l'information et de la communication SIC. Elle est présente dans la production, la diffusion et l'appropriation de l'information au sein de l'espace public [Lamizet, 1995]. Plusieurs auteurs comme Bernard Lamizet (médiation culturelle, politique, communicationnelle), Yves Jeanneret (médiation culturelle, médiation des savoirs), Isabelle Fabre et Cécile Gardiès (médiation documentaire) etc. ont étudié ce concept.

Dans un premier temps, nous abordons la notion de médiation documentaire. Cette notion est assez jeune et sa définition est en cours de développement. Étudiée par Cécile Gardiès, la notion de médiation documentaire, en s'appuyant sur des dispositifs matériels et humains, permet de « *concilier deux choses jusque là non rassemblées pour favoriser l'accès à l'information* » [Gardiès, 2012, p. 145]. En d'autres termes, il s'agit d'unir usager et information par l'intermédiaire de la technique et du travail intellectuel humain. Cette médiation nécessite donc un médiateur : le documentaliste, le bibliothécaire, le professionnel de l'information etc.

De son coté, Cécile Gardiès considère que la médiation documentaire permet de faciliter la

35 Rey, A. (dir.) (2005). *Dictionnaire culturel en langue française*. Le Robert, Paris
36 REY-DEBOVE, J. (2004). *Le Robert Brio: analyse des mots et régularités du lexique* . Robert, Paris.

communication entre celui qui produit l'information et celui qui la reçoit [Gardiès, 2011]. Que ce soit pour communiquer ou bien faciliter l'accès à l'information, la médiation documentaire passe tout abord par un traitement documentaire.

2.3.1 Le traitement documentaire

Considéré comme étant un « dispositif documentaire secondaire », le traitement documentaire consiste à extraire des informations à partir des dispositifs documentaires primaires (les documents) [Couzinet, 2011]. Il s'agit d'une médiation qui selon Jean Meyriat permet de créer une valeur ajoutée à l'information : une information sur l'information [Meyriat, 1985].

Pour comprendre le traitement documentaire exercé par les professionnels de l'information pour décrire les documents, il faut mettre en évidence le rôle des langages documentaires. Pour Caroline Courbières, la notion de langage documentaire « *renvoie au phénomène de médiation* » qui engage les professionnels de l'information [Courbières, 2011, p.131].

Un langage documentaire est utilisé pour décrire une ressource afin de faciliter l'accès à l'information qu'elle contient. L'originalité de ces langages est due à la compréhension de leur raison d'être et de leur contexte de communication [Maniez, 1987]. Autrement dit, l'enjeu des langages documentaires est, d'une part, de retrouver facilement et rapidement les documents et d'autre part, d'établir une communication entre celui qui indexe le document (le documentaliste) et celui qui le lit (l'usager). Cette communication, que nous pouvons nommer médiation documentaire, permet d'un coté un accès facile à l'information, et de l'autre, une représentation intellectuelle de cette information.

En effet, la médiation documentaire consiste à établir un lien entre le professionnel de l'information et le document par l'intermédiaire d'un langage documentaire. Le but est d'assurer une représentation intellectuelle du document ainsi qu'en faciliter l'accès.

Avec le numérique, le traitement documentaire introduit des outils techniques et des langages informatiques pour représenter le contenu numérique. Plus précisément, de nouveaux « dispositifs» commencent à apparaître pour conserver, traiter et diffuser les documents numériques. Les bibliothèques numériques considérées comme des dispositifs info-communicationnels nécessitent de repenser la médiation documentaire. Nous abordons ce sujet dans la section qui suit en introduisant la notion de métadonnées.

2.3.2 Les métadonnées en bibliothèques numériques

Les bibliothèques numériques gèrent des documents numériques qui peuvent être d'une nature variée : du texte, graphique, photo, son, vidéo et image animée, et qui sont stockés sur des supports informatique transportables et connectés sur des réseaux. Pour bien gérer et exploiter les ressources numériques, il faut pouvoir les décrire d'une manière efficace. Pour cela il faut les identifier et les structurer. La structuration des données numériques passe de plus en plus par l'utilisation des langages de balisage (notamment le langage XML).

Les professionnels de l'information ont déjà une longue pratique dans la structuration des contenus des documents qu'ils manipulent. Avant même l'informatique, les bibliothèques utilisaient des fiches cartonnées dont la structure a été normalisée en 1954 sous la référence ISBD (International Standard Bibliographic Description). Ces descriptions ont ensuite été informatisées sous la forme de notices bibliographiques et normalisées (voir par exemple les formats MARC en 1964 utilisant la norme ISO 2709 dont la conception a démarré en 1960). Ces descriptions facilitent la gestion des ressources documentaires et l'échange de données numériques entre bibliothèques.

Le terme « métadonnée » (en anglais : metadata) est apparu dans le cadre de la description des ressources sur Internet dans les années 1990 et s'est ensuite généralisé. Cependant, l'expression « une information sur l'information » a été utilisée par Jean Meyriat pour caractériser le traitement documentaire permettant non seulement l'accès à l'information dont l'usager a besoin, mais aussi à l'information nécessaire pour accéder à cette dernière [Meyriat, 1985].

Une définition formelle du terme métadonnée a été proposé par François Role. Une métadonnée est un couple <I,S>, où I représente un identifiant permettant d'accéder à une ressource et S représente une structure de données [Role, 2000]. Cette définition met en évidence deux éléments essentiels : l'identification des ressources et la structuration des données. L'identification a pour objectif de faciliter l'accès à l'information, tandis que la structuration permet de décrire les données pour les rendre exploitables par les machines. Une métadonnée peut être définie comme un ensemble structuré d'informations décrivant une ressource quelconque dans le but d'en faciliter l'accès.

Les métadonnées deviennent de plus en plus des éléments essentiels dans la gestion des contenus numériques sur le web. Elles apparaissent comme un moyen efficace pour gérer des collections hétérogènes au sein des bibliothèques numériques [Role, 2000].

Selon Christiane Borgman, le contenu d'une bibliothèque numérique comprend les données, les métadonnées qui décrivent les divers aspects des données et les métadonnées permettant de lier les données entre elles ou bien de les lier avec d'autres métadonnées [Borgman, 1999]. De ce point de

vue, les métadonnées permettent à la fois une description des ressources numériques et une liaison avec d'autres ressources. Autrement dit, l'enjeu principal des métadonnées est d'enrichir les données et de les relier avec d'autres ressources. Carl Lagoze ajoute que les bibliothèques numériques se distinguent des moteurs de recherche par la valeur ajoutée aux ressources permettant de « contextualiser l'information » [Lagoze et al., 2006]. Pour illustrer cette notion nous nous référons à Hervé Le Crosnier qui considère qu'une bibliothèque numérique travaille sur le «contexte» : « .. *les métadonnées, pour donner toute leur dynamique, doivent elles aussi être reliées et contextualisées .. ».* Il s'agit de constituer une information sur le document permettant de le relier à l'organisation de la connaissance [Le Crosnier, 2007, p.131]. Autrement dit, les métadonnées permettent non seulement l'identification et la description des ressources, mais aussi de relier les ressources disponibles sur le web entre elles.

En résumé, la médiation documentaire numérique consiste à identifier les documents numériques et les décrire pour extraire des métadonnées. Ce traitement, appelé traitement documentaire, nécessite l'intervention d'un professionnel de l'information qui va s'appuyer sur des langages documentaires et informatiques pour représenter les documents. Le but est d'offrir une représentation structurée de l'information et ses liens avec d'autres ressources sur le web.

Le schéma ci-dessous présente la médiation documentaire numérique.

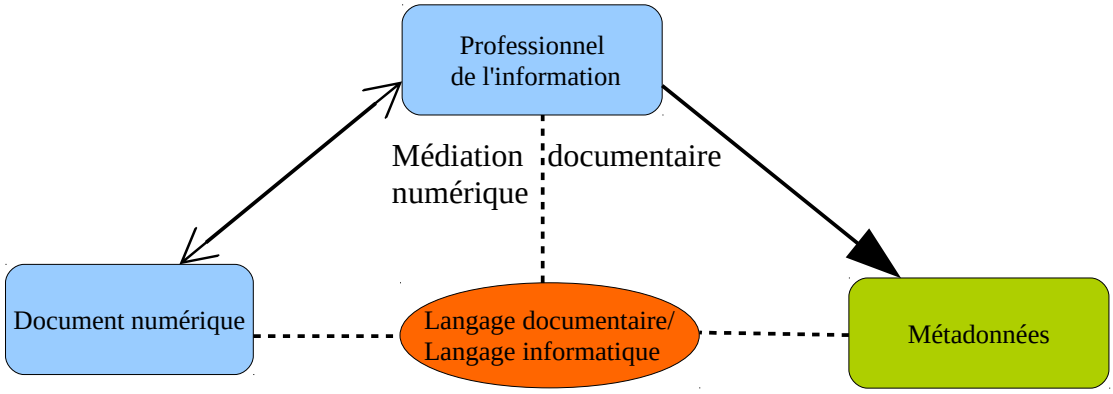


Figure 2.2 : La médiation documentaire numérique

La production, la gestion et le partage des métadonnées jouent un rôle central dans l'activité des professionnels de l'information en bibliothèques numériques. D'une part, les métadonnées sont utilisées pour identifier et structurer le contenu numérique afin d'assurer sa représentation au sein du

dispositif. D'autre part, elles jouent le rôle de médiateur entre l'usager et l'information en facilitant notamment l'accès au contenu numérique. Dans la section suivante, nous traitons ce sujet en mettant en valeur le rôle des bibliothèques numériques dans la diffusion des savoirs.

2.4 L'accès aux connaissances en bibliothèques numériques : quelle médiation ?

Les bibliothèques numériques, considérées comme des dispositifs info-communicationnels permettent une rencontre entre un utilisateur et une information. Cette rencontre oblige à penser la bibliothèque sous l'angle de la médiation. Une médiation qui se divise en deux branches complémentaires celle de la documentation et celle des savoirs. La dernière prend en considération la dimension de l'usager.

Les bibliothèques numériques sont des lieux culturels qui donnent accès aux savoirs. Autrement dit, le rôle des bibliothèques numériques n'est pas limité à numériser et à réserver des collections numériques. Leur rôle de diffusion des connaissances est au cœur de leur missions.

2.4.1 Médiation culturelle, médiation de connaissances, médiation de savoirs, quelle terminologie employer ?

Les notions médiation culturelle, de connaissances et de savoirs sont utilisées pour désigner les dispositifs mis à disposition pour faciliter l'accès à l'information. Pour comprendre la différence entre ces trois notions il faut se concentrer sur leur objet d'étude : la culture, les connaissances et les savoirs.

Pour Jean Meyriat l'information produit « *une modification de l'état de connaissance de celui qui la reçoit* » [Meyriat, 1985]. Yves Jeanneret indique que le terme connaissance est utilisé pour indiquer « *le travail productif des sujets sur eux-mêmes pour s'approprier des idées ou des méthodes* » [Jeanneret, 2000, p.85]. La connaissance est donc une information reçue qui entraîne une modification de l'état d'esprit de la personne. C'est une information assimilée permettant d'élaborer des idées, de prendre des décisions et de réaliser des actions. Elle est construite par la personne qui se l'approprie.

Le terme savoir est défini comme « *un ensemble de connaissances assez nombreuses* »[37] ou aussi « *les savants, la science* »[38]. Il est utilisé pour « *caractériser les formes de connaissance qui sont*

37 Le Robert (Dir. Alain Rey). Dictionnaire culturel en langue française
38 Ibid

reconnues par une société» [Jeannneret, 2000 p.85]. Ce sont des savoirs transmis et transcrits pour être réservés et diffusés aux individus de la société. Ils désignent les connaissances qui sont reconnues socialement et qui tendent à une certaine stabilité [Gardiès, 2011, 2012].

La notion de « culture », quant à elle, est une notion plus large et problématique dans le sens où elle est vivante et constamment ouverte à la critique et au questionnement. A ce propos, Denys Cuche indique que « *quel que soit le sens précis qui a pu être donné au mot [culture] des désaccords ont toujours subsisté sur son application à telle ou telle réalité* » [Cuche, 2010, p.7]. Notre but n'est pas de nous plonger dans une analyse de cette notion mais de chercher à la positionner par rapport aux deux notions précédentes.

Dans son sens le plus large, l'Unesco définit la culture (1982) *«comme l'ensemble des traits distinctifs, spirituels et matériels, intellectuels et affectifs, qui caractérisent une société ou un groupe social. Elle englobe, outre les arts et les lettres, les modes de vie, les droits fondamentaux de l'être humain, les systèmes de valeurs, les traditions et les croyances* »[39]. La notion de culture apporte une dimension sociale plus large.

Pour illustrer notre vision sur ces trois notions nous les représentons sous forme de graphe (voir figure 2.3 ci-dessous). La notion de culture est une notion plus large qui englobe les savoirs et les connaissances, elle caractérise une société. Les savoirs quant à eux sont des connaissances reconnues et diffusées dans la société après avoir été validés par des communautés scientifiques, littéraires etc. Enfin, les connaissances sont des informations assimilées et appropriées par des individus.

39 http://www.bak.admin.ch/themen/04117/index.html?lang=fr

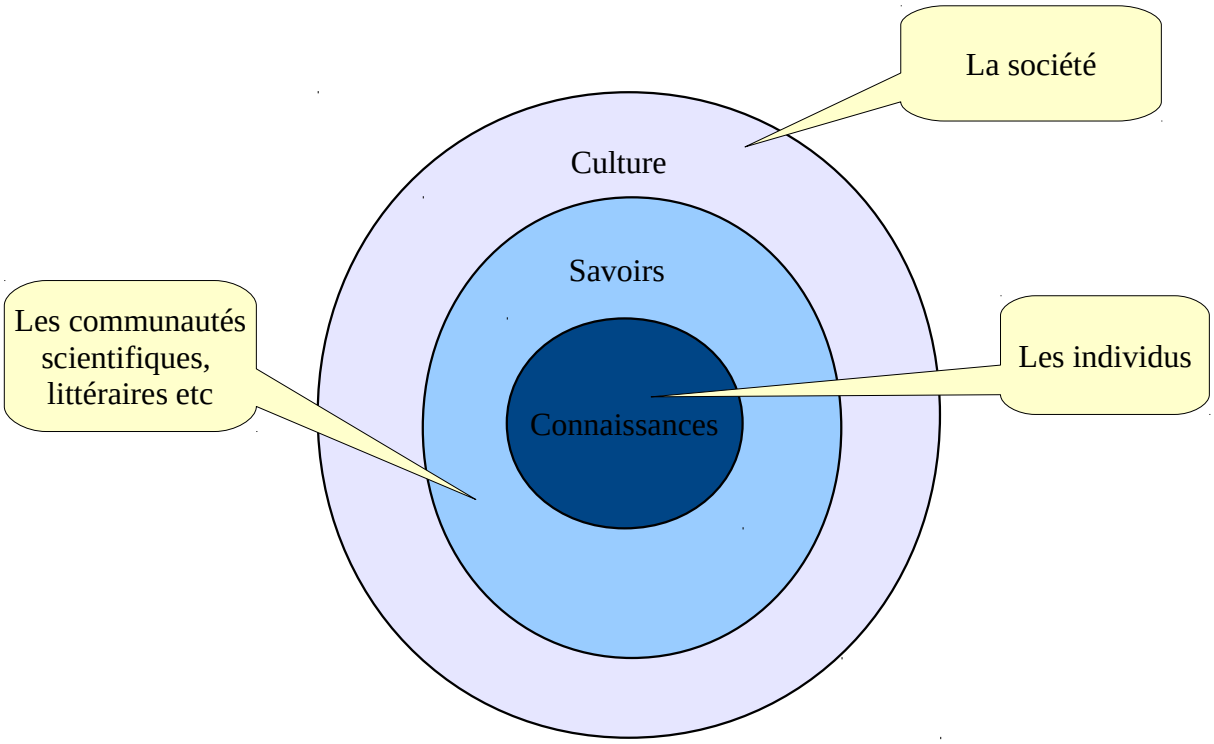


Figure 2.3: Positionnement des notions : connaissances, savoirs et culture

2.4.2 Médiation des savoirs en bibliothèques numériques

Une bibliothèque numérique est différente d'un moteur de recherche par le fait de proposer une information organisée et vérifiée. Profitant de leur présence sur le web, les bibliothèques numériques donnent accès aux collections numériques à des usagers distants. L'enjeu de la numérisation est non seulement d'assurer la réservation des collections mais aussi de les présenter pour leur valeur intellectuelle. De ce fait, les bibliothèques numériques ont une place très importante à occuper sur le web. Elles assurent un accès ouvert et libre à des collections numériques d'une valeur informative.

Nous considérons que la notion de médiation des savoirs est la plus adaptée pour désigner la médiation en bibliothèques numériques. D'une part, la bibliothèque donne accès aux savoirs, c'est-à-dire l'ensemble des connaissances transcrites et stables. L'utilisateur va décomposer cette information (savoirs) pour l'ajouter à ses connaissances. Dans ce cas là, l'information décomposée et appropriée devient des connaissances dans l'esprit des sujets qui l'assimilent.

Pour illustrer notre vision des bibliothèques numériques nous avons dessiné un schéma permettant de mettre en valeur la médiation documentaire et la médiation des savoirs au sein de la bibliothèque

numérique. (Voir figure 2.4 ci-dessous).

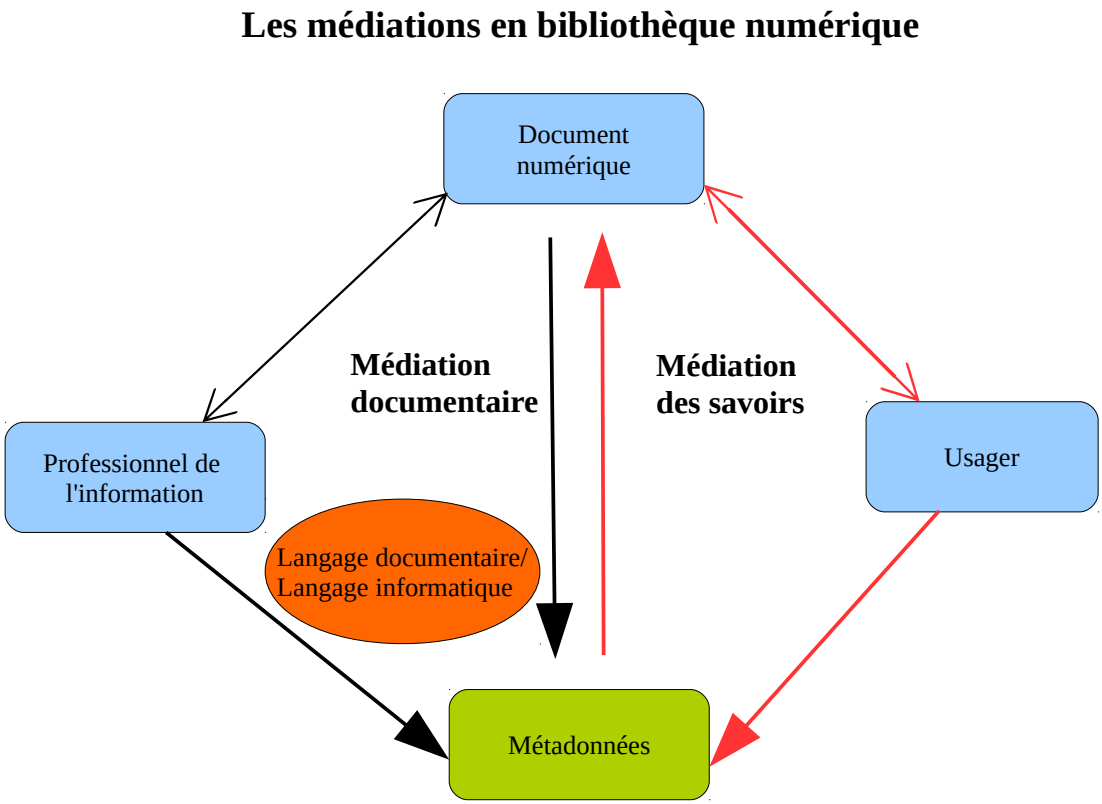


Figure 2.4: La médiation documentaire et la médiation des savoirs en bibliothèques numériques

En résumé les bibliothèques numériques assurent une double médiation : une médiation documentaire et une médiation des savoirs. La première, par le biais du traitement documentaire, offre une information accompagnée par d'autres éléments susceptibles d'en assurer une compréhension globale. Autrement dit, il s'agit de proposer des éléments d'information permettant une organisation et une contextualisation de l'information. Plus précisément, il s'agit de fournir des métadonnées permettant à la fois d'identifier et de décrire les informations, et en même temps de servir de clés d'accès à l'information. De ce fait, la médiation documentaire entraîne une médiation des savoirs dans le sens où l'objectif du traitement documentaire est de faciliter l'accès aux savoirs. Contrairement à Cécile Gardiès qui considère que la médiation documentaire est une médiation des savoirs [Gardiès, 2012], nous pensons qu'il s'agit de deux notions distinctes qui peuvent être pensées selon deux niveaux complémentaires. Le premier niveau est documentaire et permet d'unir le document et le professionnel de l'information. Le deuxième niveau est lié à l'accès aux savoirs et permet d'unir le document et l'usager. Autrement dit, le premier niveau de médiation correspond à la médiation documentaire qui permet de produire des métadonnées. Ces dernières interviennent ensuite dans une médiation des savoirs qui permet aux usagers d'accéder aux documents .

La médiation des savoirs, consiste à établir un lien entre l'usager et le document par l'intermédiaire de la médiation documentaire. En d'autres termes, la médiation des savoirs permet à l'usager d'accéder aux documents. Cet accès se fait grâce à un traitement intellectuel réalisé par le professionnel de l'information (médiation documentaire). Cela signifie que la médiation des savoirs nécessite en amont une médiation documentaire.

Conclusion du chapitre

Dans ce chapitre nous avons apporté un éclairage sur la notion de bibliothèque numérique. Après une analyse des différentes définitions, nous avons pu déterminer quelques éléments qui caractérisent les bibliothèques numériques. Ces caractéristiques peuvent être résumées ainsi : la gestion des collections numériques ; l'accès libre aux collections numériques ; le besoin d'une expertise de type documentaire ; la nécessité de la technologie pour créer, gérer et diffuser des ressources et des service en ligne.

Ensuite, nous avons cherché des fondements théoriques en sciences de l'information et de la communication permettant un cadrage conceptuel de la notion de bibliothèque numérique. Notre attention s'est tournée vers le concept de dispositif. Utilisé en sciences de l'information et de la communication, ce concept renvoie non seulement aux aspects techniques de l'outil mais aussi à ses aspects sociaux. D'où l'usage de l'expression de dispositif info-communicationnel comme concept fondamental en SIC. Mettant en relation les personnes et les objets pour échanger et communiquer l'information, les dispositifs info-communicationnels prennent en considération la dimension technique, sociale et médiatique de l'outil.

Après avoir étudié les deux notions que sont la bibliothèque numérique et le web sémantique, nous consacrons le troisième chapitre aux usages et aux applications du web sémantique en bibliothèques numériques. Nous nous intéressons plus précisément à l'évolution des pratiques des professionnels de l'information suite à l'introduction du web sémantique en bibliothèques numériques.

Chapitre 3

Les usages et les applications du web sémantique en bibliothèques numériques

Après avoir abordé dans les deux chapitres précédents les deux notions « web sémantique » et « bibliothèque numérique », ce chapitre a pour objectif de s'interroger sur les usages et les applications du web sémantique en bibliothèques numériques.

Les bibliothèques numériques ne cessent d'évoluer afin de répondre aux besoins des usagers. Un changement de paradigme a conduit l'approche orientée-utilisateur et l'approche orientée-usage à prendre peu à peu la place de l'approche orientée-système. Cela se manifeste par des développements au niveau des catalogues prenant en compte les besoins de l'usager. Les principaux changements liés à notre sujet se font au niveau de la représentation des données permettant de proposer des modèles conceptuels de références bibliographiques (FRBR, RDA, Dublin Core). Le but est de concevoir des modèles communs entre plusieurs communautés (bibliothèques, archives, musées …) afin de pouvoir échanger les données efficacement pour présenter aux usagers une information complète.

De par ces propriétés (voir chapitre 1), le web sémantique présente un fort intérêt dans l'évolution des bibliothèques numériques. Il se présente comme étant porteur de solutions à ces problèmes, conduisant les bibliothèques à s'y intéresser. Ainsi, dans ce chapitre, en introduisant les pratiques professionnelles, nous abordons les applications du web sémantique en bibliothèques numériques tout en soulignant les défis et les avantages qui les accompagnent.

Ce chapitre est divisé en deux sections. La première aborde la question des pratiques professionnelles en bibliothèques numériques. Nous commençons par une introduction théorique sur l'emploi des termes « usage » et « pratique ». Ensuite nous nous interrogeons sur les pratiques des professionnels en bibliothèques numériques. La seconde section quant à elle s'intéresse aux applications du web sémantique en bibliothèques numériques. Nous présentons les changements qui interviennent au sein des catalogues des bibliothèques en nous focalisant sur les aspects liés aux pratiques des professionnels de l'information.

3.1 Usages et pratiques en bibliothèques numériques : le professionnel en question

En SIC, les recherches portant sur les systèmes de recherche d'informations et les dispositifs info-communicationnels adoptent trois approches différentes qui sont : l'approche « orientée-système », l'approche « orientée-utilisateur » et l'approche « orientée-usage». La première approche était dominante depuis plus de trente ans pour une raison est avant tout technologique puisqu'il s'agit d'améliorer « *les performances de réponse entre la requête et la base de documents* » [Senié-Demeurisse et Fabre, 2011, p. 211]. La deuxième approche s'est développée dans les années 80, et relève du paradigme cognitif [Paganelli, 2012]. Il s'agit de considérer l'utilisateur comme un élément essentiel dans l'évaluation des systèmes de recherche et d'accès à l'information [Ibid]. La troisième approche commence à prendre plus d'importance depuis les années 90 [Senié-Demeurisse et Fabre, 2011, Paganelli, 2012]. Les travaux de recherche s'attachent à la dimension social et considèrent les usages et les pratiques comme objet de recherche [Chaudiron et Ihadjadène, 2010].

Notre approche de recherche est orientée-usage. Plus précisément, notre attention est dirigée vers les pratiques professionnelles suite à l'introduction du web sémantique en bibliothèques. Dans un premier temps, nous introduisons et discutons les deux notions d'usage et de pratique afin d'apporter plus de clarification sur leur contexte d'utilisation en SIC.

3.1.1 Usage et pratique, quelle utilisation en SIC ?

« Usage » et « pratique » deux notions utilisées dans les études qui prennent en considération la dimension humaine du système d'information. La communauté de recherche en SIC utilise parfois ces deux termes comme étant des synonymes ou bien sans prendre en compte les cadres théoriques divers qui les encadrent. Depuis quelques années, de nombreux chercheurs (Chaudiron, Ihadjadène, Paganelli, Le Coadic, Couzinet, Gardiès, Jouet, Le Marec etc.) se sont intéressés à analyser ces deux notions et à proposer des cadres théoriques permettant de les distinguer.

Le mot usage signifie « *l'action d'user quelque chose, de s'en servir, de l'appliquer pour satisfaire un besoin* »[40]. L'usage d'un outil est donc l'action de s'en servir pour répondre à un besoin spécifique. Le terme en lui même englobe trois éléments : celui (l'usager) qui va se servir de quelque chose (outil, objet, système) et l'action (l'usage) à faire pour satisfaire un besoin. Yves-François Le Coadic définit l'usage comme « *une activité sociale, l'art de faire, la manière de faire. C'est une activité que l'ancienneté ou la fréquence rend normale, courante dans une société donnée mais elle n'a pas force de loi.* » [Le Coadic, 2004, p.19]. L'auteur met l'accent sur l'effet de

40 REY-DEBOVE, J. (2004). *Le Robert Brio: analyse des mots et régularités du lexique* . Robert, Paris.

répétition d'une activité sociale permettant à cette dernière d'atteindre un statut d'usage.

La notion d'usage varie selon les cadres théoriques mobilisés et les contextes d'analyse appliqués. Elle renvoie à plusieurs définitions allant de l' « adoption » à l' « appropriation » en passant par l' « utilisation » [Breton et Preloux, 2012]. Retenue par la sociologie des usages, l'appropriation correspond, en premier lieu, à la maîtrise technique et cognitive de l'usager, ensuite à l'intégration de cette maîtrise aux pratiques quotidiennes pour enfin offrir des possibilités de réinventions et même de participation à la conception d'innovations. Tandis que l'utilisation renvoie à l'action de se mettre en contact avec la technique ; cela concerne le design des interfaces homme-machine. Enfin l' « adoption » est privilégiée par la sociologie de la consommation et de la diffusion [Ibid].

Cependant, les usages ne peuvent pas être pensés indépendamment des interactions entre les individus et les dispositifs [Chaudiron et Ihadjadène, 2010]. Cette relation décrite par ces auteurs par un « contrat de communication » qui fixe des règles indiquant les possibles et les impossibles [Ibid]. L'usage est donc un processus d'interactions complexe (Chaudiron, 2010), un phénomène complexe (Breton, 2012) mettant en relation des acteurs humains et des dispositifs, selon des règles fixant les possibilités et les contraintes liées à cet usage.

Le terme pratique implique une *« activité volontaire visant des résultats concrets »*[41], *« une action habituelle, concrète »*. Cette définition ne renvoie par forcement à l'emploi d'un objet ou d'un système. Autrement dit, elle concerne toute personne effectuant une activité habituelle sans que cela exige l'emploi d'un dispositif ou d'un système.

En SIC, la notion de pratique employée par Josiane Jouet, recouvre à la fois l'utilisation de la technique ainsi que les comportements, les attitudes et les représentations qui se rapportent à l'outil [Jouet, 1993]. Donc la notion de pratique est plus large que celle d'usage qui est limitée à l'emploi de la technique [Ibid]. En d'autre termes, la notion de pratique prend en considération les comportements de ceux qui se servent d'un système ; ces comportements seront les activités volontaires, les attitudes et les réactions observables.

Selon Stéphane Chaudiron et Madjid Ihadjadène, et contrairement à Josiane jouet, il n'y a pas d'imbrication de l'une des notions dans l'autre, au contraire « *il s'agit de deux notions distinctes qui permettent d'appréhender la réalité selon deux angles différents et complémentaires »* [Chaudiron et Ihadjadène, 2010, p.10]. Pour distinguer les deux notions, ces derniers proposent de réserver le terme d'usage pour les travaux qui portent sur les dispositifs et leurs interactions avec les usagers, et celui de pratique pour les approches orientées sur les comportements dans les différentes sphères

41 REY-DEBOVE, J. (2004). *Le Robert Brio: analyse des mots et régularités du lexique* . Robert, Paris.

informationnelles, journalistiques et culturelles [Ibid].

Après avoir abordé les deux notions d'usage et de pratique, il nous semble nécessaire de prendre position par rapport à l'emploi de ces deux notions dans un contexte professionnel, celui des bibliothèques numériques.

3.1.2 Usage et pratique dans un contexte professionnel

Nous partageons l'opinion de Stéphane Chaudiron et Madjid Ihadjadène comme quoi les deux notions impliquent deux angles d'approches différents mais complémentaires. Les auteurs suggèrent d'utiliser le terme « usage » pour désigner la façon dont nous utilisons les dispositifs techniques et le terme « pratique » pour se référer aux études qui sont centrées sur l'humain et ses comportements [Chaudiron et Ihadjadène, 2010].

En quelque sorte, les différentes analyses mobilisées pour étudier l'emploi de ces deux notions nous permettent de dire que le terme « usage » renvoie à une action exercée par une personne lui permettant de se servir d'un système afin de satisfaire un besoin d'informations.

Néanmoins, le terme « pratique » renvoie à des activités habituelles. Autrement dit, pour analyser les pratiques il faut prendre en considération la relation créée entre l'objet et l'utilisateur et ses attitudes qui se manifestent par des activités observables. Les pratiques s'intéressent, en plus de l'utilisation du système, à la façon dont les individus se comportent. Il s'agit d'une approche permettant de réclamer « *les savoirs mobilisables, les compétences immédiates, les habitus, les arts de faire, les désirs d'agir, etc.* » [Ibid, p.10]. Les auteurs rajoutent que le terme « pratique » est habituellement utilisé dans les champs professionnels [Ibid]. Ceci est lié au fait que dans un contexte professionnel, il est plus facile d'identifier et de repérer des savoirs, des compétences et des activités.

Dans notre thèse, nous nous intéressons à l'usager professionnel. Le terme pratique nous semble le plus adapté à notre approche. Dans la section suivante, nous présentons le profil de l'usager professionnel, ensuite nous examinons les pratiques des professionnels de l'information en bibliothèques numériques.

3.1.3 Usager – professionnel : quelle pratique ?

Le mot usager désignait au début du XIV siècle « *la personne qui a un droit réel d'usage* »[42] , et c'est en 1904 que le terme est employé pour désigner « *la personne qui utilise* »[43]. Quelle distinction

42 REY, A. (2005). *Dictionnaire culturel en langue française*. Le Robert, Paris.
43 Ibid

peut-on faire entre les deux termes « usager » et « utilisateur » ? A ce propos, Yves-François Le Coadic considère que l'usager de l'information *« est une personne qui fait en sorte d'obtenir, de la matière information, la satisfaction d'un besoin d'information.»* [Le Coadic, 2007, p. 59]. L'usager est, pour Le Coadic, celui qui se sert de l'objet information ou bien du système d'information pour satisfaire un besoin d'information. Tandis que l'utilisateur est « *un usager en devenir* » parce qu'il s'agit d'exercer une action permettant de se mettre en contact avec l'objet information ou le système d'information (utilisation) ; le but étant de s'en servir pour satisfaire un besoin (usure), ou pour consommer (consommation) [Ibid].

Pour Josiane Senié-Demeurisse et Isabelle Fabre, la distinction entre ces deux termes réside dans le fait que l'utilisateur est une personne qui se sert d'un outil en s'adaptant au mode d'emploi suggéré par le système. Alors que, en citant (Perriault, 1989), l'usager est en décalage avec le mode d'emploi dans le sens où « *il se met de lui même dans sa relation à l'outil.*» Il devient un partenaire des systèmes d'information [Senié-Demeurisse et Fabre, 2011, p. 214]. Dans ce sens, le terme usager englobe à la fois l'action d'utiliser et d'user. L'usager est un utilisateur régulier du système d'information, il est aussi un utilisateur actif dans le sens où son interaction avec le système peut aller jusqu'à apporter des modifications sur son fonctionnement.

Dans le contexte de notre sujet, l'usager professionnel est le professionnel de l'information. Il nous semble important de présenter son profil avant de s'interroger sur ses pratiques.

3.1.3.1 Le professionnel de l'information

La diversité des profils professionnels dans le domaine de l'information a soulevé une difficulté pour trouver un terme qui satisfasse tout le monde [Thiolon, 2003]. D'où l'adoption de l'expression « professionnel de l'information» qui permet de couvrir l'ensemble des professions relatives à la création, au traitement, à la recherche, à la conservation et à la diffusion de l'information.

Selon Jean Michel Salaun, les professions de l'information ont deux racines : les bibliothécaires et les archivistes. La profession de bibliothécaire est la plus vieille permettant de mettre en relation des lecteurs et des livres ; les archivistes quand à eux sont les gardiens des archives de l'institution dont ils dépendent, leur missions consistent à classer, indexer et conserver des documents pour pouvoir les retrouver [Salaun et Arsenault, 2009]. Au début de XXe siècle, un changement de paradigme s'est effectué, il ne s'agit plus de gérer et réserver des collections de livres mais de se préoccuper de l'utilisateur et de ses besoins d'information [Ibid]. A partir des années trente, Suzanne Briet introduit le concept de documentation à l'origine d'une nouvelle profession, les documentalistes. Et c'est depuis quelques années, que l'expression « professionnel de l'information» a commencé à être

utilisée par les chercheurs en SIC. Cela est dû à un mouvement de transition au niveau des écoles et des institutions aux états unies qui vont changer de nom vers l'utilisation du terme « science(s) de l'information » [Ibid].

L'expression « professionnel de l'information » est un terme général permettant de regrouper les autres professions, bibliothécaire, archiviste et documentaliste. Ceux-ci sont tous des professionnels de l'information dans le sens où ils font de l'objet information, quel que soit sa forme et son support, leur activité de travail.

3.1.3.2 Les pratiques et les activités informationnelles dans un contexte professionnel

Dans le cadre de l'accès à l'information, le terme de « pratiques informationnelles » est utilisé pour désigner l'ensemble des dispositifs, sources, outils et compétences intervenant dans les situations de production, d'organisation, de traitement, de recherche, d'usage, de partage et de communication de l'information ; ainsi il englobe les comportements, les représentations et les attitudes informationnelles de l'humain associés à ces situations [Chaudiron et Ihadjadène, 2010, p. 3]. Le terme de pratique implique donc une dimension sociale. Selon Yves Jeanneret, il n'y aurait pas des usages de l'information mais des pratiques informationnelles permettant à des sujets sociaux de se confronter à des dispositifs de médiation produits par d'autres [Jeanneret, 2008].

Les pratiques informationnelles peuvent être définies comme « *une manière concrète d'exercer une activité d'information visant des résultats concrets* » [Gardiès et al., 2010, p. 13]. Dans un contexte professionnel, il s'agit des procédures et des manières d'utiliser l'information (recherche, évaluation, traitement, stockage, diffusion) permettant le déroulement d'une activité [Ibid]. Selon ces auteurs l'approche des pratiques informationnelles « *permet de comprendre les besoins et l'appropriation de l'information mais aussi l'usage de l'information* » [Ibid, p.22]. Cela nous permet de dire que les pratiques informationnelles recouvrent à la fois les usages des dispositifs (information - système) et les besoins et l'appropriation de l'information (information – usager).

Dans ses recherches liées aux usages de l'information dans les organisations, Brigitte Guyot introduit la notion d'activité d'information permettant de « *définir l'information non seulement comme un ressource mais comme un liant social et organisationnel* » [Guyot, 2002, p.2].

Céline Paganelli, quand à elle, utilise la notion d'«activités informationnelles » afin de considérer, d'un coté, les usages et les pratiques et, de l'autre, la production et la conception des dispositifs. Dans un contexte professionnel, les activités informationnelles permettent d'effectuer une tâche ou de résoudre un problème, et à long terme d'enrichir et de mettre à jour les connaissances [Paganelli,

2013, p. 223].

Nous pensons que dans un contexte professionnel les deux notions « pratiques informationnelles » et «activités informationnelles » peuvent désigner la même chose. En effet, les deux notions prennent en considération la dimension sociale de l'usage d'un dispositif qui permet l'accès à l'information nécessaire pour exercer une tâche, résoudre un problème, réaliser une mission etc.

3.1.4 Evolution des pratiques professionnelles de l'information

L'évolution des pratiques professionnelles de l'information s'effectue dans un contexte de transformation de l'environnement documentaire due à l'évolution technologique. Autrement dit, l'introduction d'une nouvelle technologie dans un contexte de travail documentaire amène à s'interroger sur les pratiques professionnelles exercées pour assurer au mieux son usage et en même temps la meilleure façon de se l'approprier.

Avec l'émergence des TIC et à l'heure du numérique, les nouvelles technologies ont transformé et créé de nouveaux services dans les bibliothèques et centres de documentation. Cela a touché évidement les métiers des professionnels de l'information. Leurs pratiques et leurs compétences sont remises en question par l'appropriation des nouvelles technologies, notamment l'informatique et l'Internet. A cet égard, la directrice de l'ENSSIB[44], Anne-Marie Bertrand indique qu' « *il est dès lors indispensable d'appréhender la question de l'évolution des formations des conservateurs et bibliothécaires sous l'angle de ces nouvelles compétences exigées par le numérique.* » [Gimazane et al., 2010, p. 65]. Bertrand a distingué trois types de compétences : l'intégration des contenus dans les collections ; l'acquisition et la numérisation ; la mise en visibilité des collections dématérialisées. Elle souligne qu'il est nécessaire que les professionnels de l'information se positionnent comme des fournisseurs de ressources documentaires et qu 'il est impératif de « *s'interroger sur la valeur ajoutée, réelle et perçue, que peuvent proposer les professionnels.* » [Ibid, p.66].

Le volume 50 numéro 3 de la revue « Documentaliste-Science de l'information » intitulé « les métiers de l'information et la donnée » a été consacré aux changements intervenus au sein des professions de l'information suite aux évolutions technologiques, ainsi qu'à ceux concernant les usages et les pratiques autour des données. Pour Dominique Cotte, il faut une intelligence sociale permettant le traitement des données et des informations, et non pas seulement des traitements automatiques effectués par des algorithmes et des calculs. Cette intelligence sociale correspond bien

44 L'École nationale supérieure des sciences de l'information et des bibliothèques est un établissement public créé en 1992.

au travail des professionnels de l'information exercé depuis des générations pour gérer les documents et leurs contenus [Cotte, 2013].

La question de la gestion des connaissances, notamment le traitement du contenu des documents, est centrale pour étudier les pratiques professionnelles. Laurence Balicco, dans le chapitre intitulé « la gestion des connaissances en contexte professionnel : principes et enjeux » au sein de l'ouvrage « l'information professionnelle », évoque l'évolution de la gestion des connaissances dans un contexte professionnel. Selon Balicco, les activités professionnelles de l'information « *d'abord centrées sur les documents, puis sur leur contenu, ont maintenant évolué pour considérer les personnes (concevant, transmettant ou utilisant les documents), leurs compétences (et leur transmission) afin de les accompagner efficacement dans leur processus d'accès à l'information en environnement professionnel »* [Balicco, 2013, p.119]. Balicco souligne l'évolution des activités professionnelles prenant en compte la dimension sociale du métier. Autrement dit, l'usager est maintenant au centre des préoccupations des professionnels de l'information, il s'agit d'offrir des services lui permettant un meilleur accès à l'information. Elle ajoute que les métiers sont « *plus proches d'un rôle de médiateur des connaissances contenues dans le document*. » [Ibid]. Autrement dit, le professionnel de l'information joue le rôle de médiateur, d'une part, entre le document et sa représentation, et d'autre part, entre le document et l'usager.

Éric de Grolier (1996) considère que le métier de documentaliste est un métier d'intermédiaire permettant de mettre en contact celui qui a besoin du savoir et celui qui sait [Fondin, 2001]. Autrement dit, le professionnel de l'information peut être considéré comme un médiateur entre celui qui produit l'information et celui qui va l'utiliser au travers de l'objet informationnel, le document.

Maintenant que nous avons abordé la question de l'évolution des pratiques professionnelles de l'information, nous examinons dans la section suivante ces pratiques dans le contexte des bibliothèques numériques.

3.1.5 Les pratiques professionnelles en bibliothèques numériques

Dans le contexte des bibliothèques numériques, William Y. Arms considère que le défi est dans la capacité des individus et des organisations à utiliser efficacement la technologie. Il rajoute que les bibliothèques numériques ne peuvent pas évoluer plus vite que se déroule l'appropriation et l'adoption de la technologie par ses usagers [Arms, 2001, p. 18].

Il y a peu d'études sur les pratiques professionnelles en bibliothèques numériques. Certaines s'intéressent aux compétences et aux connaissances des professionnels de l'information. Une étude

faite par Youngok Choi et Edie Rasmussen interroge les professionnels en bibliothèques numériques académiques afin d'identifier des activités, des compétences et des savoirs communs et nécessaires pour « les bibliothécaires numériques ». Leurs résultats ont permis de regrouper les compétences et les connaissances des professionnels de l'information en trois catégories : la technologie, la bibliothéconomie et les autres compétences [Choi et Rasmussen, 2006] (voir tableau 3.1 ci-dessous).

Le tableau 3.1 présente le classement des compétences et des connaissances des professionnels de l'information au sein des bibliothèques numériques.

Technologie	**Bibliothèque**	**Autres**
• Architecture et logiciel de bibliothèque numérique (4.52) • Standards techniques et de qualité (4.33) • Langages de balisage web (4.04) • Développement des bases de données et gestion des systèmes (3.77) • Compétences en conception web (3.71)	• Les besoins des utilisateurs (4.42) • Archivage numérique et préservation (4.52) • Catalogage, métadonnées (4.00) • Indexation (3.71) • Développement de la collection (2,75)	• Communication et compétences interpersonnelles (4,60) • Gestion de projet / leadership (4.56) • Questions juridiques (3,90) • Compétences en écriture de subvention / proposition (3.88) • Enseignement et formation en groupe (3.52)

Tableau 3.1 - Classement des compétences et des connaissances des professionnels de l'information : étude réalisée par [Choi et Rasmussen, 2006]

Selon cette étude, les compétences et les connaissances à mobiliser concernant la technologie implique de maîtriser l'architecture et les logiciels utilisés en bibliothèques numériques, les normes et les standards techniques et de qualité, les langages du balisage sur le web, le développement des bases de données et des systèmes de gestion et la conception web.

Tandis que les compétences et les connaissances liées à la bibliothèque se concentrent autour des besoins des usagers, de l'archivage et la préservation numérique, du catalogage et des métadonnées, de l'indexation et du développement de collections. Les autres compétences sont quant à elles, la communication, la gestion de projet, les questions juridiques, les compétences en écriture et en enseignement.

Les participants à cette étude ont été invités à évaluer leurs compétences et leur connaissances dans

trois domaines (la technologie, la bibliothèque et autre), cela suivant un classement basé sur une échelle de 5 points. Les résultats montrent que les compétences et les connaissances des professionnels de l'information les «mieux classées», sont celles liées, d'une part, à l'architecture de la technique (logiciel, norme, langage), et d'autre part, celles qui sont en lien avec l'usager (les besoins, la communication).

Une autre étude de Pussadee Nonthakumjane intitulée « *Key skills and competencies of a new generation of LIS professionals* » [Nonthacumjane, 2011] , étudie les compétences des professionnels de LIS[45] dans les bibliothèques universitaires en Afrique du Sud. Pour se faire, il a fait un état de l'art des compétences et des connaissances des professionnels de l'information en bibliothèques académiques à l'ère du numérique. Les données sont collectées après une analyse et une classification des compétences exigées dans la revue de la littérature.

Ces compétences sont classées en trois catégories : génériques, professionnelles et personnelles. Le tableau 3.2 ci-dessous montre les compétences professionnelles :

Connaissances spécifique à la discipline	Tennant (1999)	Kwasik (2002)	Lovato-Grassman (2003)	Partridge & Hallam (2004)	Choi & Rasmussen (2006,2009)	Howard (2009)	Nontha cumjane (2010)
1. Métadonnées	✓	✓	✓	✓	✓	✓	✓
2.Développement des bases de données	✓	✓	✓	✓	✓	✓	✓
3.Besoins des usagers	✓	✓	✓	✓	✓	✓	✓
4.Archivage numérique et réservation	✓	✓	✓	✓	✓	✓	✓
5.Développement des collections	✓	✓	✓	✓	✓	✓	✓
6.Système de gestion de contenu	✓	✓	✓	✓	✓	✓	✓

Tableau 3.2 - Connaissances spécifiques à la discipline (LIS) pour la nouvelle génération des professionnels [Nonthacumjane, 2011].

Ce tableau regroupe des connaissances professionnelles liées à la discipline des sciences de l'information et des bibliothèques. Les compétences mises en lumière sont : la gestion et la création de métadonnées, le développement des bases de données, la considération des besoins des usagers, l'archivage numérique, le développement des collections et la gestion des systèmes d'information.

Les métadonnées prennent une place importante dans la gestion des bibliothèques numériques comme nous l'avons mentionné dans le chapitre 2. Cette étude vient renforcer cette idée, et aussi pour montrer que les besoins de l'usager doivent être au centre de préoccupations des professionnels

45 LIS : Library and information science

de l'information.

Selon [Sreenivasulu, 2000], le bibliothécaire numérique, « digital librarian » en anglais, est le professionnel de l'information qui gère une bibliothèque numérique associant des fonctionnalités liées à l'organisation des connaissances, la diffusion de l'information numérique, l'extraction des données et des connaissances à partir des nouveaux entrepôts de connaissances, la numérisation, le stockage numérique, l'accès universel aux connaissances, le catalogage et l'indexation des documents numériques. Il considère que le rôle de « bibliothécaire numérique » ne se résume pas à faciliter l'accès à l'information numérique mais il s'agit de créer et de développer des « canaux de connaissances », des sources de connaissances numériques avec d'autres partenaires permettant de partager et d'enrichir le domaine de connaissance numérique [Ibid, p.18].

Pour résumer, ces différentes études montrent que les pratiques professionnelles en bibliothèques numériques nécessitent des connaissances et des compétences professionnelles, personnelles et générales. Certaines compétences sont liées au développement technique qui accompagne l'infrastructure de la bibliothèque numérique (les logiciels, les langages informatiques, le développement des bases de données et des systèmes de gestion et la conception web). D'autres compétences sont liées aux pratiques des professionnels en bibliothèques numériques considérant les besoins de l'usager parmi leurs priorités. La gestion des métadonnées et le développement des collections numériques, se positionnent comme étant des compétences indispensables en bibliothèques numériques.

En effet, les professionnels de l'information en bibliothèques numériques peuvent être considérés comme étant des médiateurs entre l'usager et les savoirs. D'une part, ils facilitent l'accès à l'information numérique, et d'autre part, ils contribuent à la création d'une valeur ajoutée autour de cette information. Autrement dit, les professionnels de l'information peuvent devenir des acteurs dans la création et la gestion des connaissances numériques, et cela en élaborant des métadonnées permettant à la fois de décrire les ressources numériques et de les lier à l'ensemble des connaissances.

Après avoir examiné dans la première section de ce chapitre les pratiques des professionnels de l'information en bibliothèques numériques, nous présentons dans la section suivante les applications du web sémantique en bibliothèques numériques afin de nous interroger sur les pratiques des professionnels de l'information dans ce contexte.

3.2 Les applications du web sémantique en bibliothèques numériques: repenser les pratiques professionnelles

Selon Yves Jeanneret, il y a trois définitions de la nouveauté : « *la nouveauté technique, celle qui définit les fonctionnalités des dispositifs ; la nouveauté sociale des usages, celle qui fait que de nouveaux acteurs réalisent des actes culturels nouveaux avec le réseau ; enfin la nouveauté médiatique des annonces* » [Jeanneret, 2000, p. 54]. L'effet social d'une nouvelle technologie est aussi important que ses fonctionnalités techniques et impose de se questionner sur l'utilité de cette technologie. Pour sa part, Stéphane Chaudiron indique que « *l'évaluation des technologies et des applications suscite un intérêt nouveau correspondant à l'enjeu de l'acceptabilité sociale de la technique »* [Chaudiron, 2004, p.17].

Étant considéré comme une nouvelle technologie, le web sémantique nécessite des études sur sa dimension technique, c'est-à-dire les outils et les standards utilisés pour représenter le contenu numérique ; sur son impact social, c'est-à-dire les usages et les pratiques des acteurs du web sémantique.

Les technologies du web sémantique concernent les acteurs intéressés par la gestion et la publication de données structurées sur le web. Les bibliothèques ont donc vocation à s'intéresser à l'utilisation de ces technologies. D'une part, le fonctionnement du web sémantique dépend de son appui sur les données structurées. Or, les bibliothèques numériques se basent sur l'identification et la description des données (catalogage, indexation) pour faciliter l'accès à l'information. L'enjeu est donc de permettre aux bibliothèques numériques de s'intégrer dans l'architecture du web. Autrement dit, il s'agit de mettre les données dans le web et non pas sur le web, de *«s'efforcer de faire partie intégrante de cet espace global d'information, et ne pas se contenter d'y positionner un point d'entrée comme l'était le formulaire de recherche des OPAC en ligne ».* [Bermès et al., 2013, p. 63].

Le web sémantique vise à lier les données numériques entre elles en utilisant un langage informatique tel que le triplet RDF. Pour se faire, il faut déterminer des terminologies nécessaires pour construire des données structurées (des graphes conceptuels, des ontologies) [Illien et al., 2013]. Ces terminologies ne peuvent pas être construites tout simplement en utilisant un langage informatique, il faut s'appuyer sur un langage de vocabulaire et sur une expertise de type documentaire. De ce fait, les langages documentaires utilisés par les professionnels de l'information pour décrire les ressources peuvent être utilisés dans le contexte du web sémantique pour assurer une base terminologique sur laquelle s'appuyer afin de construire des graphes conceptuels et ainsi de lier les données entre elles.

Cela nous permet de dire que la bibliothèque peut profiter du web sémantique et ainsi rendre plus accessibles ses données. Réciproquement, le web sémantique a d'une certaine façon « besoin » de l'expertise documentaire afin de lier de façon judicieuse les données.
Nous présentons dans la section suivante les catalogues en bibliothèque numérique, considérés comme étant les moyens par lesquels les utilisateurs accèdent au contenu numérique. Cela nous permet de nous interroger sur les apports du web sémantique pour les bibliothèques numériques.

3.2.1 Les catalogues en bibliothèques numériques : vers une approche orientée usager du web

Le moyen permettant à un usager d'accéder aux contenus d'une bibliothèque est la consultation du catalogue. Le catalogue est l'outil documentaire qui permet de repérer un document dans une bibliothèque. Il permet de recenser le contenu d'une bibliothèque et constitue donc un outil de recherche indispensable pour y accéder. Initialement sous forme de fiches cartonnées imprimées, les premiers OPAC[46] voient le jour avec l'utilisation de l'informatique en bibliothèque au début des années 80 et étaient consultables uniquement en bibliothèques. Depuis les années 90, ces derniers ont une interface public accessible depuis le web.

Selon Emmanuelle Bermès, le catalogue est un outil conçu pour économiser le temps de l'usager passé à chercher une information primaire, « *il doit fournir toutes les informations utiles permettant au lecteur de juger, à partir de la seule notice bibliographique si le livre trouvé répond effectivement à son besoin* » [Bermès et al., 2013, p. 18]. Ceci renvoie donc au travail exercé par le bibliothécaire pour représenter le contenu physique et intellectuel des documents (le catalogage et l'indexation). Le catalogue permet d'enrichir la collection d'une bibliothèque de descriptions fiables et structurées assurant au mieux sa gestion par le professionnel et sa découverte par le public [Illien et al., 2013].

Les catalogues ont subi des évolutions pour pouvoir s'adapter aux besoins des usagers. Les catalogues dits de « nouvelle génération » se sont développés dans les années 2000 pour prendre en compte les usages du web : la recherche fédérée, le classement par pertinence et l'intégration des ressources en ligne [Bermès et al., 2013]. Cela a donc conduit les bibliothécaires à se préoccuper des formats favorisant les échanges, et ce faisant à l'interopérabilité.

46 Online public access catalog

3.2.1.1 Les formats de description des ressources numériques : structuration et échange des métadonnées

Au début, des modèles d'échanges étaient réalisés pour favoriser les échanges des notices bibliographiques entre les bibliothèques, notamment le format MARC[47] en utilisant le protocole Z39.50[48]. Le format MARC se présente comme un format de description et de structuration des notices bibliographiques permettant à la machine de reconnaître « *les éléments constitutifs d'une notice bibliographique et de savoir comment les traiter* »[49]. En 2001, la Bibliothèque du Congrès a développé un ensemble d'outils permettant d'utiliser des données structurées en format MARC dans un environnement XML. « *Ce cadre est destiné à être flexible et extensible pour permettre aux utilisateurs de travailler avec les données MARC de manière spécifique à leurs besoin*s »[50]. Exposer les données MARC en XML permet d'échanger les données bibliographiques selon le protocole OAI-PMH. Originaire du mouvement de l'accès libre (Open access), le protocole OAI-PMH[51] a été utilisé par plusieurs institutions, y compris les bibliothèques, pour échanger sur Internet des métadonnées afin de multiplier l'accès aux documents numériques. Il permet la convergence des données de la recherche (les archives ouvertes) et des données du domaine culturel et des bibliothèques [Bermès et Poupeau, 2012]. L'avantage de l'OAI-PMH est qu'il permet d'envoyer des requêtes au serveur en HTTP qui, à son tour, répond par un flux de données en XML[52]. Il a donc la particularité de s'inscrire dans les technologies du web. Étant un protocole qui permet la communication entre des institutions variées, il était nécessaire de penser à un format de description commun entre ces instituions.

Du format MARC/XML dérive deux autres formats, MODS[53] et METS[54]. Le premier permet de

47 MARC (*machine readable cataloging*, catalogage lisible par machine) est une famille de formats informatiques conçus dès la fin des années 1960 pour les premiers d'entre eux (MARC 1 et MARC 2) dans le but de diffuser et d'échanger des notices bibliographiques. Les formats MARC sont implémentés dans les SIGB et donc les catalogues informatisés. Les formats MARC les plus utilisés aujourd'hui dans le monde sont le MARC21 - maintenu par la *Library of Congress* - et l'UNIMARC - géré par l'IFLA et très largement utilisé par les bibliothèques françaises, et INTERMAC utilisé à la bibliothèque nationale de France. Il existe également des formats MARC pour les autorités, les exemplaires et les classifications. [source: ENSSIB]

48 Z39.50 est un protocole d'interrogation de bases de données bibliographiques, il permet la récupération des données en format MARC

49 Cacaly, Serge et al. (2004). Dictionnaire de l'information. Paris : Armans Colin

50 http://www.loc.gov/standards/marcxml/

51 Open Archive Initiative Protocol for Metadata Harvesting

52 Xtensible Markup Language

53 MODS (Metadata Object Description Schema) est un modèle destiné au traitement de données bibliographiques, en particulier dans le contexte des bibliothèques, mais peut être élargi à d'autres usages. Il est particulièrement intéressant dans le cadre de projets de description de documents numérisés. [Source site de la BNF http://www.bnf.fr/fr/professionnels/f_mods/s.mods_presentation.html]

54 METS, créé à l'initiative de la Digital Library Federation, tente d'enrichir le travail de MOA2 (Making of America II) et de fournir un format XML pour encoder des métadonnées nécessaires à la fois pour la gestion des objets d'une bibliothèque numérique à l'intérieur d'un entrepôt et pour l'échange de tels objets entre différents entrepôts (ou entre des entrepôts et leurs utilisateurs). [Source http://www.loc.gov/standards/mets/METSOverview.v2_fr.html]

renseigner des informations bibliographiques de façon standardisée et de traiter les relations entre les documents originaux et leurs versions numérisées [Moirez et Stuzmann, 2013]. Le second permet de fournir un mécanisme souple pour l'encodage des métadonnées des objets numériques en exprimant les liens entre les différentes métadonnées.

Le Dublin Core[55], format de données minimal, a ainsi été utilisé pour assurer une description à la fois simple, structurée et commune des ressources numériques. L'objectif étant d'améliorer le signalement et la recherche de ressources au-delà des diverses communautés et des nombreux formats descriptifs propres à chaque spécialité, tout en restant suffisamment structuré.

Ensuite, en raison de la diversification des formats de production et pour s'adapter aux usages des différents types d'utilisateurs (bibliothécaire, usagers, éditeurs …), un nouveau modèle a été conçu pour la description des données bibliographiques, il s'agit du modèle FRBR[56]. Ce modèle orienté usager répond aux besoins des utilisateurs qui se résument en quatre tâches : trouver, identifier, choisir et obtenir [Bermès et al., 2013]. L'avantage du modèle FRBR, selon l'IFLA, est double. Concernant les utilisateurs, il permet une recherche plus facile en regroupant les résultats autour de l'œuvre. Tandis que pour les bibliothécaires, il s'agit de profiter de ce modèle pour assurer une meilleure organisation des métadonnées. [Aalberg et al, 2011].

Issues d'une tradition de partage, les bibliothèques sont les pionnières à penser la structuration de l'information afin d'en assurer sa communication avec d'autres institutions. Dans le contexte des bibliothèques numériques, la description des ressources numériques se fait en utilisant les formats que nous avons évoqués, MARC, FRBR, RDA ou Dublin Core ainsi qu'en utilisant les protocoles Z39.50[57] ou le protocole OAI-PMH. Nous parlons donc des métadonnées pour désigner un langage structuré décrivant les données, et suffisamment standardisé pour être partageable par les machines.

Cependant, l'approche orientée usager dans l'environnement du web doit prendre en compte les pratiques des internautes qui ne sont pas bornées par les frontières symboliques des sites des différentes bibliothèques [Bermès et Poupeau, 2012]. Il s'agit d'attirer l'attention des usagers qui cherchent de l'information sur le web et non pas seulement des usagers qui sont déjà sur le site de la bibliothèque.

Aujourd'hui, le problème majeur des catalogues est leur visibilité. Non référencés par les moteurs

55 Le Dublin Core est un format descriptif à la fois simple et générique, comprenant 15 éléments différents, qui a été créé en 1995 à Dublin (Ohio) par OCLC et le NCSA (National Center for Supercomputing Applications).

56 Le modèle FRBR (Functional Requirements for Bibliographic Records / Fonctionnalités requises des notices bibliographiques) est un modèle conceptuel de données bibliographiques élaboré par un groupe d'experts de l'IFLA de 1992 à 1997.

57 Z39.50 est un protocole d'interrogation de bases de données bibliographiques, il permet la récupération des données en format MARC

de recherche, les notices bibliographiques et les contenus des bibliothèques numériques ne sont accessibles que par un passage par le site de la bibliothèque en interrogeant le catalogue. Les notices bibliographiques ne sont pas structurées selon les standards du web mais selon des standards spécifiques à la communauté des bibliothèques. A ce sujet, Gildas Illien et al indiquent que « *la plupart des notices bibliographiques restent enfouies dans les profondeurs du Web invisible* » [Illien et al, 2013, p. 10]. Les références des bibliothèques ont peu de chance d'être consultées puisqu'elles ne sont pas référencées par les moteurs de recherche qui sont considérés comme les points d'entrées pour les usages de navigation sur l'internet [Ibid].

Plusieurs initiatives ont émergé au sein de la communauté des professionnels d'information en bibliothèques pour développer des catalogues prenant en compte les besoins de l'usager et la gestion des métadonnées. Cela a permis, sans doute, d'améliorer la recherche et l'accès à l'information pour les utilisateurs faisant usage du catalogue de la bibliothèque. Cependant, l'internaute évolue dans un monde plus intuitif qu'une salle de référence bibliographique. A la recherche des données et non pas nécessairement d'ouvrages, il passe à coté des outils qui ne lui permet pas un accès rapide et facile à l'information dont il a besoin [Illien et al, 2013].

Face à ce problème, il devient alors judicieux pour les bibliothèques d'adopter l'architecture du web. Plus précisément, il s'agit d'utiliser les trois principes clés du web : le protocole de communication http[58], le système d'identifiants et le principe de l'hypertexte pour relier les ressources [Bermès, Poupeau, 2012]. Le web sémantique semble offrir des solutions pour améliorer la visibilité des données notamment en se basant sur les principes du web et sur la structuration des données. Dans l'idée d'une intégration des bibliothèques numériques vers le web sémantique, il faut donc faire évoluer les formats de diffusion en attribuant aux notices des identifiants pérennes (URI) adaptés aux pratiques des moteurs, et tirer parti de la forte structuration des catalogues en exprimant leurs liens internes (entre notices bibliographiques et notices d'autorité, en particulier) au moyen de RDF [Illien et al, 2013].

3.2.1.2 La représentation des connaissances au sein du web sémantique

Le web sémantique peut être considéré comme une évolution pour les langages documentaires. Il doit être vu comme une entreprise pluridisciplinaire qui englobe des techniques documentaires, de l'intelligence artificielle et de la gestion des connaissances [Menon, 2003]. Il s'inscrit dans la continuité du travail des professionnels de l'information notamment dans l'approfondissement des notions de vocabulaires et de métadonnées [Bachimont et al., 2011].

58 Hypertext Transfer Protocol

L'organisation des connaissances est d'ores et déjà au cœur du métier des professionnels de l'information. Ainsi le changement se fait au niveau des techniques utilisées pour organiser l'information. Dans le cadre du web sémantique, la représentation des connaissances se fait en utilisant des langages informatiques comme RDFS et OWL afin de construire des ontologies. Pour [Aussenac-Gilles et al., 2006] les ontologies sont vues comme une version élaborée et formelle des langages documentaires. Cependant, les concepts et les relations d'une ontologie ne peuvent pas être créés automatiquement, ce qui nécessite le travail de concepteurs d'ontologie. Par ailleurs, les concepts et les propriétés (les attributs et les relations) sont définis par des experts du domaine [Rhee et al., 2008]. Les ontologies se basent ainsi sur une tradition terminologique et documentaire qui consiste à déterminer les notions d'un domaine donné. Cela implique un retour des documentalistes dans la gestion des ontologies au sein du web sémantique.

Le groupe d'incubation W3C Library Linked Data Incubator Group[59] a identifié les ressources utiles pour l'utilisation du web sémantique en bibliothèques. Elles sont classées en trois catégories :

- Les jeux de données (Datasets) : collections de métadonnées structurées selon des modèles de description (comme Dublin Core) en utilisant des éléments de vocabulaire contrôlés (comme Rameau ou LCSH[60]). Il s'agit d'exposer ses modèles de description en utilisant le triplet RDF. (Voir annexe II, pour les jeux de données publiés)
- Les vocabulaires d'autorité (Value vocabularies) : ils permettent de décrire le contenu des ressources comme les listes des vedettes matières de la bibliothèque du Congrès, Rameau , VIAF[61] pour les fichiers autorité auteur, le thesaurus d'art et d'architecture (AAT), Geonames pour définir des locations géographiques, la classification décimale Dewey, la classification décimale universelle. (Voir annexe III, pour les vocabulaires d'autorité publiés)
- Les éléments de description des métadonnées (metadata element set): ils permettent de définir des attributs et des classes pour décrire les entités en utilisant le modèle de graphe RDFS ou les ontologies OWL, tels que les éléments du Dublin Core, FRBR, RDA, MARC 21 et FOAF. (Voir annexe IV, pour les éléments de description publiés).

Pour représenter les connaissances dans le web sémantique, le w3c a conçu un modèle de données SKOS[62] permettant de développer des standards pour soutenir les systèmes d'organisations des

59 Un groupe d'incubation du W3C s'est consacré aux bibliothèques et au web de données (Library Linked Data), de mai 2010 à août 2011, visant à améliorer l'interopérabilité globale des données de bibliothèques sur le web grâce au web sémantique et au web de données.
60 Library of Congress Subject Headings
61 Virtual Internationl Authority File
62 Simple knowledge Organization System (2009)

connaissances comme les thésaurus, les classifications, les listes de vedettes matière, etc. En s'appuyant sur le triplet RDF, le modèle de données SKOS assure le partage et la liaison entre les systèmes d'organisation des connaissances sur le Web. Ce modèle a la même structure que plusieurs systèmes d'organisations de connaissances utilisés en bibliothèques (Thésaurus, listes de vedettes matières).

La figure 3.1 ci-dessous représente une notice « Economic cooperation » issue du thesaurus représentée en SKOS.

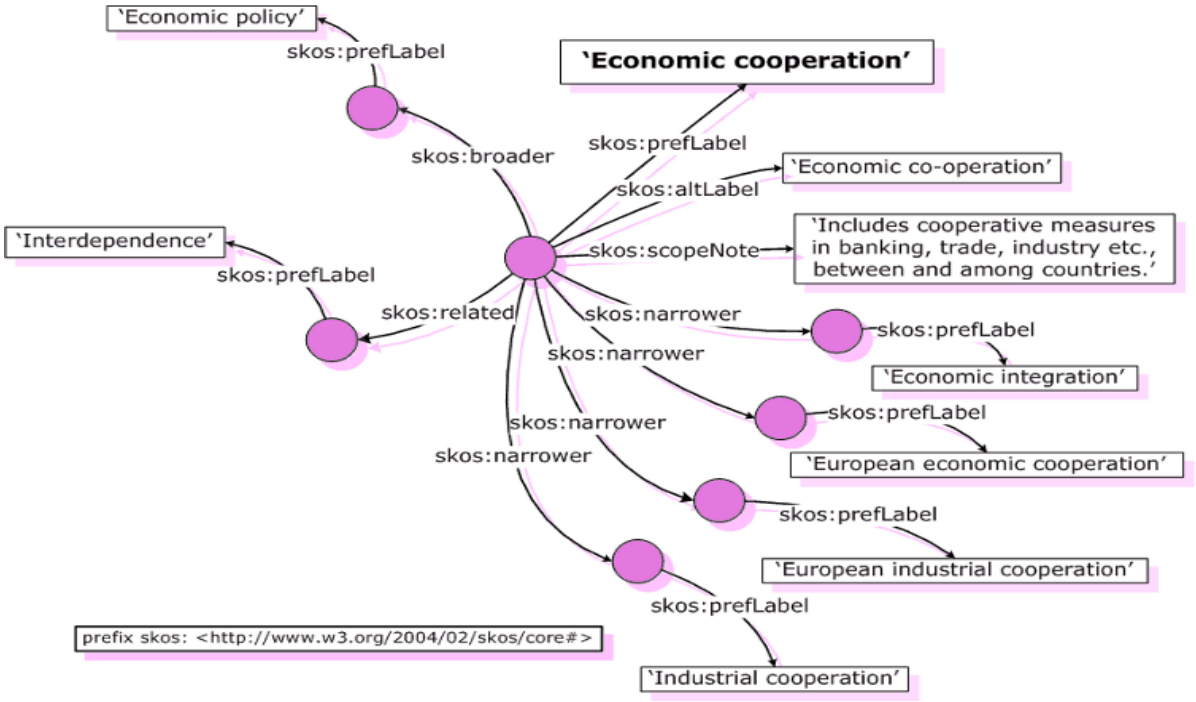


Figure 3.1 : Représentation en SKOS d'une notice issue d'un thesaurus

Utilisé en bibliothèque numérique, le modèle de données SKOS permet d'établir des relations qui peuvent être utilisées comme des passerelles entre des systèmes utilisant des ontologies différentes et provenant de contextes applicatifs ou linguistiques variés. Cela favorise donc les échanges et le partage des données entre plusieurs institutions. Ainsi, les liens entre les documents et leur sujet sont laissés aux modèles de description de ressources documentaires comme Dublin Core [Ibid]. Pour décrire des personnes physiques ou morales et leur relations à des documents, certaines bibliothèques utilisent un modèle de graphe RDF appelé FOAF[63], qui fut destiné à l'origine pour décrire des personnes vivantes dans le cadre du web social. FOAF est un modèle de description indépendant de tout organisme de normalisation « *ce qui lui assure une souplesse et une évolutivité plus grandes* » [Bermès et al., 2013].

63 Friend Of A Friend

3.2.1.3 La recherche d'information au sein du web sémantique

La recherche d'information est historiquement liée aux sciences de l'information et à la bibliothéconomie qui visent à représenter des documents dans le but d'en récupérer des informations. C'est un domaine ancien qui ne cesse d'évoluer dans le but de « *rationaliser le processus complexe permettant l'identification, au sein de volumes de plus en plus important d'informations, de celles qui sont potentiellement intéressantes pour l'utilisateur »* [Boughanem et Savoy, 2008]. Selon Boughanem et Savoy, la tâche principale d'un système de recherche d'information est de sélectionner des documents pertinents susceptible de répondre aux besoins en information de l'utilisateur. Selon Jacques Maniez et Widad Moustafa El-Hadi, le but de tout système informatique documentaire est de raccourcir le chemin entre l'usager et l'information. Ils ajoutent que toute technologie répondant à cet objectif s'inscrit dans le processus de traitement de l'information [Maniez et El-Hadi, 1999].

L'interopérabilité des métadonnées et des langages documentaires est considérée comme un des problèmes les plus exigeants auxquels il faut réfléchir, notamment dans le domaine de la recherche d'information en bibliothèques numériques. Selon Emmanuelle Bermes, le web sémantique propose une forme d'interopérabilité qui repose sur *«la création d'un espace global d'information, utilisant les liens pour permettre de naviguer de manière transparente d'une ressource à l'autre* » [Bermès, 2011, p. 5]. De ce fait, le web sémantique, en proposant une interopérabilité basée sur les standards du web et sur des liens entre les ressources, permet à plusieurs systèmes de communiquer entre eux, d'échanger et de partager l'accès à l'information sans difficulté. L'enjeu n'est pas seulement de proposer une interopérabilité technique mais aussi sémantique. L'objectif de l'interopérabilité sémantique est d'assurer une compréhension commune de la signification des données et des services échangés [Heiler, 1995]. Cette compréhension commune est possible grâce aux graphes conceptuels, « les ontologies », qui établissent des liens explicites entre les différents concepts provenant des différentes sources et systèmes d'information. Selon Boughanem et Savoy, la recherche d'information utilisant le web sémantique peut « *dépasser les limites de recherches basées sur des requêtes formulées par des mots-clé*s » et cela en s'appuyant sur des « concepts » qui prennent en compte le sens exprimé derrière les mots [Boughanem et Savoy, 2008, p. 237]. Pour expliciter le sens des données au sein du web sémantique Bruno Bachimont considère qu'il s'agit d'une réécriture dans les termes de la logique et cela se fait grâce aux langages formalisés RDF et OWL [Bachimont et al., 2011].

Offrant une interopérabilité technique et sémantique, la recherche d'information utilisant le langage de requête SPARQL permet d'interroger les bases des triples RDF. L'intérêt de la recherche en

SPARQL, selon Emmanuelle Bermès, « *réside dans la possibilité d'enchaîner des triplets dont certains éléments ne sont pas connus* » [Bermès et al., 2013, p.144]. Autrement dit, ce parcours permet de découvrir des éléments éloignés qui, au début de la recherche, étaient inconnus. Cela permet aux usagers, d'une part, de découvrir des nouvelles informations, et d'autre part, de se décharger d'une partie de leur tâche d'exploitation et de recherche. Cette tâche devient plus difficile en navigant dans un grand espace de ressources : l'utilisateur risque de perdre ses repères et d'être désorienté. Dans cette situation, l'interopérabilité proposée par le web sémantique permet l'accès à l'information et sa découverte éventuelle.

Pour Petar Ristoski et Heiko Paulheim, les informations du web sémantique peuvent être utilisées de manière bénéfique dans le processus de découverte d'informations, en particulier en utilisant des données liées et ouvertes (linked open data). Cependant, leur étude a montré que l'exploration de données et la découverte de connaissances ne permettent pas encore d'exploiter tout le potentiel du Web sémantique. [Ristoski et Paulheil, 2016].

Dans ce contexte, Thomas Baker indique que les bibliothèques ont un rôle central à jouer dans la gestion des données liées, notamment dans l'ouverture de leurs données. Ceci peut enrichir la description des ressources documentaires avec des données fournies par d'autres communautés favorisant la découverte de l'information [Baker et al., 2011].

3.2.2 Le web sémantique en bibliothèques numériques, quel(s) usage(s) ?

Le rapport final de la mission du groupe d'incubation du W3C "Bibliothèques et web de données"[64] indique que l'utilisation des standards du Web sémantique en bibliothèques :

- rend les données bibliographiques plus visibles et réutilisables ;
- favorise le développement de fonctionnalités multilingues ;
- réduit la redondance existant sur le web au sein des descriptions bibliographiques grâce à l'identification claire d'entités clés qui seront partagées sur le web de données ;
- favorise la recherche interdisciplinaire car il crée de la connaissance au travers des liens établis entre de multiples bases de connaissance spécifiques aux disciplines. [Baker et al., 2011].

Une étude sur les usages du web de données en bibliothèques numériques [Hallo et al., 2016] a été réalisée afin de repérer les vocabulaires et les ontologies, et de souligner les avantages et les

64 Le rapport montre comment utiliser les standards du web sémantique pour rendre les données des bibliothèques plus visibles et réutilisables sur le web.

problèmes qui accompagnent ces usages au sein de cinq bibliothèques (la bibliothèque nationale de France, Europeana, la bibliothèque de Congrès, la bibliothèque nationale d'Espagne et la bibliothèque nationale de Royaume-Uni). (Voir tableau 3.3 ci-dessous).

Leurs résultats montrent que les principaux avantages de l'utilisation du web de données en bibliothèques numériques sont : l'amélioration de la visibilité, la possibilité d'établir des liens avec d'autres services en ligne, l'amélioration de la récupération des données ouvertes, l'interopérabilité, la possibilité d'interroger des métadonnées de plusieurs institutions et la transformation facile des autorités en SKOS.

Tandis que les inconvénients auxquels sont confrontés les bibliothèques sont : la difficulté de catalogage (plusieurs vocabulaires pour la même métadonnée), le désaccord pour fournir des données, la difficulté de transférer les données dans de nouveaux modèles, la nécessité de développer des outils pour la transition vers le web de données, le manque d'expertise, le peu d'applications utilisant le web de données, les problèmes de cartographie, les besoins de liens utiles pour les jeux de données et le contrôle de la qualité de jeux de données.

Le Tableau 3.3 présente une comparaison des fonctionnalités liées au web de données dans cinq bibliothèques nationales :

Bibliothèque	Avantages	Vocabulaires et ontologies	Problèmes	Future
Bibliothèque nationale de France	- Améliorer la visibilité des données - Liens vers d'autres services en ligne - Publication de sujets dans SKOS - Récupération des données ouvertes	FRBR, SKOS, InterMarc, XMLEAD, Dublin Core, RDF	- Difficulté de cataloguer les ressources	- Participation communautaire au catalogage et au contrôle de la qualité
Europeana	- Permet l'interopérabilité sans affecter les modèles de données - Fournit des requêtes à plusieurs métadonnées liées provenant des institutions européennes	RDF, OAI_ORE, SKOS, FOAF, Dublin Core	- Le manque d'accords pour fournir des données, la difficulté de la migration des données à EDM	- Une grande implication de la communauté avec de nouveaux jeux de données, en précisant les liens, en améliorant la qualité des données - Améliorer la visibilité en intégrant des initiatives telles que Schema.org
Bibliothèque de Congrès	- Interconnexion à d'autres sources de données - Modèle de personnes, lieux et événements	BIBO, BIO, Dublin Core, ISBD, ORG, SKOS, RDF SCHEMA, OWL, FOAF, WGS84, RDA	- Besoin de développer des outils pour transformer les métadonnées en données liées - Manque d'experts dans différents domaines pour les transformations	- L'utilisateur peut vérifier la qualité des sources d'information - Il est nécessaire de développer des navigateurs sur BIBFRAME - Amélioration du Marc21 aux outils de migration BIBFRAME
Bibliothèque britannique	- Données liées à d'autres sources de données - Des points d'intérêt liés aux ressources bibliographiques, telles que les personnes, les lieux, les événements, les thèmes - Amélioration de la visibilité	BIBO, BIO, Dublin Core, ISBD, ORG SKOS, RDF Schema, OWL, FOAF, Ontologı´a de eventos WGS84	- Manque d'applications compatibles avec le web de données	- Identifier les besoins concernant les liens - Améliorer les outils de liaison de données - Obtenir des commentaires sur l'usage des données
Bibliothèque nationale de l'Espagne	- Liens avec d'autres sources de données. Liste d'autorités sujet en SKOS - NSL scholar avec annotations et recommandations	FRBR, FRAD, FRSAD, IFLA, ISBD, RDF, DC, RDA, MADS/ RDF, Dublin Core	- Problèmes de cartographie: toutes les relations de base de FRBR ne pouvaient pas être extraites des enregistrements MARC	- Développer des outils de visualisation - Affiner la cartographie - Obtenir de nouveaux liens - Obtenir de nouvelles sources de données

Tableau 3.3 - Comparaison des fonctionnalités liées au web de données dans cinq bibliothèques nationales [Hallo etal, 2016]

3.2.3 Comparaison entre les bibliothèques numériques et le web sémantique

Les similitudes entre la représentation des connaissances au sein des bibliothèques et le web sémantique sont évidentes dans la mesure où les deux permettent une représentation conceptuelle des ressources numériques. En effet, la frontière entre l'emploi des normes de représentation dans ces deux environnements (les bibliothèques et le Web sémantique) est « artificielle » [Greenberg, 2007]. Alors que ce qui fait défaut dans le contexte du web sémantique sont les principes et les objectifs d'utilisation des schémas de métadonnées et du système ontologique. Greenberg s'interroge sur le choix de tels schémas et sur le niveau de représentation nécessaire pour décrire correctement l'information afin qu'un agent puisse la manipuler avec succès et fournir un service utile [Ibid].

Selon l'auteur, la bibliothèque et le web sémantique ont beaucoup de points communs :

- Ils se sont développés, en partie, en réponse à une abondance d'informations ;
- Ils ont des missions fondées sur : les services, l'accès à l'information et la découverte de connaissances ;
- Ils s'appuient sur des normes et standards internationales pour avancer ;
- Ils sont nés et se sont développés dans un esprit de partage et de collaboration [Ibid, p.207].

Pour la communauté des bibliothèques, il y a une similitude entre les intérêts des bibliothèques (construction de vocabulaires, description des propriétés des ressources, identification des ressources, échange et agrégation des métadonnées), et les intérêts à l'origine du développement du web sémantique [Heery, 2004]. La principale différence entre les deux approches est que les métadonnées produites dans les bibliothèques sont faites pour être comprises par les humains tandis que celles produites dans le contexte du web sémantique sont destinées à la fois à l'homme et à la machine.

Le tableau suivant présente un comparatif entre la bibliothèque numérique et le web sémantique (voir tableau 3.4 ci-dessous)

	Bibliothèques numérique	**Web sémantique**
Services	• accès depuis les catalogues des bibliothèques • accès aux documents numériques • recherche d'information • destinés aux utilisateurs	• accès aux données depuis le web en utilisant les moteurs de recherche • accès aux données numériques • recherche et découverte des données

	humains	• destinés aux humains et aux machines
Technologie	• systèmes de gestion des bases de données • langages documentaires (Rameau, LCSH etc.) • normes et standards (IFLA, ISO, OCLC etc.)	• langages informatiques (XML, RDF) • ontologies (OWL) • normes et standards compatibles avec le web (W3C)
Acteurs	• professionnels de l'information • informaticiens • éditeurs	• informaticiens • ingénieurs • développeurs • professionnels de l'information
Données	• données libres • données structurées	• données libres • données structurées
Pratiques	• identifier les documents • décrire les documents (Indexation, catalogage) • partager, collaborer, et réutiliser des métadonnées entre bibliothèques	• identifier les données (URI) • lier les données (RDF) • développer des ontologies (RDFS, OWL) • partager et réutiliser des métadonnées sur le web

Tableau 3.4 - Comparaison entre la bibliothèque numérique et le web sémantique

Maintenant que nous avons examiné l'évolution des catalogues en bibliothèques numériques en mettant en lumière certaines contributions du web sémantique dans ce domaine, nous nous interrogeons dans la section suivante le rôle du professionnel de l'information dans le contexte du web sémantique.

3.2.4 Le rôle du professionnel de l'information dans le contexte du web sémantique

Toute innovation technique engendre des modifications au sein des métiers en rapport avec cette innovation. Les métiers des professionnels de l'information se trouvent confrontés de manière régulière à des changements liés à l'introduction d'une nouvelle technique dans leur pratiques professionnelles, comme c'est le cas récemment avec le web sémantique .

Sylvie Dalbin considère que le travail avec et dans le web de données nécessite de se placer dans des nouvelles perspectives favorisant : l'ouverture des données, l'utilisation d'une technique plus importante que l'informatique documentaire classique et la participation à l'élaboration des supports d'information [Dalbin et al., 2011]. Selon Dalbin, le web sémantique ne supprime pas les

fondements des métiers de la gestion de l'information qui sont le fait de structurer, normaliser et qualifier l'information. Au contraire, il vient remettre en avant scène ces compétences « *que les évolutions du Web tout court (et les illusions informatiques en général) avaient un temps semblé reléguer à l'arrière-plan* » [Ibid, p. 58]. Pour Dominique Cotte, les données ne peuvent pas être traitées uniquement de façon automatique par les calculs et les algorithmes, il faut une intelligence sociale qui correspond aux pratiques des professionnels de l'information [Cotte, 2013]. Ainsi pour Bruno Menon, « *les professionnels de l'information s'attendent à être parties prenantes dans la mise en œuvre du web sémantique* » [Menon, 2003, p. 387] . Selon lui, les professionnels de l'information, de par leur expérience de description des contenus, peuvent devenir des acteurs dans l'infrastructure du web sémantique. De plus, leurs connaissances en termes de besoins d'information et de stratégies de recherche leur permettent d'envisager des nouvelles fonctionnalités aux outils du web sémantique [Menon, 2003].

Gildas Illien s'interroge sur l'avenir du catalogueur en tant que gestionnaire de données. Selon lui, le catalogage prend une place de plus en plus importante dans le traitement de l'information numérique au sein du web sémantique, en raison de la nécessité d'utiliser des données structurées [Illien et al, 2013]. D'un côté, le web sémantique offre le langage informatique nécessaire pour identifier et lier les données numériques. Tandis que, les professionnels de l'information assurent la détermination des terminologies nécessaires pour construire les graphes conceptuels.

La différence entre la science de l'information et l'informatique est que cette dernière étudie les algorithmes et les calculs traitant l'information, alors que la science de l'information étudie la nature même de l'information et ses usages par les humains. Ces deux champs ne sont pas en compétition mais au contraire sont complémentaires [Saracevic,1999]. C'est pourquoi, il est nécessaire que les professionnels de l'information et les informaticiens travaillent ensemble pour assurer la réussite du web sémantique et son usage adéquat dans la société.

Pour résumer, les professionnels de l'information tentent de relever les défis portés par les récents bouleversements technologiques, liés principalement à l'essor du numérique, ainsi que ceux provenant de l'apparition de nouveaux usages et pratiques autour de l'information. Cela nécessite de s'adapter à de nouveaux outils et de transmettre le savoir faire documentaire pour assurer un bon fonctionnement de la technique dans la société. Le web sémantique offre de nombreux avantages visant à valoriser le travail documentaire, il apparaît ainsi comme une opportunité afin de prendre une place fondamentale dans l'organisation des connaissances.

3.2.5 Les défis du web sémantique

Le web sémantique et la communauté des bibliothèques ne cohabitent pas affirme Jane

Greenberg. D'une part, les membres de la communauté du Web sémantique ne sont pas pleinement conscients des compétences et des connaissances dont disposent les bibliothécaires, alors même que ceux-ci peuvent contribuer significativement à l'avancement du Web sémantique. De l'autre, les bibliothécaires peinent à adopter les technologies du web sémantique en comparaison des informaticiens et ingénieurs. Cela pose problème dans un monde où la technologie avance très rapidement [Greenberg, 2007]. Le manque de conscience et de sensibilisation des bibliothécaires constitue un obstacle majeure pour le développement du web sémantique dans les bibliothèques [Goddard et Byrne, 2010 ; Gonzales, 2014].

De son coté, Emmanuelle Bermès indique que la création d'un espace informationnel global permettant de relier les données des bibliothèques n'est pas sans difficultés. D'une part, la conversion des données pour permettre leur publication dans le web de données questionne les formats existants. Les données stockées en masse dans les bibliothèques seront compliqués à exposer selon les modèles de triplet RDF, et de plus, certaines ne sont pas compatibles avec ces modèles. D'autre part, un autre obstacle apparaît concernant l'attribution des identifiants URI pérennes et compatibles avec les protocoles du web. Il s'agit d'un problème lié à la maintenance et à la pérennité des ces identifiants [Bermès, 2011].

Les résultats d'une étude de cas norvégienne sur les défis organisationnels de l'usage du web sémantique en bibliothèques numériques (le cas de la bibliothèque nationale de Norvège) indique que:

- sur le plan stratégique, l'impact du web sémantique dans les bibliothèques numériques est modéré. « *Cette conclusion est basée sur le fait que les choix technologiques stratégiques ont déjà été faits lorsque la bibliothèque a décidé de passer d'une bibliothèque traditionnelle à une bibliothèque numérique* ». Cependant, les problèmes se posent du côté des métadonnées en raison de l'absence d'un cadre de développement d'ingénierie ontologique et d'un outil d'annotation de métadonnées sémantiques capables de réduire le coût de l'annotation des métadonnées.
- au niveau organisationnel, l'impact de la technologie Web sémantique sera élevé, car le processus d'ingénierie ontologique et la production de métadonnées sémantiques affecteront l'ensemble de l'organisation de la bibliothèque.
- sur le plan technologique, l'impact sera relativement faible parce que la technologie nécessaire pour soutenir la bibliothèque numérique traditionnelle est déjà en place et le passage à la technologie web sémantique n'est pas compliqué [Bygstad et al., 2009].

Cette étude de cas montre que les contraintes majeures de mise en œuvre du web sémantique en bibliothèques numériques reposent sur les niveaux stratégique et organisationnel plutôt que sur les aspects technologiques. Cette étude de cas ne peut pas être généralisée parce qu'il s'agit d'un cas particulier lié à la bibliothèque nationale de Norvège. Néanmoins, elle peut être comparée avec des études similaires portant sur l'usage du web sémantique en bibliothèques nationales (la bibliothèque nationale de France par exemple).

Un autre défi est lié à la protection de la vie privée. Plus les données sont liées, plus il est possible que des informations personnelles soient exposées. Les bibliothécaires, avec leur longue tradition de protection de la vie privée, devront jouer un rôle actif dans le développement de données liées afin de garantir la protection des droits [Goddard et Byrne, 2010]. Ainsi, la gestion des droits pose des problèmes pour les données liées dans les bibliothèques. Un mélange de contenu autorisé et gratuit dans un environnement de données liées serait extrêmement difficile à gérer [Ibid].

Ross Singer considère que l'obstacle majeur à l'adoption des données liées par les bibliothèques serait l'absence de confiance [Singer, 2009]. Les bibliothécaires dans leurs pratiques professionnelles s'appuient sur des langages contrôlés pour décrire les ressources. Autrement dit, ils font confiance aux catalogueurs et à ceux qui développent les langages documentaires, ce qui pourrait ne pas être le cas pour les données provenant d'autres acteurs de l'information. Pour cette raison, l'auteur indique que les bibliothécaires doivent avoir l'esprit ouvert et profiter des ontologies existantes qui répondent à leurs besoins même si elles sont faites par d'autres communautés [Ibid].

Pour résumer, les principaux défis et obstacles liés à l'adoption du web sémantique en bibliothèques numériques sont les suivants:

- méconnaissance de la part de la communauté du web sémantique des compétences et des connaissances des bibliothécaires
- manque de sensibilisation des bibliothécaires envers les technologies du web sémantique
- conversion des données vers le web sémantique
- attribution des identifiants pérennes
- confiance en métadonnées produites par d'autres acteurs de l'information
- gestion des droits
- protection de la vie privée
- production et gestion des métadonnées sémantiques

Bien que certains des défis soient importants, il est intéressant de noter que la majorité des obstacles sont de nature non technique. La technologie est prête, il s'agit maintenant de préparer les bibliothèques et les bibliothécaires à l'intégrer et à la mettre en place. [Goddard et Byrne, 2010]

Conclusion du chapitre

Nous avons essayé dans ce chapitre d'analyser les usages du web sémantique en bibliothèques numériques en nous intéressant aux pratiques professionnelles des professionnels de l'information. Comme le disait Suzanne Briet, les métiers de professionnel de l'information dépendent de la technique, « *le documentaliste est principalement un technicien* » et « *les techniques qu'il utilise dépendent, pour leur maintenance et leur renouvellement, des acquisitions d'une connaissance fondamentale*» [Meyriat, 1994 cité par Couzinet, 2001]. Le cœur de métier des professionnels de l'information reste toujours le même. Ce qui change c'est la technique utilisée pour pratiquer leurs connaissances et leurs savoirs faire.

L'introduction du web sémantique en bibliothèques numériques doit être vue comme une nouvelle technique permettant aux professionnels de l'information de pratiquer leurs connaissances et leurs savoirs faire, et surtout de jouer un rôle important dans le développement de cette technique. Le web sémantique permet l'identification et la représentation des données afin d'en faciliter l'accès et le partage entre humains et machines. Autrement dit, le web sémantique offre la technique nécessaire pour identifier et relier les données sur le web. Tandis que, la représentation des connaissances ne se fait pas seulement en utilisant les langages informatiques du web sémantique, il faut une intervention humaine permettant de fournir à ces langages les données et les modèles conceptuels nécessaires pour représenter les connaissances. Étant experts dans le domaine de l'organisation des connaissances, les professionnels de l'information peuvent jouer un rôle important dans le web sémantique. D'une part, ils peuvent se servir du web sémantique pour proposer de nouveaux services plus adaptés aux usagers du web en rendant les bibliothèques plus visibles sur le web. Et d'autre part, ils peuvent devenir des acteurs important dans le développement du web sémantique, et ceci en introduisant leur expertise de type documentaire permettant la gestion des métadonnées.

Conclusion

Cette première partie, dédiée à l'état de l'art de la question des usages du web sémantique en bibliothèques numériques, nous a permis d'examiner les concepts et les notions traitant ce sujet.

Dans le premier chapitre, l'analyse de différentes définitions et théories du web sémantique, nous a permis de le considérer comme une nouvelle approche de la représentation des connaissances. Le web sémantique offre le cadre technique permettant d'identifier, de structurer, de décrire et de lier les données en assurant leur compréhension par l'homme et la machine. Ceci permet de favoriser l'interopérabilité et l'échange des données sur le web.

Dans le deuxième chapitre, nous avons examiné la notion de bibliothèque numérique. Celle-ci peut être considérée comme un dispositif info-communicationnel permettant d'établir un lien entre un usager et une information par le biais d'une double médiation complémentaire ; celle qui sert à identifier et à représenter l'information , la médiation documentaire ; et celle qui permet d'offrir un accès aux savoirs, la médiation des savoirs.

Dans le contexte des bibliothèques numériques, les métadonnées jouent un rôle d'intermédiaire, d'une part, entre le document et le professionnel de l'information, et d'autre part, entre le document et l'usager. Elles permettent à la fois d'identifier et de décrire les données, et en même temps de servir de clés d'accès à ces dernières.

Nous avons vu dans le dernier chapitre de cet état de l'art que le web sémantique peut être utile pour les bibliothèques numériques, notamment pour accroître leur visibilité sur le web. Il constitue le domaine technique permettant aux professionnels de l'information d'exercer leurs pratiques documentaires et de les faire évoluer pour répondre aux besoins des usagers du web. Ainsi, les professionnels de l'information peuvent jouer un rôle important dans le développement du web sémantique. Un travail intellectuel de nature documentaire nous semble nécessaire pour définir les relations entre les données dans les graphes conceptuels qui permettent la représentation des connaissances au sein du web sémantique.

En effet, il s'agit dans cette thèse de repenser la médiation documentaire dans le contexte de l'usage du web sémantique en bibliothèques numériques. Pour cela, il convient d'analyser les pratiques documentaires des professionnels de l'information et de comprendre les enjeux et les intérêts à relier les données des bibliothèques numériques avec d'autres données sur le web.

Partie II: Méthodologie de recherche et analyse des résultats

Introduction

Notre démarche se divise en trois étapes. Premièrement, nous avons analysé trois recensements des bibliothèques numériques en France. Cette étape nous permet de constituer une sélection des bibliothèques sur laquelle nous appuyer pour réaliser notre enquête via un questionnaire et un entretien.

Ensuite, nous nous intéressons aux usages et aux applications du web sémantique en bibliothèques numériques. Il s'agit, tout d'abord, de faire une étude quantitative sous la forme d'un questionnaire envoyé aux bibliothèques sélectionnées. Cela permet de récolter des données sur l'usage et le non usage du web sémantique dans ces bibliothèques. Cette étude est suivie d'une enquête qualitative sous forme d'entretiens semi-directifs avec des professionnels de l'information. Le but général de ces entretiens est d'étudier les pratiques des professionnels de l'information suite à l'introduction du web sémantique en bibliothèques numériques.

Cette partie est divisée en deux chapitres (chapitre 4 et chapitre 5). Dans le chapitre 4, nous décrivons les méthodes de recherche que nous avons mobilisées dans notre enquête de terrain. Il s'agit de décrire chaque méthode : la sélection, le questionnaire et les entretiens.

Dans le chapitre 5, nous présentons les résultats de notre enquête. Notre sélection est constituée de 98 bibliothèques numériques. Trente trois bibliothèques ont répondu à notre questionnaire et dix ont participé aux entretiens. Dans cette étude, nous testons nos hypothèses de recherche. Nous effectuons une analyse thématique afin d'examiner le discours des enquêtés qui nous informe sur leurs représentations, leurs expériences et leurs attentes de l'usage du web sémantique en bibliothèques numériques.

Chapitre 4

Enquête de terrain : méthodes et déroulements

Ce chapitre a pour objectif de présenter et de décrire les méthodes de recherche mobilisées dans le cadre de cette thèse.

« *Par définition une enquête de terrain est limitée, étroitement circonscrite dans le temps et dans l'espace* » affirment Stéphane Beau et Florence Weber [Beau et Weber, 2010, p. 22]. D'une part, il faut que l'enquête soit réalisable en pratique, et d'autre part, elle doit répondre à une problématique provisoire et à des questions préalables [Ibid]. Guidée par une question de départ, celle des usages et des applications du web sémantique en bibliothèques numériques, le terrain de notre recherche est celui des bibliothèques numériques. Plusieurs institutions s'intéressent à ce domaine : états, universités, laboratoires de recherche, entreprises, associations, etc. De même, les acteurs qui interviennent dans ce domaine viennent d'horizons très différents : professionnels du patrimoine, responsables politiques et administratifs, professionnels de l'information, informaticiens, industriels, amateurs, etc.. Face à cette réalité mouvante, l'intérêt de limiter le sujet à un domaine géographique et à des acteurs bien définis nous paraît indispensable.

Le choix de la France comme zone géographique est motivé par notre intérêt d'explorer ce terrain. A notre connaissance et d'après l'état de l'art, aucune enquête sur les usages et les applications du web sémantique n'a encore été réalisée sur un terrain semblable. Cela permet, d'une part, d'enrichir la recherche dans ce domaine, et d'autre part, d'aboutir à une analyse critique sur le développement de ces applications dans les bibliothèques numériques en France.

Ce chapitre s'articule en trois sections. Dans la première nous présentons la construction de notre sélection, celle-ci est basée sur une analyse de trois recensements des bibliothèques numériques en France.

Ensuite, dans la deuxième section, nous présentons le déroulement de notre enquête quantitative sous la forme d'un questionnaire envoyé aux bibliothèques sélectionnées. Le questionnaire a pour objectif de s'interroger sur les usages et les applications du web sémantique en bibliothèques numériques.

Enfin, nous présentons notre enquête qualitative sous la forme d'entretiens semi-directifs. L'analyse de ces derniers doit nous permettre de mettre en évidence les représentations qu'ont les professionnels de l'information du web sémantique et de son usage en bibliothèques.

4.1 La sélection des bibliothèques numériques

4.1.1 Objectif de recherche

Notre objectif de recherche est d'étudier les usages et les applications du web sémantique en bibliothèques numériques en France. Dans un premier temps, il s'agit de construire notre sélection, celle-ci est basée sur une analyse de trois recensements des bibliothèques numériques en France.

Il s'agit d'une sélection non-exhaustive des bibliothèques numériques en France. L'exhaustivité en cette recherche n'est ni poursuivie, ni souhaitée, cela parce que le flux d'information concernant ces projets est mouvant et rapidement obsolète. La période de recherche est limitée entre le 1er Juin et le 15 Juillet 2016 .

4.1.2 Les critères de sélection des bibliothèques numériques

Les critères sur lesquels nous nous appuyons pour sélectionner les bibliothèques numériques, relèvent de l'ensemble des caractéristiques que nous avons pu dégager à partir d'une analyse des différentes définitions de bibliothèque numérique (voir chapitre 2). Celles-ci considèrent que la bibliothèque numérique :

- gère des collections numériques (documents numérisés et numériques)
- permet l'accès à distance aux documents
- nécessite la technologie pour lier et mettre les ressources et les services en ligne
- contient des informations de qualité
- exige une expertise de type documentaire
- met à disposition une information organisée pour en faciliter l'accès
- assure la conservation des informations sur le long terme.

Les critères retenus pour sélectionner les bibliothèques numériques sont :

- Un accès à distance aux ressources numériques
- Un traitement documentaire (information organisée, indexée , classée, etc.)
- Des outils de recherche et/ ou de navigation
- Une information de qualité

4.1.3 Les bibliothèques qui n'entrent pas dans le cadre notre recherche

Les portails qui renvoient à des bibliothèques numériques ont été exclus du recensement. Ces portails ne gèrent pas des collections numériques mais ils proposent des liens pour y accéder. Ils peuvent être considérés comme des bibliothèques virtuelles qui proposent un accès aux bibliothèques numériques (voir chapitre 2 section 2.1.1).

Un deuxième critère pris en compte est celui lié au domaine scolaire. D'une part, ce domaine est assez large et en lien étroit avec le domaine de l'éducation. Les recherches en SIC traitant ce sujet se concentrent essentiellement autour des études portant sur les professeurs documentalistes et la maîtrise de l'information. D'autre part, les recensements qui sont faits sur les bibliothèques numériques en France ne prenaient pas en compte ce domaine. C'est pourquoi les bibliothèques scolaires ont été exclues du recensement du fait qu'elles ne relèvent pas directement du cadre de notre recherche.

Enfin, nous avons exclu les «bibliothèques numériques internationales». Le numérique a facilité la coopération et l'échange de données et de métadonnées entre plusieurs institutions (bibliothèques, musées, etc). D'où l'apparition des bibliothèques numériques qui sont gérées par plusieurs institutions mettant en ligne des documents numériques provenant de plusieurs pays et en langues différentes. Il n'existe pas de définition de ce qu'est une bibliothèque numérique internationale mais dans les signets de la bibliothèque nationale de France, nous trouvons une catégorie intitulée « bibliothèques numériques internationales » recouvre quatre bibliothèques : la bibliothèque européenne en ligne Europeana, la bibliothèque numérique mondiale, le réseau francophone numérique RFN et les programmes internationaux de numérisation partagée.

Une question se pose alors : qu'est-ce qu'une « bibliothèque internationale » dans le contexte du numérique ? Est-ce que le qualificatif «international» concerne le contenu, les institutions, les usagers ?

Le numérique a permis d'ouvrir les portes des bibliothèques vers le monde entier. Une bibliothèque numérique, quel que soit son type, est accessible via le web à n'importe qui et n'importe où. Donc, une des propriétés d'une telle bibliothèque est son accès universel. Cependant, nous avons remarqué que ces bibliothèques gèrent des collections numériques provenant de plusieurs institutions situés dans des pays différents.

Ces bibliothèques ont été exclues de l'enquête parce qu'elles dépassent notre terrain d'étude celui de la France. Elles constituent en elles mêmes un terrain de recherche intéressant à explorer dans des études ultérieures.

4.1.4 Les outils de recensement

Il n'existe aucun recensement exhaustif des bibliothèques numériques en France. La nature mouvante des sites de ces bibliothèques rend en effet difficile leur identification. Plus précisément, l'instabilité des sites web des bibliothèques numériques (changement d'adresse URL, maintenance, mise à jour, etc.) implique un suivi régulier pour mettre à jour le recensement.

Nous nous sommes appuyés sur des recensements déjà réalisés par la bibliothèque nationale de France et l'unité régionale de formation à l'information scientifique et technique (URFIST) de Lyon. Le choix de ces outils est justifié par le fait que les recensements ont été produits par des professionnels de l'information (la communauté de pratique). Ensuite, nous avons choisi l'inventaire multilingue du patrimoine numérique MICHAEL[65] pour compléter les bibliothèques numériques. Nous avons utilisé le catalogue permettant de recenser les collections numérisées en France. Ce dernier, qui depuis 2004 bénéficie du soutien de la Commission européenne dans le cadre du projet MICHAEL, a pour vocation de valoriser les collections numérisées de France et à favoriser l'accès aux ressources numérisées[66].

1- Recensement réalisé par la bibliothèque nationale de France :[67]

Le site recense exclusivement les bibliothèques numériques généralistes ou pluridisciplinaires. Les bibliothèques numériques sont classées par ordre géographique : Répertoires et bibliothèques numériques internationales, Outils et projets, France, Europe (hors France), Amérique du nord, Amérique latine, Asie, Afrique et Autres parties du monde (Australie, Nouvelle-Zélande, îles du Pacifique et de l'Atlantique, etc.). Chaque bibliothèque est décrite par un court résumé et quelques métadonnées (Pays, langue, thème(s), type, auteur, liens).

Notre terrain de recherche étant limité par une zone géographique celle de la France, nous avons donc sélectionné les bibliothèques recensées sous la catégorie « France ». Ensuite, chaque site a été parcouru pour vérifier s'il fonctionne bien et s'il répond à nos critères de recensement.

Résultats :

La bibliothèque nationale de France recense 20 bibliothèques numériques dans la zone géographique « France ». Après avoir parcouru chaque site, nous en avons rejeté quatre parce qu'ils pointaient vers des pages inaccessibles. Les 16 autres bibliothèques répondant à nos critères ont été retenues. Les résultats figurent dans un tableau de l'annexe V.

65 Multilingual Inventory of Cultural Heritage in Europe
66 http://www.numerique.culture.fr/pub-fr/a_propos.html
67 http://signets.bnf.fr/html/categories/c_011textes_num.html

2- Recensement réalisé par Urfist de Lyon[68]

Le site recense des fonds numérisés provenant de différents organismes. Ces fonds sont communément appelés "bibliothèques numériques" ("digital librairies"). « *Les bibliothèques numériques présentées ici sont des collections organisées de documents électroniques en accès libre et généralement gratuit sur Internet, associées à une interface permettant la recherche et la consultation de ces documents*. »[69].

Les bibliothèques sont classées par quatre accès différents :

- Fonds d'établissements publics et privés
- Œuvres originales de scientifiques et d'écrivains
- Numérisation du patrimoine
- Archives ouvertes et libre accès

Ce recensement est fait par un organisme universitaire qui a pour mission principale de « *développer la connaissance et la maîtrise de l'information scientifique numérique dans l'enseignement supérieur et la recherche* » [Ibid]. Cependant, les informations risquent d'être anciennes parce que la dernière mise à jour a été réalisée en décembre 2009.

C'est pourquoi une actualisation a été nécessaire pour examiner l'état des sites recensés par l'URFIST. Ensuite, une analyse a été réalisée pour sélectionner les bibliothèques répondant à nos critères de sélection. Les résultats figurent dans le tableau ci-dessous. Une présentation détaillée est affichée dans l'annexe VI.

Résultats :

Bibliothèques recensées	77
Bibliothèque rejetées	52
Bibliothèques retenues	**25**

L'URFIST de Lyon recense 77 bibliothèques numériques classées par thème et non par zone géographique. Après avoir parcouru et analysé chaque site, nous avons rejeté 52 bibliothèques et retenu 25. 19 bibliothèques ne répondent pas aux critères de sélection : des archives, des portails, des sites qui proposent des collections numérisées sans aucun outil de recherche. Huit bibliothèques

68 URFIST de Lyon (2009). [En ligne] http://urfist.univ-lyon1.fr/fonds-numerises-570805.kjsp?RH=1216648187047. Consulté le 19/03/2014

69 Ibid

ont été déjà recensées par la BnF (des doublons). Dix sites sont inaccessibles. Onze bibliothèques ne se trouvent pas dans la zone géographique étudiée, celle de la France. Quatre bibliothèques sont des bibliothèques internationales (la bibliothèque numérique mondiale, Europeana, Gutenberg et Biodiversity heritage library).

3- Inventaire multilingue du patrimoine numérique MICHAEL :[70]

Le portail MICHAEL propose un accès aux collections numériques d'Europe. Les collections décrites rassemblent des contenus numérisés à partir d'objets tels que des photographies, des documents d'archive, des cartes, des objets, des monuments archéologiques. Pour chaque collection numérisée, nous accédons notamment aux informations suivantes :

- son titre,
- sa description,
- le(s) sujet(s) abordé(s,
- la/les période(s) et la/les civilisation(s) concernée(s),
- la/les zone(s) géographique(s) concernée(s),
- des informations concernant la/les institution(s) chargée(s) de sa conservation.

Nous avons choisi le catalogue qui recense les institutions à l'origine des projets de numérisation en France. Ce site propose un classement par type d'institutions :

Archives (50) ; Archives nationales (7) ; Archives départementales (99) ; Archives municipales (206) ; Bibliothèques (37) ; Bibliothèques municipales (141) ; Bibliothèques universitaires (19) ; Centres de documentation (87) ; Centres et laboratoires de recherche (33) ; Conservations des, SDAP (19) ; Filmothèques (30) ; Institutions d'art contemporain (FNAC, FRAC, CAC) (15) ; Musées (181) ; Photothèques (16) ; Services de l'archéologie (15) ; Services de l'inventaire (29).

Résultats :

Étant intéressés par les bibliothèques nous avons donc sélectionné les trois catégories : bibliothèques (37), bibliothèques municipales (141) et bibliothèques universitaires (19). Chaque site a été parcouru et analysé pour vérifier s'il répond à nos critères de sélection des bibliothèques numériques. Une représentation détaillée des résultats figure dans un tableau de l'annexe VII.

70 http://www.michael-culture.org/about.html

Bibliothèques recensées	197
Bibliothèques rejetées	122
Bibliothèques répondant à nos critères de sélection	75
Doublons BNF et MICHAEL	2
Doublons Urfist et MICHAEL	10
Triplons BNF, Urfist et Michael	6
Bibliothèques retenues	75– 18 = 57

Résultats des trois recensements

Les résultats des trois recensements figurent dans l'annexe VIII. Le tableau ci-dessous indique le nombre de bibliothèques recensées, rejetées et retenues.

Bibliothèques recensées	294
Bibliothèques rejetées	178
Doublons	18
Bibliothèques retenues	98

Au final, notre sélection est construite de 98 bibliothèques numériques, soit 33,33 % des bibliothèques recensées par les trois outils étudiés. Cette sélection constitue le corpus sur lequel nous nous appuyons pour mener l'enquête quantitative.

Dans la section suivante nous présentons l'enquête quantitative visant à répertorier et identifier les usages et les applications du web sémantique en bibliothèques numériques en France.

4.2 Enquête quantitative - le questionnaire : les usages du web sémantique en bibliothèques numériques en France

Réalisée suivant une démarche inductive, cette enquête quantitative s'articule en quatre étapes.

La première consiste à délimiter un objet d'étude et à choisir une méthode d'investigation. Notre objet d'étude est le web sémantique en bibliothèques numériques. Nous avons choisi le questionnaire comme technique d'investigation.

La deuxième porte sur la collecte des données. Il s'agit tout d'abord de définir notre terrain d'étude et de construire notre questionnaire. Nous nous sommes appuyés sur la sélection que nous avons réalisée pour définir notre terrain d'étude. Ensuite, le questionnaire a été envoyé aux bibliothèques cibles par le biais de la messagerie électronique.

La troisième consiste à traiter les résultats, tandis que la dernière correspond à l'interprétation des résultats afin d'expliquer et de comprendre les différents usages du web sémantique en bibliothèques numériques.

4.2.1 Le questionnaire

Le questionnaire relève de la démarche quantitative permettant de recueillir des données exploitables à partir des analyses statistiques et des analyses de contenus. Les données portent sur trois catégories : des faits (données factuelles), des jugements subjectifs et des cognitions [Javeau, 1992]. Les faits relèvent du comportement des individus enquêtés. Alors que les jugements subjectifs se manifestent par leurs opinions, leurs attitudes, leurs motivations et leurs attentes. Enfin, les cognitions portent sur leur niveau de connaissances [Ibid].

Un questionnaire est envisagé lorsqu'il s'agit de montrer des phénomènes sociaux « *à partir de la mise en évidence de régularités statistiques* » [Guibert et Jumel, 1997, p. 105]. Il s'agit d'établir un lien entre une pratique étudiée et un milieu social [Ibid]. Notre enquête porte sur les usages du web sémantique dans un milieu professionnel, celui des bibliothèques numériques.

4.2.2 L'objectif de l'enquête

L'enquête porte sur les usages et les applications du web sémantique en bibliothèques numériques en France. L'objectif est de repérer les bibliothèques numériques qui exploitent le web sémantique. Nous cherchons à comprendre les raisons et les obstacles qui accompagnent la mise en place du web sémantique dans ces bibliothèques.

4.2.3 La sélection

En s'appuyant sur la sélection déjà réalisée (voir section 4.1), le questionnaire a été envoyé à la totalité des bibliothèques retenues (98 bibliothèques numériques).

4.2.4 Les moyens matériels

Le questionnaire a été envoyé aux bibliothèques cibles par le biais de la messagerie

électronique ou directement via le site de la bibliothèque qui propose une rubrique « contact ». Le questionnaire a été envoyé pour la première fois le 11 août 2016. Étant en période de vacances d'été, le taux de réponse était très faible. Pour cette raison, une relance a été réalisée le 12 septembre 2016, à la rentrée des vacances, tenant compte des adresses erronées ou hors d'usage. Du fait d'un temps de réponse important, deux relances ont été effectuées auprès des bibliothèques numériques qui ne s'étaient pas manifestées, cela jusqu'au 21 novembre 2016.

4.2.5 Descriptif du questionnaire

Le questionnaire débute par une introduction précisant le cadre de l'étude, les objectifs, l'anonymat, le temps de passation, les remerciements et les coordonnées de l'enquêteur. Ce questionnaire est articulé autour de cinq volets : (Le questionnaire figure dans l'annexe XIII)

1. Le profil des bibliothécaires / bibliothèques
2. La connaissance des outils du web sémantique par les bibliothécaires
3. Les applications du web sémantique en bibliothèques
4. Les pratiques documentaires
5. Les avantages et les obstacles à l'utilisation du web sémantique

La première partie porte sur le profil des professionnels de l'information et des bibliothèques. Elle permet de collecter des données factuelles sur la nature de leur fonction ainsi que des informations sur le statut, le site web de la bibliothèque et le courrier électronique permettant de les contacter.

La deuxième partie vise à s'interroger sur les connaissances des professionnels de l'information au sujet du web sémantique. Plus précisément, nous nous intéressons aux degrés de compréhension des professionnels de l'information de certaines expressions liées au web sémantique : « le web sémantique », « le web de données », « RDF », « OWL » et « SKOS ». Il s'agit de questions fermées dont les réponses ont été fixées à l'avance en utilisant une échelle graduée ou « échelle de Likert » : la personne interrogée exprime son degré d'accord ou de désaccord à la question [Le roy et Pierrette, 2012]. L'échelle contient en générale de quatre à plusieurs choix de réponse permettant de nuancer le degré d'accord. Notre choix s'est porté sur une échelle de quatre réponses en évitant une position intermédiaire :

Pas du tout 1 2 3 4 Tout à fait

Cela parce que la position médiane est souvent le refuge des non réponses [Ibid]. Autrement dit, nous avons évité d'avoir des réponses intermédiaires résultant d'une position qui veut dire souvent « je n'ai pas compris » ou « je ne sais pas ». Cependant cette méthode oblige la personne à se positionner favorablement ou défavorablement. Les avantages des questions fermées résident dans la facilité à les analyser et à y répondre. Néanmoins, elles sont réductrices et elles risquent d'aiguiller vers une réponse conforme à l'attente de réalisateurs de l'enquête et non pas aux pensées de la personne interrogée [Javeau, 1992]. C'est pourquoi une étape de « pré-test » a été indispensable pour valider notre questionnaire. (Nous aborderons ce point dans la section suivante).

La troisième partie porte sur les applications du web sémantique en bibliothèques numériques. Dans un premier temps, nous nous intéressons à repérer les bibliothèques qui utilisent le web sémantique. Nous avons utilisé une question fermée à choix binaire (Oui / Non). Ensuite, cette question a été complétée par deux questions ouvertes permettant de savoir les raisons pour lesquelles les bibliothèques utilisent ou non le web sémantique. Le choix des questions ouvertes est motivé, d'une part, par notre incertitude à pouvoir proposer l'intégralité des réponses possibles, et d'autre part, par l'intérêt à obtenir des informations plus qualitatives sur ce sujet. La personne interrogée est libre de s'exprimer comme elle le veut en utilisant ses propres mots. Cela permet donc d'obtenir des informations plus détaillées et des idées auxquelles nous n'avons pas pensé. Cependant, ce genre de questions est plus compliqué à coder et à analyser.

Enfin, nous nous sommes intéressées à identifier les applications (outils, langages, standards) du web sémantique utilisées dans les bibliothèques. Une question semi-ouverte a été utilisée afin de recueillir le plus d'information possible sur ces applications. Une liste de réponses a été prévue (URI, RDF, OWL, JSON, SPARQL et autres) en permettant en même temps aux personnes interrogées la possibilité d'ajouter des informations qui n'étaient pas recensées dans cette liste.

La quatrième partie a pour vocation de s'interroger sur les pratiques documentaires des professionnels de l'information. Une question fermée à choix binaire (Oui/ Non) suivie de deux questions ouvertes ont été employées. Le but étant de comprendre si les pratiques documentaires ont été touchées ou non par le web sémantique. Si non, pourquoi, et si oui, comment ces pratiques ont été influencées. Une enquête sous forme d'entretiens sera présentée dans la section qui suit permettant de mieux comprendre les changements intervenus en terme de pratiques documentaires des professionnels d'informations en bibliothèques numériques.

La cinquième et dernière partie porte sur les avantages et les inconvénients du web sémantique en bibliothèques numériques. Des questions semi-ouvertes ont été mobilisées permettant aux personnes interrogées de choisir dans une liste de réponses avec la possibilité d'ajouter d'autres

informations. Pour la question qui porte sur les avantages du web sémantique en bibliothèques numériques, nous avons proposé les réponses suivantes : faciliter l'échange des notices bibliographiques entre plusieurs institutions ; faciliter l'interopérabilité entre les données ; faciliter la publication en ligne ; valoriser le fonds numérique ; améliorer la visibilité des collections numériques ; faciliter la recherche d'information ; trouver des résultats plus pertinents ; je ne sais pas et autres. Pour la question portant sur les obstacles de l'usage du web sémantique en bibliothèques numériques, nous avons proposé les réponses suivantes : la complexité technique ; la diversité des standards ; le manque d'expertise des professionnels d'information ; les contraintes juridiques ; je ne sais pas et autres. Nous avons également proposé la possibilité de choisir plusieurs réponses. Ensuite, une question fermée à choix binaire (Oui/Non), suivie de deux questions ouvertes portant sur l'importance du web sémantique pour le futur des bibliothèques, a été posée afin de connaître les opinions des professionnels de l'information sur ce sujet.

Enfin une question portant sur la volonté de suivre une formation permettant de maîtriser les standards du web sémantique en bibliothèque a été posée afin de percevoir l'intérêt des professionnels de l'information par le web sémantique.

4.2.6 Le pré-test : mise à l'œuvre du projet de questionnaire

Cette étape est indispensable pour vérifier que le questionnaire est suffisamment clair et cohérent. Il s'agit de le tester auprès des personnes ayant les mêmes caractéristiques de la population de l'enquête [Javeau, 1992]. Ce questionnaire a été testé auprès de quatre professionnels de l'information afin de savoir si les termes utilisés sont compréhensibles et si l'ordre des questions et la forme utilisés permettent de bien recueillir les informations souhaitées. Des modifications ont été réalisées suite aux retours et remarques des professionnels permettant d'améliorer la qualité de questionnaire.

4.2.7 La rédaction et la réalisation matérielle de l'enquête

La rédaction définitive du questionnaire a été faite en tenant compte des enseignements du pré-test. Nous avons utilisé « Google form » pour créer ce questionnaire permettant aux personnes interrogées de le remplir en ligne. Le principale avantage de cette technique est qu'elle permet d'interroger un nombre important de personnes à un moindre coût. Mais un des inconvénients principaux est le faible taux de réponse [Le Roy, Pierrette, 2012]. Ayant besoin d'interroger un nombre important de personnes réparties sur toute la France, le questionnaire en ligne nous paraît toutefois le plus adapté à nos besoins.

Nous avons donc envoyé un courrier électronique aux bibliothèques cibles. Un texte a été rédigé permettant d'exposer clairement et brièvement à l'organisme qui répond à l'enquête, les objectifs de l'enquête et la garantie de l'anonymat et le remerciement. Le courrier contenait le texte et le lien pointant vers le questionnaire est affiché dans l'annexe IX. 33 bibliothèques sur 98 ont répondu au questionnaire, soit 33,67 % des bibliothèques sélectionnées.

4.2.8 L'analyse des données

Une analyse qualitative a été mobilisée afin d'analyser les données du questionnaire. Une grille d'analyse primaire a été réalisée à partir de la grille du questionnaire. Celle-ci a pour objectif d'identifier les thèmes principaux, les sous thèmes du questionnaire et les objectifs souhaités de chaque thème (Voir tableau 4.1 ci-dessous). Ensuite, une analyse thématique a été réalisée pour les questions ouvertes. Cette méthode consiste à repérer des thèmes principaux et des thèmes secondaires. Les premiers permettent d'exprimer le contenu du segment de texte analysé, et les seconds servent à spécifier les premiers dans ses différents aspects [Ghiglione et Matalon, 1985]. Les grilles d'analyses thématiques seront abordées dans le chapitre suivant.

Le tableau 4.1 présente la grille d'analyse du questionnaire :

Thèmes	Sous-thèmes	Objectifs
Profil des professionnels de l'information	• Statut	Catégorisation
Profil de la bibliothèque	• Statut	Catégorisation
Web sémantique	• Connaissances générales • Usage du web sémantique • Apports du web sémantique • Obstacles à l'usage du web sémantique	Analyse thématique et vérification des hypothèses
Pratiques documentaires	• Influence du web sémantique • Formations • Avenir des bibliothèques	Analyse thématique et vérification des hypothèses

Tableau 4.1 – Grille d'analyse du questionnaire

L'enquête quantitative par le biais d'un questionnaire permet de recueillir des données facilement exploitables. Cependant cette technique ne permet pas de traiter un sujet d'une manière profonde à partir de réponses plus détaillées. De plus, elle ne permet pas d'être en contact direct avec la

personne interrogée donnant la possibilité d'échanger avec celle-ci. Ce questionnaire nous a donc permis de recueillir des données sur les bibliothèques qui utilisent ou non le web sémantique, les standards, les avantages et les obstacles accompagnant cette utilisation. Cependant, nous souhaitons traiter ce sujet de manière plus approfondie en abordant la question des pratiques documentaires. C'est pourquoi, une enquête qualitative par le biais d'entretiens nous a parus pertinente afin de répondre à nos besoins d'information. Nous abordons dans la section suivante le déroulement de cette enquête.

4.3 Enquête qualitative – Entretiens semi-directifs : les pratiques documentaires dans le contexte du web sémantique

Une méthode qualitative est un arrangement de techniques de recueil et de techniques de traitement permettant de résoudre une problématique de recherche [Mucchielli, 1996]. La spécificité fondamentale de cette méthode vient de son inscription dans le paradigme compréhensif considérant les phénomènes humains comme des phénomènes de sens [Ibid]. Une des techniques utilisée dans les méthodes qualitatives est l'entretien. Cette dernière a pour objectif de recueillir de l'information auprès de la personne interviewée. Il s'agit pour l'interlocuteur de bien réussir son entretien et pour l'enquêteur de se faire comprendre [Guibert et Jumel, 1997].

4.3.1 Objectif de recherche et hypothèses

Notre précédente enquête nous a permis de fournir des éléments essentiels à la compréhension de notre objet de recherche. Une enquête plus profonde sous forme d'entretien était nécessaire afin de se mettre en relation avec les professionnels de l'information pour mieux comprendre leur points de vue sur l'utilisation du web sémantique en bibliothèques numériques.

Cette enquête va nous permettre de nous interroger sur les pratiques documentaires des professionnels de l'information en bibliothèques numériques avec pour objectif de tester nos hypothèses.

Les hypothèses posées s'ancrent dans les sciences de l'information et de la communication. Elles s'inscrivent dans une approche info-communicationnelle visant à présenter les bibliothèques numériques comme des dispositif info-communicationnels et le web sémantique comme une approche de la représentation des connaissances. Cette approche permet d'étudier les usages sous l'angle de la médiation. Une médiation documentaire qui réintroduit des pratiques professionnelles, et une médiation des savoirs qui s'inscrit dans une politique d'ouverture des données sur le web.

L'hypothèse principale sur laquelle se base notre recherche est que l'utilisation du web sémantique en bibliothèques numériques nécessite une médiation documentaire. Une donnée ne peut pas être traitée seulement de façon automatique, c'est-à-dire uniquement par l'utilisation des langages normalisés du web sémantique (XML, RDF...). La mise en contexte des données nécessite une intelligence sociale (Cotte, 2013). Autrement dit dans le cadre du web sémantique, l'enrichissement des données par des métadonnées repose sur un traitement documentaire dont l'objectif est d'associer les données à des contextes d'usage précis et qu'un algorithme seul ne peut réaliser.

H1

Les professionnels de l'information peuvent jouer un rôle dans la mise en œuvre et l'évolution du web sémantique en introduisant leur expertise documentaire relative à la gestion des métadonnées.

H2

La représentation des connaissances ne se fait pas seulement en utilisant les langages informatiques et standardisés du web sémantique, il faut une intervention humaine permettant de fournir à ces langages les données et les modèles conceptuels nécessaires pour représenter les connaissances.

H3

Les bibliothèques numériques peuvent se servir du web sémantique pour proposer de nouveaux services plus adaptés aux usagers du web et pour rendre les bibliothèques plus visibles.

4.3.2 Méthode : entretiens semi-directifs

L'entretien semi-directif se situe dans un niveau intermédiaire entre l'attitude directive qui donne priorité à la liberté d'expression de l'interlocuteur et l'attitude non directive qui vise à obtenir des réponses à une série des questions bien définies dans un ordre précis [Guibert et Jumel, 1997]. Cette technique nécessite la préparation d'une grille de thèmes servant de guide d'entretien.

Notre choix de mener des entretiens semi-directifs est motivé par notre intérêt à orienter l'entretien vers des thèmes jugés prioritaires pour l'étude en laissant en même temps une certaine autonomie aux interlocuteurs. Cela nous permet d'obtenir, d'une part, des réponses donnant la possibilité de tester nos hypothèses, et d'autre part, des nouvelles informations auxquelles nous n'avions pas pensé avant.

4.3.3 Les participants

Notre terrain de recherche a été déjà délimité aux bibliothèques numériques sélectionnées. Dans l'étape précédente, nous avons envoyé un questionnaire aux bibliothèques cibles. Les bibliothèques qui ont répondu au questionnaire constituent notre terrain de recherche concerné par l'enquête qualitative.

La population de cette enquête est constituée des professionnels de l'information travaillant pour les bibliothèques numériques qui ont répondu à notre questionnaire (33 bibliothèques sur 98, soit 33,67 % des bibliothèques sélectionnées) (Voir tableau 4.2 ci-dessous).

Le tableau 4.2 présente les bibliothèques numériques qui ont répondu au questionnaire :

Bibliothèque nationale	**Bibliothèque de lecture publique**	**Bibliothèque universitaire**	**Bibliothèque spécialisée**
• Bibliothèque nationale de France • Bibliothèque nationale et universitaire de Strasbourg • Bibliothèque publique d'étude	• CIRDOC – Médiathèque occitane • Médiathèque intercommunale André Malraux -Lisieux • Médiathèque Astrolabe (Numel) • Direction des bibliothèques (Bibliothèque Vendée) • Bibliothèques municipales de Grenoble • Bibliothèques municipales de Chalons-en-Champagne • BMVR Louis Nucéra • Bibliothèque municipale de Dijon • Bibliothèque de Rennes Métropole • Bibliothèque humaniste de	• Univerisité Jean Monnet • BIU santé • Bibliothèque de l'école des Mines de Paris • Centre d'études supérieures de la renaissances (Univesité Tours) • Paris 1 Pantheon Sorbonne • Bibliothèque Clermont université • Bibliothèque Diderot de Lyon • SCD- Université de Poitiers • SCD – Université Toulouse 3 • Université Lille (Nordnum) • Centre de recherche bretonne et celtique • Université Lille Sciences et technologies • SCD – Université	• Textesrares • Institut national d'histoire de l'art • Mathdoc • CNDSMDP, médiathèque Hector Berlioz

	Sélestat • Médiathèque Georges-Delaw (Sedan) • Médiathèque départementale du Puy-de-Dome. • Médiathèques Valence Romans Agglomération	Nice Sophia Antipolis	

Le tableau 4.2 - Les bibliothèques numériques qui ont répondu au questionnaire

4.3.4 L'échantillon de l'enquête

Nous avons envisagé de réaliser une dizaine d'entretiens. Dune part, la question concernant le nombre d'entretiens à réaliser n'est pas pertinente car elle met dans le même plan des entretiens qui ont des statuts différents [Beau et Weber, 2010]. Autrement dit, un entretien approfondi avec un professionnel ciblé n'est pas comparable à un entretien réalisé avec une personne quelconque. Stéphane Beau et Florence Weber affirment que « *les entretiens approfondis ne visent pas à produire des données et n'ont donc pas besoin d'être nombreux* » [Ibid, p. 156]. D'autre part, nous pensons qu'une dizaine d'entretiens est un échantillon suffisant pour représenter la population enquêtée (10 bibliothèques sur 33, soit 33,3 % des bibliothèques qui ont répondu au questionnaire).

Dans un premier temps, nous avons accordé une attention particulière aux bibliothèques qui utilisent le web sémantique. C'est pourquoi, toutes ces bibliothèques ont été sollicitées pour participer à notre enquête, sauf une qui a répondu assez tard au questionnaire. Par manque de temps, nous n'avons pas pu contacter cette bibliothèque afin de réaliser l'entretien. Huit bibliothèques ont été contactées pour participer à l'enquête : six bibliothèques ont accepté (deux bibliothèques nationales (trois entretiens) et 4 bibliothèques universitaires) ; une bibliothèque s'est excusée et une n'a pas répondu.

Ensuite, le non usage du web sémantique en bibliothèques numériques nous semble aussi intéressant surtout pour comprendre les raisons pour lesquelles ces bibliothèques ne s'y 'intéressent pas. Afin de compléter la dizaine d'entretiens, nous avons donc contacté les bibliothèques qui n'utilisent pas le web sémantique. Nous avons choisi au hasard trois bibliothèques de type différent (une bibliothèque spécialisée, une bibliothèque publique et une bibliothèque universitaire). Étant donné que le taux de réponses n'est pas de 100 %, il a été nécessaire de contacter au final 19

bibliothèques[71]. Certains professionnels de l'information se sont excusés parce qu'ils pensaient ne pas être compétents pour réaliser l'entretien, d'autres n'ont pas répondu malgré les relances.

Les bibliothèques cibles ont été contactées via la messagerie électronique. Un texte a été rédigé remerciant la personne ayant répondu au questionnaire en lui demandant de participer à un entretien dans le cadre de ce travail de recherche (le message envoyé aux bibliothèques figure dans l'annexe X). Dans ce message nous avons clairement présenté les objectifs de l'entretien, l'organisation et les modalités de réalisation (face à face, par téléphone et par Skype).

Finalement, nous avons réalisé 10 entretiens : 7 entretiens avec des bibliothèques qui utilisent le web sémantique et 3 entretiens avec celles qui ne l'utilisent. (voir tableau 4.3 ci-dessous)

	Nombre d'entretiens	Utilise le web sémantique
Bibliothèque nationale	3 dont 2 à la BnF	Oui
Bibliothèque de lecture publique	1	Non
Bibliothèque universitaire	5	4 bibliothèques
Bibliothèque spécialisée	1	Non

Tableau 4.3 : Le nombre d'entretiens réparti sur les quatre types de bibliothèques

4.3.5 La grille d'entretien

La grille d'entretien débute par une présentation du travail de recherche. Elle comprend 4 parties sur: le profil du professionnel de l'information, la gestion des collections numériques en bibliothèques, l'usage du web sémantique en bibliothèques numériques et les pratiques documentaires.

Présentation du travail de recherche

Avant de poser les questions nous avons présenté brièvement le sujet de notre recherche, les objectifs de l'entretien (s'interroger sur leurs pratiques documentaires suite à l'introduction du web sémantique en bibliothèques numériques), l'anonymat, la permission d'enregistrer l'entretien et la durée prévue pour le réaliser.

Le profil du professionnel de l'information

Dans cette partie, nous avons demandé aux enquêtés de se présenter (décrire brièvement leur cursus universitaire, leur parcours professionnel, leurs activités et leurs missions à la bibliothèque

71 27 demandes ont été envoyées aux bibliothèques cibles dans une période allant du 14 septembre 2016 au 15 février 2017. 8 demandes aux bibliothèques qui utilisent le web sémantique (6 ont accepté de faire l'entretien) dont, et 19 aux bibliothèques qui ne l'utilisent pas, (3 ont accepté de faire l'entretien).

et les formations qu'ils ont suivies dans le cadre de leur travail).

Le but de cette première partie est d'identifier les différents profils des professionnels de l'information et leur domaine d'activité en lien avec des bibliothèques numériques. Cela nous permet d'effectuer des comparaisons entre les participants, à propos de l'usage du web sémantique en bibliothèques. Elle permet également de favoriser l'échange en laissant aux participants la liberté de s'exprimer.

La gestion des collections numériques en bibliothèques

L'objectif de cette partie est de cerner les représentations qu'ont les professionnels de l'information de la bibliothèque numérique, et de discuter autour des langages documentaires utilisés pour la gestion de collections numériques dans ces bibliothèques. Cette partie se compose de trois questions :

- *Qu'est-ce qu'une bibliothèque numérique pour vous ?* Dans l'état de l'art nous avons fait une synthèse de différentes définitions de bibliothèque numérique. Nous accordons de l'importance à la façon dont les professionnels de l'information définissent une bibliothèque numérique. L'objectif est de pouvoir comparer et regrouper les différentes définitions afin d'en retirer des caractéristiques communes, basées sur les opinions des professionnels de l'information (la communauté de pratique).
- *Pouvez-vous parler des langages documentaires utilisés dans votre bibliothèque pour décrire les collections numériques ?* L'objectif est de discuter autour des langages documentaires que les participants utilisent pour gérer les collections numériques en bibliothèques. Cela nous permet, d'une part, de repérer les différents langages documentaires, et d'autre part, de comprendre le rôle de ces langages dans la gestion des collections numériques.
- *Pouvez-vous me parler de métadonnées produites dans votre bibliothèques ?* Un des éléments essentiels dans une bibliothèque numérique est la production des métadonnées. Le but est de savoir si les bibliothèques produisent ou non des métadonnées. Si oui, quels sont les différents types de métadonnées produites. Sinon, est ce qu'il y a des pistes stratégiques pour la production des métadonnées dans l'avenir.

Le web sémantique en bibliothèques numériques

Le but de cette partie est de pouvoir identifier les usages et les applications du web sémantique en bibliothèques numériques. Cette partie est composée de quatre questions principales :

- *Quelle expression préférez vous utiliser « le web sémantique » ou « le web de données » ?* L'objectif de cette question est de connaître la différence faite par les professionnels de l'information entre ces deux expressions. Cela nous paraît intéressant pour comprendre l'usage de ces deux expressions dans un milieu professionnel, celui des bibliothèques numériques.
- *Quels sont selon vous les enjeux / les apports du web sémantique en bibliothèques numériques ?* L'objectif est de connaître ce que le web sémantique peut apporter de plus à la bibliothèque.
- *Est-ce que votre bibliothèque utilise le web sémantique ?* Le but est de savoir si la bibliothèque utilise ou non le web sémantique. Si oui, il s'agit de repérer les différents outils / standards utilisés pour identifier et lier les données. Sinon, il s'agit de connaître les raisons pour lesquelles les bibliothèques ne s'intéressent pas à l'utilisation du web sémantique en bibliothèques.
- *Si la bibliothèque venait à évoluer et à introduire le web sémantique quels seraient selon vous les principaux défis ?* L'objectif est de pouvoir mettre en lumière d'un point de vue professionnel les principaux défis liés à l'utilisation du web sémantique. Il s'agit ensuite de pouvoir comparer les différentes opinions afin de retirer des éléments propres à chaque type de bibliothèque.

<u>Les pratiques documentaires</u>

Dans cette dernière partie, nous nous intéressons à la relation entre la technique et les pratiques professionnelles en général, et le web sémantique et les pratiques documentaires en particulier. Quatre questions composent cette partie :

- *Que pensez vous de l'influence de la technologie sur vos pratiques professionnelles ?* Nous cherchons à savoir à quel degré les pratiques professionnelles peuvent être touchées par la technologie.
- *Quels types de compétences mobiliser en termes de pratiques documentaires pour accompagner l'utilisation du web sémantique en bibliothèques numériques* ? L'objectif implicite de cette question est de savoir si les participants pensent que les pratiques documentaires ont une place à occuper dans le contexte du web sémantique. Nous avons évité de poser la question d'une manière directe pour ne pas orienter les enquêtés vers des réponses comme (oui ou non).

- *Penser vous qu'il soit nécessaire de suivre une formation sur le web sémantique avant de l'intégrer en bibliothèques ? Pourquoi ?* L'objectif est de pouvoir repérer les motivations des participants pour suivre une formation sur le web sémantique.
- *Comment voyez vous l'avenir de votre profession ?* Nous avons clôturé notre entretien par une question large en laissant l'enquêté s'exprimer d'une manière libre. Le but est de connaître le discours des professionnels de l'information sur le futur de leur profession et leur positionnement par rapport aux différents changements / évolutions intervenus sur leur métier.

4.3.6 Le déroulement de l'entretien

Les entretiens se sont déroulés entre le 26 septembre 2016 pour le premier d'entre eux et le 1er mars 2017 pour le dernier.

Les personnes rencontrées sont des professionnels de l'information qui ont des pratiques documentaires liées aux collections numériques.

Deux entretiens ont été réalisés face à face et les autres par téléphone, à la demande de l'interlocuteur ou pour des raisons pratiques de faisabilité. Les entretiens ont été enregistrés à l'aide d'un magnétophone pour ceux réalisés en présentiel et d'une application spécifique dans le cas des entretiens téléphoniques. La durée de chaque entretien varie entre 45 minutes et 2h15.

4.3.7 Le mode de saisie

Les entretiens ont été intégralement enregistrés puis entièrement retranscrits. Nous avons gardé la forme des propos des enquêtés en corrigeant quelques fautes comme la négation et en enlevant les phrases répétées ainsi que les phrases qui n'ont pas de sens. Cependant, dans certains enregistrements téléphoniques la qualité de la ligne n'était pas très bonne et il était difficile de comprendre certains mots ou certaines phrases. L'ensemble des entretiens représente 12 heures d'enregistrement, dont les transcriptions sont regroupées dans l'annexe XII.

4.3.8 L'analyse des données

Le traitement des données des entretiens relève de l'analyse de contenu. Plus précisément il s'agit d'une analyse thématique.

Dans un premier temps, une grille d'analyse générale a été faite à partir de la grille d'entretien. Le

but étant de repérer les thèmes principaux, les sous thèmes et de les associer aux hypothèses qu'y correspondent.

Le tableau 4.4 ci-dessous synthétise la grille d'analyse :

Thèmes	Sous-thèmes	Hypothèses
Bibliothèques numériques	• Sur la définition des bibliothèques numériques • Sur la gestion des collections numériques • Sur les métadonnées	
Web sémantique	• Sur les nominations du web sémantique • Sur les enjeux du web sémantique en bibliothèques numériques • Sur les défis à l'utilisation du web sémantique	**H3** Les bibliothèques numériques peuvent se servir du web sémantique pour proposer de nouveaux services plus adaptés aux usagers du web et pour rendre les bibliothèques plus visibles. **H2** La représentation des connaissances ne se fait pas seulement en utilisant les langages informatiques et standardisés du web sémantique, il faut une intervention humaine permettant de fournir à ces langages les données et les modèles conceptuels nécessaires pour représenter les connaissances.
Pratiques professionnelles	• Sur l'influence de la technologie • Sur les pratiques documentaires • Sur l'évolution du métier	**H1** Les professionnels de l'information peuvent jouer un rôle dans la mise en œuvre et l'évolution du web

		sémantique en introduisant leur expertise documentaire relative à la gestion des métadonnées.

Tableau 4.4 – Grille d'analyse des entretiens

Ensuite, des analyses thématiques plus spécifiques liées à des thèmes / questions précises ont été réalisées en s'appuyant sur une lecture des données d'entretiens. Pour chaque question, nous avons repéré les thèmes principaux, les sous-thèmes et les textes qu'y correspondent. Les grilles de ces analyses seront présentées dans le chapitre suivant.

Conclusion du chapitre

Ce chapitre a été consacré à la présentation des méthodes de recherche que nous avons mobilisées pour mener notre étude sur le web sémantique en bibliothèques numériques en France.

Dans un premier temps, notre objectif était d'effectuer une sélection des bibliothèques numériques en France. Nous nous sommes appuyées sur trois recensements déjà faits. Ces derniers ont été analysés afin de vérifier si les bibliothèques numériques répondaient à nos critères de sélection. Ces critères ont été choisis suite à une analyse des différentes définitions des bibliothèques numériques que nous avons effectuée dans le deuxième chapitre de cette thèse. Selon ces critères, les bibliothèques numériques doivent proposer un accès à distance aux ressources numériques. Ensuite, elles doivent offrir des outils de recherche permettant aux utilisateurs de retrouver facilement les informations qui répondent à leurs besoins. Enfin, un traitement documentaire est nécessaire pour assurer l'organisation des connaissances numériques.

Les recensements utilisés sont ceux réalisés par la bibliothèque nationale de France, l'unité régionale de formation à l'information scientifique et technique (URFIST) de Lyon et l'inventaire multilingue du patrimoine numérique MICHAEL. Les trois recensements ont été vérifiés afin de valider le bon fonctionnement des liens qu'ils proposent. L'analyse de ces trois recensements nous a permis de sélectionner 98 bibliothèques, qui ont été classées en quatre catégories : les bibliothèques nationales, les bibliothèques universitaires, les bibliothécaires de lecture publique et les bibliothèques spécialisées. Ce classement s'appuie essentiellement sur les types d'institutions à l'origine de ces bibliothèques.

Ensuite, nous avons commencé notre enquête par une approche quantitative sous la forme d'un questionnaire envoyé aux bibliothèques sélectionnées dans l'étape précédente. L'objectif étant de

s’interroger sur les usages et les applications du web sémantique en bibliothèques numériques. A savoir quelles sont les bibliothèques qui utilisent le web sémantique, les raisons pour lesquelles elles l'utilisent et les obstacles auxquelles elles sont confrontées. Ainsi, ce questionnaire avait pour objectif d’interroger les professionnels de l'information sur leur compréhension du web sémantique et sur leurs pratiques documentaires touchées ou non par ce dernier.

Nous avons eu 33 réponses, soit 33,67 % de l'ensemble des bibliothèques cibles. Le tableau ci-dessous indique leur répartition sur les quatre types de bibliothèques :

Bibliothèque nationale	**Bibliothèque universitaire**	**Bibliothèque de lecture publique**	**Bibliothèque spécialisée**
3	13	13	4

Par la suite, une enquête qualitative sous la forme d'entretiens semi-directifs a été réalisée afin de mener une étude approfondie permettant de se mettre en contact avec les professionnels de l'information des bibliothèques ayant répondu au questionnaire. Les entretiens avaient pour objectif d'apporter des réponses à nos questionnements portant sur le web sémantique, les bibliothèques numériques et les pratiques documentaires. Nous avons pu réalisé dix entretiens répartis sur des professionnels de l’information travaillant dans les différents types des bibliothèques et dans celles qui utilisent le web sémantique et celles qui ne l'utilisent pas.

Nous avons réalisé 10 entretiens, soit 33,3 % des bibliothèques qui ont répondu au questionnaire. Le tableau ci-dessous indique leur répartition sur les quatre types de bibliothèques :

Bibliothèque nationale	**Bibliothèque universitaire**	**Bibliothèque de lecture publique**	**Bibliothèque spécialisée**
3	5	1	1

Dans le chapitre suivant, nous présentons et nous analysons les résultats de notre enquête quantitative et qualitative. L'analyse des résultats doit nous permettre de tester nos hypothèses et d'apporter une clarification à nos questionnements.

Chapitre 5

Présentation et analyse des résultats

Ce chapitre a pour objectif de présenter les résultats de notre enquête sur les usages et les applications du web sémantique en bibliothèques numériques en France. Le but est de valider nos hypothèses de recherche et d'apporter des éclaircissements sur l'évolution des pratiques professionnelles suite à l'introduction du web sémantique en bibliothèques numériques.

Dans la première section, nous présentons notre sélection des bibliothèques numériques. Le but est d'identifier les caractéristiques spécifiques à chaque catégorie de bibliothèques numériques.

Nous exposons, dans la deuxième section, les résultats qui découlent de l'analyse du questionnaire. L'objectif est d'apporter des données sur les usages du web sémantique en bibliothèques numériques en France. Ensuite, nous cherchons à souligner les apports et les obstacles liés à l'utilisation du web sémantique dans ces bibliothèques.

Enfin, dans la troisième section, nous présentons les résultats des entretiens. Une analyse thématique est appliquée afin d'examiner le discours des enquêtés. Nous souhaitons apporter plus de clarifications sur les représentations qu'ont les professionnels de l'information (la communauté de pratique) sur l'usage du web sémantique en bibliothèques numériques. Plus précisément, nous cherchons à montrer que les professionnels de l'information peuvent jouer un rôle important dans la mise en place du web sémantique en bibliothèques numériques.

5.1 Analyse de la sélection

Nous introduisons dans cette section une analyse des bibliothèques retenues. Ces dernières ont été sélectionnées suite à une analyse de trois recensements des bibliothèques numériques en France. La sélection a porté sur 98 bibliothèques numériques, soit 33,33 % des bibliothèques recensées par les trois outils étudiés.

5.1.1 Classement des bibliothèques numériques

Notre objectif de départ était de pourvoir proposer une typologie des bibliothèques numériques. Pour cela, nous avons tout d'abord effectué une recherche dans la littérature de revues sur des typologies des bibliothèques numériques. Nous avons pu repéré deux études traitant ce sujet. Une étude de [Sharon et Frank, 2000] classe les bibliothèques numériques en trois catégories :

- Bibliothèque numérique autonome « Stand-alone Digital Library (SDL) » : elle ressemble à une bibliothèque classique en bénéficiant de l'informatique pour gérer des documents numérisés ou numériques. (Elle produit, numérise, achète, diffuse ses propres ressources et services).
- Bibliothèque numérique fédérée « Federated Digital Library (FDL) » (comprend plusieurs bibliothèques autonomes, des dépôts hétérogènes, connecté via un réseau, elle dispose d'une interface unifiée ; son problème majeur est l'interopérabilité).
- Bibliothèque numérique de récolte « Harvested Digital Library (HDL) » est une bibliothèque virtuelle (récolte des objets numériques non complets, surtout des métadonnées et des résumés) et les liens qui pointent vers ces objets numériques.

Et une autre étude de [Berroneau, 2005, p.37] classe les bibliothèques numériques en trois groupes selon leurs ressources :

- les ressources produites et mises en ligne par la bibliothèque elle-même (documents numérisés documents numériques, ou renseignements généraux par exemple),
- les ressources commerciales auxquelles la bibliothèque est abonnée (comme des périodiques électroniques),
- les ressources extérieures mises à disposition de l'usager (comme un répertoire de sites internet).

La première typologie distingue les bibliothèques qui gèrent elles-mêmes des collections

numériques (création, traitement, gestion, diffusion), celles qui gèrent l'accès aux collections numériques en collectant des métadonnées et des liens pointant vers les documents numériques. Celles-ci ne disposent pas de collections numériques, elles gèrent seulement l'accès à ces collections. Enfin, celles qui proposent un accès unifié à des dépôts hétérogènes de collections numériques.

La deuxième distingue trois types des bibliothèques classés selon leurs ressources. Les ressources peuvent être donc produites par la bibliothèque elle-même, ou bien collectées depuis des ressources extérieures ou des ressources payantes.

Le tableau 5.1 présente une comparaison entre les deux typologies :

	Typologie 1	**Typologie 2**
Bibliothèque numérique autonome	Gère ses propres ressources	Les ressources sont produites et mises en ligne par la bibliothèque elle-même
Bibliothèque numérique de récolte	Gère l'accès aux collections numériques (métadonnées, liens)	Les ressources sont récoltées de l'extérieur (répertoire des sites)
Bibliothèque numérique fédérée	Permet un accès unifié à des dépôts hétérogènes	
Bibliothèque disposant des ressources payantes		Les ressources sont payantes (sous la forme des abonnements)

Tableau 5.1- Une comparaison entre les deux typologies de bibliothèques numériques

Nous pouvons constater qu'il y a une ressemblance entre les deux typologies. Une bibliothèque qui produit elle-même ses ressources peut être considérée comme une bibliothèque numérique autonome (typologie 1), c'est-à-dire une bibliothèque qui crée et diffuse elle-même ses ressources. Une bibliothèque numérique de récolte (typologie 1) peut être considérée comme une bibliothèque qui dispose de ressources externes (typologie 2), c'est-à-dire une bibliothèque qui collecte des métadonnées en proposant des liens vers des ressources localisées dans d'autres bibliothèques. Cependant, la typologie 2 distingue les bibliothèques qui disposent de ressources payantes sous forme d'abonnements et la typologie 1 identifie une catégorie de bibliothèques numériques qui proposent un accès unifié à des dépôts hétérogènes de ressources numériques.

Ces deux typologies s'opposent aux critères que nous avons retenus pour définir les bibliothèques numériques. Premièrement et avant de commencer notre étude sur les bibliothèques numériques, nous avons parlé de trois types de bibliothèques (numérique, virtuelle et électronique) (voir chapitre

2 section 2.1.1). La bibliothèque numérique de récolte (typologie 1) et son équivalent dans la typologie 2 seraient plus proches de la bibliothèque virtuelle. Celle-ci gère l'accès aux collections numériques en collectant des métadonnées et des liens vers d'autres sites. Ces bibliothèques ont été différenciées des bibliothèques numériques que nous avons retenues dans notre étude. Ces dernières ont été étudiées de manière approfondie dans le chapitre 2 section 2.1.2.

Nous pensons que les bibliothèques numériques telles que nous les avons définies ressemblent à l'ensemble des bibliothèques numériques autonomes et fédérées, c'est-à-dire celles qui gèrent leurs propres ressources numériques en y proposant un accès libre ou/et un accès unifié.

De ce fait, nous avons finalement renoncé à l'idée de proposer une typologie des bibliothèques numériques. D'une part, les typologies des bibliothèques numériques étudiées ne correspondent pas à notre définition d'une bibliothèque numérique, d'autre part, nous pensons que les bibliothèques numériques sélectionnées peuvent être classées selon le type d'institutions à l'origine de leur création.

Afin de proposer un classement des bibliothèques numériques françaises, nous nous sommes appuyées sur le type d'institutions qui les gèrent : bibliothèque nationale, bibliothèque de lecture publique, bibliothèque universitaire et bibliothèque spécialisée.

Les bibliothèques nationales sont des établissements publics qui dépendent du ministère de la Culture. Leurs fonctions principales résident dans leur dépôt légal qui entraîne la conservation à long terme des publications d'un pays. Ensuite, elles maintiennent et élaborent les bibliographies nationales, les identifications normalisées et les normes de catalogage et de l'indexation [Cacaly, 2001]. En se référant au dictionnaire encyclopédique de l'information et de la documentation[72], la bibliothèque publique d'information et la bibliothèque nationale et universitaire de Strasbourg ont été classées « bibliothèque nationale » [Ibid].

Les bibliothèques de lecture publique sont des établissements publics ouverts au grand public dont les services sont gratuits, elles couvrent : les bibliothèques municipales, les bibliothèques territoriales, les bibliothèques intercommunales et les bibliothèques départementales de prêt. Les bibliothèques universitaires sont rattachées à des universités pour répondre aux besoins de la pédagogie et de la recherche. Les bibliothèques spécialisées, quant à elles, recouvrent les bibliothèques axées sur des disciplines et des spécialités précises. Elles sont généralement situées au sein d'un centre de recherche, d'une institution, d'une entreprise etc. [Cacaly, 2004].

72 Cacaly, Serge (dir.) 2001. Dictionnaire encyclopédique de l'information et de la documentation. Paris : Nathan

La figure 5.1 ci-dessous indique la répartition de nombre de bibliothèques sélectionnées selon leur type. Le résultat détaillé de cette sélection figure dans l'annexe XI.

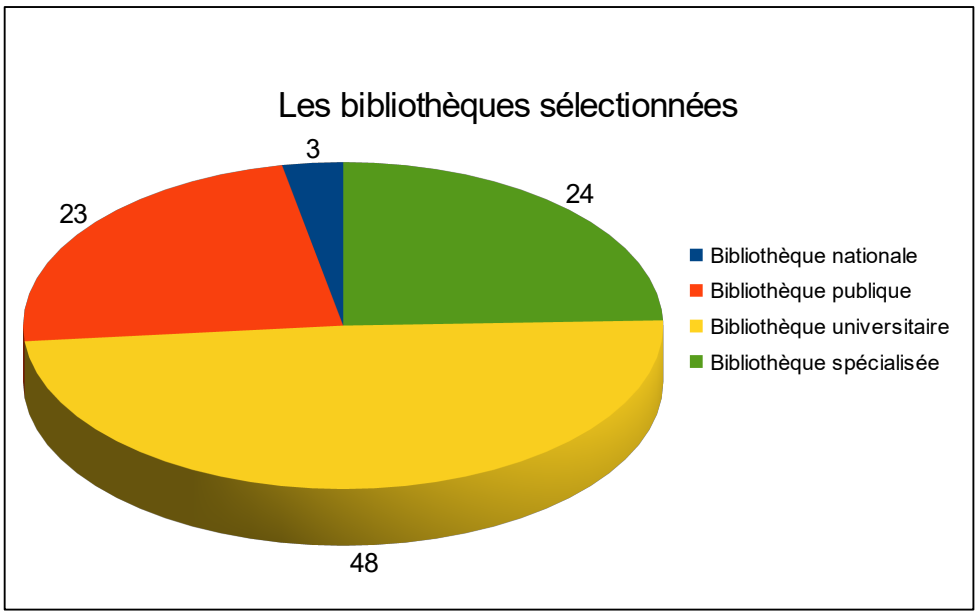


Figure 5.1 : Répartition de nombre de bibliothèques sélectionnées selon leur type

5.1.2 Caractéristiques des bibliothèques numériques

Les bibliothèques numériques nationales

Ces bibliothèques sont créées et gérées par les bibliothèques nationales. Elles mettent en ligne des collections numérisées ou numériques appartenant à la bibliothèque nationale. En France, en 1997 le Premier ministre Lionel Jospin prononce un discours où il exprime la nécessité de développer une présence culturelle sur les réseaux d'information et cela en adaptant des politiques de numérisation du patrimoine culturel [Claerr et Westeel, 2013].

Le but est de proposer un accès gratuit aux collections de ces bibliothèques, et d'assurer une diffusion numérique du patrimoine sur Internet. Des enjeux culturels, économiques et politiques accompagnent la construction des bibliothèques numériques nationales. Le rapport sur « la diffusion numérique du patrimoine, dimension de la politique culturelle » remis par Bruno Ory-Lavollée en 2002 indique que « *les réseaux sont devenus des lieux de compétition économique et stratégique entre nation*s » [Ory-Lavollée, 2002, p.25].

Possédant des fonds patrimoniaux considérables et profitant des moyens techniques, humains et financiers importants, les bibliothèques numériques nationales sont les pionnières dans la mise en œuvre des projets de numérisation, des politiques documentaires numériques, des normes et des

standards permettant la création, la gestion, la conservation et la diffusion des collections numériques. L'analyse de la sélection indique que les trois bibliothèques nationales en France disposent des bibliothèques numériques. Cela confirme leur engagement et leur investissement dans les projets numériques et la valorisation du patrimoine culturel.

Les bibliothèques numériques de lecture publique

Ces bibliothèques sont créées et gérées par des bibliothèques municipales, territoriales intercommunales et départementales. Elles bénéficient du soutien du ministère de la Culture et de la Communication qui a mis en place des dispositifs de co-financement pour accompagner la transition de ces bibliothèques vers le numérique [Claerr et Westeel, 2013].

Ces bibliothèques proposent un accès gratuit aux collections numériques qu'elles possèdent. Elles visent à diffuser et à valoriser des fonds patrimoniaux locaux.

Les bibliothèques numériques universitaires

Ces bibliothèques sont créées et gérées par les bibliothèques universitaires. Elles disposent d'une autonomie renforcée des universités. Cependant la politique du ministère de l'Enseignement supérieur et de la Recherche en matière de construction des bibliothèques numériques s'inscrit dans des objectifs visant :

- à favoriser l'accès à des fonds documentaires principalement de recherche structurés selon des standards ouverts et des outils d'exploitation des contenus.
- à améliorer la visibilité de ces fonds en favorisant l'interopérabilité avec d'autres dispositifs d'accès aux contenus numériques [Claerr et Westeel, 2013, p. 34].

Les bibliothèques numériques universitaires disposent des fonds et des corpus documentaires de recherche et d'étude. Leur but est de faciliter l'accès à ces corpus et les rendre visibles en favorisant l'utilisation des standards ouverts. Elles ont comme objectif de valoriser l'information scientifique et technique en la rendant accessible.

Les bibliothèques numériques spécialisées

Ces bibliothèques sont gérées par des associations, des fondations, des centres de recherche etc. Elles disposent des fonds des disciplines ou des spécialités précises. Des programmes nationaux pris en charge par l'état ont été mobilisés pour soutenir la numérisation du patrimoine scientifique surtout les corpus des centres de recherche, par exemple les programmes de numérisation concertée et de coopération numérique [Ibid]. Ces bibliothèques visent à valoriser leurs collections numériques en facilitant notamment leur accès et la coopération avec d'autres dispositifs d'accès à

l'information.

Le tableau 5.2 présente les caractéristiques des quatre catégories de bibliothèques numériques :

Bibliothèques numériques nationales	**Bibliothèques numériques de lecture publique**	**Bibliothèques numériques universitaires**	**Bibliothèques numériques spécialisées**
- gérer par les bibliothèques nationales - permettre l'accès à distance des collections numériques - numériser, gérer et conserver des fonds patrimoniaux nationaux - valoriser des fonds patrimoniaux nationaux - élaborer des politiques documentaires numériques, des normes, des standards etc.	- gérer par les bibliothèques de lecture publique comme les bibliothèques municipales, territoriales, intercommunales et départementales - permettre l'accès à distance des collections numériques - numériser, gérer et conserver des fonds locaux - valoriser des fonds patrimoniaux locaux	- gérer par les bibliothèques universitaires - permettre l'accès à distance des collections numériques - numériser, gérer et conserver des fonds documentaires - valoriser des fonds documentaires de recherche et d'étude	- gérer par des centres de recherche, des associations, des entreprises etc. - permettre l'accès à distance des collections numériques - numériser, gérer et conserver des fonds patrimoniaux scientifiques, des fonds privés etc. - valoriser des fonds scientifiques

Tableau 5.2 – Caractéristiques des quatre catégories des bibliothèques numériques

5.1.3 Les limites de l'analyse de sélection

L'analyse de la sélection présente des limites. Tout d'abord, cette sélection comme nous l'avons indiqué précédemment ne recense pas la totalité des bibliothèques numériques en France pour les raisons mentionnées auparavant. Il est donc possible qu'il y ait de bibliothèques numériques répondant à nos critères de sélection mais qui n'ont pas été étudiées.

Les catégories que nous avons utilisées pour classer les bibliothèques numériques recensées ne prennent pas en compte « les bibliothèques numériques internationales ». Ces bibliothèques ont été exclues de la recherche, mais nous pensons qu'elles représentent une catégorie importante des bibliothèques numériques.

Cependant, malgré les limites évoquées, cette sélection des bibliothèques numériques apparaît suffisante pour mener notre enquête quantitative et qualitative sur l'usage du web sémantique en bibliothèques numériques.

5.2 Présentation et analyse des résultats tirés du questionnaire

Après avoir construit notre sélection des bibliothèques numériques, nous avons mené une enquête quantitative sous la forme d'un questionnaire envoyé aux bibliothèques numériques sélectionnées. Dans cette section, nous présentons et nous analysons les résultats de cette enquête.

5.2.1 Le profil des bibliothèques

Trente trois bibliothèques ont répondu au questionnaire. La figure 5.2 ci-dessous montre leur répartition sur les quatre catégories mentionnées auparavant : bibliothèque numérique nationale, bibliothèque numérique de lecture publique, bibliothèque numérique universitaire et bibliothèques numérique spécialisée.

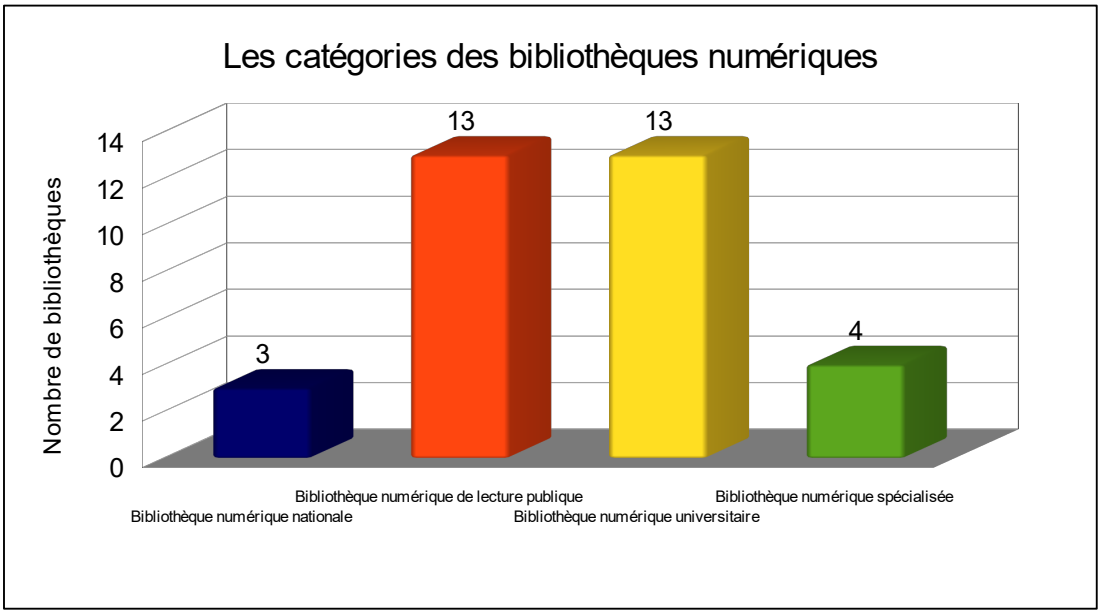


Figure 5.2 : les catégories des bibliothèques numériques

Après avoir présenté les profils des bibliothèques qui ont répondu au questionnaire, la figure 5.3 ci-dessous montre le taux de réponses de chaque type de bibliothèques par rapport au nombre des bibliothèques sélectionnées.

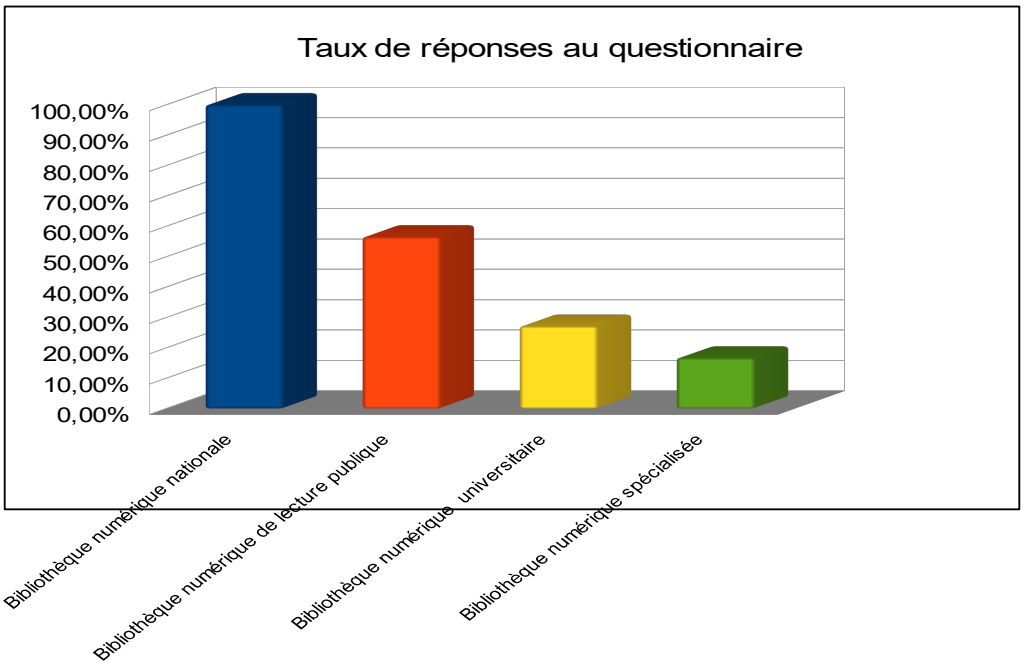


Figure 5.3 : Taux de réponses au questionnaire selon le type de bibliothèque

5.2.2 Profil des professionnels de l'information

Dans cette section nous présentons les profils des professionnels de l'information qui ont répondu au questionnaire envoyé aux bibliothèques cibles.

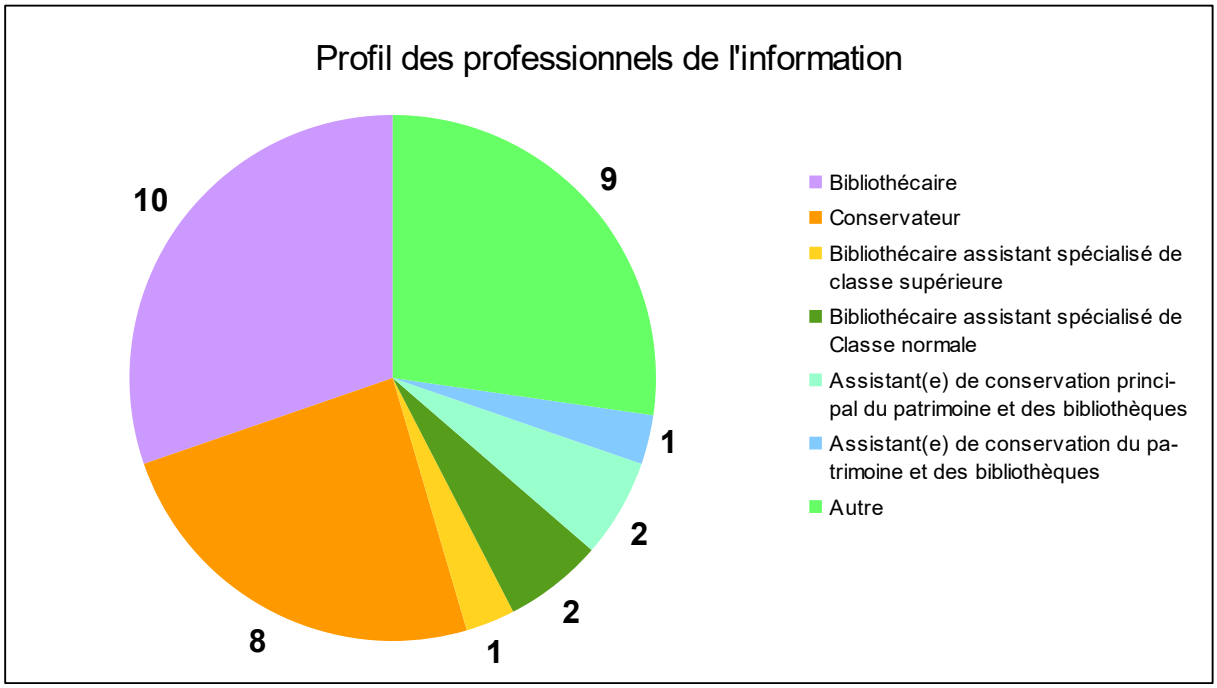


Figure 5.4 : Répartition de nombre de professionnels de l'information selon leur profil

Les neuf profils de la catégorie « autre » varient entre : ingénieur d'études (documentaire, chargé de ressources documentaires, en documentation) ; chef de service ; chef de projet ; ingénieur en analyse et traitement de bases de données ; adjoint du patrimoine, technicienne de bibliothèque et CHED principal.

5.2.3 Connaissances des professionnels de l'information sur le web sémantique

Afin de connaître le degré de connaissances des professionnels de l'information sur le web sémantique, nous avons posé quelques questions permettant d'identifier les niveaux de compréhension de certaines expressions telles que « web sémantique », « web de données / linked data», « Ressource Desciption Frame, RDF », « Web Ontology Language OWL » et « Simple Knowledge Organization System SKOS ». Les diagrammes ci-dessous montrent les résultats obtenus :

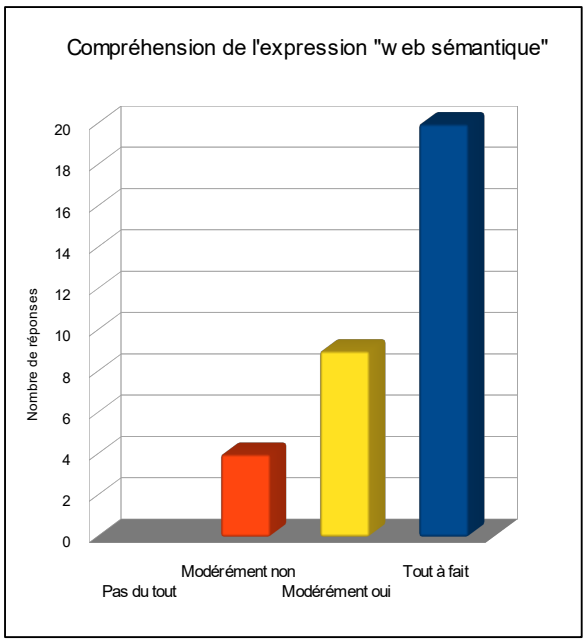


Figure 5.5 : Compréhension de l'expression "web sémantique" par les enquêtés

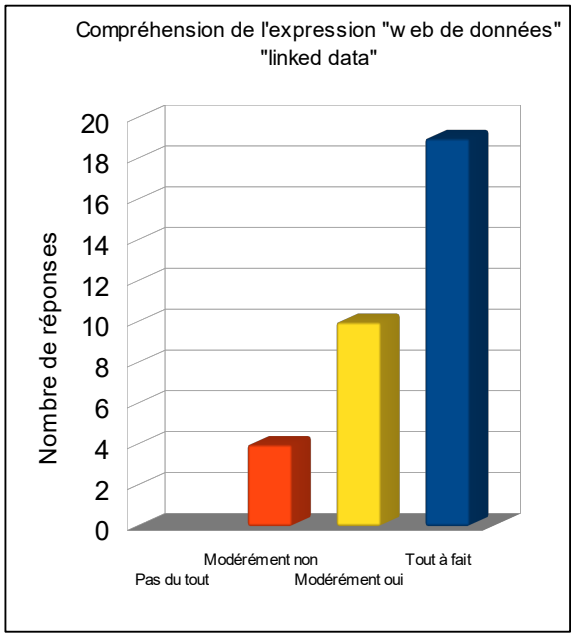


Figure 5.6 : Compréhension de l'expression "web de données" par les enquêtés

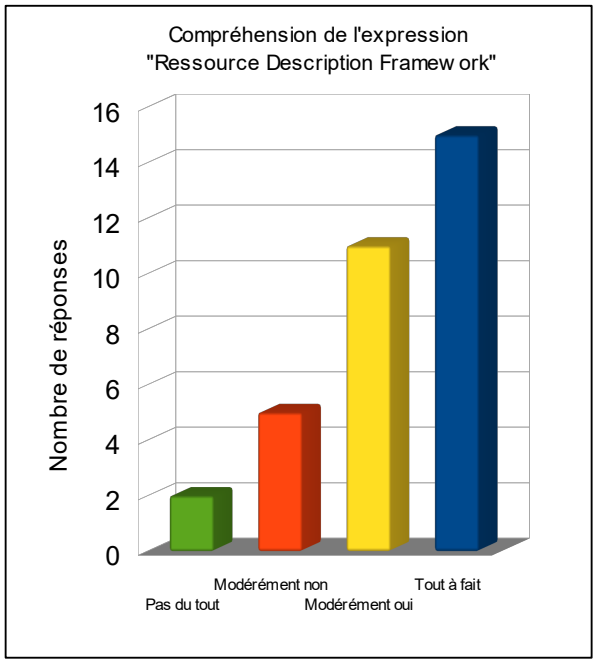


Figure 5.7 : Compréhension de l'expression "Ressource Description Framework" par les enquêtés

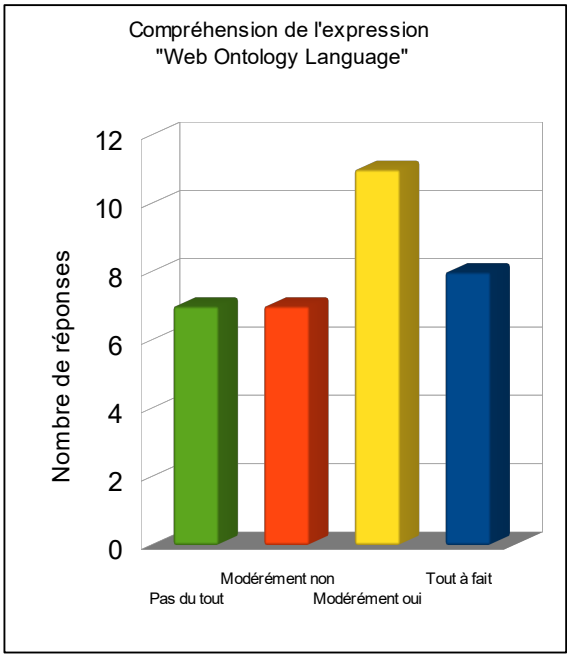


Figure 5.8 : Compréhension de l'expression "Web Ontology Language" par les enquêtés

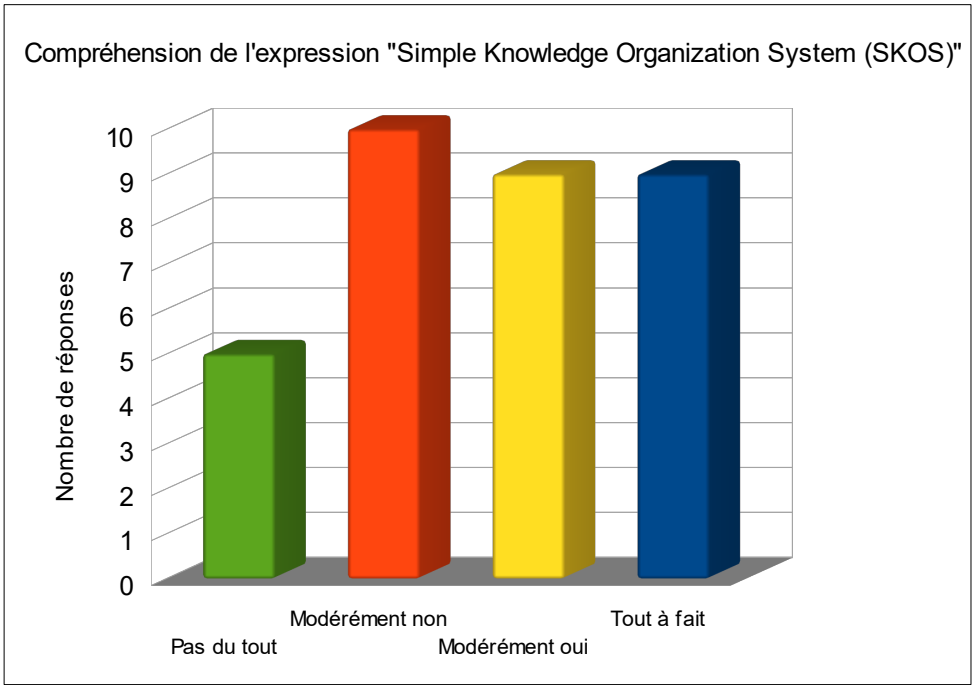


Figure 5.9 : Compréhension de l'expression "Simple Knowledge Organization System (SKOS)"

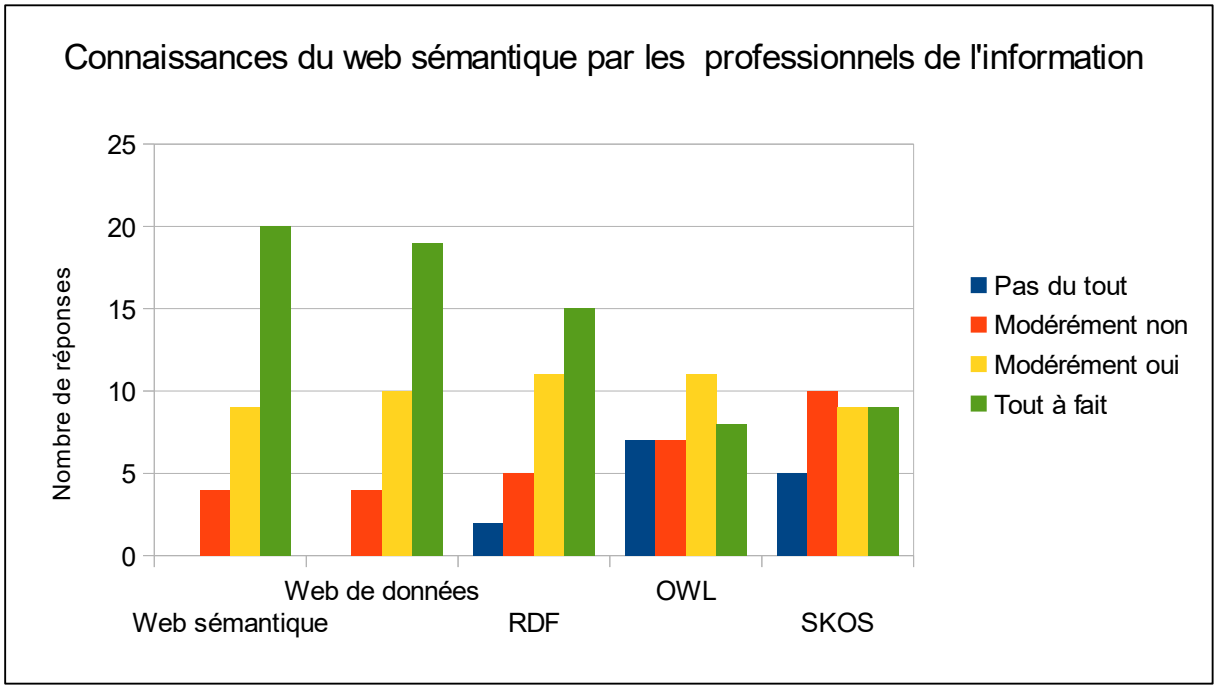


Figure 5.10 : Connaissances de quelques notions liées au web sémantique par les professionnels de l'information

Les résultats montrent que les enquêtés ont une bonne compréhension de la notion exprimée par les deux expressions web sémantique et web de données. Cependant pour la notion RDF qui représente le standard permettant de lier les données entre elles, 7 enquêtés [(2) pas du tout et (5) modérément non] ont indiqué une incompréhension de cette notion. Tandis que pour les notions OWL et SKOS, les ontologies utilisées pour représenter les données dans le cadre du web sémantique, les résultats montrent qu'une quinzaine des enquêtés ne comprennent pas ces notions. Cela est dû au fait que le web sémantique n'est utilisé que dans 9 bibliothèques ayant répondu au questionnaire. L'usage du web sémantique dans ces bibliothèques se résume dans quelques unes à des initiatives pour identifier et lier les données.

Nous nous sommes intéressées d'examiner s'il existe un lien entre le niveau de connaissances sur le web sémantique et le type de bibliothèques dans lesquelles les professionnels de l'information travaillent. Nous avons remarqué que les quatre réponses « modérément non » sur la question de compréhension des notions web sémantique et web de données proviennent des enquêtés travaillant en bibliothèques de lecture publique. Et cinq réponses sur sept [(1) pas du tout et (4) modérément non] sur la question de la compréhension de la notion RDF proviennent aussi des enquêtés travaillant en bibliothèques publiques. Nous pouvons dire que les professionnels de l'information travaillant en bibliothèques numériques de lecture publique semblent être moins intéressés par les technologies du web sémantique que les autres bibliothèques .

5.2.4 Usage du web sémantique en bibliothèques numériques

Après avoir présenté les profils des bibliothèques et des professionnels de l'information qui ont répondu au questionnaire, nous présentons dans cette sous section les résultats de la question sur l'usage et le non usage du web sémantique dans ces bibliothèques.

Suite aux réponses sur l'usage du web sémantique en bibliothèque, nous pouvons tester notre hypothèse:

H3

Les bibliothèques numériques peuvent se servir du web sémantique pour proposer de nouveaux services plus adaptés aux usagers du web et pour rendre les bibliothèques plus visibles.

La figure 5.11 présente le nombre de bibliothèque utilisant ou non le web sémantique :

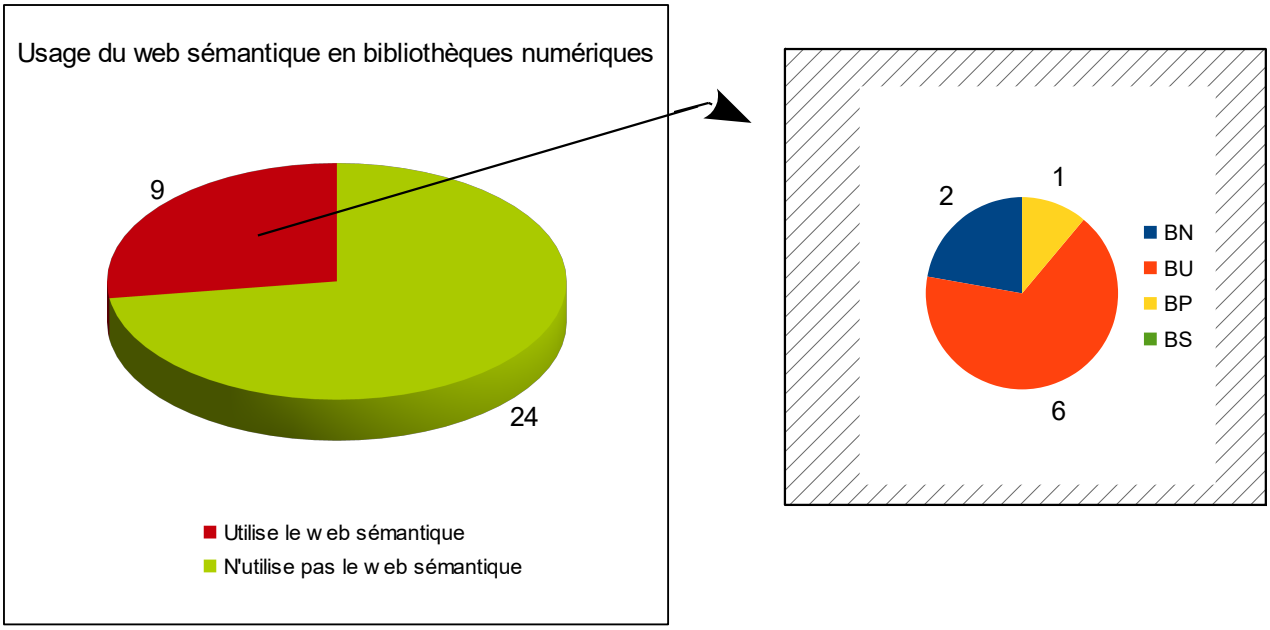


Figure 5.11 : Nombre de bibliothèques utilisant ou non le web sémantique

Les 9 bibliothèques qui utilisent le web sémantique sont majoritairement des bibliothèques universitaires (6), deux bibliothèques nationales et une bibliothèque publique. Six bibliothèques universitaires sur 13, deux bibliothèques nationales sur 3, une bibliothèque publique sur 13 et aucune bibliothèque spéciale n'utilisent le web sémantique. Les bibliothèques nationales et les bibliothèques universitaires semblent être les plus impliquées dans l'usage du web sémantique en

France.

Ensuite, afin de savoir pourquoi les bibliothèques utilisent ou non le web sémantique, nous avons posé deux questions ouvertes (pourquoi oui et pourquoi non). Pour étudier ces questions, nous avons utilisé la méthode de l'analyse thématique.

Le tableau 5.3 présente une analyse thématique de la question « pourquoi la bibliothèque utilise le web sémantique » :

Pourquoi la bibliothèque utilise le web sémantique ?(8 réponses)		
Thème principale	**Thème secondaire**	**Texte**
Visibilité	• Référencement • Web	1. de mieux nous inscrire dans le web des données liées, afin d'améliorer notre visibilité sur le Web.[Q03] 2. Pour améliorer le référencement des données [Q16] 3. Pour une meilleure visibilité de nos données [Q17] 4. d'améliorer le référencement.[Q25] 5. Nous avons en projet de rendre compatibles nos outils numériques avec les standards du Web sémantique dans les années à venir afin d'améliorer la visibilité de nos ressources et d'apporter de nouvelles fonctionnalités [Q07]
Evolution des catalogues	• Bases de données bibliographiques • Formats	1. Nous travaillons actuellement au développement d'applications basées sur le web sémantique afin d'accompagner l'évolution de nos catalogues et de nos bases de données bio-bibliographiques. [Q03] 1. Les catalogueurs s'adaptent peu à peu au modèle FRBR, lequel est censé (artificiellement puisque le modèle FRBR est antérieur à la notion de web sémantique) permettre de faire la transition vers le web sémantique. [Q24] 2. La bibliothèque suit les recommandations de l'Abes concernant l'adaptation du format Unimarc. [Q23] 3. pour une base de données des auteurs. Nous alignons également cette base avec IdRef et Wikidata pour désambiguïser le contenu de la base,

		aligner nos données avec nos propres données (des inventaires d'archives, par exemple) ou récupérer des données (en projet : récupérer les coordonnées spatiales des villes à partir de Wikidata par exemple). [Q27]
Réutilisation des données	• Différents acteurs	1. faciliter leur réutilisation par des acteurs autres que ceux des bibliothèques [Q16]
Interopérabilité	• Données • Systèmes	1. afin de permettre une meilleure opérabilité entre nos données, [Q25] 2. pour une meilleure interopérabilité de notre système [Q17]

Tableau 5.3 - Analyse thématique de la question « pourquoi la bibliothèque utilise le web sémantique »

Les raisons pour lesquelles les bibliothèques enquêtées utilisent le web sémantique se concentrent autour de 4 thèmes principaux. La visibilité sur le web et l'évolution des catalogues constituent les atouts majeurs à l'utilisation du web sémantique dans ces bibliothèques. Cela valide notre hypothèse (H 3) portant sur l'utilisation du web sémantique pour proposer des services plus adaptés aux usagers du web en rendant les bibliothèques plus visibles.

Ensuite, ces bibliothèques considèrent que le web sémantique permet l'interopérabilité entre les données et les différents systèmes et la réutilisation des données par des différents acteurs.

Le tableau 5.4 présente une analyse thématique de la question « pourquoi la bibliothèque n'utilise pas le web sémantique :

Pourquoi la bibliothèque n'utilise pas le web sémantique ? (19 réponses)		
Thème	**Sous-thème**	**Texte**
Moyens humains	• Compétences • Manque d'intérêt	1. Manque de compétences internes. [Q09] 2. problème de compétences [Q10] 3. Manque de moyens et de compétences pour le mettre en place [Q13] 4. Pas assez de temps pour utiliser ces standards.[Q01] 5. Peu d'intérêt pour la chose [Q05]
Moyens techniques	• Complexité • Compatibilité • Logiciels	1. Les logiciels utilisés par la bibliothèque ne sont pas encore adaptés au RDF [Q04] 2. Pas encore d'outils adéquats et compréhension technique encore trop faible [Q14] 3. Complexité technique, [Q15]

		4. difficulté de transition vers ces nouveaux standards [Q28] 5. Ils ne sont pas proposés par notre éditeur de SIGB [Q10]
Orientation stratégique	• Futur projet • Migration des données • Evolution • Budget	1. Nous avons en projet de rendre compatibles nos outils numériques avec les standards du Web sémantique dans les années à venir afin d'améliorer la visibilité de nos ressources et d'apporter de nouvelles fonctionnalités [Q07] 2. données structurées en XML et bases de données en cours de migration, le web sémantique sera une prochaine étape (travail de longue haleine pour l'alignement des données avec les réservoirs d'autorités bien que partiellement déjà fait) [Q09] 3. Projet évoqué et à mettre en place [Q12] 4. SIGB et métadonnées en cours de refonte pour aller vers l'usage du web sémantique [Q32] 5. La bibliothèque numérique que je gère existe depuis 2001. Il n'y a pas eu d'évolution depuis.[Q21] 6. Inquiétude sur les besoins de financement [Q11]

Tableau 5. 4 -Analyse thématique de la question « pourquoi la bibliothèque n'utilise pas le web sémantique »

Les résultats montrent que les bibliothèques utilisent le web sémantique pour les raisons suivantes :

- Accroître la visibilité de la bibliothèque sur le web
- Évoluer les catalogues des bibliothèques
- Améliorer l'interopérabilité des données et des systèmes
- Favoriser la réutilisation des données

En revanche, les raisons pour lesquelles les bibliothèques n'utilisent pas le web sémantique se résument en :

- Manque des compétences en interne
- Manque des moyens techniques pour la mise en place du web sémantique
- Manque d'intérêt

Certaines bibliothèques n'utilisent pas le web sémantique pour le moment, mais elles ont des projets

pour le mettre en place dans le futur.

5.2.4.1 Les standards du web sémantique utilisés dans les bibliothèques enquêtées

Le diagramme ci-dessous présente les standards du web sémantique utilisés dans les bibliothèques enquêtées :

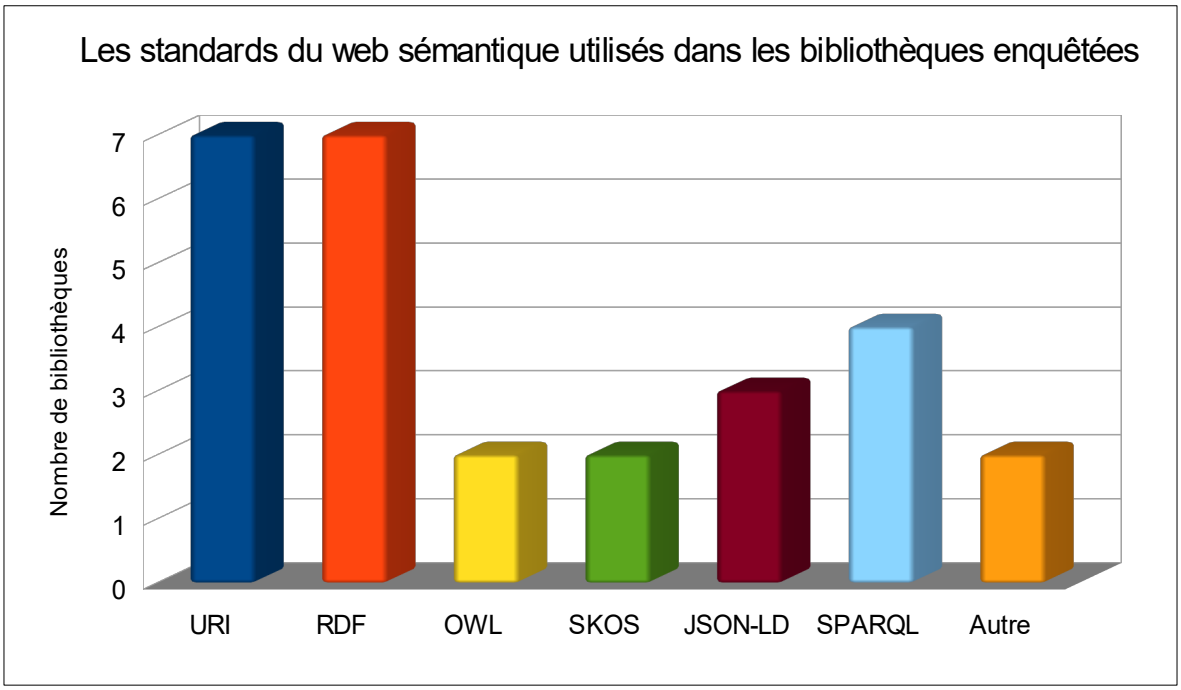


Figure 5.12 : Répartition des standards du web sémantique utilisés en bibliothèques

Les standards les plus utilisés dans les bibliothèques enquêtées sont les identifiants uniques (les URIs) et les triplets RDF permettant de lier les données entre elles. Cela pourrait justifier les réponses des professionnels de l'information sur leur connaissance des standards du web sémantique. Ces connaissances qui étaient plus marquées au niveau des standards URI et RDF.

Dans la catégorie « autre », les enquêtés ont mis « FRBR[73] / RDA[74] » et « Dublin Core[75] ». Ces deux modèles de description peuvent être utilisés dans le cadre du web sémantique pour décrire

73 Le modèle FRBR (Functional Requirements for Bibliographic Records / Fonctionnalités requises des notices bibliographiques) est un modèle conceptuel de données bibliographiques élaboré par un groupe d'experts de l'IFLA de 1992 à 1997. (Source : http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html)

74 RDA se présente comme « une nouvelle norme pour la description des ressources et l'accès à celles-ci, conçue pour le monde numérique » (*a world digital for resource description and access, designed for the digital world*). Adaptation, ouverture et flexibilité constituent les trois caractéristiques essentielles de cette nouvelle norme. (Source : http://www.bnf.fr/fr/professionnels/rda/s.rda_objectifs.html)

75 Le Dublin Core est un format descriptif à la fois simple et générique, comprenant 15 éléments différents, qui a été créé en 1995 à Dublin (Ohio) par OCLC et le NCSA (National Center for Supercomputing Applications). (Source : http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html)

d'une manière simple et commune les documents numériques (Dublin Core), et pour décrire les entités et les relations des données bibliographiques (FRBR /RDA). Grâce à l'utilisation des métadonnées structurées en XML, ces modèles peuvent facilement être intégrés dans les standards du web sémantique. Ils permettent de fournir des modèles conceptuels permettant d'établir des relations entre les données bibliographiques, les données d'autorité et les sujets.

Les deux bibliothèques qui utilisent tous les standards mentionnés sont la bibliothèque nationale de France et la bibliothèque universitaire Nice Sophia Antipolis.

5.2.5 Le web sémantique et les pratiques documentaires

Dans cette section nous abordons la question de la relation entre le web sémantique et les pratiques documentaires des professionnels de l'information. Cela nous permet de tester notre hypothèse :

H 1

Les professionnels de l'information peuvent jouer un rôle dans la mise en œuvre et l'évolution du web sémantique en introduisant leur expertise documentaire relative à la gestion des métadonnées.

Nous considérons qu'une donnée ne peut pas être traitée d'une façon automatique, uniquement en utilisant des langages normalisés du web sémantique (XML, RDF...). La mise en contexte des données nécessite une intelligence sociale (Cotte, 2013), qui peut être d'une nature documentaire. Ainsi, la représentation des connaissances ne se fait pas seulement en utilisant les langages standardisés du web sémantique, il faut une intervention humaine permettant de fournir à ces langages les données et les modèles conceptuels nécessaires pour représenter les connaissances.

Concernant la question « Est-ce que l'utilisation du web sémantique en bibliothèque a changé vos pratiques documentaires ? » destinée aux bibliothèques qui utilisent le web sémantique, nous avons eu 9 réponses [sept professionnels de l'information ont répondu « oui » et deux ont répondu « non »]. Cette question a été complétée par deux autres ouvertes afin de mieux comprendre les enjeux du web sémantique sur les pratiques documentaires.

Le tableau 5.5 présente les réponses des enquêtés sur les raisons pour lesquelles le web sémantique a changé ou non leurs pratiques documentaires :

Pourquoi oui ?	Pourquoi Non ?
• Ouverture vers d'autres communautés professionnelles [Q30]	• Les répercussions des applications que nous sommes en train de mettre en place sont encore à évaluer. [Q03]
• Sans commentaire [Q26]	• Une des pages les plus consultées dans data.bnf.fr est celle qui rend compte des auteurs liés entre eux via des contributions communes, hors pour l'instant, ces liens entre auteurs ne sont pas exprimés dans le RDF.[Q16]
	• Nous étions déjà très attentifs à l'interopérabilité des données [Q17]
	• Parce qu'il s'agit toujours de suivre des normes de description.[Q23]
	• il s'agit surtout d'un balisage des données, transparent pour le documentaliste[Q25]
	• En l'état, il s'agit plutôt d'une perspective qui est devant nous. Nous savons que nous devons évoluer vers ces standards, mais nos outils n'évoluent pas nécessairement au même rythme. [Q28]
	• RDA est en cours de déploiement [Q24]

Tableau 5.5 - Les raisons pour lesquelles le web sémantique a changé ou non les pratiques documentaires des professionnels de l'information enquêtés

Une analyse thématique des réponses liées à la question (Pourquoi non?) est représentée dans le tableau 5.6 ci-dessous :

Thème	Sous-thème	Texte
Evolution du métier	➢ Normes ➢ Standards	• En l'état, il s'agit plutôt d'une perspective qui est devant nous. Nous savons que nous devons évoluer vers ces standards, mais nos outils n'évoluent pas nécessairement au même rythme. [Q28] • Parce qu'il s'agit toujours de suivre des normes de description.[Q23] • RDA est en cours de déploiement[Q24]
Pratiques au cœur du métier	➢ Interopérabilité ➢ Liens entre les données	• Nous étions déjà très attentifs à l'interopérabilité des données[Q17] • Une des pages les plus consultées dans data.bnf.fr est celle qui rend compte des auteurs liés entre eux via des contributions communes, hors pour

		l'instant, ces liens entre auteurs ne sont pas exprimés dans le RDF.[Q16] • il s'agit surtout d'un balisage des données, transparent pour le documentaliste[Q25]

Tableau 5.6 – Analyse thématique de la question « pourquoi le web sémantique n'a pas changé vos pratiques documentaires »

Les enquêtés qui pensent que le web sémantique n'a pas changé leurs pratiques documentaires, considèrent, d'une part, que leur métier est en évolution permanente, et d'autre part, leurs pratiques documentaires rejoignent les objectifs du web sémantique. Plus précisément, le cœur du métier des professionnels de l'information se concentre autour de l'identification, la description, la gestion, la diffusion et le partage des données, ainsi que les notions de l'interopérabilité et les liens entre les données ne sont pas étrangers à l'univers des bibliothèques. L'évolution du métier se fait au niveau de la technique utilisée pour exercer ces tâches tout en répondant aux besoins des usagers. Le web sémantique semble être donc une technique permettant aux professionnels de l'information d'exercer leurs pratiques documentaires et de les faire évoluer.

Un enquêté a répondu que le web sémantique a changé ses pratiques documentaires, cela en s'ouvrant vers d'autres communautés professionnelles. Ceci est dû au fait que le web sémantique permet aux bibliothèques de collaborer et de partager les données avec d'autres communautés sur le web.

5.2.6 Les apports du web sémantiques en bibliothèques

Toutes les bibliothèques enquêtées ont répondu à cette question, c'est-à-dire celles qui utilisent ou non le web sémantique. Le diagramme ci-dessous représentent les résultats obtenus :

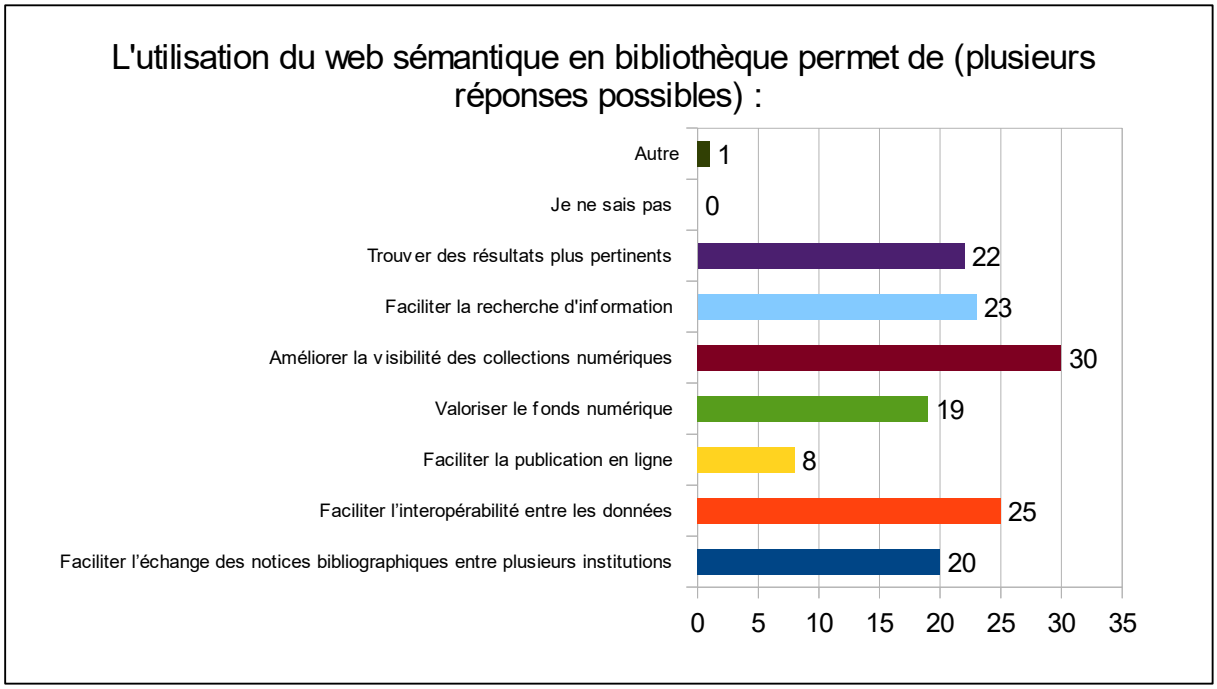


Figure 5.13 : Les apports du web sémantique en bibliothèques numériques

La réponse de la catégorie « autre » est : enrichissement des métadonnées, exigence de structuration et de modélisation.

Les résultats montrent que le web sémantique permet d'améliorer la visibilité des données sur le web selon 90 % des bibliothèques enquêtées (le résultat le plus élevé des réponses suggérées). Cela valide notre hypothèse :

H 3

Les bibliothèques numériques peuvent se servir du web sémantique pour proposer de nouveaux services plus adaptés aux usagers du web et pour rendre les bibliothèques plus visibles.

5.2.7 Les obstacles à l'utilisation du web sémantique en bibliothèques

Les réponses proviennent de la totalité des bibliothèques enquêtées. Le diagramme ci-dessous synthétise les réponses :

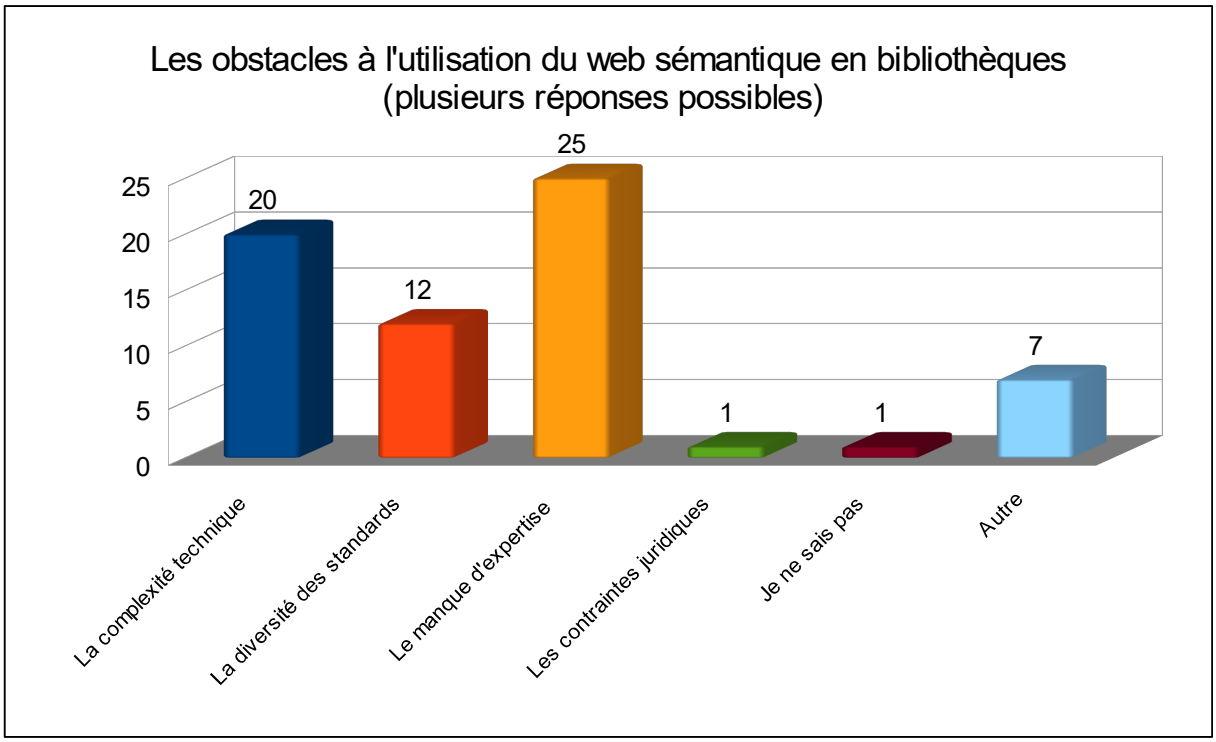


Figure 5.14 : Les obstacles à l'utilisation du web sémantique en bibliothèques numériques

Les réponses de la catégorie « autre » sont :

- Les grandes agences bibliographiques nationales sont encore à une phase d'étude et les autres établissements n'ont pas encore sauté le pas.
- Absence de logiciels de bibliothèque intégrant le Web sémantique
- outils en permettant la pleine utilisation et migration/transformation des données
- Les outils ne sont pas matures et complexes à utiliser. Les outils libres sont peu ou mal documentés. Il est également difficile d'identifier les vocabulaires ou ontologies utilisés. La plupart des ontologies sont des usines à gaz.
- la difficulté à effectuer cette transition dans des établissements de petite taille avec des moyens limités (en particulier RH informatique)
- Contraintes administratives et difficultés à fédérer tous les acteurs
- Le retard des fournisseurs de SIGB ou des fournisseurs de ressources électronique

Les résultats montrent que les obstacles à l'utilisation du web sémantique se concentrent autour de trois éléments: le manque d'expertise, la complexité technique du web sémantique et la complexité des standards.

5.2.8 Le web sémantique et le futur des bibliothèques

La plupart des enquêtés (32 sur 33) pensent que le web sémantique est important pour le futur des bibliothèques. Le tableau 5.7 ci-dessous analyse leurs réponses portant sur l'importance du web sémantique pour le futur des bibliothèques numériques.

Pourquoi le web sémantique est important pour le futur des bibliothèques ?		
Thème	**Sous-thème**	**Texte**
Visibilité	• Web • référencement • catalogues • données	1. Il constitue une occasion essentielle d'améliorer la visibilité des données des bibliothèques sur le web [Q03] 2. Mise en valeur de nos données et collections en ligne[Q05] 3. Ainsi une plus grande visibilité des catalogues.[Q06] 4. visibilité du catalogue[Q08] 5. Cela évitera que nos données ne deviennent complètement invisibles [Q17] 6. Améliorer le référencement et la visibilité des catalogues et bibliothèques numériques, directement dans le Web[Q19] 7. nécessaire destiné à permettre aux bibliothèques d'être plus visibles sur Internet[Q20] 8. Parce que cette notion est au cœur de l'identification de l'information sur Internet.[Q23] 9. Cela permettra le décloisonnement des données produites et leur inclusion dans le web [Q14]
Acteur du web	• ouverture • données fiables • évolution du web	1. Le web sémantique permettra aux bibliothèques de sortir de l'univers qui leur est propre et d'être un acteur déterminant du web. Mettre en avant la fiabilité des données des bibliothèques sur le web[Q03] 2. Il est important pour les bibliothèques de suivre les dernières évolutions techniques du Web, le Web sémantique étant certainement amené à devenir un standard quasi obligatoire dans les prochaines années.[Q07] 3. Sortir des outils et formats propriétaires, optimiser les possibilités d'internet[Q15] 4. Devenir acteur du Web sémantique est la voie naturelle pour les bibliothèques, la seule en tout cas qui permettrait aux

		bibliothèques d'exister et de peser dans la société[Q30]
Partage de données	• Interopérabilité • réutilisation des données • échange des données • standards	1. Interopérabilité[Q02] 2. Permet une meilleure interopérabilité et échange des données, leur enrichissement. [Q06] 3. dissémination des notices, partage de qualité sur les réseaux sociaux, etc. [Q08] 4. Société de partage et de communication, les bibliothèques doivent évoluer avec. [Q12] 5. Permettre l'interopérabilité des silos de données[Q24] 6. la donnée sera au cœur de l'échange d'information et elle doit être standardisée au maximum pour entrer en interaction avec les autres données du web et être correctement exploitée[Q25] 7. Faciliter la réutilisation des données[Q03]
Recherche d'information	• Qualité des réponses • pertinence	1. Qualité des réponses[Q08] 2. Pour une information mieux qualifiée et des recherches plus pertinentes[Q10] 3. Le web sémantique offre ces potentialités avec des données de bonnes qualités et justes. C'est à mon avis l'une des compétences que peut apporter une bibliothèque.[Q21] 4. Pour passer d'un catalogue centré sur le document à un catalogue centré sur l'information[Q32]

Tableau 5.7 – Analyse thématique de la question sur l'importance du web sémantique pour les bibliothèques numériques

L'enquêté qui a répondu que le web sémantique n'est pas important pour le futur des bibliothèques, a donné cette explication:

« Ce qui importe principalement, c'est de rendre l'information bibliographique ergonomique du point de vue des usages du web. Le web sémantique peut être un outil pour cela, mais toute autre technologie du web (schema.org, knowledge graph, etc.) peut l'être également. »

L'analyse thématique des réponses montre que la plupart des enquêtés pensent que le web sémantique est important pour le futur du web sémantique. Leurs réponses ont été réparties sur 4 thèmes principaux : pour améliorer la visibilité sur le web, pour devenir acteur du web, pour partager les données et pour améliorer la recherche d'informations. Tandis qu'un seul enquêté a répondu le contraire en disant que le web sémantique est un outil parmi d'autres technologies du

web que la bibliothèque peut utiliser pour rendre l'information plus ergonomique.

Les résultats obtenus montrent, d'une part, que les bibliothèques numériques peuvent se servir du web sémantique pour proposer des services adaptés aux usages du web en améliorant la visibilité de leurs données sur la web. D'autre part, les bibliothèques peuvent devenir des acteurs actifs dans la mise en œuvre du web sémantique, cela grâce aux pratiques documentaires exercées par les professionnels de l'information pour gérer les données et les métadonnées.

5.2.9 Formation des professionnels de l'information à l'usage du web sémantique

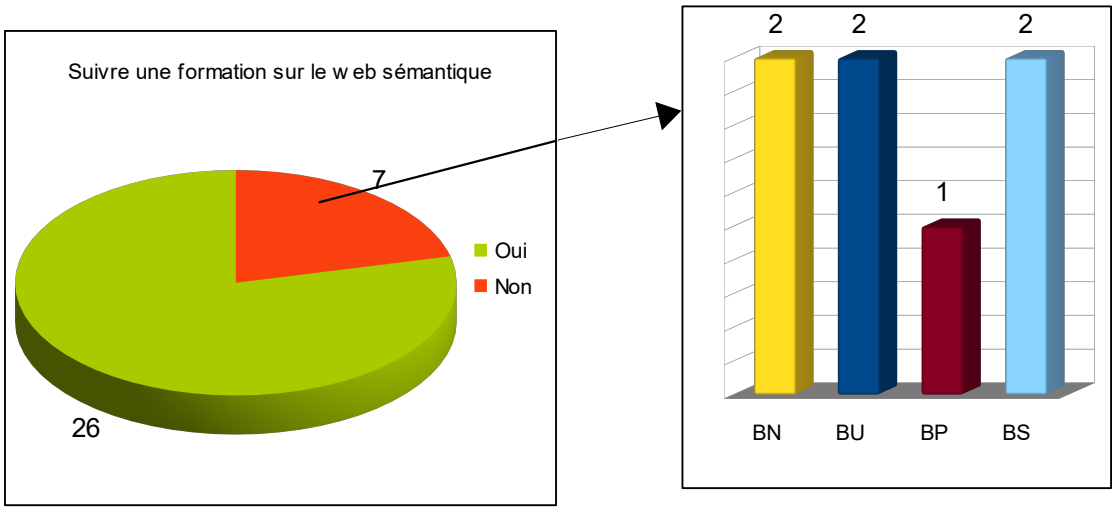


Figure 5.15: Répartition de nombre d'enquêtés souhaitant suivre ou non une formation sur le web sémantique

Les résultats indiquent que 79 % des professionnels de l'information souhaitent suivre une formation pour maîtriser les standards du web sémantique. Les professionnels de l'information sont intéressés par le web sémantique pour faire évoluer leurs compétences et leurs pratiques documentaires.

5.2.10 Limites du questionnaire

La phase de questionnaire présente des limites :

- La première limite réside dans la petite taille de notre échantillon. Les résultats pourraient être plus représentatifs avec un échantillon plus grand.
- La deuxième difficulté vient de l'analyse des réponses ouvertes. Celles ci sont parfois trop courtes pour pouvoir les analyser d'une manière approfondie.
- La dernière vient des réponses qui ne contiennent pas d'éléments significatifs, et n'apportent

aucune valeur aux résultats.

Ce questionnaire a été complété par une étude qualitative sous la forme d'entretiens. Le but étant d'approfondir notre enquête sur l'usage du web sémantique, et d'avoir des réponses plus détaillées permettant une analyse plus significative. Dans la section suivante, nous présentons les résultats de cette enquête.

5.3 Présentation et analyse des résultats des entretiens

5.3.1 Informations sur les enquêtés

Le tableau 5.8 ci-dessous présente les profils des enquêtés :

Numéro d'entretien	**Titre**	**Poste à la bibliothèque**
E 1	Bibliothécaire	Responsable du service informatique
E 2	Conservateur	Adjointe responsable du service informatique
E 3[76]	Conservateur	Chef du projet et son adjointe
E 4	Conservateur	Chef du service: ingénierie des métadonnées
E 5	Bibliothécaire	Responsable de gestion des ressources électroniques
E 6	Bibliothécaire	Responsable de catalogage centralisé et des métadonnées
E 7	Conservateur	Responsable des services des ressources informatiques documentaires
E 8	Ingénieur de recherche au CNRS	Responsable de la bibliothèque
E 9	Conservateur	Responsable du service
E 10	Bibliothécaire	Chef du projet de développement numérique

Tableau 5.8 - Les profils des enquêtés

Les résultats montrent que les titres des enquêtés varient entre deux groupes principaux les conservateurs (E2, E3, E4, E7 et E9) et les bibliothécaires (E1, E5, E6 et E10), à part un seul (E8) qui est « ingénieur de recherche au CNRS ». La plupart d'entre eux occupent des postes de « responsable de service » ou bien de « chef de projet » au sein de la bibliothèque. Ces postes se concentrent autour des services liés à la gestion des ressources numériques et des métadonnées.

76 Deux personnes ont été présentes lors de cet entretien

Le tableau 5.9 présente les missions des enquêtés au sein des bibliothèques :

Enquêtés	Missions
E1	• Gestion des moyens informatiques • Gestion des bibliothèque numériques • Gestion des ressources et des supports numériques
E2	• Coordination de la service informatique • Intervention au niveau des applications et des logiciels • Contact avec les prestataires
E3a, E3b	• Gestion de projet • Coordination des activités des développeurs et les activités des personnes • Développement de l'interface • Médiation autour des données [E3b]
E4	• Evolution des règles de catalogage • Gestion des formats • Normalisation documentaire • Coordination des identifiants
E5	• Coordination des services au public • Traitement bibliographique • Documentation électronique • Formation aux ressources en ligne des usagers
E6	• Gestion des métadonnées • Mise en ligne des thèses • Catalogage
E7	• Informatique documentaire • Mise en ligne des thèses • Signalement • Catalogage • Outils de découverte
E8	• Gestion de la bibliothèque • Gestion des moyens humains et financiers
E9	• Harmonisation des données bibliographiques et les autorités • Administration de la base de SIGB et de l'interface de recherche
E10	• Prise en charge de l'analyse des processus de traitements de données • Mise en place d'une plate-forme de service pour les professionnels des bibliothèques (projet)

Tableau 5.9 - Les missions des enquêtés au sein des bibliothèques

5.3.2 Définition d'une bibliothèque numérique

Dans l'état de l'art, nous avons présenté et analysé les différentes définitions des bibliothèques numériques proposées à la fois par la communauté de recherche et la communauté de pratique. Nous nous sommes intéressés aussi à la façon dont les professionnels de l'information (la communauté de pratique) définissent une bibliothèque numérique. Le tableau 5.10 ci-dessous regroupe les définitions proposées par les enquêtés[77] :

Enquêtés	Définitions des bibliothèques numériques
E1	« une bibliothèque numérique patrimoniale avec des images numérisées très classiques, une interface de recherche, feuilletage en ligne etc. uniquement sur des documents libres de droits […] ça c'était considéré comme la bibliothèque numérique pendant longtemps par opposition à la bibliothèque physique patrimoniale ou pas » « la notion de la bibliothèque numérique a commencé à s'accroître pour englober des documents qui n'étaient pas forcement libres, pas forcement patrimoniaux et pas forcement des livres »
E2	«au départ, il y a les collections numérisées et les collections numériques natives et aussi il y a tous les services associés à ces collections numériques. C'est d'abord l'outil de mise en ligne [...] ça fait vraiment la chose brute pour pouvoir présenter les documents sur le web, mais il y a aussi tout le site internet et il y a aussi tous les services autour de la dissémination des collections. Comment on peut faire disséminer les contenus de la bibliothèque numérique ? comment on peut les partager selon les fonctionnalités du web collaboratif ? »
E3a E3b	« je regrouperais sous ce terme toutes les traces de la présence des bibliothèques de façon numérique, mais c'est une question que je trouve difficile. Après quand on s'intéresse à faire la médiation c'est vrai qu'il faut s'intéresser à la façon dont les gens se représentent la bibliothèque numérique. Je pense que dans l'esprit des usagers c'est avoir accès aux documents numérisés. Je pense que dans l'esprit commun c'est ça, après je pense qu'ils ne se rendent pas compte de l'ingénierie qui a derrière pour accéder à ces documents numérisés, et c'est finalement beaucoup plus large que seulement les documents numérisés. » [E3a] « la bibliothèque numérique recouvre beaucoup la notion d'accès à distance [...] pour moi c'est paradoxale, c'est-à-dire qu'il y a d'une part la bibliothèque numérique et il y a d'autre part la documentation électronique, pour autant ce sont des objets qui ne devraient pas être séparés [..] finalement ce qui caractérisera la documentation électronique que c'est un document électronique natif contrairement à la bibliothèque numérique où le document était numérisé [..] Il y a un problème de signalement qui ne se pose pas de la même manière entre la documentation électronique et les bibliothèques numériques, au sens où en fait quand on numérise, on a pas forcement en tête les métadonnées, quand on produit un document nativement numérique au

77 Les entretiens ont été intégralement retranscrits, nous avons gardé la forme des propos des enquêtés en corrigeant quelques fautes comme la négation et en enlevant les phrases répétées et les phrases qui n'ont pas de sens.

	contraire souvent il est structuré, donc il y a une possibilité de produire des métadonnées, qui est quand même beaucoup plus facile. Tout l'enjeu aujourd'hui c'est d'arriver à transformer un gigantesque dépôt de versions numérisées de documents en quelque chose qui soit assez accessible que la documentation électronique, ça c'est le grand défi qui s'oppose à nous et le web sémantique peut être une solution pour ça, et je pense que c'est très lié, mais ça ne fait pas tout, il y a aussi toute la part de la médiation […]. Ce qu'il y a aussi derrière c'est la notion de conservation, c'est-à-dire par exemple tout ce qui est tombé dans le domaine public aujourd'hui, c'est vrai que les bibliothèques numériques permettent finalement la consultation de ces documents, encore il faut les trouver, c'est toujours le même problème.[..] je pense au départ c'est souvent une numérisation au km qui est plus une numérisation de prévention [..] qui n'est pas pensée dans la communication, elle est pensée du coté de la conservation alors que à la base, moi, ce que j'ai en tête, une bibliothèque numérique c'est quand même l'accès distant, c'est la communication, mais en fait on s'aperçoit que c'est un domaine qui est complètement déchiré entre cet aspect de conservation et cet aspect de communication » [E3b]. « c'est-à-dire que nous par exemple notre expérience c'est qu'on diffuse les données qui ont été faites par des bibliothécaires pour des bibliothèques, donc quand on diffuse sur le web disons qu'on leur donne une portée qu'elles n'avaient pas auparavant, une portée notamment d'ouverture qui n'est pas inhérente à la dimension d'initiation propre au monde des bibliothèques […] il y a une médiation à faire autour des données il y a une médiation à faire avec les gens qui doivent les réutiliser, c'est-à-dire pour pouvoir les réutiliser correctement il faut avoir conscience que ces données ont un caractère archivistique, celles qui ont été produites dans le cadre d'une activité, elles ne prennent sens que du point de vue de cette activité. »
E4	« alors il y a plein de manière de la définir, en fait ce qui vient tout de suite à l'esprit [...] c'est un dispositif d'accès aux collections numériques, je le vois vraiment sous l'aspect accès [...] si j'étais rigoureux, je dirais que c'est l'ensemble des dimensions de la bibliothèque [..] je pense systématiquement à l'accès, aux fonctions liées à l'accès et à la diffusion sur le web, mais ce n'est pas que la numérisation [.. .] ça peut s'élargir à toute optique de collections, en particulier des collections numériques comme les acquisitions des publications scientifiques qui se trouvent dans les archives ouvertes, ça pour moi aussi, ça correspond à une bibliothèque numérique, il y a des contenus amenés à faire partie de bibliothèques numériques. »
E5	« il y a deux cotés à voir il y a le coté stratégique de la bibliothèque numérique et il y a le coté technique. Au niveau stratégique, je pense qu'une bibliothèque numérique […] c'est l'action de se rendre visible sur le web […] c'est l'action de rendre sur le web des contenus et des savoirs qui sont conservés physiquement dans la bibliothèque, ou qui sont nativement numériques, voilà donc ça c'est au niveau stratégique […] techniquement quand on parle de bibliothèques numériques, dans le métier on a souvent l'idée d'un portail derrière, qui soit en fait une adresse unique dans laquelle on va mettre tout le contenu, qui sera accessible par une page d'accueil, une interface de recherche etc., et techniquement une bibliothèque numérique, je trouve que sa définition elle est physiquement sur le web, sur une infrastructure web. Il faut avoir un endroit avec un site web qui soit le site qui permet à la fois de stocker, de consulter. Mais ce qui est souvent n'est pas pris en compte dans les projets des bibliothèques numériques, c'est l'exposition des données qui sont dans ce site web vers les moteurs de recherches » « ça fait parti des critères c'est l'accès libre au contenu je pense que c'est

	incontournable, même si rien empêche dans une bibliothèque numérique de proposer un accès complètement libre et un accès à des ressources sous forme d'abonnement [..] qui sont accessibles uniquement sur l'identification pour telle ou telle catégorie d'usager »
E6	« C'est un terme que je n'utilise pas du tout » « On parle de ressources électroniques, on ne parle pas de bibliothèques numériques parce que bibliothèque numérique, on va penser à Gallica, on va penser à quelque chose qui a été numérisée etc […] En tout cas, nous, on va proposer un catalogue qui permet l'accès à des ressources électroniques et des ressources papier, c'est un accès uniquement à distance»
E7	« ça va être la mise à disposition des documents qui ne sont pas physiques, donc des documents numériques, aussi bien les documents que l'on peut produire [..] ainsi que les documents que l'on s'abonne ou que l'on achète, les ebook et les bases de données et les documents numérisés qui sont issus de la numérisation des collections patrimoniales, notamment qui sont mis en valeur soit par exposition, soit par la bibliothèque virtuelle, voilà c'est tout ça la bibliothèque numérique. »
E8	« c'est un ensemble de documents […]il y a une grande volonté d'assurer une pérennité de cette documentation et d'assurer cette pérennité au sens premier du terme, dans le temps, mais aussi sur les modalités d'accès avec des outils qui ne soient pas propriétaires [...] la bibliothèque numérique c'est quelque chose vers lequel on va c'est évident, qu'on prépare assez longtemps dans le domaine, notamment la numérisation [...] nou,s on travaille à gérer cette bibliothèque numérique comme une bibliothèque, c'est-à-dire en créant des collections et en faisant en sorte d'indexer d'identifier […] .
E9	« C'est une bibliothèque qui donne accès à des documents numériques, accès à une base bibliographique. Une bibliothèque traditionnellement est un ensemble organisé de collections. Une bibliothèque numérique doit avoir une interface qui ressemble à celle d'un site web, du coup pour moi voilà ça dépend beaucoup de la nature du site, le site est organisé comme une interface prioritairement pour donner accès à des documents. Pour moi il faut qu'il soit organisé. »
E10	« J'ai toujours milité pour l'insertion du numérique comme un support, un nouveau média parmi d'autres. Après la diversification des bibliothèques en médiathèques (CD, DVD), il m'a semblé que la migration vers la bibliothèque en ligne n'est pour moi qu'une poursuite de la diversification des modalités d'accès à l'information en bibliothèque. Donc je ne mettrai pas forcément la bibliothèque numérique comme un espace autonome complètement scindé de la bibliothèque dans ces différents services et ces différentes composantes. Le fait de pouvoir rendre des services à distance est aujourd'hui indispensable pour les administrations. Et dans la mesure où les bibliothèques donnent accès à des collections, une partie de leurs collections doit pouvoir être accessible en ligne. Et d'autre part c'est en phase avec l'évolution des pratiques culturelles puisqu'on constate, dans le domaine de la musique évidemment mais aussi dans celui de la vidéo et dans d'autres également, que l'usage de collections numérisées est déjà bien répandu donc c'est aussi une manière pour la bibliothèque de prendre acte de l'évolution des pratiques culturelles tout simplement. »

Tableau 5.10 – Les définitions de bibliothèque numérique selon les enquêtés.

L'analyse de ces différentes définitions montre que, d'une part, certains enquêtés font la distinction entre les « bibliothèques numériques patrimoniales » et les « ressources électroniques ». Les premières sont celles qui proposent l'accès à des documents numérisés libres de droits et les deuxièmes offrent l'accès à des documents numériques (libres de droits ou sous droits), des publications scientifiques, des abonnements etc.. D'autre part, certains enquêtés pensent que les bibliothèques numériques évoluent pour englober des nouveaux supports et des documents non patrimoniaux (sous droits) afin de répondre aux besoins et aux pratiques des usagers.

Leurs rôles varient entre la conservation pérenne et l'accès distant à l'information. Le tableau 5.11 ci-dessous regroupe trois catégories permettant de caractériser une bibliothèque numérique selon les enquêtés.

Les bibliothèques numériques		
Collections	**Organisation**	**Services**
• Documents numérisés • Documents numériques • Documents libres de droits • Documents sous droits • Documents patrimoniaux • Documents multisupports • Publications scientifiques • Abonnements • Collections organisées	• Dissémination • Partage • Identification • Conservation • Communication • Gestion des métadonnées • Diffusion sur le web • Visibilité sur le web • Numérisation • Indexation • Ouverture des données • Enrichissement des données	• Interface de recherche • Outil de mise en ligne • Consultation des documents • Dispositif d'accès aux collections numériques • Portail • Accès libre • Accès à distance

Tableau 5.11 – Les caractéristiques d'une bibliothèque numérique selon les enquêtés.

5.3.3 Les enjeux du web sémantique en bibliothèques numériques

Avant de parler des enjeux du web sémantique en bibliothèques numériques, nous avons demandé aux enquêtés lesquelles des expressions « web de données » ou « web sémantique » ils préfèrent utiliser et pour quelles raisons.

Les résultats indiquent qu'un seul enquêté préfère utiliser l'expression « web sémantique », deux considèrent que les deux expressions désignent la même chose, un enquêté utilise les deux expressions et cela selon le contexte et les autres préfèrent l'expression « web de données ». Une analyse thématique a été faite pour identifier les raisons pour lesquelles les enquêtés préfèrent utiliser telle ou telle expression. Les résultats figurent dans le tableau 5.12 ci-dessous :

	Thème	Sous-thème	Texte
Web de données / *Linked data/ Open Linked data*	Evolution du web	• Relier les données • Ouverture • Identifier les données	« quand on dit web de données on insiste sur le fait qu'on ne relie plus des documents mais on relie des données, et pour moi c'est ça l'intérêt de la nomination web de données, que ça informe de manière que je trouve assez clair sur qu'est ce que c'est l'enjeu par rapport au web traditionnel » [E4] « dans linked open data on a la notion d'ouverture et de vocabulaire contrôlé. Donc c'est plus simple à mon avis de partir de là pour expliquer aujourd'hui, qu'il n' y a pas de différence sémantique mais il y a une différence vraiment de point de vue du traitement de l'objet »[E3a]
	Clarté de terme	• Précision	« je préfère le terme de linked open data parce que c'est moins trompeur » [E3a] « parce ce que je le trouve effectivement plus précis, et à la fois plus circonscrit, même si peut être que ça donne l'impression d'être moins ambitieux dans un premier temps, mais du coup ça me semble plus réaliste, et ça me semble être plus en phase avec justement ce que vous évoquiez, c'est à dire les missions et les activités des bibliothécaires »[E10] « C'est vrai que le web de données c'est beaucoup plus parlant ».[E6]
	Documentation	• Métadonnées • Organisation • Diffusion	« par rapport aux bibliothécaires dont la mission est finalement la modélisation, l'organisation et la mise à disposition des savoirs sous toutes leurs formes, je trouve que la question de présenter ça sous l'angle de linked open data me semble être un pas plus facile à franchir et plus opérant, parce qu'on remet au cœur la question de la métadonnée et des relations qu'on construit entre ces métadonnées. »[E10] « Et c'est ces relations qui font sens, et donc sans aller chercher l'adjectif sémantique, en parlant de relation, on rejoint tout à fait la modélisation FRBR etc., et en ajoutant le terme « open » ça nous remet dans une posture d'agent public constituant une connaissance au service de tous. Donc c'est aussi pour derrière ces valeurs là que je

			privilégie le terme de « linked open data ». [E10]
Web sémantique	Ambiguïté du terme	• Intelligence artificielle • Linguistique	« Web sémantique c'est trompeur, c'est-à-dire on a l'impression qu'on fait de l'intelligence artificielle et que c'est que linguistique »[E3a] « parce que le terme du web sémantique a donné lieu à des interprétations différentes » « le web sémantique un peu confus, parce qu'on va penser à la traduction de langage naturel qui n'est pas le cas »[E10]
	Données structurées		« le fait de dire sémantique on va insister derrière qu'il y a des données structurées »[E4]
	Fondement historique	• Tradition scientifique	«quand j'écris un article sur le web sémantique, j'emploie cette expression là, parce qu'elle regroupe à mon avis une tradition scientifique qui est sémantique pour le coup, qui est une tradition du signe scientifique et de ce point de vue là, je trouve le web sémantique est une bonne expression »[E3a] « historiquement on a appelé ça web sémantique dans la fin des années 90 »[E6] « la différence n'est pas trop clair, en gros le web sémantique c'est plus les standards et les formats, et le web de données c'est un peu le résultat de ce qu'on veut en faire, je trouve que le coté chronologique est plus intéressant de dire on est parti de ça» [E6]
	Accès aux savoirs	• Accès universel	« parce qu'il est plus ambitieux, sachant qu'on est loin, c'est une belle utopie en général, je trouve que les bibliothèques de toute façon depuis les antiquités fonctionnent sur des utopies, la bibliothèque d'Alexandrie a fonctionné sur l'utopie de donner accès à tous les savants dans un espace unique, ça été aussi le cas pour tout ce qui a essayé de faire Otlet et l'institut bibliographique universel, c'est un peu toujours vers quoi le bibliothécaire essaye d'avoir un jour, un endroit où les savoirs humains soient accessibles à tous voilà Internet c'est un outil qui pourrait permettre ça »[E7]

Tableau 5.12 – Analyse thématique des réponses à la question concernant l'utilisation de l'expression « web sémantique » et « web de données »

En résumé, l'expression «web de données » est utilisée parce qu'elle est plus précise, elle indique d'une manière claire une évolution d'un web où nous pouvons lier les documents vers un web où nous pouvons relier les données entre elles. En plus, cette expression est plus parlante aux professionnels de l'information parce qu'elle touche le cœur de leur métier : l'organisation et la diffusion des données et des métadonnées. En revanche, l'expression «web sémantique », bien qu'elle soit plus ambitieuse, est employée pour privilégier le coté historique et l'évolution chronologique du terme.

Après avoir analysé les différents points de vue liés à la terminologie utilisée par les enquêtés, nous présentons par la suite les enjeux liés à l'utilisation du web sémantique en bibliothèques numériques.

Les résultats figurent dans le tableau 5.13 ci dessous :

Les enjeux du web sémantique en bibliothèques numériques		
Thème	**Sous-thème**	**Texte**
Visibilité	• Signalement • Référencement • Web	« la visibilité via les moteurs de recherche, parce que actuellement tous les catalogues de bibliothèques sont dans le web profond»[E10] « le web sémantique apporte avant tout une visibilité supplémentaire aux données »[E2] « l'exposition de nos données, parce que si on n'expose pas nos données, les gens ne savent pas qu'elles existent, un fond qui n'est pas signalé c'est un fond qui est mort »[E2] « rendre visible les données le plus possible sur le web, pour les bibliothèques universitaires c'est vraiment un enjeu crucial »[E2] « il y a un enjeu de signalement […] ça répond à un grand enjeu de visibilité des données et des documents des bibliothèques »[E3b] « la modélisation RDF aide beaucoup à faire remonter les résultats dans les moteurs de recherche »[E3a] « le web sémantique c'est l'occasion de redonner aux bibliothèques et aux établissements culturels une visibilité importante sur le web »[E7] « on a la possibilité aujourd'hui de passer à une dimension supérieure au niveau de la diffusion, de signalement de nos ressources avec celles du web et puis participer à ces mouvements d'ouverture des données publiques » [E7] « La visibilité je pense que c'est le premier aspect […] mais une visibilité intéressante à l'exploitation de nos

		données et à la réutilisation des données […] voilà à la fois qu'on soit visible et que ça soit intéressant à l'utilisateur final » [E6]
Recherche d'information	• Interface de recherche	« créer des interfaces qui mêlent à la fois les ressources qu'on peut trouver dans la bibliothèque, dans le service de documentation et les données qu'on peut trouver dans les laboratoires ... » [E7] « faciliter la recherche d'information […] les enjeux pour moi c'est tout ce défi d'enrichissement, de faciliter la recherche de l'information »[E8] « enrichir les interfaces de recherche […] de créer des données qui n'existent pas [..] à partir des données extraites ailleurs »[E9]
Evolution	• Métier • Ouverture • Langages documentaires • Catalogue	« gagner de la qualité sans mobiliser des ressources humaines, toujours pareil et c'est là où ça devient extrêmement complexe, parce qu'il faut arriver à passer du marc, des outils fermés propriétaires etc. à une logique complètement ouverte, complètement orientée web »[E1] « Introduction des nouveaux formats dans les bibliothèques […] pour créer des métadonnées pérennes »[E2] « Les professionnels sont intéressés par l'évolution des catalogues […] parce que la modélisation RDF se fait autour d'entités et l'évolution de catalogue va vraiment vers ce modèle d'entités. »[E3a] « si on rate le virage on va avoir de moins en moins du poids dans la société » [E7] « le web sémantique est ainsi une boite à outils parmi d'autres qui participent à l'évolution de notre métier »[E10] « l'enjeu est aussi de passer d'un paradigme où on était centré sur le document à un paradigme où on est centré sur l'information et sur la donnée […] donc l'enjeu est de passer du catalogue centré sur le document au catalogue centré sur l'information»[E10]
Communication	• Communautés • Partage • Réutilisation • Ouverture • Interopérabilité	« Les gens qui étaient intéressés par les données en RDF n'étaient pas forcement des gens du monde du livre et des bibliothèques, c'étaient plutôt des gens des laboratoires extérieurs »[E3a] « ça permet d'ouvrir la communauté des bibliothèques sur d'autres communautés, c'est-à-dire faire parler des communautés entre elles, qui ne parlaient pas auparavant »[E3a] « ça permet de faire communiquer des communautés qui ne parlent pas, parce que les formats ne parlent pas entre eux, comme les musées les archives les bibliothèques, c'est l'interopérabilité effectivement »[E6]

		« dessiloter les applications, on voulait pouvoir avoir un modèle qui soit organisé autour d'entités et qui puisse faire communiquer les bases de données entre elles » [E3a] «un style d'interopérabilité qui va permettre d'ouvrir beaucoup le champs, et avoir des nouveaux types de partenariat autour des données. »[E4] « les enjeux peuvent être justement la réutilisation des métadonnées, et de pouvoir réadapter les métadonnées qui sont fournies par d'autres, pour reproduire de l'information différente, et de pouvoir mettre en valeur certaines informations qui sont disponibles, parce qu'elles sont conformes aux standards du web sémantique »[E5] « c'est de décloisonner nos données […] une mise en relation de nos contenus, à la fois pour nous enrichir nous même, pour profiter de ce qui est fait dans wikipédia par exemple, mais aussi redistribuer ce qu'on produit me semble essentiel »[E10] « si ces informations qu'on construit sont disséminées et qu'elles servent à enrichir d'autres pratiques ailleurs alors l'objectif est atteint »[E10]
Acteur du web	• Enrichir le web • Valeur ajoutée • Contenus rares	« on a un rôle de devenir acteur du web sémantique, ce n'est pas simplement être présent sur le web et être visible, à mon avis c'est être acteur du web sémantique, essayer de faire entendre notre voix auprès de ceux qui font l'architecture du web, auprès W3C, essayer de peser sur les choix stratégiques pour essayer d'améliorer la société, tout simplement ça fait partie aussi des missions des bibliothèques, je pense au contraire l'enjeu si on n'arrive pas à relever le défi, on risque de devenir des établissements de seconde zone, de moins en moins fréquentés »[E7] « commencer par une bibliothèque numérique en mettant les corpus pour lesquels on a une vraie valeur à rajouter, on va venir enrichir le web parce qu'on est les seuls à avoir ces contenus là »[E7]
Usager	• Enrichir les pratiques	« enrichir l'expérience de l'utilisateur »[E9] « l'enjeu qui est derrière c'est de mettre d'une part nos collections sur les parcours des internautes, parce que actuellement on est bien trop souvent dans le web profond dans certains cas, ou bien très mal référencés dans d'autres »[E10] «Pour avoir une vision de la bibliothèque qui est centrée sur le fait de construire et de produire de l'information et de la richesse au service des citoyens quels que soient leurs pratiques »[E10]

Tableau 5.13 – Analyse thématique de la question sur les enjeux du web sémantique en

bibliothèques numériques

Les résultats nous permettent de valider l'hypothèse **H3** en montrant que la visibilité sur le web constitue un enjeu important de l'utilisation du web sémantique en bibliothèques numériques. Plus précisément, les professionnels de l'information pensent que le web sémantique permet d'apporter plus de visibilité aux données des bibliothèques numériques grâce au signalement et au référencement via les moteurs de recherche.

Ensuite, les résultats ont montré que la communication de la bibliothèque avec l'extérieur, que ce soit au niveau technique (standards, formats, systèmes) et au niveau partage et réutilisation des données avec d'autres communautés, est un enjeu essentiel à l'utilisation du web sémantique en bibliothèques numériques.

Le web sémantique est considéré comme une évolution du métier des professionnels de l'information surtout au niveau des formats et des langages utilisés pour identifier et lier les données entre elles afin de les diffuser et de les partager avec d'autres communautés. Ensuite, le web sémantique facilite la recherche d'information en améliorant les interfaces de recherche permettant de regrouper les données autour des entités et en valorisant la découverte des nouvelles informations.

Enfin, le web sémantique offre aux bibliothèques une opportunité de devenir des acteurs sur le web. Possédant des collections numériques de valeur, les bibliothèques numériques doivent continuer à jouer un rôle important dans la diffusion des savoirs dans la société.

5.3.4 Les obstacles à l'utilisation du web sémantique en bibliothèques numériques

L'utilisation du web sémantique en bibliothèques numériques n'est pas sans obstacles. Une analyse thématique a été faite pour examiner les réponses des enquêtés.

Les résultats figurent dans le tableau 5.14 ci-dessous :

Thème	Sous-thème	Texte
Niveau technique	• Logiciels	« le seul élément commun à tous les collègues, c'est le logiciel, tant que le logiciel n'est pas mûr pour le web sémantique, ça sert à rien. On a un logiciel très marc très métier, voilà qui est costaud, mais qui ne passera pas la transition bibliographique, c'est un des problèmes qu'on aura à gérer d'ici quelques années, et ensuite il est incapable de fournir des triplets rdf, il est incapable d'intégrer des notices xml, c'est compliqué

		donc voilà il faudra évoluer » [E1]
Niveau stratégique	• Coût • Référentiels • Interface • Méconnaissances des compétences des professionnels de l'information • Identifiants pérennes	« il faut avoir certains moyens financiers [..] on a contacté des sociétés et effectivement les sociétés qui proposent des outils puissants d'alignements basés sur les technologies du web sémantique les vendent très cher, il faut avoir un budget conséquent et après il faut qu'on puisse maintenir le projet sur la durée »[E9] « il y a tout un problème d'arbitrage de sujet, et on s'est dit, est ce qu'on se fonde sur la manière dont fonctionne les référentiels de la bnf, est ce qu'on utilise d'autres référentiels, chez nous par exemple, une bibliothèque spécialisée dans le domaine de la santé, on a d'autres vocabulaires contrôlés, on a le thesaurus MeSH» [E2] « il y a toutes ces questions d'affichage aussi qui sont importantes, mais il y a quelque chose que je trouve pour lesquelles les universités ont beaucoup de retard, c'est la création d'interface de type data.gouv.fr ou data.bnf, c'est-à-dire une interface qui mêle à la fois les ressources qu'on peut trouver dans la bibliothèque, dans le service de documentation, qu'on peut trouver dans les laboratoires, et là je pense que les services communs de documentation doivent être pilotes de ces projets là, et devraient initier des projets d'interfaces, qu'on soit porteur, mais malheureusement ce n'est pas la cas, soit parce que les bibliothèques se heurtent à l'organisation administrative comme c'était le cas ici, à partir du moment où on a évoqué ça, on s'est pris la réaction tout de suite, c'est pas à vous de faire ça, et nous, on est juste partenaires »[E7] « Il y a un vrai défi institutionnel de mutualisation, de l'exposition des données, je pense à celles qui ont vocation à être mutualisées, donc pas les données locales mais justement la question de la notion d'œuvre etc. » [E10] « le défi va toujours être les identifiants»[E1]
Niveau métier	• Compétences à évoluer • Manque de conscience et de sensibilisation • Confiance en métadonnées produites par d'autres acteurs	« les compétences sont très largement distribuées, je ne sais pas combien il y a des catalogueurs, mais je pense qu'il y a une centaine, tous ces gens là ont été formés par la filière habituelle du métier du livre, dans laquelle le web sémantique était à peine abordé même les tous derniers qui sortent des formations »[E1] « c'est les compétences des collègues pour moi, c'est vraiment le gros défi, c'est vraiment de prendre conscience que le métier évolue »[E8] « j'ai envie de dire que les bibliothécaires sont souvent en manque de confiance en elles, et tous ces aspects l'informatique ça leur fait vraiment peur, aujourd'hui le web de données c'est quelque chose qui leur fait

		peur»[E8] « je pense qu'il faut acquérir des compétences sur le sujet, les compétences en interne ne suffiront pas … la difficulté ça consiste déjà pour pouvoir mobiliser de temps de travail pour se pencher sur ces problématiques, pour essayer d'en tirer des leçons en interne, deuxièmement joindre des projets ou des groupes de travail qui travaillent sur ces problématiques, il faut avoir du temps et des moyens »[E9] « Il y en a un qui concerne la formation des collègues, pour essayer de repérer quels peuvent être les enrichissements sémantiques et les alignements qui peuvent être utiles de faire, c'est-à-dire qu'il faut apprendre aux bibliothécaires à devenir non plus seulement des catalogueurs mais aussi des personnes capables de mettre en relation des données qui font sens, mais qui sont issues de silos différents et qui ne sont pas produits au sein de notre communauté métier, donc c'est une nouvelle habitude à prendre, comment les juger fiables et comment ensuite les mettre en relation »[E10]

Tableau 5.14 – Analyse des réponses à la question sur les obstacles à l'utilisation du web sémantique en bibliothèques numériques

Les résultats montrent que les obstacles à l'utilisation du web sémantique sont peu de nature technique. C'est au niveau métier et au niveau stratégique, que les bibliothèques et les professionnels se sont confrontés à des difficultés. Au niveau stratégique, la mise en place du web sémantique nécessite de prendre en compte le coté financier du projet, les interfaces de recherche, les choix des référentiels sur lesquels il faut s'appuyer pour représenter les données et la reconnaissance des compétences des professionnels de l'information dans ce domaine.

Au niveau métier, les professionnels de l'information ne sont pas formés à utiliser les standards du web sémantique. C'est pourquoi il faut qu'ils prennent des initiatives permettant de suivre ces évolutions, et de collaborer avec d'autres communautés.

5.3.5 Le web sémantique et les pratiques des professionnels de l'information

Nous nous intéressons dans cette thèse aux pratiques des professionnels de l'information travaillant en bibliothèques numériques. Plus précisément, nous cherchons à savoir comment les pratiques professionnelles sont influencées par les technologies en général et le web sémantique en particulier. Et réciproquement, nous nous intéressons aux pratiques des professionnels de l'information qui peuvent être mobilisées afin de mieux accompagner la mise en place du web

sémantique.

Une analyse thématique a été faite pour examiner les réponses des enquêtés. Dans un premier temps, nous avons classé les réponses liées à la question de l'influence de la technologie sur les pratiques professionnelles en trois catégories : positive (7 réponses), négative (une réponse) et neutre (2 réponses).

Nous donnons ci-dessous des extraits des réponses positives :

> « Si on prend le métier de bibliothécaire sans l'informatique et bibliothécaire avec l'informatique c'est rien à voir. C'est arrivé quand même tôt les premiers langages MARC […] Après il y a eu la logique de réseau avec l'arrivée d'internet [qui a permis] d'échanger des notices bibliographiques. »[E2]

> « c'est aspect de traitement de données qui change beaucoup […] finalement le bibliothécaire n'a plus le monopole de ses données, donc il n'est pas forcement amené à produire, mais il est amené à vérifier à croiser les références, il devient plus expert que initiateur […] le bibliothécaire avait un rôle traditionnel, il avait quasiment le monopole institutionnel sur l'accès à l'information, et le fait d'ouvrir l'accès à l'information, oblige les gens vraiment à avoir des réflexions sur ces cadres de production et d'accès à l'information [..] »[E3a]

> « elle est importante très importante elle implique qu'on fonctionne différemment […] les bibliothécaires ne se retrouvent plus seuls isolés dans leurs coins. »[E8]

> « Pour un service comme le mien c'est essentiel, on est censé suivre l'évolution des formats des langages des logiciels, pour moi c'est central c'est le cœur de mon métier » [E9]

> « Je la trouve très positive, c'est-à-dire si je compare 20 ans auparavant, aujourd'hui on a des technologies qui nous permettent d'être un peu à la hauteur de nos ambitions. C'était quand même plus délicat de travailler sur l'échange des données il y a 20 ans, alors qu'internet venait tout juste de naître et qu'on avait des capacités d'échange et de traitement de données très limitées, donc je trouve que c'est très positif »[E10]

> « La technologie va avoir une grosse influence sur nos pratiques parce qu'on s'ouvre à l'extérieur […] on va travailler en collaboration avec un autre collègue […]. »[E8]

> « avant la description était le plus important, aujourd'hui c'est le lien […] on passe moins de temps à décrire, on passe beaucoup de temps à vérifier et à lier un auteur à un livre etc. à s'assurer la fiabilité de ce lien […] aujourd'hui on récupère, on passe de catalogueurs au catalinkeurs »[E6]

Une réponse négative :

> «L'omniprésence de Google le moteur de recherche aurait tendance à renvoyer les catalogues dans la préhistoire […] les pratiques documentaires sont balayées par les moteurs de recherche. Là il faut qu'on évolue sinon nos pratiques documentaires seront obsolètes»[E1].

Deux réponses neutres :

> « on a grandi avec la technologie, donc elle n'a pas d'influence, ça existait toujours dans nos pratiques » [E3b]

> « Je n'ai pas d'opinion, je suis ni pour ni contre »[E5].

Les résultats montrent que les enquêtés trouvent que l'influence de la technologie sur les pratiques professionnelles est plutôt positive. D'une part, elle a aidé les professionnels de l'information à sortir de leur isolement et à communiquer avec l'extérieur, ceci en collaborant et en échangeant les données avec d'autres institutions. D'autre part, étant au cœur du métier, elle est nécessaire pour évoluer les pratiques afin de répondre aux besoins des usagers.

Ensuite, nous avons demandé aux enquêtés quelles sont les pratiques professionnelles à mobiliser pour accompagner la mise en place du web sémantique en bibliothèques numériques.

Les réponses figurent dans le tableau 5.15 ci-dessous :

Les pratiques professionnelles mobilisées pour accompagner la mise en place du web sémantique en bibliothèques numériques		
Thème	**Sous-thème**	**Texte**
Compétences documentaires classiques	• Indexation • Description • Catalogage • Métadonnées • Normes	« ce sont des compétences documentaires classiques de type description, indexation catalogage »[E1] « C'est un enjeu pour les bibliothèques de proposer des métadonnées de qualité, justement les données des bibliothèques sont des données établies selon un consensus confiant rationnel, qui ont été vraiment sélectionnées et validées au niveau de leur qualité, et c'est vraiment là où les bibliothèques ont une carte à jouer, et d'autant plus l'intérêt de rendre visibles leur métadonnées, et là on voit l'efficacité du facteur humain face aux algorithmes »[E2] « pour moi c'est toute la partie indexation, c'est clairement sur ces compétences qu'on va apporter […] le cœur de notre métier qui va basculer je dirais, être même mis en valeur, j'espère que dans le web sémantique, c'est toute cette partie enrichissement par le biais de l'indexation de la gestion des référentiels, qui va être mis en valeur »[E8] « le web sémantique n'est pas quelque chose d'automatisable, si on n'a pas d'abord les données qui sont prêtes »[E7] « ça nous touche, parce que c'est lié à des projets des agents bibliographiques nationaux et de l'évolution des normes » [E9] « D'abord les compétences de base sur la question du traitement du document, le bagage technique de documentaliste et du bibliothécaire est indispensable […]

		donc je dirais ce bagage il ne faut pas le négliger et parfois il faut même en remettre une couche pour mettre à jour les collègues »[E10] « on a quelque chose qui est fait nul part ailleurs, c'est-à-dire on a l'habitude de cataloguer, on a l'habitude de créer de fichiers d'autorité, on a l'habitude de faire ça à un niveau très poussé surtout en France, et data.bnf a vraiment montré tout ce qui pouvait apporter le travail du bibliothécaire»[E7]
Compétences à acquérir	• Ouverture sur le web • Maîtriser les langages de balisages (Html, Xml) • RDF	« Si on dit web sémantique on dit web, et ça pour beaucoup de collègues c'est la révolution […] rédiger un article qui va être visible sur le web c'est une autre dimension, et pour l'instant ce n'est pas acquis, ça fait partie de la profession ça s 'apprend, il y en a qui ont été formés, il y en a qui refusent et puis il y en a qui sont entre les deux »[E1] « initiation aux bases des autres langages mobilisés pour le web (html, xml, etc) […] ils offrent un terrain, une couche très souvent utilisée dans le cadre du web sémantique qui me semble important de maîtriser […] il faut comprendre que c'est une manière de structurer l'information aussi »[E10] « je pense que ce sont des compétences informatiques notamment au niveau RDF pour accompagner vraiment le web sémantique en bibliothèques numériques »[E2]
Compétences informatiques	• Informatique • Informatique documentaire	« alors là on est dans l'évolution du métier, de l'apprentissage […] donc ce qui est sûr dans l'apprentissage du métier à mon avis, il y a une part d'informatique beaucoup plus conséquente, ça me paraît en fait inévitable […] Je pense qu'un bibliothécaire doit pouvoir avoir la main sur la technologie dans le sens où il doit pouvoir comprendre la technologie et l'infuser c'est-à-dire être à l'origine »[E3a] « connaissances en informatique documentaire »[E6]

Tableau 5.15 – Analyse thématique des réponses à la question sur les pratiques professionnelles mobilisées pour accompagner la mise en place du web sémantique en bibliothèques numériques

Les résultats montrent que les pratiques documentaires « classiques » viennent en premier pour accompagner la mise en place du web sémantique en bibliothèques numériques. Il s'agit des compétences sur le traitement des documents : l'identification, la description et l'indexation. Ce bagage documentaire pourrait être mis en valeur avec le web sémantique. Ce dernier se base sur des données structurées, or les bibliothèques ont l'habitude de structurer leurs données pour y faciliter l'accès. Cela valide notre hypothèse :

H1

Les professionnels de l'information peuvent jouer un rôle dans la mise en œuvre et l'évolution du web sémantique en introduisant leur expertise documentaire relative à la gestion des métadonnées.

Utilisées toutes seules, les pratiques documentaires ne suffisent pas pour accompagner l'usage du web sémantique en bibliothèques numériques. Il faut acquérir des compétences en informatique, plus précisément, il faut des connaissances de base sur les langages de balisage (Html, XML) et les standards du web sémantique (URI, RDF etc).

Pour comprendre la représentation des données dans le cadre du web sémantique, nous avons demandé aux enquêtés de nous expliquer, d'après leur expérience, comment les liens sont établis entre les données.

Une analyse des réponses figurent dans le tableau 5.16 ci-dessous :

Représentation des données dans le cadre du web sémantique		
Thème	**Sous-thème**	**Texte**
Médiation documentaire	• Travail intellectuel • Langages documentaires • Rameau • Formats de production • Formats de diffusion • FRBR • Catalogage • Vocabulaire contrôlée • Organisation des connaissances	« le web sémantique n'arrive pas tout seul, il arrive dans un monde documentaire où on a vu à un moment le FRBR » [E1] « rien que sur la modélisation, rien que le mapping entre les formats, ça demande un travail intellectuel, on a deux équipes ici qui ont participé à cette modélisation, en fait deux modélisateurs, des gens qui ont fourni le travail intellectuel de modélisation et de mapping dans le rdf » [E3a] « Rameau est déjà en SKOS, tout Rameau est publié en SKOS, Rameau est un vocabulaire contrôlé et en même temps joue le rôle d'ontologie, il définit des domaines [..] il continue d'être produit en Marc en Intermarc, le format de base de production de Rameau reste l'Intermarc, et à priori il va rester comme ça, et de fait on a pas dans l'idée qu'il faut reproduire nativement du RDF, c'est-à-dire on continuera à produire dans des formats d'autre que RDF, et il y aura toujours ce mapping qui sera fait entre les formats de production et les formats de diffusion, ce qui est sûr c'est que la production et la diffusion d'un point de vue format ça sera toujours distinct en tout cas dans notre politique » [E3a] «[Établir des liens] est un travail intellectuel, c'est-à-dire que c'est le catalogueur qui va faire le lien avec un auteur, qui va faire le lien avec une vedette rameau etc etc, le lien est fait dans la production, et même de plus en plus on commence à se demander si on ne pourrait pas demander

		[aux catalogueurs] de lier carrément à de URLs et non pas lier juste à un mot qui existe dans un référentiel, d'aller chercher l'URL, qui n'est pas forcement évident d'un point de vue de pratique de catalogage […] par ailleurs il faut savoir que nous quand on a commencé à frbriser le catalogue, c'est-à-dire à lier des auteurs à des œuvres ou des œuvres à des éditions, il y avait quand même une tradition de lier les identifiants dans leur production même, c'est-à-dire que dans MARC il y avait une zone [..] il y avait cette possibilité d'insérer des identifiants dans les zones du MARC et de lier les entités en elles » [E3a] « c'est plus une question d'architecture qu'une question de format, on peut très bien faire de linked data avec des données en MARC dans l'écosystème de production interne, parce que depuis 87 les catalogueurs font déjà des liens entre les notices bibliographiques et les notices d'autorités, ce que j'appelle de vintage linked data, les notices ont des identifiants et non pas des URIs, mais des identifiants qui sont bien gérés, derrière il y a des données en MARC, mais on a déjà des entités qui sont liées entre elles, et de ce point de vue là on a déjà une logique de production en entités, [..] c'est un travail de catalogueur, qui est non seulement de la production des données mais l'administration des liens » [E4] « au niveau métier, il faut un modèle qui peut rester du MARC, c'est ce que les gens ont l'habitude de faire, c'est sur ce format là que s'opère toute la coordination bibliographique, on peut s'appuyer la dessus et faire évoluer, le modèle physique ça peut être une base de données, ça peut être une base de données relationnelle du sql [..] on a plus de 300 personnes qui travaillent dans ce format là et qui le maîtrisent, et plutôt que de le changer d'une manière très destructive, le mieux est de le faire évoluer pour accorder de nouveaux besoins»E4 « on s'est dit qu'avec l'arrivée de la recherche par texte c'est la fin de vocabulaire contrôlé, et même c'est la fin des catalogues [..] alors que pour la simple raison que la recherche par texte ne donne pas d'indice sur la manière dont les connaissances sont organisées et liées entre elles. c'est un enjeu pour les bibliothèques de proposer des métadonnées de qualité » E2 « le web ressuscite les pratiques de la profession , en tout cas la profession telle qu'elle a été pensée au début de l'époque moderne, et à mon avis le RDF est un langage qui sort tout droit de la tradition scientifique de 17eme siècle »E3a

Tableau 5.16 – Analyse thématique des réponses liées à la représentation des données dans le cadre du web sémantique

Les résultats nous ont permis de valider notre hypothèse :

H 2

La représentation des connaissances ne se fait pas seulement en utilisant les langages informatiques et standardisés du web sémantique, il faut une intervention humaine permettant de fournir à ces langages les données et les modèles conceptuels nécessaires pour représenter les connaissances.

D'une part, l'analyse des résultats a montré que la représentation des données dans le cadre du web sémantique nécessite une médiation documentaire. Un travail intellectuel est nécessaire pour créer les liens entre les données, plus précisément, il s'agit du travail exercé par les catalogueurs. Les langages documentaires comme Rameau constituent une base conceptuelle sur laquelle les bibliothèques peuvent s'appuyer pour exposer leur données dans le web sémantique. Cela montre un retour aux langages contrôlés pour décrire et lier les données.

D'autre part, les enquêtés trouvent qu'au lieu de changer les formats utilisés en bibliothèques comme le format Marc, il faut mieux de les faire évoluer pour pouvoir diffuser les données sur le web.

5.3.6 Evolution du métier des professionnels de l'information

Nous avons clôturé l'entretien par une question sur l'avenir du métier des professionnels de l'information. Le but est de savoir quelles visions ont les professionnels de l'information sur l'évolution de leur métier en relation avec les développements techniques.

Le tableau 5.17 présente une analyse thématique des réponses à la question de l'évolution du métier des professionnels de l'information.

Comment voyez-vous l'avenir de votre métier ?		
Thème	**Sous-thème**	**Texte**
En évolution	• Continuité • Défis • Technologie • Informatique • Catalogage	« les outils de catalogage vont évoluer, les méthodes vont évoluer, c'est-à-dire qu'avant on récupérait des notices sur Electre, peut être dans quelques années on va les récupérer directement de databnf, qui eux même vont les récupérer de Oclc, ça sans doute va évoluer [...] mais il faudra toujours être capable de décrire correctement un document et de le rattacher à la bonne notice»[E1] « je pense qu'ils auront besoin toujours de nos compétences, dans notre équipe il y a des gens qui font du catalogage depuis 30 ans, et il reste de catalogage à faire, donc je pense comme tout le monde que

		l'informatique va prendre de plus en plus de place, ce sont des compétences qu'il va falloir les acquérir, qu'on va les demander de plus en plus aux jeunes, et d'une certaine manière dans nos pratiques de tous les jours on utilise de plus en plus le web, la technologie a une place très importante »[E3b] « le métier évolue vers le bibliothécaire polytotechnicien, c'est-à-dire quelqu'un qui est complètement ancré dans la synthèse politique entre ce que veulent les usagers et entre ce que veut lui [..] c'est-à-dire comment est ce qu'on infuse tout ça dans la technologie et on l'adapte à l'environnement technologique d'aujourd'hui, c'est vraiment infuser l'aspect politique documentaire du métier dans l'environnement technologique »[E3a] « je le vois encore une fois un métier très ancien la bibliothèque mais qui a toujours réussi à relever le défi des évolutions techniques ou technologiques »[E7] « c'est la continuité de tout ce qui s'est passé depuis toutes ces années [...] on est en perpétuelle remise en cause on est en perpétuelle évolution [...] il faut suivre ce qui se passe pour pouvoir s'y préparer »[E8]
Médiation	• autour des données : - réutilisation - gestion documentaire - gestion des métadonnées - qualité des données - visualisation • autour du public : - ateliers - accueil - animations - services • des savoirs - diffusion des savoirs - références pour la recherche d'information	«ils font beaucoup d'animations, beaucoup d'activités, la bibliothèque en train de devenir dans certains endroits vraiment un troisième lieu »[E9] « on sera toujours aussi multi casquettes, on sera toujours autant orienté donnée qu'orienté service aux lecteurs, ça sera toujours une question de gestion de réservoir des métadonnées et d'exposition des données sur le web que de rendre service à des lecteurs »[E2] « on va avoir un développement du contrôle de la qualité des données, le but de bibliothécaire sera de produire des données fiables, des données de confiance qu'elles soient interopérables et visibles et qu'elles puissent être aussi facilement réutilisables »[E2] «on a vraiment besoin d'un nouveau métier qui est vraiment la médiation autour de cet objet documentaire particulier qui est la donnée »[E3a] « Je pense que les bibliothèques peuvent aller dans deux directions entre les collections et les usagers. Du point de vue d'usager, il y a tout un mouvement qui va de la bibliothèque comme troisième lieu à la proposition des ateliers etc, qui je pense devrait se développer si on veut que la bibliothèque continue à prendre une place dans l'espace public, ça c'est du coté usager. Et coté collection [..] qu'on puisse valoriser l'expérience très importante qu'on a dans la description des collections, c'est-à-dire que les bibliothèques ont des années d'expertise en description, on a élaboré des normes et des formats extrêmement riches, et l'enjeu est d'arriver à

		conserver cette richesse et à l'améliorer en partenariat avec d'autres filières professionnelles»[E9] «Il me semble qu'on va aller, mais c'était déjà le cas, vers une plus grande spécialisation entre les bibliothèques d'un côté qui seront tournées vers la médiation et l'accueil du public et d'autre part ceux qui seront centrés sur la donnée et sur la gestion documentaire » [E10] « on a une place dans la société que personne ne la remet en cause [..] je pense que l'avenir de notre métier dépend de notre capacité à relever le défi du web sémantique, [..] je pense qu'il faut garder une place importante dans la diffusion des savoirs, il faut être des établissements de références pour la recherche d'informations et la recherche de connaissances »[E7]

Tableau 5.17 - Analyse thématique des réponses à la question de l'évolution du métier des professionnels de l'information.

Les résultats montrent que les professionnels de l'information pensent que leur métier est toujours en évolution permanente. Le cœur du travail est le même, il s'agit toujours de décrire les documents pour en faciliter l'accès. Les bibliothèques ont réussi à relever les défis techniques en adaptant les pratiques documentaires aux différents environnements technologiques.

Ensuite, les professionnels de l'information considèrent que leur métier va vers deux directions, celle de la médiation autour des données et celle de la médiation autour du public. Enfin, les bibliothèques doivent garder une place importante dans la société d'aujourd'hui, cela en continuant à diffuser des savoirs et en se positionnant comme des établissements de références pour la recherche d'information.

5.3.7 Limites des entretiens

Les entretiens présentent certaines limites :

- La première limite réside dans la petite taille de notre échantillon. Les résultats pourraient être plus représentatifs avec un échantillon plus grand.
- L'analyse thématique laisse une grande latitude au chercheur dans le codage du discours. Il s'agit « d'un métadiscours émis par le codeur sur le discours » [Ghiglione et Matalon, 1985, p.185]. De ce fait, cette méthode pourrait conduire le codeur à des interprétations subjectives dans la définition des thèmes.
- Le discours des enquêtés peut être influencé durant l'entretien. Les personnes interrogées répondent d'une manière à satisfaire l'enquêteur et non pas à ce qu'elles pensent. Cela se manifeste par une hésitation à répondre à certaines questions, ou bien par des commentaires : «est ce que à cela que vous attendez » « est ce que cette réponse vous convient » etc.

Conclusion du chapitre

Dans ce chapitre nous avons présenté les résultats de notre enquête sur les usages et les applications du web sémantique en bibliothèques numériques en France. Après avoir constitué notre sélection des bibliothèques numériques, nous avons mené une enquête par questionnaire suivi d'une enquête par entretien afin de tester nos hypothèses de recherche. Cette enquête nous a permis de collecter des données auprès des professionnels de l'information sur des problématiques liées à l'usage du web sémantique en bibliothèques.

L'analyse de la sélection nous a permis de classer les bibliothèques recensées dans quatre catégories : bibliothèques numériques nationales, bibliothèques numériques universitaires, bibliothèques numériques de lecture publique et bibliothèques numériques spécialisées. Ce classement était utile pour pouvoir mener des comparaisons permettant de mieux comprendre l'usage du web sémantique en lien avec les différentes catégories de bibliothèques.

Ensuite, nous avons analysé le questionnaire envoyé aux bibliothèques sélectionnées. Nous avons constaté que les bibliothèques numériques qui utilisent le web sémantique sont majoritairement des bibliothèques universitaires et de bibliothèques nationales[78].

La plupart des professionnels de l'information semblent être conscients des fondements principaux du web sémantique. En revanche, ils ont montré une méconnaissance des standards liés à la représentation des connaissances et des ontologies (OWL, SKOS).

Les tableaux ci-dessous synthétisent l'analyse du questionnaire :

Les bibliothèques numériques qui utilisent le web sémantique			
Type de bibliothèque	**Nombre**	**Standards**	**Pourquoi ?**
Bibliothèque numérique nationale	2	URI ; RDF ; OWL ; SKOS ; JSON-LD ; SPARQL ; Dublin CORe	• Améliorer la visibilité • Réutiliser des données • Faciliter l'interopérabilité
Bibliothèque numérique de lecture publique	1	RDF	• Améliorer la visibilité • Faciliter l'interopérabilité
Bibliothèque numérique universitaire	6	URI ; RDF ; JSON-LD ; SPARQL	• Améliorer la visibilité • Évoluer les catalogues
Bibliothèque numérique spécialisée	0		

Tableau 5.18 – Les bibliothèques qui utilisent le web sémantique (Résultats du questionnaire)

78 Six bibliothèques universitaires sur neuf, deux bibliothèques nationales et une bibliothèque publique utilisent le web sémantique

Les bibliothèques qui n'utilisent pas le web sémantique		
Type de bibliothèque	**Nombre**	**Pourquoi ?**
Bibliothèque numérique nationale	1	• Pas de réponse
Bibliothèque numérique de lecture publique	12	• Manque de compétences • Manque d'intérêt • Futur projet
Bibliothèque numérique universitaire	7	• Manque de compétences • Manque de moyens • Complexité technique • Futur projet
Bibliothèque numérique spécialisée	4	• Manque de compétences • Moyens financiers

Tableau 5.19 – Les bibliothèques qui n'utilisent pas le web sémantique (Résultats du questionnaire)

Importance du web sémantique pour le futur des bibliothèques	
Bibliothèque numérique nationale	• Interopérabilité • Échange des données
Bibliothèque numérique de lecture publique	• Mise en valeur des données • Visibilité des catalogues • Interopérabilité • Partage • Qualité des réponses
Bibliothèque numérique universitaire	• Visibilité des données • Acteur du web • Sortir des outils et formats prioritaires • Partage • Interopérabilité • Réutilisation des données
Bibliothèque numérique spécialisée	• Qualité des données

Tableau 5.20 – L'importance du web sémantique pour le futur des bibliothèques (Résultats du questionnaire)

Le tableau 5.21 ci-dessous résume l'analyse des entretiens :

Le web sémantique en bibliothèque				
Type de bibliothèque	**Les enjeux**	**Les obstacles**	**Les pratiques mobilisées**	**L'avenir du métier**
Bibliothèque nationale	- Améliorer la visibilité - Enrichir les interfaces de recherche - Évoluer les catalogues - Partager et réutiliser les données - S'ouvrir vers d'autres communautés - Enrichir l'expérience des usagers	- Moyens financiers - Manque de compétences en interne	- Pratiques informatiques - Compétences documentaires	- Nous aurons besoin de nos compétences - La technologie va prendre plus de place - Vers le bibliothécaire « polytotechnicien » - Médiation autour du public - Médiation autour des données - Valoriser les expertises de description
Bibliothèque de lecture publique	- Améliorer la visibilité - S'ouvrir sur le web - Faire évoluer le métier - Enrichir les pratiques des usagers	- Logiciels non adaptés au web sémantique - Identifiants pérennes - Formation des professionnels de l'information - Fiabilité des métadonnées produites par d'autre communautés - Manque de compétences - Mutualisation des données	- Compétences documentaires - Langages de balisage (html, xml)	- En évolution (récupération des données) - Médiation autour du public - Médiation autour des données
Bibliothèque universitaire	- Améliorer la visibilité - Avoir une interface unifiée - Avoir des métadonnées pérennes - Avoir une place dans la société - Améliorer l'interopérabilité - S'ouvrir vers d'autres communautés - Réutiliser les données - Devenir acteur du web	- Référentiels - Méconnaissances des compétences des professionnels de l'information	- Données et métadonnées de qualité - Efficacité du facteur humain - Informatique documentaire - Compétences en informatique	- Capacité de relever le défi technique - Gestion des métadonnées - Contrôle de qualité des données - Diffusion des savoirs - Établissements de références
Bibliothèque spécialisée	-Faciliter la recherche d'information	- Manque de confiance - Manque de compétences	- Compétences d'indexation	- En continuité avec nos pratiques - En évolution perpétuelle

Tableau 5.21 – L'usage du web sémantique en bibliothèques numériques (Résultats des entretiens)

Nous avons testé notre hypothèse de recherche **H 1** portant sur le rôle des professionnels de l'information dans la mise en place du web sémantique. Les résultats auprès des enquêtés qui utilisent le web sémantique ont montré que sept enquêtés sur neuf trouvaient que le web sémantique n'a pas changé leurs pratiques parce que d'un coté, cela fait partie de l'évolution du métier, et de l'autre, les notions d'interopérabilité et les liens entre les données relèvent de leur pratiques documentaires. De même, l'analyse du discours des enquêtés (les entretiens) a montré que les pratiques professionnelles qui seraient mobilisées pour accompagner la mise en place du web sémantique en bibliothèques relèvent surtout des pratiques documentaires classiques (l'indexation, la description, les métadonnées).

Nous avons, par la suite, testé l'hypothèse **H 2** portant sur la nécessité d'un travail intellectuel pour représenter les connaissances. Les résultats ont montré que l'utilisation du web sémantique nécessite une médiation documentaire. Les langages documentaires constituent la base des modèles conceptuels permettant de construire des ontologies dans le cadre du web sémantique.

Enfin, les résultats nous ont permis de tester notre troisième hypothèse **H 3** portant sur la visibilité des bibliothèques sur le web. L'analyse des résultats a montré que les bibliothèques numériques qui utilisent le web sémantique le font principalement pour améliorer leur visibilité sur le web. Aussi, les résultats du questionnaire ont indiqué que le web sémantique permet d'améliorer la visibilité des données sur le web pour 90 % des bibliothèques enquêtées. De même, les résultats des entretiens ont montré que la visibilité sur le web est un enjeu crucial pour les bibliothèques.

Chapitre 6

Les bibliothèques numériques : des nouveaux acteurs sur le web

Ce dernier chapitre vise à synthétiser, discuter et mettre en perspectives les résultats obtenus dans notre recherche.

Les bibliothèques numériques ont un rôle à jouer dans la société d'aujourd'hui. Possédant des collections numériques structurées et de qualité, elles doivent occuper une place plus importante sur le web.

Dans un premier temps, l'accent est porté sur les pratiques documentaires mises en valeur par notre recherche dans le cadre de l'usage du web sémantique en bibliothèques numériques. Il s'agit, d'une part, de souligner la convergence entre les pratiques documentaires et le web sémantique, et d'autre part, de montrer que les langages documentaires peuvent être utilisés dans le cadre du web sémantique pour déterminer les concepts et les relations qui les lient au niveau des graphes conceptuels et des ontologies.

Dans un deuxième temps, nous nous intéressons à la valorisation des collections numériques. Nous nous concentrons plus particulièrement sur la question de la visibilité des bibliothèques numériques sur le web et de l'interopérabilité, qui peuvent être améliorées en exposant leurs données au sein du web sémantique.

Ensuite, nous abordons le sujet de l'ouverture des données en bibliothèques numériques et leurs réutilisations par d'autres communautés, ceci à travers la notion de « Linked open data ».

Enfin, nous terminons ce chapitre en présentant les principaux défis et obstacles à l'usage du web sémantique en bibliothèques numériques.

6.1 Valoriser les pratiques documentaires

L'objectif est de montrer que les pratiques documentaires s'accordent avec les objectifs du web sémantique en terme de traitement des données (identification, structuration, description et relation entre les données) pour en faciliter l'accès et le partage sur le web.

6.1.1 Le web sémantique : une technique pour la documentation

Nous avons abordé dans le chapitre 2 la notion de médiation documentaire. Nous avons indiqué que cette médiation permet d'établir un lien entre un document et un professionnel de l'information par l'intermédiaire de la technique et du traitement documentaire. Autrement dit, il s'agit de représenter d'une manière intellectuelle le contenu des documents afin de faciliter leur accès. Cela entraîne une médiation des savoirs qui permet de mettre en relation un document et un utilisateur.

Dans le contexte des bibliothèques numériques, il s'agit d'une médiation « cachée » dont l'objectif est de mettre en lien un document et un utilisateur par l'intermédiaire des métadonnées. Ces dernières permettent, d'une part, d'identifier et de structurer les données, et d'autre part, d'établir un lien entre l'usager et l'information en facilitant notamment l'accès au contenu numérique.

Nous parlons d'une médiation cachée parce que l'utilisateur ne se rend pas compte du travail exercé derrière l'information pour y accéder, et parce que le professionnel de l'information n'est pas en contact direct avec l'utilisateur pour répondre à ses besoins.

Cette médiation cachée continue d'être présente et même mise en valeur dans le contexte du web sémantique en bibliothèques numériques. Pour Jean-Michel Salaun et le collectif Roger T. Pédauque, c'est un mouvement de redocumentarisation, dans le sens où documentariser signifie traiter un document comme le font traditionnellement les professionnels de l'information [Pédauque, Salaun, 2005, 2006, 2007, 2008]. Ce mouvement implique un retour aux techniques documentaires et aux méthodes de classification et d'indexation exercées par les professionnels de l'information pour structurer l'information [Dalbin et al., 2011] et pour optimiser l'usage du document en permettant un meilleur accès à son contenu et une meilleure mise en contexte [Salaun, 2008].

Dans le tableau 6.1 ci-dessous, nous présentons les techniques de traitement utilisées dans le cadre du web sémantique et dans celui des bibliothèques numériques.

	Web sémantique	Bibliothèques numériques
Techniques d'identification	Des identifiants uniques (URI)	Des identifiants uniques (ISBN, ISSN, DOI etc.)
Techniques de description	Indexation structurée en XML (métadonnées)	Catalogage ; Indexation contrôlée (vocabulaire contrôlée, récemment en XML)
Techniques de liaisons	Le triplet RDF	Des modèles conceptuels (FRBR, FRAD etc.)

Tableau 6.1 – Les techniques de traitement des données utilisées dans le cadre du web sémantique et des bibliothèques numériques

Comme le montre le tableau 6.1, les techniques utilisées pour traiter les documents en bibliothèques numériques et dans le cadre du web sémantique sont très similaires. Dans les deux cas, il s'agit d'identifier les documents et les données pour mieux les repérer. Ensuite, il s'agit de les décrire pour structurer et représenter le contenu des documents par des métadonnées afin de faciliter leur accès. Enfin, il s'agit de relier les données entre elles. Le web sémantique s'appuie sur le triplet RDF pour lier les données, tandis qu'en bibliothèques numériques, les relations entre les données peuvent se faire grâce à des modèles conceptuels comme le FRBR. Ce dernier permet de créer des relations entre les données bibliographiques et les données d'autorités (sujet, auteur).

La différence essentielle réside dans l'élément de base utilisé. Avec le web sémantique, le traitement documentaire se concentre autour de la donnée et non pas autour du document comme c'est le cas en bibliothèques numériques. Autrement dit, il s'agit en bibliothèques numériques d'identifier et de décrire les documents (un livre, un article, etc.), et dans le cadre du web sémantique d'identifier et de décrire les données (le nom d'un auteur, un lieu, etc.). Cela a pour effet de décomposer le document et par conséquent d'accroître ses possibilités de traitement ainsi que son utilisation. Chaque donnée identifiée et décrite peut être facilement liée à d'autres données sur le web, ce qui permet de multiplier les points d'accès à ce document ainsi que son usage.

Dans l'état de l'art, nous avons évoqué la question de l'évolution des métiers de la documentation. Selon Suzanne Briet et Jean Meyriat, la documentation dépend de la technique et du travail intellectuel (Briet, 1951 ; Meyriat, 1981), dans le sens où son développement est étroitement lié aux protocoles et standards utilisés pour identifier, décrire, échanger et faciliter l'accès à l'information. Dans le contexte des bibliothèques numériques, une approche orientée usager du web a permis de développer des formats de description des ressources numériques (Dublin Core, FRBR, MODS etc.) et des protocoles (Z39.50 et OAI-PMH) afin de favoriser les échanges et l'interopérabilité entre les

différents acteurs de l'information (bibliothèques, archives, musées etc.). L'objectif est d'offrir aux usagers un accès facile aux différentes ressources numériques provenant de ces diverses communautés. Cependant, cette approche orientée usager ne prend pas en compte les pratiques des internautes qui cherchent l'information sur le web en utilisant les moteurs de recherche [Bermès et Poupeau, 2012], qui sont considérés comme les points d'entrées pour naviguer sur l'internet [Illien et al, 2013]. Non référencés par les moteurs de recherche, les contenus des bibliothèques numériques restent cachés dans le web invisible [Ibid]. Afin d'améliorer leur visibilité, les bibliothèques numériques peuvent intégrer l'architecture du web. Cela nécessite donc d'adopter les standards du web permettant d'identifier, de décrire et de diffuser les ressources numériques tout en répondant aux besoins des utilisateurs du web.

Dans cet esprit, nous considérons le web sémantique comme une technique offrant aux bibliothèques des standards qui font évoluer leurs services tout en préservant les pratiques professionnelles de type documentaire. Autrement dit, les standards du web sémantique permettent aux bibliothèques numériques de diffuser leurs données sur le web ainsi que d'améliorer leur visibilité [Baker et al, 2011 ; Illien et al, 2013, Bermès et al, 2013]. Les bibliothèques numériques offrent à leur tour des données structurées parfaitement adaptées à un usage au sein du web sémantique.

De cette façon, les bibliothèques numériques peuvent devenir des acteurs importants dans le développement du web sémantique, et ceci en introduisant leur expertise de type documentaire à travers la gestion des métadonnées. Plus précisément, les professionnels de l'information, grâce à leurs expériences de description des contenus et leurs connaissances en termes de besoins d'informations, peuvent envisager de nouvelles fonctionnalités aux outils du web sémantique [Menon, 2003].

Les résultats de la première enquête ont montré que les professionnels de l'information (7 réponses sur 9)[79] considèrent que leur métier est en évolution permanente et que leurs pratiques documentaires rejoignent les objectifs du web sémantique. De plus, les résultats ont montré que les professionnels de l'information pensent que le web sémantique est important pour le futur des bibliothèques numériques (32 réponses sur 33). Certains enquêtés considèrent que le web sémantique permet aux bibliothèques numériques de se positionner comme des acteurs importants sur le web. Ci-dessous quelques extraits tirés du questionnaire et illustrant ces points de vue :

> « Devenir acteur du web sémantique est la voie naturelle pour les bibliothèques, la seule

79 Résultats de la question : Est-ce que l'utilisation du web sémantique a changé vos pratiques documentaires ? (Questionnaire)

> en tout cas qui permettrait aux bibliothèques d'exister et de peser dans la société [Q30] »
> « Le web sémantique permettra aux bibliothèques de sortir de l'univers qui leur est propre et d'être un acteur déterminant du web.[Q03] »

De même, les résultats de la seconde enquête (6 réponses sur 10)[80] ont montré que les professionnels de l'information considèrent que le web sémantique fait partie de l'évolution du métier et qu'en adoptant ses standards, les bibliothèques numériques peuvent devenir des acteurs sur le web. Possédant des collections numériques de valeur et d'intérêt social et culturel, les bibliothèques numériques ont un rôle important à jouer dans la diffusion des savoirs dans la société. Ci-dessous quelques extraits tirés des entretiens et illustrant ces points de vue :

> « On a un rôle de devenir acteur du web sémantique, ce n'est pas simplement être présent sur le web et être visible, à mon avis, c'est être acteur du web sémantique et essayer de faire entendre notre voix auprès de ceux qui font l'architecture du web, auprès de W3C. C'est essayer de peser sur les choix stratégiques pour essayer d'améliorer la société, tout simplement, ça fait aussi partie des missions des bibliothèques. Je pense qu' au contraire, si on n'arrive pas à relever ce défi, on risque de devenir des établissements de seconde zone et de moins en moins fréquentés »[E7]

> « L'enjeu est aussi de passer d'un paradigme où on était centré sur le document à un paradigme où on est centré sur l'information et sur la donnée […] donc l'enjeu est de passer du catalogue centré sur le document au catalogue centré sur l'information»[E10]

Ensuite, les résultats de la deuxième enquête (7 réponses sur 10)[81] ont montré que les pratiques documentaires sont les plus mobilisées dans la mise en place du web sémantique en bibliothèques numériques. Il s'agit des compétences liées au traitement des documents : l'identification, la description et l'indexation. Ci-dessous quelques extraits tirés des entretiens et illustrant ce point de vue :

> « Pour moi toute la partie indexation, c'est clairement sur ces compétences là qu'on va apporter, [..] le cœur de notre métier va être même mis en valeur. J'espère que dans le web sémantique, cette partie d'enrichissement, par le biais de l'indexation et de la gestion des référentiels, va être mis en valeur»[E8]

> «ce sont des compétences documentaires classiques de type description, indexation, catalogage »[E1]

> « C'est un enjeu pour les bibliothèques de proposer des métadonnées de qualité, justement les données des bibliothèques sont des données établies selon un consensus confiant et rationnel. Elles ont été vraiment sélectionnées et validées au niveau de leur qualité, et c'est vraiment là où les bibliothèques ont une carte à jouer, et d'autant plus l'intérêt de rendre visible leur métadonnées, et là on voit l'efficacité du facteur humain face aux algorithmes »[E2]

80 Résultats de la question portant sur les enjeux du web sémantique en bibliothèques numériques (Entretiens)
81 Résultats de la question liée aux pratiques mobilisées pour accompagner la mise en place du web sémantique en bibliothèques numériques

Les bibliothèques numériques doivent évoluer pour répondre aux besoins des usagers du web. Il s'agit d'adopter des techniques permettant aux bibliothèques numériques de réaliser leurs objectifs tout en s'appuyant sur des pratiques professionnelles liées au traitement des documents et à la gestion des métadonnées. Le web sémantique offre le cadre technique permettant aux professionnels de l'information de pratiquer leurs savoirs faire et leurs connaissances dans le traitement des documents dans l'objectif d'améliorer la diffusion des ressources numériques sur le web. A travers leurs collections numériques de grande valeur, les bibliothèques numériques peuvent enrichir le web avec des données structurées, vérifiées et parfois rares. C'est pourquoi, les bibliothèques numériques peuvent devenir des acteurs puissants du web en donnant aux informations qui circulent sur le web davantage de fiabilité et de pertinence.

Nous avons présenté le web sémantique comme une technique pour la documentation qui permet aux bibliothèques d'améliorer la diffusion de leurs contenus sur le web. Mais cette technique a elle-même besoin de modèles conceptuels pour représenter les connaissances. C'est pourquoi dans la section suivante, nous abordons le rôle des langages documentaires dans le web sémantique.

6.1.2 Les langages documentaires comme modèles conceptuels

Dans l'état de l'art, nous avons présenté le web sémantique comme une nouvelle approche de représentation des connaissances qui permet, à l'aide d'un ensemble de standards, d'identifier et de lier des données entre elles.

Cependant, le web sémantique n'est pas un outil de l'intelligence artificielle lié au traitement automatique des langues [Berners-Lee, 2006]. Au contraire, le web sémantique s'appuie sur des données structurées. Et pour pouvoir extraire du sens, le web sémantique a besoin de décrire les données de manière à ce qu'elles soient compréhensibles par les machines et par les hommes [Bernes-Lee, 2001, 2006]. De plus, les liens entre les données ne se font pas de manière automatique. Le web sémantique offre le cadre technique permettant de lier les données entre elles sur le web, mais cela nécessite une base référentielle sur laquelle s'appuyer afin de déterminer les terminologies nécessaires pour construire les liens entre les données dans le cadre du web sémantique [Dalbin et al., 2011].

Étant structurés et liés, les langages documentaires constituent un terrain intéressant pour appliquer les techniques du web sémantique. A cet égard, Abderrazak Mkadmi et Imad Saleh indiquent que l'exposition des données bibliographiques en RDF représente un élément intéressant permettant au web sémantique de travailler sur des données fiables et contrôlées par des professionnels [Saleh et

Mkadmi, 2008, p.201]. Ainsi, Madjid Ihadjhadene et Laurence Favier considèrent que le web sémantique est un exemple de conception de langages documentaires basée sur la description et le contrôle grammatical et sémantique des termes de cette description [Ihadjadène et Favier, 2008]. De leur côté, Laurence Favier et Widad Mustafa El Hadi pensent que l'avenir du web sémantique dépend de l'offre de SOC (Système d'organisation des connaissances) fondé sur des langages contrôlés et interopérables [Favier et EL-Hadi, 2013]. Pour Bruno Bachnimont, le web sémantique s'inscrit dans le travail des professionnels de l'information notamment dans l'approfondissement des notions de vocabulaires et de métadonnées [Bachimont et al., 2011].

Nous pensons que la représentation des connaissances dans le cadre du web sémantique nécessite une intervention humaine afin de fournir des modèles conceptuels fondés sur des langages contrôlés comme les langages documentaires. Autrement dit, les graphes et les ontologies du web sémantique s'appuient, d'une part, sur un cadre technique, celui qui assure l'identification et les liens entre les données sur le web (Les URIs, RDF, RDFS, OWL), et d'autre part, sur un cadre conceptuel permettant de déterminer les concepts et les liens entre eux (les langages documentaires).

L'analyse des résultats de la deuxième enquête nous a permis d'éclaircir certains points communs entre le web sémantique et les langages documentaires. Ci-dessous des extraits de l'entretien réalisé avec deux professionnels de l'information à la BnF :

> « A mon avis, il faut bien voir une chose, c'est qu'un des premiers catalogues en France a été fait par le bibliothécaire Clément au 17 siècle. Ce bibliothécaire a institué ce qu'on appelle la cotation Clément. Or qu'est-ce que c'est que la cotation Clément ? C'est un petit code qui permet à la fois d'identifier et de localiser dans un ensemble scientifique de description de collection un document, et c'est spécifiquement ce que fait un URI aujourd'hui sur le web. Je pense qu'il y a quand même deux domaines qui se parlent complètement, cette tradition de catalogage qu'on a historiquement et le langage utilisé sur le web, qui est un langage finalement scientifique, et de toute façon le bibliothécaire est un scientifique, c'est un scientifique de collection mais qui est aussi un bibliothécaire en tant que tel dans cette volonté de description, de décrire et de localiser » [E3b].
>
> «Le web, en fait, il ressuscite les pratiques de la profession, en tout cas la profession telle qu'elle a été pensée au début de l'époque moderne, et à mon avis c'est le RDF. En tous cas le web sémantique c'est un langage qui sort tout droit de la tradition scientifique du 17eme siècle » [E3b].
>
> « Rien que la modélisation, rien que le mapping entre les formats, ça demande un travail intellectuel. Là on a deux équipes ici qui ont participé à cette modélisation, en fait deux modélisateurs, des gens qui ont fourni le travail intellectuel de modélisation et de mapping dans le RDF ça c'est sûr » [E3b].
>
> «C'est un travail intellectuel, c'est-à-dire que c'est le catalogueur qui catalogue et qui va faire le lien avec un auteur, qui va faire le lien avec une vedette RAMEAU etc. Le lien

> est fait dans la production, et même de plus en plus on commence à se demander si on ne pourrait pas leur demander de lier carrément à des url et non pas lier juste à un mot qui existe dans un référentiel ailleurs, d'aller chercher l'url qui n'est pas forcément évident d'un point de vue de pratique de catalogage. » [E3b]

L'enquêté [E3b] affirme clairement que le web sémantique réintroduit des pratiques professionnelles propres aux bibliothécaires, celles qui permettent d'identifier et de localiser les documents. Il s'agit de deux domaines (le web et les bibliothèques) qui communiquent et qui ont les mêmes objectifs : faciliter l'accès à l'information.

Ensuite, pour savoir quel rôle les langages documentaires occupent dans l'usage du web sémantique en bibliothèques, l'enquêté [E3b] a indiqué que les vocabulaires contrôlés jouent le rôle d'ontologies permettant de définir des domaines de connaissances. Par exemple, le langage documentaire Rameau[82] est maintenant exposé en format SKOS, ceci veut dire qu'il est dans un format de représentation standardisé exprimé en XML et qu'il utilise le triplet RDF pour relier les concepts. Rappelons que Rameau est composé d'un vocabulaire de termes reliés entre eux, qui permet d'indiquer les règles de construction des vedettes matières relatives à l'indexation des documents. En exposant Rameau en format SKOS, il y a un profit mutuel entre les bibliothèques et le web sémantique. D'une part, les données de la bibliothèque gagnent plus de visibilité sur le web tout en valorisant ce langage traditionnel d'indexation. D'autre part, la représentation des connaissances dans le cadre du web sémantique devient réalisable grâce aux langages contrôlés comme celui de Rameau.

A ce propos, l'enquêté [E3b] a précisé que Rameau est en SKOS mais qu'il est produit en Intermarc :

> «Rameau est déjà en SKOS, tout rameau est publié en SKOS. C'est un vocabulaire contrôlé et en même temps, il joue le rôle d'ontologie, il définit des domaines. Il est publié dans 'databnf' en SKOS, mais il continue d'être produit en Intermarc. Le format de base de production de Rameau reste l'Intermarc, et à priori, il est appelé à rester comme ça, et de fait on n'est pas dans l'idée qu'il faut reproduire nativement du RDF, c'est-à-dire on continuera à produire dans des formats autres que RDF, et il y aura toujours ce mapping qui sera fait entre ces formats de production et les formats de diffusion, ce qui est sûr c'est que la production et la diffusion d'un point de vue format sera toujours distincte, en tout cas dans notre politique » [E3b].

Cela montre qu'il y a une distinction entre les formats de production et les formats de diffusion au

82 RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) est le langage d'indexation matière utilisé, en France, par la Bibliothèque nationale de France, les bibliothèques universitaires, de nombreuses bibliothèques de lecture publique ou de recherche ainsi que plusieurs organismes privés. Il est élaboré depuis 1980, de façon autonome, en relation avec le « Répertoire de vedettes-matière » de l'Université Laval à Québec, et avec la liste de vedettes-matière de la Bibliothèque du Congrès (Library of Congress Subject Headings). [Source BnF].

sein de la bibliothèque. Les premiers sont des formats classiques basés sur MARC ou Intermarc permettant de décrire les données, et les deuxièmes sont des formats du web sémantique comme le RDF permettant la diffusion de ces données. A ce sujet, l'enquêté [E4] de la BnF indique :

> « Le rdf est un format de diffusion et non pas un format de production, d'ailleurs c'est actuellement le cas. Je pense que ce n'est pas forcement la peine de changer le format de production pour passer à quelque chose en rdf. Pour moi la valeur ajoutée du rdf c'est plutôt dans l'exposition. La valeur ajoutée de rdf est plutôt d'exposer les données de manière à ce qu'elles soient compatibles avec les standards du web » [E4]

L'enquêté [E4] signale qu'au lieu de changer les formats de production basés sur MARC, il faut plutôt les faire évoluer, parce que ces formats sont utilisés dans la plupart des bibliothèques en France, et parce qu'ils permettent aussi de créer des liens entre les données. Ci-dessous un extrait de l'entretien illustrant son point de vue :

> « On peut très bien faire du linked data avec des données en MARC dans l'écosystème de production interne. Parce que depuis 87, pour être précis, les catalogueurs font déjà des liens entre les notices bibliographiques et les notices d'autorité. Ce que j'appelle du vintage linked data. Parce que les notices ont des identifiants mais pas des URIs, mais des identifiants qui sont bien gérés. Derrière il y a des données en marc, mais on a déjà des entités qui sont liées entre elles, et finalement de ce point de vue là, on a déjà une logique de production qui pense à linker en entités »[E4]

A ce propos, l'enquêté [E3b] ajoute:

> « Il faut savoir que nous, quand on a commencé à frbriser le catalogue, c'est-à-dire à lier des auteurs à des œuvres ou des œuvres à des éditions, il y avait quand même une tradition de lier les identifiants dans leur production même. C'est-à-dire que dans Marc il y avait une zone qui permettait, quand vous étiez en train de cataloguer un document d'insérer l'identifiant de l'auteur, la-dessus on pouvait s'appuyer justement pour lier des auteurs à des documents » [E3b].

Les enquêtés de la BnF [E3b et E4] ont signalé que les formats de catalogage et de description basés sur MARC peuvent continuer à être des formats de production des données bibliographiques en bibliothèques. Ce sont des formats riches et bien maîtrisés par les bibliothécaires, ils permettent de structurer et de décrire les données rendant possible leur intégration dans le web sémantique. Le RDF est de son côté un format de liaison, de diffusion et d'exposition des données sur le web. Mais il semble qu'il y ait une possibilité de faire évoluer les formats de production, notamment le format MARC, pour lier les auteurs à des documents.

Cela montre la richesse des formats et des langages documentaires utilisés en bibliothèques pour structurer et lier les données bibliographiques. De ce fait, l'usage du web sémantique ne doit être introduit que dans un contexte de continuité et d'évolution des bibliothèques pour améliorer la

diffusion de leurs données et pour assurer leur présence dans le web.

6.1.3 Evolution des pratiques professionnelles

Dans l'état de l'art, nous avons abordé la question de l'évolution des pratiques professionnelles en bibliothèques numériques. Nous avons cité une étude de Youngok Choi et Edie Rasmussen portant sur les activités, les compétences et les savoirs nécessaires en bibliothèques numériques. Les résultats ont montré que les compétences et les connaissances des professionnels de l'information les « mieux classées » sont celles liées à l'architecture de la technique (logiciel, norme, langage), et à l'usager (les besoins, la communication) [Choi et Rasmussen, 2006].

Une autre étude de Pussadee Nonthakumjane, a montré que les compétences des professionnels de l'information à l'ère numérique se concentrent autour de la gestion, la création de métadonnées, le développement des bases de données, la considération des besoins des usagers, l'archivage numérique, le développement des collections et la gestion des systèmes d'information [Nonthacumjane, 2011] .

Les pratiques professionnelles en bibliothèques numériques nécessitent des compétences liées à la gestion des métadonnées, au développement des collections numériques, au développement technique qui accompagne l'infrastructure de la bibliothèque numérique et enfin des compétences permettant de répondre aux besoins des usagers.

Dans le contexte de l'usage du web sémantique en bibliothèques numériques, notre enquête a permis d'identifier les pratiques, les compétences et les connaissances nécessaires pour accompagner l'usage du web sémantique en bibliothèques numériques.

Le tableau 6.2 ci-dessous synthétise les résultats :

Pratiques documentaires classiques	**Pratiques et connaissances à acquérir**
• Indexation • Catalogage • Description • Gestion des métadonnées • Informatique documentaire	• Maîtriser les langages de balisage (XML, HTML) • Maîtriser les standards du web sémantique (URI, RDF, OWL etc.) • Coopérer et partager les données et les métadonnées entre des différents acteurs de l'information sur le web

Tableau 6.2 – Les pratiques, les compétences et les connaissances nécessaires pour accompagner l'usage du web sémantique en bibliothèques numériques

Les résultats montrent que les compétences et les connaissances qui accompagnent l'usage du web sémantique en bibliothèques numériques sont avant tou celles liées aux pratiques documentaires « classiques ». Cependant, il est aussi nécessaire d'acquérir des connaissances et des compétences permettant de maîtriser les langages de balisage et les standards du web sémantique.

Les professionnels de l'information tentent de relever les défis liés à l'introduction des nouvelles technologies en bibliothèques numériques. Cela implique de s'adapter aux nouveaux outils permettant de faire évoluer leurs pratiques et de transmettre leurs savoirs faire documentaires afin de gérer les métadonnées et d'organiser des connaissances. En effet, le web sémantique peut être considéré comme une nouvelle technique qui permet de valoriser les pratiques documentaires, et d'offrir aux professionnels de l'information une opportunité d'occuper une place fondamentale dans l'organisation des connaissances.

Nous abordons dans la section suivante la question de la valorisation des collections numériques en bibliothèques. Nous discutons des enjeux du web sémantique sur la visibilité des bibliothèques numériques et la réutilisation de leurs données.

6.2 Valoriser les collections numériques

Les études empiriques nous ont permis de tester les hypothèses de recherche. Nous avons constaté que la visibilité sur le web constitue un enjeu crucial pour les bibliothèques numériques. Intégrer les données des bibliothèques dans le web permet d'accroître leur visibilité ainsi que leur usage.

Dans cette section, nous abordons la question de la visibilité des bibliothèques numériques sur le web ainsi que l'interopérabilité technique et sémantique des données.

6.2.1 Améliorer la visibilité des bibliothèques numériques sur le web

Depuis les années 90, la valorisation du patrimoine culturel préoccupe les politiques publiques numériques des états. En France, le Premier ministre de l'époque Lionel Jospin a affirmé en 1997 la nécessité d'entrer dans la société de l'information, notamment, grâce à une politique de numérisation du patrimoine culturel permettant son accès libre au public.

En 2002, Bruno Ory-Lavollée indique que la diffusion du patrimoine contribue « au rayonnement international de la France » [Ory-Lavollée, 2002]. En 2009, dans son rapport intitulé « partager notre patrimoine culturel » Bruno Ory-Lavollée encourage à maximiser la présence des données publiques culturelles sur les réseaux et l'interopérabilité entre les projets de numérisation [Ory-

Lavollée, 2009].

En 2010, Marc Tessier signale dans son rapport sur la numérisation du patrimoine écrit que « *rien ne sert d'être disponible si l'on n'est pas visible* » [Tessier, 2010]. Autrement dit, la numérisation et la publication sur le web du patrimoine numérisé ne suffisent pas pour y accéder. Il ne s'agit pas d'être simplement existant sur le web, mais avant tout d'être visible.

Dans le contexte des bibliothèques numériques, la visibilité implique une réflexion sur les usages actuels du web. Selon Pauline Moirez et Dominique Stutzmann, l'outil principal de communication des bibliothèques numériques est formé par le catalogue et les métadonnées. Ces dernières sont indispensables pour le signalement et le référencement des collections numériques [Moirez et Stutzmann, 2013]. Dans le chapitre 2, nous avons indiqué qu'il y a eu un changement de paradigme conduisant les bibliothèques numériques vers une approche orientée-usager qui se manifeste par des développements au niveau des catalogues afin de prendre en compte les besoins de l'usager. Ensuite, des changements se font au niveau de la conception de modèles communs entre plusieurs communautés (bibliothèques, archives, musées …) afin de pouvoir échanger les données et les métadonnées efficacement tout en enrichissant l'expérience de recherche des utilisateurs.

Les notices bibliographiques sont structurées selon des standards spécifiques à la communauté des bibliothèques et non pas selon ceux du web. Face à ce problème, les bibliothèques ont commencé à adopter des protocoles et des formats pour échanger sur Internet des métadonnées et des données structurées en XML comme le protocole OAI-PMH et le format Dublin Core par exemple, qui permet la communication entre des institutions variées. L'objectif est d'améliorer le signalement et le référencement des collections numériques au-delà des diverses communautés, tout en restant suffisamment structuré.

Aujourd'hui, le problème majeur des bibliothèques numériques est leur visibilité. Les notices bibliographiques et les contenus des bibliothèques numériques ne sont accessibles que par un passage par le site de la bibliothèque. Il faut donc connaître le site de la bibliothèque pour pouvoir accéder à ses contenus. Or, l'internaute selon Brigitte Simonnot, n'est plus l'usager des systèmes documentaires [Simonnot, 2012]. Les utilisateurs du web, ceux qui font usage des moteurs de recherche pour accéder à l'information, ne peuvent pas accéder aux contenus des bibliothèques numériques puisqu'ils ne sont pas référencés par ces moteurs.

Adopter l'architecture du web demeure une priorité pour les bibliothèques numériques qui souhaitent valoriser leurs contenus en les rendant visibles sur le web. En se basant sur les principes du web et sur la structuration des données, le web sémantique offre aux bibliothèques numériques

une opportunité pour améliorer la visibilité de leurs données. Les identifiants URIs assurent une identification pérenne des données sur le web ainsi qu'un chemin d'accès pour obtenir une représentation de ces données en s'appuyant sur le protocole HTTP [Bachimont et al., 2011]. Une fois que les données sont identifiées et localisées sur le web, il est alors possible de les lier à d'autres données en utilisant le triplet RDF. Ce mécanisme permet de multiplier les points d'accès aux données des bibliothèques, puisqu'elles sont liées à d'autres données disponibles sur d'autres sites web. Cela a pour effet d'augmenter le nombre de liens pointant sur les ressources numériques des bibliothèques et ainsi que d'améliorer leurs référencements par les moteurs de recherche.

L'objectif est de mettre les données des bibliothèques numériques sur le chemin de l'internaute. Autrement dit, il faut que celui-ci puisse les retrouver facilement en faisant une simple recherche sur des moteurs de recherche sans passer forcément par le site web de la bibliothèque.

Les résultats de l'enquête empirique nous ont permis de tester nos hypothèses de recherche. Nous avons considéré que l'usage du web sémantique en bibliothèques numériques permet d'améliorer leur visibilité (**H 3**).

Les résultats du questionnaire ont montré que les bibliothèques qui utilisent le web sémantique (9 bibliothèques sur 33) ont comme objectifs : d'améliorer leur visibilité sur le web (5 réponses sur 8)[83], de faire évoluer leurs catalogues (4 réponses sur 8), d'améliorer l'interopérabilité (2 réponses sur 8) et enfin de réutiliser les données (1 réponse). De plus, les résultats de la question liée aux apports du web sémantique en bibliothèques ont montré que le web sémantique permet d'améliorer la visibilité des données sur le web selon 90 % des bibliothèques enquêtées (30 bibliothèques sur 33)[84]. Enfin, les résultats de la question : pourquoi le web sémantique est important pour le futur des bibliothèques ?[85] ont indiqué que les bibliothèques numériques peuvent se servir du web sémantique pour proposer des services adaptés aux usages du web en améliorant leur visibilité sur le web (9 réponses sur 24)[86], pour partager les données (6 réponses), pour devenir des acteurs du web (4 réponses) et pour améliorer la recherche d'information (4 réponses) .

De même, les résultats des entretiens ont montré que la visibilité sur le web constitue un enjeu essentiel de l'utilisation du web sémantique en bibliothèques numériques. Pour la question liée aux enjeux du web sémantique en bibliothèques numériques[87], 6 enquêtés sur 10 ont indiqué que le web sémantique améliore le signalement et le référencement des données ainsi que leur visibilité.

83 Résultats de la question : pourquoi la bibliothèque utilise le web sémantique ? (Questionnaire : 8 réponses)
84 Voir figure 5.13 page 141
85 Voir tableau 5.7 page 143
86 Le nombre de réponses le plus élevé qui concerne cette question
87 Voir tableau 5.13 page 154

Parmi les bibliothèques interrogées, une seule (la BnF) expose ses données sur le web en utilisant les standards du web sémantique. Les autres bibliothèques ont commencé à intégrer le web sémantique, mais leurs expériences sont limitées à identifier et tester les liens entre les données. La BnF affirme que certaines données s'affichent désormais dans les résultats du moteur de recherche Google.

Ci-dessous des extraits tirés de l'entretien réalisé avec les deux enquêtés de la BnF [E3a at E3b] illustrant ce point de vue :

> « il y a quand même un enjeu de signalement. Nous, en tant que responsables de databnf on se demande quel est l'intérêt du RDF, qu'est-ce qu'il fait en plus par rapport aux autres formats [...] je pense que le signalement, le fait que ça remonte dans les moteurs de recherche [...] ça répond à un grand enjeu de visibilité des données des bibliothèques, [...] Je pense que le fait que « databnf » remonte dans les résultats Google, c'est un enjeu important » [E3b].
>
> **Pourquoi certaines données sont référencées par le moteur de recherche Google et d'autres non ?**
>
> « C'est mieux que zéro, avant on était à zéro, on arrive rarement sur un contenu de la BnF depuis les moteurs de recherche» [E3b]
>
> « surtout ce qui fait qu'on remonte, c'est qu'on a du contenu rare, c'est-à-dire qu'on a un contenu que même Wikipédia n'a pas, donc on a cette exhaustivité qui vient du fait qu'on a le dépôt légal [...] Mais le rdf en tant que tel n'est pas forcément ce que va voir le moteur, en revanche le fait d'avoir organisé et modélisé les données selon un modèle d'entités qui est plutôt un modèle rdf de graphe, oui cette organisation là parle au moteur de recherche, du moins elle parle aux usagers sur internet, parce que les usagers ont tendances à chercher par titre, par auteur, etc., donc de ce point de vue là, la modélisation rdf aide beaucoup à faire remonter même si ce n'est pas forcément la seule solution, il y a aussi d'autres formats qui utilisent le modèle d'entités » [E3a]

L'analyse de ces réponses montre que l'usage du web sémantique a permis d'afficher des résultats provenant de la BnF dans le moteur de recherche Google. Mais la question se pose sur les raisons pour lesquelles certaines données s'affichent dans Google et d'autres non. Le fait que la BnF dispose de contenus rares organisés autour d'entités selon le modèle RDF aide à améliorer leur référencement dans les moteurs de recherche. Cela veut dire que les autres données (non rares) ne seraient pas forcément référencées par le moteur de recherche.

6.2.2 Favoriser l'interopérabilité technique et sémantique

Les enjeux principaux qui conditionnent l'évolution des bibliothèques numériques sont la visibilité et l'interopérabilité qui permettent d'échanger rapidement et sans difficulté des données

sur le web.

Les bibliothèques ont une longue tradition dans l'échange et le partage des données bibliographiques avec d'autres bibliothèques. Pour pouvoir communiquer, les bibliothèques ont élaboré des protocoles et des formats qui leur sont propres et qui ne sont pas ouverts à d'autres communautés. Par la suite, les bibliothèques ont commencé à adopter des formats et des protocoles plus ouverts permettant de communiquer avec d'autres institutions comme les archives et les musées. Dans le chapitre 3 de l'état de l'art, nous avons présenté les différents langages et formats utilisés par les bibliothèques pour décrire les ressources numériques, MARC, FRBR, RDA et Dublin Core, ainsi que les protocoles pour échanger des données comme le protocole Z39.50 ou le protocole OAI-PMH.

Selon Pauline Moirez et Dominique Stutzmann, les bibliothèques doivent positionner leurs données dans un écosystème plus large permettant de communiquer avec des acteurs différents (bibliothèques, éditeurs, institutions culturelles, moteurs de recherche etc.). Cela nécessite selon elles une interopérabilité double, celle qui assure le transport des données (interopérabilité technique) et celle qui garantit la compréhension des informations échangées (interopérabilité sémantique) [Moirez et Stutzmann, 2013].

Le défi aujourd'hui est de pourvoir exposer les données des bibliothèques dans le web, ce dernier est basé sur des protocoles et des standards normalisés et interopérables. Plutôt qu'utiliser des protocoles et des formats spécifiques adaptés qu'aux bibliothèques, il faut commencer à adopter des standards ouverts pour pouvoir communiquer et échanger les données efficacement et sans barrière sur le web.

Pour Emmanuelle Bermès, le web sémantique propose une forme d'interopérabilité qui repose sur des liens permettant de naviguer de manière transparente sur le web [Bermès , 2011]. De ce fait, le web sémantique, en proposant une interopérabilité basée sur les standards du web, permet aux bibliothèques de communiquer avec d'autres communautés, d'échanger et de partager l'accès à l'information sans difficulté. L'enjeu est de proposer une interopérabilité technique et sémantique. Grâce aux données structurées et aux liens explicites établis entre les concepts des graphes et des ontologies, le web sémantique garantit la communication entre des sources et des systèmes d'information variés tout en assurant une compréhension commune de la signification des données. Cette compréhension commune est possible grâce aux graphes conceptuels et aux ontologies du web sémantique [Bachimont, 2007, 2011 ; Bouganem et Savoy, 2008]. Ainsi pour Abderrazak Mkadmi et Imad Saleh, les cadres communs offerts par OWL ou SKOS assurent une interopérabilité plus facile et compréhensible entre les concepts issus de vocabulaires contrôlés

hétérogènes [Saleh et Mkadmi, 2008]. En effet, la représentation des vocabulaires de cette façon facilite l'identification, l'exploitation et les relations d'équivalences entre des concepts issus de vocabulaires contrôlés hétérogènes [Ibid].

Les résultats de la première enquête ont montré que l'interopérabilité constitue un enjeu important de l'usage du web sémantique en bibliothèques numériques. Elle apparaît en deuxième position (25 réponses sur 33) à la question portant sur les enjeux et les apports du web sémantique en bibliothèques numériques. Et concernant la question portant sur le futur des bibliothèques numériques, les enquêtés pensent que le web sémantique pourrait améliorer l'interopérabilité et le partage des données sur le web (7 réponses sur 24)[88].

Dans la deuxième enquête, les résultats (5 enquêtés sur 10)[89] ont montré que l'interopérabilité et la communication entre les différentes communautés sur le web, constituent un enjeu essentiel de l'utilisation du web sémantique en bibliothèques numériques. Cela confirme les différents points de vue discutés dans l'état de l'art [Bachimont, 2007, 2011 ; Bouganem et Savoy ; Saleh et Mkadmi, 2008 ; Bermès, 2015] et les résultats obtenus dans une étude sur les usages du web de données en bibliothèques numériques [Hallo et al, 2015].

En offrant une interopérabilité technique et sémantique, le web sémantique peut faciliter la découverte d'informations. Dans l'état de l'art, nous avons indiqué que, les liens entre les données permet de découvrir des éléments éloignés qui, au début de la recherche, étaient inconnus [Bermès et al., 2013]. L'interopérabilité et le partage des données peuvent enrichir la description des ressources documentaires avec des données fournies par d'autres communautés ce qui favorise la découverte de l'information [Baker et al, 2011]. Cela permet aux usagers, d'une part, de découvrir des informations nouvelles auxquelles ils n'ont pas pensé au début, et d'autre part, de les décharger d'une partie de leur tâche d'exploitation et de recherche. Dans cette situation, l'interopérabilité proposée par le web sémantique permet l'accès à l'information et sa découverte éventuelle.

Cependant, la question de la découverte d'informations et de connaissances en lien avec l'usage du web sémantique n'a pas été développée dans le discours des enquêtés. En revanche, les enquêtés du questionnaire et ceux des entretiens ont indiqué que l'usage du web sémantique permet d'améliorer la recherche d'information (23 réponses sur 33)[90] et (3 enquêtés sur 10)[91] en proposant des informations de qualité et plus pertinentes (22 réponses sur 33)[92].

88 Voir tableau 5.7 page 143
89 Voir tableau 5.13 page 154
90 Résultats de la question des apports du web sémantique en bibliothèques numériques (voir figure 5.13 page 141)
91 Résultats de la question des enjeux du web sémantique en bibliothèques numériques (voir tableau 5.13 page 154)
92 Ibid

6.3 Favoriser l'ouverture des données

Dans l'état de l'art, nous avons analysé les différentes expressions utilisées dans le cadre de l'évolution du web : « web sémantique », « web de données », « linked data », « web 3.0 » « web of data » et « linked open data ». L'expression linked data a été utilisée par Tim Berners pour enlever l'ambiguïté contenue dans l'expression web sémantique et pour éclaircir les enjeux et les objectifs derrière cette évolution du web. En 2010 Tim bernes Lee a ajouté une nouvelle étoile à son modèle avec la notion de « linked open data ». Il ajoutait le terme « open » pour designer qu'il s'agit aussi de diffuser des données ouvertes pour favoriser la réutilisation et l'ouverture des données.

6.3.1 Ouverture et partage des données avec d'autres communautés

L'ouverture des données des bibliothèques numériques sur la toile donne la possibilité à de multiples et innovantes réutilisations des données et permet de positionner les bibliothèques dans le flux informationnel du web et dans l'économie de la connaissance indiquent Pauline Moirez et Dominique Stutzmann [Moirez et Stutzmann, 2013].

Les résultats des entretiens (5 enquêtés sur 10)[93]ont montré que le partage et la réutilisation des données sur le web constituent un enjeu essentiel dans l'utilisation du web sémantique en bibliothèques numériques. Les enquêtés ont signalé que l'usage du web sémantique permet à plusieurs communautés de communiquer et de partager les données de manière simple sur le web.

La réutilisation des données, contrairement aux habitudes des bibliothèques, se fait par d'autres communautés comme les centres de recherche, les laboratoires, les sites privés etc.

Pour illustrer ce point de vue, un extrait de l'entretien réalisé avec l'enquêté [E3a] montre que la réutilisation des données de la BnF ne se fait pas seulement par des bibliothèques mais aussi par des laboratoires et des sites privés.

> « On s'est aperçu que les gens qui étaient intéressés par les données en RDF n'étaient pas forcément des gens du monde du livre et des bibliothèques, c'était plutôt des gens des laboratoires extérieurs, comme l'institut national de géographie, comme des gens qui s'intéressaient à la culture de façon amateur. On a un site qui réutilise nos données qui est « libre théâtre », c'est un site de publication de texte de théâtre amateur dans le domaine public, on a eu éventuellement le moteur de recherche Isidore. C 'est vraiment des laboratoires scientifiques qui sont particulièrement intéressés par ce format » [E3a].

Pour favoriser la réutilisation des données, la BnF a mis en œuvre une politique de médiation autour des données de la bibliothèque. Cette médiation consiste à mener un dialogue avec ceux qui réutilisent les données tout en adoptant une technique de visualisation permettant de mieux comprendre leur contexte. Ci-dessous un extrait de l'entretien avec l'enquêté [E3a] illustrant ce

93 Voir tableau 5.13 page 154

point de vue :

> «On pourrait être plus didactique sur le fonctionnement de sparql, pour expliquer comment faire des requêtes simples et trouver des choses assez précises, ça pourrait être intéressant pour des projets de recherche en master. On pourrait montrer comment utiliser des outils de visualisation, pour une visualisation simple des lieux, il y a des choses à trouver de ce point de vue-là, on s'adresse vraiment à des publics très différents, c'est pour ça qu'il faut mettre en place la médiation» [E3a].

En posant la question sur les retours des « réutilisateurs », de leur expérience, l'enquêté [E3b] a répondu :

> « On a des retours seulement lorsque ça ne marche pas, alors on est content quand ça ne marche pas. Généralement l'idée, c'est quand même qu'un utilisateur de moteur de recherche puisse arriver aux ressources de la BNFsans connaître la BNF [...] mais du coup ça fait qu'on n'a pas vraiment de regard sur ce qui se fait, on ne sait pas vraiment si nos données ont convaincu ou pas ceux qui ont voulu les utiliser, est-ce qu'ils ont pu l'utiliser et si ils n'ont pas pu, pourquoi ? etc. La-dessus on essaye de construire des espaces de dialogue » [E3b].

Cette réponse montre qu'il est important de dialoguer avec les utilisateurs des données. Il s'agit donc d'une médiation autour des données qui sert à comprendre et à analyser les contextes d'usage de ces données dans le but de favoriser leur réutilisation. C'est pourquoi, la médiation autour les données demeure une nécessité pour les bibliothèques qui veulent exposer leurs données sur le web. Ci-dessous un extrait de l'entretien avec l'enquêté [E3b] illustrant ce point de vue :

> «On a vraiment besoin d'un métier nouveau qui est vraiment la médiation autour de cet objet documentaire particulier qui est la donnée» [E3b].

Nous considérons les bibliothèques numériques comme des dispositifs info-communicationnels, permettant d'établir un lien entre un usager et une information par le biais d'une double médiation complémentaire ; celle qui sert à identifier et représenter l'information, la médiation documentaire ; et celle qui permet d'offrir un accès universel aux savoirs, la médiation des savoirs. Nous ajoutons à ces médiations une nouvelle dimension due à l'ouverture des données des bibliothèques sur le web de données, celle de la médiation liée à la réutilisation des données.

L'usage du web sémantique en bibliothèques numériques nécessite trois niveaux de médiation :

- Niveau 1 : la médiation documentaire permet de représenter le contenu des documents numériques par des métadonnées tout en s'appuyant sur des pratiques documentaires (catalogage, indexation, langages documentaires) et les standards du web sémantique (URI, XML, RDF). Chaque donnée identifiée et décrite peut être utilisée de manière indépendante dans des contextes variés, ainsi elle peut être liée à d'autres données sur le web.

- Niveau 2 : la médiation des savoirs permet de faciliter l'accès aux documents numériques grâce aux métadonnées. Étant plus visibles sur le web, les collections des bibliothèques numériques ont plus de chance d'être retrouvées par les usagers. Ceci repositionne les bibliothèques numériques comme étant des institutions de références et de diffusion des savoirs.
- Niveau 3 : la médiation autour de la donnée permet de favoriser la réutilisation et la récupération des données et des métadonnées par différents acteurs (bibliothèques et autres).

La réutilisation des données multiplie leurs contextes d'usage. Cela a pour effet de créer de nouvelles données et par conséquent de favoriser la production des savoirs et l'avancement des recherches grâce à la valeur ajoutée des données réutilisées.

Dans la section suivante, nous prolongeons notre réflexion de recherche en abordant la question du «linked open data».

6.3.2 Le linked open data

En 2010, Tim Berners Lee a introduit le système d'annotations « 5 stars linked open data », les 5 étoiles de Linked open data, en indiquant qu'il s'agit d'encourager les gouvernements à aller vers le web de données. Les données liées et ouvertes (Linked open Data) sont donc des données liées publiées sous licence ouverte, et réutilisables gratuitement.

La figure ci-dessous représente le système d'annotations de « 5 stars linked open data »[94] :

Figure 6.1 : Le système d'annotations de « 5 stars linked open data »

Certains auteurs pensent que le web sémantique est un outil pour favoriser l'ouverture de l'information. Selon Bruno Bachimont, les ontologies du web sémantique représentent une opportunité dans le cadre de l'ouverture des données publiques, permettant aux organismes publics de développer de nouveaux services [Bachimont et al., 2011]. Cette ouverture n'est pas limitée aux données publiques mais elle englobe aussi les données scientifiques. A ce propos, Bruno Bachimont ajoute que l'ouverture des données scientifiques permet de mieux diffuser non seulement les

94 Berners-Lee, Tim (2010). Linked data [En ligne]. https://www.w3.org/DesignIssues/LinkedData.html

résultats de recherche mais aussi les données scientifiques à l'origine de ces résultats [Ibid].

De son coté, Fabien Gadon indique que l'ouverture des données désigne un ensemble de mouvements technologiques, culturels et politiques permettant l'accès libre aux données. Pour Gandon, une « donnée ouverte » est « *une donnée publique, parfois brute, et librement accessible, le terme « open » étant défini comme la qualité de la donnée à être réutilisable, reproductible, redistribuable, librement et sans discrimination*. » [Gandon et al., 2012, p.74].

Une des recommandations du rapport du groupe d'incubation du W3C: "Library Linked Data Incubator Group Final Report" (2011) incite, dans un premier temps, les bibliothèques à identifier de jeux des données susceptibles d'être exposés dans le web de données, et ensuite, à encourager le débat sur l'ouverture des données et les aspects juridiques qui l'accompagne.

L'ouverture des données en bibliothèques numériques implique des réflexions sur la nature des données ouvertes et sur la question du droit d'auteur. Cette ouverture ne peut pas être réalisée sans que les conditions techniques, économiques et juridiques soient prêtes et mises en place.

Le web sémantique, considéré comme un projet pour lier les données numériques et les rendre visibles sur le web, constitue le moyen technique pour favoriser l'ouverture des données liées (Open Linked Data). Grâce aux standards ouverts et compatibles avec le web, et à l'interopérabilité des données, le web sémantique favorise l'ouverture et la réutilisation des données.

Dans la deuxième enquête, nous avons demandé aux professionnels de l'information leur avis par rapport à l'utilisation de différentes expressions pour désigner le web sémantique. L'analyse des résultats a montré que certains professionnels de l'information préfèrent utiliser l'expression «web de données » parce qu'elle est plus précise, et touche le cœur de leur métier : l'organisation et la diffusion des données et des métadonnées. En revanche, l'expression «web sémantique », bien qu'elle soit plus ambitieuse, est employée pour privilégier le coté historique et l'évolution chronologique du terme.

Cependant, quelques enquêtés préfèrent utiliser l'expression « linked open data » parce qu'elle est moins « trompeuse » et comprend la notion de l'ouverture des données. Ci-dessous des extraits tirés de l'entretien avec l'enquêté [E10] et illustrant ce point de vue :

> « Par rapport aux bibliothécaires dont la mission est finalement la modélisation, l'organisation et la mise à disposition des savoirs sous toutes leurs formes, je trouve que la question de présenter ça sous l'angle de linked open data me semble être un pas plus facile à franchir et plus opérant parce qu'on remet au cœur la question de la métadonnée et des relations qu'on construit entre ces métadonnées. »[E10]

> « Et ce sont ces relations qui font sens, et donc sans aller chercher l'adjectif sémantique, en parlant de relation on rejoint tout à fait la modélisation FRBR etc. et en ajoutant le terme « open » ça nous remet dans une posture d'agent public constituant une

> connaissance au service de tous. Donc c'est aussi pour ces valeurs là que je privilégie le terme de « linked open data ».[E10]

Nous pensons que l'utilisation de l'expression web de données ou linked data est plus adaptée que celle de web sémantique dans le contexte des bibliothèques. Elle exprime bien l'objectif de cette évolution qui consiste à traiter les données et non pas les documents. Par contre, l'expression web sémantique semble être plus loin des pratiques des professionnels de l'information. Cela a pour effet de ne pas sensibiliser ces derniers aux enjeux de cette évolution du web.

Nous pensons que l'expression linked open data correspond bien aux objectifs des bibliothèques numériques, dans le sens où ces dernières visent, à décrire et lier les données, et à assurer un accès libre et gratuit à leurs données.

L'ouverture des données en bibliothèques numériques concerne les données et les métadonnées. Les premières couvrent les documents libres de droits et les données publiées par les bibliothèques (des rapports, des statistiques etc.). Et les deuxièmes sont les métadonnées utilisées par les bibliothèques pour identifier, décrire et relier les données entre elles sur le web.

Dans ce contexte, Sylvie Dalbin indique que travailler dans le cadre du web sémantique suppose de rompre avec la logique de fermeture des données [Dalbin et al., 2011]. Appliquer le web sémantique en bibliothèques numériques nécessite de travailler avec des données ouvertes. Pour relier les données entre elles, il faut être capable de communiquer sans difficulté technique ni juridique. Le partage des données et des métadonnées et leur réutilisation exigent que ces dernières soient sous licence ouverte.

Le linked open data est un projet à la fois technique, économique, culturel et politique. Notre recherche ne couvre pas tous ces aspects. Nous avons présenté brièvement ce sujet pour souligner les enjeux liés à la réutilisation des données en bibliothèques et pour ouvrir des perspectives de recherches qui aborderaient ces différents aspects.

6.4 Les défis et les obstacles à l'usage du web sémantique en bibliothèques numériques

Dans l'état de l'art, nous avons abordé la question des défis et des obstacles liés à l'adoption du web sémantique en bibliothèques numériques. Certains auteurs pensent que ces obstacles sont dus à une méconnaissance des compétences et des connaissances des bibliothécaires par la communauté du web sémantique [Greenberg, 2007]. D'autres trouvent qu'il y a un manque de sensibilisation des bibliothécaires envers les technologies du web sémantique [Byrne, 2010; Gonzales, 2014]. Pour Emmanuelle Bermès, ce sont des difficultés liées à la conversion des données dans le web des

données. Cela pose des problèmes de compatibilité avec les formats existants en bibliothèques ainsi que des problèmes liés à la maintenance des identifiants URI [Bermès, 2011]. Deux autres obstacles concernant la protection de la vie privée [Goddard et Byrne, 2010] et l'absence de confiance dans les données provenant d'autres disciplines [Singer, 2009].

Malgré tous les défis cités, il faut noter que la majorité des obstacles ne sont pas d'ordre technique. Il s'agit maintenant de préparer les bibliothèques et les bibliothécaires à intégrer le web sémantique et le mettre en place dans les bibliothèques [Goddard et Byrne, 2010] .

Dans la première enquête, les résultats[95] ont montré que les obstacles à l'utilisation du web sémantique se concentrent autour de trois éléments : le manque d'expertise (25 réponses sur 33), la complexité technique (20 réponses sur 33) et la diversité des standards (12 réponses sur 33). Les principales causes impliquées dans le non usage du web sémantique en bibliothèques numériques[96] sont le manque d'intérêt des professionnels de l'information et le manque des moyens techniques pour sa mise en place.

Dans la deuxième enquête, les résultats[97] ont montré que les obstacles à l'utilisation du web sémantique se situent davantage au niveau métier et stratégique qu'au niveau technique. Au niveau stratégique, les principaux obstacles sont d'ordre financiers ou liés au choix des référentiels sur lesquels s'appuyer pour représenter les données. Au niveau métier, il y a le manque d'expertise et la non reconnaissance des compétences des professionnels de l'information dans ce domaine. Les obstacles techniques, quant à eux, sont liés aux logiciels existants en bibliothèques, généralement pas adaptés aux standards du web sémantique.

En résumé, les résultats de notre enquête confirment celles de l'état de l'art. Les principaux obstacles à l'usage du web sémantique ne sont pas de nature technique, mais plutôt liés au manque d'expertise des professionnels de l'information dans ce domaine et au manque de confiance en leurs compétences. Cependant, les résultats de la deuxième enquête ont montré que les principales compétences pour accompagner l'usage du web sémantique en bibliothèques numériques sont des compétences documentaires.

Pour mieux interpréter les résultats obtenus nous pensons qu'il y a deux facteurs qui influencent les représentations qu'ont les professionnels de l'information de l'usage du web sémantique en bibliothèques numériques : le type de bibliothèque où ils travaillent et leurs connaissances et expertises dans le domaine.

95 Voir figure 5.14 page 157
96 Voir tableau 5.4 page 135
97 Voir tableau 5.14 page 157

Les résultats du questionnaire ont montré que les bibliothèques publiques sont les moins impliquées dans l'usage du web sémantique[98]. Ceci est dû au manque de compétences et d'intérêt des professionnels de l'information. Par contre les bibliothèques nationales (3 sur 2) sont les plus engagées dans ce mouvement, surtout la bibliothèque nationale de France. Ces bibliothèques possèdent les moyens financiers, techniques et humains permettant d'investir et d'essayer de nouveaux standards comme ceux du web sémantique. Pour les bibliothèques universitaires, les résultats ont montré que sept bibliothèques sur treize n'utilisent pas le web sémantique pour des raisons liées au manque de compétences, de moyens et à la complexité des standards. Enfin, les quatre bibliothèques spécialisées qui ont répondu au questionnaire n'utilisent pas le web sémantique à cause d'un manque de compétences et de moyens financiers.

La deuxième enquête (les entretiens) nous a permis de construire une vision plus profonde permettant de mieux comprendre les propos des enquêtés. Nous avons remarqué que les professionnels de l'information qui travaillent à la bibliothèque nationale de France ont apporté plus d'informations sur le sujet traité en répondant de manière développée sur les aspects techniques et surtout de façon critique. Cela s'explique par la décision prise par la bibliothèque d'utiliser le web sémantique, qui par conséquent engage les professionnels de l'information à avoir une bonne connaissance et une maîtrise des standards du web sémantique. Les professionnels de la BnF ont suivi des formations et ils ont eux-mêmes donné des formations sur le web sémantique. Ces enquêtés pensent que les langages documentaires et les formats utilisés en bibliothèques peuvent continuer à exister et même être développés pour accompagner l'usage du web sémantique en bibliothèques numériques.

Les enquêtés [E2, E5, E6, E7 et E9][99] ont un discours restreint sur les aspects techniques du web sémantique. D'une part, l'usage du web sémantique dans ces bibliothèques se limite à des réflexions et des expérimentations portant sur un volume de données réduit, et d'autre part, il s'agit pour certains d'entre eux de suivre l'évolution des normes de descriptions bibliographiques. Ces enquêtés pensent que le passage vers le web sémantique est nécessaire pour que les bibliothèques se positionnent comme des acteurs déterminants sur le web. Cependant, certains d'entre eux pensent que malgré leurs compétences en terme de traitement des données structurées, les communautés du web sémantique semblent avoir méconnu leurs compétences.

Enfin, les enquêtés [E1, E8 et E10][100] ont un discours encourageant l'usage du web sémantique en

98 Les résultats de questionnaire : 24 bibliothèques sur 33 n'utilisent pas le web sémantique. 12 bibliothèques publiques, 7 bibliothèques universitaires, 4 bibliothèques spécialisées et une bibliothèque nationale. 9 bibliothèques utilisent le web sémantique : 2 bibliothèques nationales, 6 bibliothèques universitaires et une bibliothèque publique

99 Ces enquêtés ont répondu que leurs bibliothèques utilisent le web sémantique.

100 Ces enquêtés travaillent dans des bibliothèques qui n'utilisent pas le web sémantique.

bibliothèques. Les professionnels de l'information ne doivent pas avoir peur des nouvelles technologies, au contraire, cela fait partie de l'évolution de leurs métiers, bien que leur rôle sera toujours centré sur la médiation autour du public et des données.

Conclusion du chapitre

L'objectif de ce chapitre était d'examiner les résultats obtenus en les mettant en perspective. Nous avons pu montrer les enjeux de l'utilisation du web sémantique en bibliothèques numériques :

- Nous avons remarqué que certains enquêtés ont une représentation du web sémantique plus claire que les autres, surtout les professionnels de la BnF. Ils ont un regard critique de l'usage du web sémantique en bibliothèques numériques. Cela est dû à une expérience plus avancée et une maîtrise des outils du web sémantique.
- Nous avons montré que la visibilité sur le web constitue un enjeu principal de l'usage du web sémantique en bibliothèques numériques. Ces dernières ont comme objectif de valoriser leurs contenus en les intégrant dans le web.
- La réutilisation des données des bibliothèques ne se fait pas essentiellement par des bibliothèques mais par d'autres communautés comme les centres de recherche, les sites privés etc. Cela a pour effet de créer de nouveaux contextes d'usage permettant d'enrichir les données ainsi que les savoirs.
- Les objectifs des pratiques documentaires convergent avec ceux du web sémantique en terme de traitement des données structurées. Selon les professionnels de l'information, les pratiques documentaires sont nécessaires pour accompagner l'usage du web sémantique en bibliothèques numériques.
- Les langages documentaires ont un rôle à jouer dans le web sémantique. Ce dernier offre le cadre technique permettant de relier les données en utilisant le triplet RDF, mais la construction des concepts et des liens entre eux ne se fait pas de manière automatique, et implique l'usage des langages documentaires. En d'autres termes, un modèle conceptuel sur lequel s'appuyer pour déterminer ces liens.
- L'utilisation du web sémantique implique de travailler avec des données ouvertes afin de communiquer et de partager les données librement, et ainsi que pour pouvoir relier les données entre elles. C'est pourquoi la notion de linked open data semble être la plus adaptée

pour désigner cette évolution du web permettant de relier les données ouvertes.

- Les principaux obstacles à l'usage du web sémantique en bibliothèques numériques ne sont pas de nature technique, ils sont liés au manque d'expertise et au manque de confiance des professionnels de l'information en leurs compétences et leurs connaissances.

Les bibliothèques numériques peuvent devenir des acteurs importants du web. D'une part, elles contribuent à l'enrichissement du web actuel en exposant des données structurées et fiables. Et d'autre part, les professionnels de l'information grâce à leurs pratiques documentaires peuvent jouer un rôle dans la mise en œuvre et dans l'évolution du web sémantique.

Conclusion générale

Dans ce travail de thèse, nous avons étudié les usages et les applications du web sémantique en bibliothèques numériques en France. L'objectif était de s'interroger sur les pratiques professionnelles et leurs évolutions suite à l'introduction du web sémantique en bibliothèques.

La première partie était consacrée à l'état de l'art que nous avons articulé en trois chapitres. Dans le premier chapitre, nous avons discuté de la notion de web sémantique, de son histoire et de son évolution, à travers les approches de représentation des connaissances : l'approche des réseaux sémantiques et l'approche des réseaux conceptuels. Dans le deuxième chapitre, nous avons analysé la notion de bibliothèque numérique à travers les différentes définitions proposées par la communauté de pratique et la communauté de recherche. Une approche info-communicationnelle a été mobilisée afin de déterminer un cadre conceptuel permettant de comprendre la nature des bibliothèques numériques. Cela nous a permis de nous intéresser à la notion de dispositif info-communicationnel en prenant en considération la notion de médiation. Le troisième chapitre a été consacré à l'analyse des usages du web sémantique en bibliothèques numériques. Plus précisément nous nous sommes intéressés aux pratiques professionnelles mobilisées afin d'accompagner la mise en place du web sémantique en bibliothèques.

Dans la deuxième partie, nous avons présenté notre méthodologie de recherche. Le chapitre quatre a été consacré à la présentation des méthodes mobilisées pour étudier l'usage du web sémantique en bibliothèques numériques françaises. Nous avons d'abord expliqué la sélection de bibliothèques numériques basée sur une analyse de trois recensements. Cette sélection a donnée 98 bibliothèques répondant à nos critères de bibliothèques numériques. Ensuite, nous avons présenté le questionnaire envoyé aux bibliothèques sélectionnées. Nous avons reçu 33 réponses, ce qui nous a permis d'identifier les bibliothèques qui utilisent le web sémantique et celles qui ne l'utilisent pas. Enfin, nous avons présenté la méthodologie des entretiens qui ont été réalisés avec des professionnels de l'information travaillant en bibliothèques numériques. Le chapitre cinq a été consacré à la présentation et à l'analyse des résultats. Enfin, le dernier chapitre nous a permis de discuter et de mettre en perspective les résultats obtenus.

Dans cette conclusion générale, nous commençons par présenter les principales notions étudiées et les apports de notre travail. Ensuite, nous abordons les limites de l'étude effectuée et les

perspectives de recherche.

Les notions abordées dans cette recherche

Les principales notions que nous avons fréquemment mobilisées dans cette thèse sont les bibliothèques numériques, la médiation sous ces différentes formes et le web sémantique.

Les bibliothèques numériques

Les bibliothèques numériques peuvent être considérées comme des dispositifs info-communicationnels permettant d'établir un lien entre un usager et une information par le biais d'une double médiation complémentaire ; la première sert à identifier et représenter l'information, la médiation documentaire ; tandis que la seconde permet d'offrir aux usagers un accès aux savoirs, la médiation des savoirs.

La médiation documentaire, la médiation documentaire numérique et la médiation des savoirs

La médiation documentaire consiste à établir un lien entre le professionnel de l'information et le document par l'intermédiaire d'un langage documentaire. Le but est d'assurer une représentation intellectuelle du document ainsi qu'en faciliter l'accès.

La médiation documentaire numérique consiste à identifier les documents numériques et à les décrire pour extraire des métadonnées. Ce traitement, appelé traitement documentaire, nécessite l'intervention d'un professionnel de l'information qui va s'appuyer sur des langages documentaires et informatiques pour représenter les documents. Le but est d'offrir une représentation structurée de l'information et de ses liens avec d'autres ressources sur le web.

La médiation des savoirs, consiste à établir un lien entre l'usager et le document par l'intermédiaire de la médiation documentaire. En d'autres termes, la médiation des savoirs permet à l'usager d'accéder aux documents. Dans le contexte des bibliothèques numériques, cet accès se fait grâce à un traitement intellectuel réalisé par le professionnel de l'information. Cela veut dire que la médiation des savoirs nécessite en amont une médiation documentaire.

Le web sémantique

Le web sémantique peut être considéré comme une nouvelle approche de la représentation des connaissances. Il offre un cadre technique permettant d'identifier, de structurer, de décrire et de relier des données sur le web en assurant leur compréhension par l'homme et par la machine. Ce qui compte dans le web sémantique n'est pas la donnée elle-même mais les métadonnées utilisées pour décrire et relier les données afin de les rendre exploitables par les machines.

Validation des hypothèses de recherche

Pour une mise en valeur des pratiques documentaires

La première hypothèse de ce travail concernait le rôle des professionnels de l'information dans la mise en œuvre du web sémantique en bibliothèques numériques. Nous pensions que les pratiques documentaires se rencontrent avec les objectifs du web sémantique en termes de structuration des données pour en faciliter l'accès. Plus précisément, nous pensions que le web sémantique peut être considéré comme une nouvelle technique permettant aux professionnels de l'information d'exercer leurs pratiques documentaires et de les faire évoluer. D'une part, les données récoltées lors des entretiens ont permis de valider notre première hypothèse puisque 70 % des enquêtés pensent que les pratiques documentaires sont les plus mobilisées pour accompagner la mise en œuvre du web sémantique en bibliothèques. D'autre part, les données récoltées du questionnaire et des entretiens ont montré que le web sémantique fait partie de l'évolution du métier.

Cependant, les résultats de notre questionnaire ont montré que les principaux obstacles à l'usage du web sémantique sont liés au manque d'expertise des professionnels de l'information dans ce domaine, et au manque de confiance en leurs compétences. Malgré la considération des pratiques documentaires dans l'usage du web sémantique en bibliothèques, certains professionnels de l'information pensent qu'il est nécessaire de suivre des formations permettant de maîtriser les techniques et les standards du web sémantique. Cela montre que l’usage du web sémantique implique tout de même des spécificités techniques qui ne sont pas nécessairement maîtrisées par les professionnels de l’information travaillant au sein des bibliothèques.

En s'appuyant sur leurs pratiques documentaires et sur leurs compétences, les professionnels de l'information peuvent devenir des acteurs du web sémantique, dans le sens où leurs rôles ne se limitent pas à l'utilisation du web sémantique mais aussi au développement de ses standards pour assurer une meilleure organisation des connaissances.

Pour une mise en valeur des langages documentaires

Notre deuxième hypothèse concernait la représentation des connaissances dans le cadre du web sémantique. Nous avons pu montrer que les langages documentaires ont un rôle à jouer dans la représentation des données dans le contexte du web sémantique. Plus précisément, nous pensions que les graphes et les ontologies du web sémantique s'appuient, d'une part, sur un cadre conceptuel permettant de déterminer les concepts et les liens entre eux (les langages documentaires), et d'autre part, sur un cadre technique permettant d'exploiter ces données et leurs liens (les standards du web sémantique URIs, RDF, RDFS, OWL).

Les données récoltées ont permis de valider notre deuxième hypothèse puisque les enquêtés pensent que les liens entre les données ne peuvent pas être établis en utilisant seulement les standards du web sémantique. De ce fait, les langages documentaires constituent la base conceptuelle sur laquelle s'appuyer pour déterminer les liens entre les données.

Nous avons pu confirmer les résultats des travaux précédemment réalisés en considérant les langages documentaires indispensables pour rendre le web sémantique réalisable [Favier et EL-Hadi, 2013]. Autrement dit, les langages documentaires, en étant structurés et liés, facilitent l'élaboration des ontologies permettant d'organiser les connaissances en bibliothèques numériques.

Pour une mise en valeur des données des bibliothèques numériques

Notre troisième hypothèse concernait la visibilité des bibliothèques numériques sur le web. Dans l'état de l'art, nous avons constaté que le problème majeur des bibliothèques numériques est leur visibilité. Nous avons pu montré que le web sémantique peut améliorer la visibilité des bibliothèques numériques sur le web. En adoptant les standards du web sémantique, les bibliothèques numériques permettent à leurs données d'intégrer l'architecture du web. Ceci a pour effet de mettre les données sur le chemin de l'internaute sans que ce dernier soit obligé de passer par le site de la bibliothèque.

Nous avons pu validé notre hypothèse puisque 90 % des bibliothèques interrogées pensent que l'utilisation du web sémantique permet d'améliorer leur visibilité sur le web. Cependant nous n'avons pas pu vérifier la visibilité des données des bibliothèques sur les moteurs de recherche parce que parmi toutes les bibliothèques interrogées, seule la bibliothèque nationale de France exposait ses données sur le web. Cette dernière a indiqué que ses données remontaient dans les résultats du moteur de recherche Google. Néanmoins, nous avons pu constaté que seuls les *« contenus rares »*[101]

101 Expression utilisée par l'enquêté E3a pour désigner les contenus qui s'affichent dans le moteur de recherche Google

s'affichaient dans le moteur de recherche, cela veut dire qu'il y a un travail de référencement à faire pour que les autres données soient elles aussi prises en compte.

Nous avons pu confirmer les résultats des travaux précédemment réalisés portant sur les apports du web sémantique en bibliothèques numériques. Ceux-ci considéraient que l'un des avantages de l'utilisation du web sémantique en bibliothèques est de les rendre plus visibles sur le web [Baker et al, 2011, Hallo et al, 2015].

Autres résultats mis en avant par le travail empirique

Le travail empirique a pu mettre en avant plusieurs caractéristiques relatives à l'usage du web sémantique en bibliothèques numériques. Les résultats de notre recherche ont montré que l'interopérabilité et la communication entre les différentes communautés sur le web constituent un enjeu essentiel de l'utilisation du web sémantique en bibliothèques numériques. Cela a pour effet de faciliter l'échange et la récupération des données et des métadonnées entre plusieurs communautés sur le web. En effet, la réutilisation des données semblait être une des caractéristiques du web sémantique. Les résultats de notre recherche ont montré que le web sémantique facilite la réutilisation des données par des acteurs autres que ceux des bibliothèques. Une médiation autour des données nous semblait avantageux pour améliorer la récupération des données. En effet, l'usage du web sémantique en bibliothèques numériques nécessite trois niveaux de médiation ; la médiation documentaire permettant de représenter les données par des métadonnées ; la médiation des savoirs permettant de faciliter l'accès aux données grâce aux métadonnées ; la médiation autour des données favorisant la réutilisation et la récupération des données et des métadonnées par différents acteurs.

Offrant une interopérabilité technique et sémantique, le web sémantique pourrait faciliter la découverte de l'information. Notre étude ne nous a pas permis de conclure sur cette thématique. Cependant, il serait intéressant d'apporter plus de clarifications sur ce point en faisant des expérimentations (expériences, études de cas etc.) permettant de vérifier le lien entre le web sémantique et la découverte de connaissances.

Notre étude nous a permis de classer les bibliothèques numériques en quatre catégories : bibliothèques numériques nationales, bibliothèques numériques de lecture publique, bibliothèques numériques universitaires et bibliothèques numériques spécialisées. Ainsi, nous avons pu déterminer les enjeux et les obstacles liés à l'utilisation du web sémantique dans chaque catégorie.

L'utilisation du web sémantique implique de travailler avec des données ouvertes, ceci pour pouvoir

communiquer et partager les données librement, aussi pour relier les données entre elles. C'est pourquoi la notion de linked open data semble être la plus adaptée pour désigner cette évolution du web permettant de relier les données ouvertes.

Limites du travail

Le travail de recherche que nous avons mené présente certaines limites.

Tout d'abord les études relatives à l'usage du web sémantique en bibliothèques numériques sont en nombre restreint. Certains travaux cités dans l'état de l'art ne peuvent pas être généralisés, ce qui rend parfois difficile la comparaison entre les résultats des travaux abordés dans l'état de l'art et ceux de notre enquête.

Ensuite, étant donné qu'il n'était pas envisageable dans le cadre de cette thèse de réaliser un recensement exhaustif de toutes les bibliothèques numériques en France, nous nous sommes focalisées sur des recensements réalisés par trois organismes : la bibliothèque nationale de France, l'URFIST de Lyon et l'inventaire multilingue du patrimoine numérique MICHAEL. Au début de notre recherche et avant de délimiter notre terrain d'étude en France, nous avons commencé à recenser les bibliothèques numériques au niveau de l'Europe, c'est pour cette raison que l'inventaire MICHAEL a été choisi. Il aurait cependant été plus judicieux de choisir directement le catalogue des collections numérisées du patrimoine culturel français[102] qui est le volet national du projet européen MICHAEL. La sélection ne recense pas la totalité des bibliothèques numériques en France. Il est donc possible qu'il y ait de bibliothèques numériques répondant à nos critères de sélection mais qui n'ont pas été étudiées.

Malgré le fait que certains professionnels de l'information ont mentionné dans le questionnaire que leurs bibliothèques utilisent les standards du web sémantique, l'analyse de leurs entretiens a montré que cet usage est limité à des réflexions ou à des expérimentations restreintes. Par conséquent, les résultats de ces quelques entretiens ne permettent pas de tirer des conclusions générales et n'apportent pas non plus un regard critique et constructif sur l'usage du web sémantique dans ces bibliothèques.

Enfin, la taille réduite de l'échantillon de bibliothèques et de personnes interrogées ne peut refléter de manière représentative l'ensemble des pratiques professionnelles en bibliothèques numériques de

102 http://www.numerique.culture.fr/pub-fr/index.html

même que l'ensemble des représentations que se font les professionnels de l'information au sujet du web sémantique.

Perspectives de recherche

Les résultats obtenus dans cette thèse font ressortir de nouvelles pistes de recherche à explorer.

- Concernant la visibilité des bibliothèques numériques sur le web

Notre recherche a montré que les bibliothèques numériques souhaitent améliorer leur visibilité en utilisant les standards du web sémantique. Il serait important d'étudier et de mieux comprendre les liens entre l'usage du web sémantique et la visibilité des données des bibliothèques sur le web. De même, il importe de s'interroger sur les différents types d'informations qui seraient affichées dans les moteurs de recherche (par exemple des contenus rares etc.). Il s'agirait de réaliser des expérimentations permettant de vérifier si les données provenant d'une bibliothèque qui utilise les standards du web sémantique (la BnF par exemple) s'affichent dans les moteurs de recherche (Google par exemple). Ensuite, il serait nécessaire d'aborder les autres techniques (comme les techniques de référencement) qui permettent d'améliorer la visibilité des données et de les comparer avec celles du web sémantique.

- Concernant la découverte de l'information :

Afin de déterminer si le web sémantique permet de découvrir de nouvelles informations grâce aux liens entre les données, il serait intéressant de faire des expérimentations sur des systèmes d'informations permettant d'interroger des données liées. Il s'agirait d'examiner l'efficacité des liens entre les données pour la recherche d'information, mais également de s'interroger sur l'interopérabilité technique et sémantique rendant la découverte de l'information possible.

Il serait intéressant de compléter ces expérimentations par une enquête auprès des personnes qui utilisent ces dispositifs. L'objectif serait d'évaluer, d'un point de vue usager, la pertinence des résultats obtenus. Plus précisément, il s'agirait de vérifier si les résultats obtenues répondent aux besoins des usagers et à leurs attentes.

- Concernant l'évaluation des «bibliothèques numériques sémantiques » ou « linked open library »

Il serait possible de mener des recherches plus poussées permettant de dégager des critères pour évaluer les « bibliothèques numériques sémantiques ». Il s'agirait de mener des études orientées

système et orientées usage pour déterminer des modèles d'évaluation pour ces bibliothèques basés sur des critères comme la pertinence des résultats, l'interopérabilité, la découverte de l'information, la réutilisation des données, la visibilité, les standards utilisés etc. Ces modèles serviront à évaluer les bibliothèques qui utilisent les standards du web sémantique dans le but de s'interroger sur la performance de ces dispositifs (l'approche orientée système) et de comprendre les pratiques mises en place pour identifier, traiter et accéder à l'information (l'approche orientée usage).

- Concernant la réutilisation des données

Il serait aussi intéressant de mener des recherches permettant de déterminer les enjeux de la réutilisation des données culturelles et scientifiques. Il s'agirait d'identifier les différents acteurs qui seraient intéressés par ces données ainsi que les contextes d'usages de ces dernières. Dans un premier temps, il serait nécessaire de s'interroger sur les statuts juridiques de ces données ainsi que sur le cadre législatif et réglementaire permettant leur réutilisation. Ensuite, il s'agit de comprendre la manière dont ces données circulent entre les différents acteurs afin de créer une valeur ajoutée à l'information réutilisée. Cette recherche pourrait tenter d'examiner l'effet de la réutilisation des données sur l'économie des connaissances. Par exemple, en quoi la réutilisation des données permettrait de créer de nouvelles connaissances ?

Bibliographie

Bibliographie

[Aalberg et al. 2011] AALBERG, T., MERČUN, T., et ŽUMER, M. (2011). Coding FRBR-structured bibliographic information in MARC. *In International Conference on Asian Digital Libraries,* pages 128-137. Springer Berlin Heidelberg.

[Abascal et al., 2005] ABASCAL, R., RUMPLER, B. et PINON, J-M. (2005). Conception d'une ontologie dans le contexte d'une bibliothèque numérique. *In* PALERMITI, R., HENNERON, G. et POLITY, Y. *:L'organisation des connaissances : approches conceptuelles,* pages 191-200. L'Harmattan, Paris.

[Amar et Menon, 2011] AMAR, M., et MENON, B. (2011). Bienvenue dans la «gigantesque base de données». *Documentaliste-Sciences de l'Information*, *48*(4) : 22-23.

[Arms, 2001] ARMS, W.Y. (2001). *Digital Libraries: Universal Access to Human Knowledge.* Massachusetts Institue of Technology, Massachusetts.

[Aussenac-Gilles et al., 2006] AUSSENAC-GILLES, N., CONDAMINES, A. et SÈDES, F. (2006). Evolution et maintenance des ressources termino-ontologiques : une question à approfondir. *Information - Interaction - Intelligence*, Vol. Hors-série, p. 7-14.

[Bachimont et al., 2011] BACHIMONT, B., GANDON, F., POUPEAU, G., VATANT, B., TRONCY, R., POUYLLAU, S. et ZACKLAD, M. (2011). Enjeux et technologies : des données au sens. *Documentaliste-Sciences de l'Information*, 48(4):24-41.

[Bachimont, 2007] BACHIMONT, B. (2007). *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents.* Hermès-lavoisier, Paris.

[Baker et al., 2011] BAKER, T. et al. (2011). Library Linked Data Incubator Group Final Report. https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/ [En ligne].

[Balicco, 2013] BALICCO, L. (2013). La gestion des connaissances en contexte professionnel : principes et enjeux. *In* CLAVIER, V. et PAGANELLI, C. *L'information professionnelle,* pages 107-127. Lavoisier, Paris.

[Bara, 1995] BARA, B.G. (1995). *Cognitive Science: A Developmental Approach to the Simulation of the Mind.* Erlbaum, Hove, UK.

[Beau et Weber, 2010] STÉPHANE, B. et WEBER, F. (2010). *Guide de l'enquête de terrain.*

La Découverte, Paris.

[Beau, 2017] BEAU, F. (2017). Facettes et système d'information : Une approche de la classification focalisée sur un besoin de savoir pour agir. *Les Cahiers du numérique*, 13(1):15-142.

[Bellinger et al., 2004] BELLINGER, G., CASTRO, D. et MILLS, A. (2004), Data, Information, Knowledge and Wisdom. http://www.systems – thinking.org/dikw/dikw.htm [En ligne].

[Béquet, 2014] BÉQUET, G. (2014). *Trois bibliothèques européennes face à Google : aux origines de la bibliothèque numérique*. Ecoles des Chartes, Paris.

[Bermès et al., 2013] BERMÈS, E., ISAAC, A. et POUPEAU, G. (2013) *Le Web sémantique en bibliothèque*. Éditions du cercle de la Librairie, Paris.

[Bermès et Poupeau, 2012] BERMÈS, E. et POUPEAU, G. (2012). Les technologies du Web appliquées aux données structurées. *In* CALDERAN, L., LAURENT, P., LOWINGER, H. et MILLE , J. : *Le document numérique à l'heure du web de données,* pages 41-84. ADBS, [Paris].

[Bermès, 2011] BERMÈS, E. (2011). Convergence et interopérabilité : l'apport du web de données. http://conference.ifla.org/ifla77. [En ligne Consulté le 14/9/2012].

[Berners-Lee et al., 2001] BERNERS-Lee, T., HENDLER, J. et LASSILA, O. (2001). The Semantic Web [en ligne]. *Scientific American*, 284(5):29-37.

[Berners-Lee, 1989] BERNERS-Lee, T. (1989). Information Management: A Proposal. https://www.w3.org/History/1989/proposal.html n.d.Web. [En ligne].

[Berners-Lee, 1994] BERNERS-LEE, T. (1994). Universal Resource Identifiers used in the WWW. https://tools.ietf.org/html/rfc1630. [En ligne].

[Berners-Lee, 1996] BERNERS-LEE, T. (1996). The World Wide Web: Past, Present and Future. https://www.w3.org/People/Berners-Lee/1996/ppf.html. [En ligne].

[Berners-Lee, 1998] BERNERS-LEE, T. (1998). Semantic web road map. https://www.w3.org/DesignIssues/Semantic.html [En ligne].

[Berners-Lee, 2006] BERNERS-LEE, T. (2006). Linked data. https://www.w3.org/DesignIssues/LinkedData.html. [En ligne].

[Berroneau, 2005] BERRONEAU, D. (2005). Les bibliothèques numériques, la transmission

d'un savoir. Mémoire Master 2. Universite de Poitiers. http://classiques.uqac.ca/contemporains/Berroneau_Delphine/biblio_numeriques/biblio_numeriques.html[En ligne]

[Bertrand et Alix, 2015] BERTRAND, A. et ALIX, Y. (2015). *Les bibliothèques*. La Découverte, Paris.

[Beuscart et Peerbaye, 2006] BEUSCART, J-S, et Peerbaye, A. (2006). Histoires de dispositifs. *Terrains & travaux*, 11(2):3-15.

[Bishop et al., 2003] BISHOP, A. P., VAN HOUSE, N. A., et BUTTENFIELD, B. P. (2003). *Digital library use: Social practice in design and evaluation*. Mit Press.

[Borgman, 1999] BORGMAN, C. L. (1999). What are digital libraries? Competing visions. *Information processing & management*, *35*(3).

[Borgman, 2000] BORGMAN, C. L. (2000). Digital libraries and the continuum of scholarly communication. *Journal of documentation*, *56*(4):412-430.

[Boughanem et Savoy, 2008] BOUGHANEM, M. et SAVOY, J. (2008). *Recherche d'information : état des lieux et perspectives*. Hermès science, Paris.

[Brachman et Levesque, 2004] BRACHMAN, R.J., LEVESQUE, H.J. (2004). *Knowledge representation and reasoning*. Elsevier, Amsterdam.

[Brahaj et al., 2013] BRAHAJ A., RAZUM, M. et HOXHA, J. (2013). Defining Digital Library. In *International Conference on Theory and Practice of Digital Libraries,* pages 23-28. Springer, Berlin, Heidelberg.

[Breton et Proulx, 2012] BRETON, P., et PROULX, S. (2012). 11/Usages des technologies de l'information et de la communication. *In L'explosion de la communication. Introduction aux théories et aux pratiques de la communication,* pages 263-287. La Découverte, Paris.

[Briet, 1951] BRIET, S. (1951). *Qu'est ce que la documentation*. Edit, Paris.

[Brost, 1997] BORST, W.N. (1997). *Construction of engineering ontologies for knowledge sharing and reuse*. Universiteit Twente, Enschede.

[Brunet et Xuan, 2010] Brunet, N., Truong Vu, Xuan, (2010). Le web sémantique : en quoi le web semantique permet-il d'aborder le sens ? [en ligne] http://vuxuantr.free.fr/res/rapport_web_semantique-Nicolas_Brunet-Xuan_Truong_Vu.pdf

[Bygstad et al., 2009] BYGSTAD, B., GHINEA, G., et KLÆBOE, G. T. (2009). Organisational

challenges of the Semantic Web in digital libraries: A Norwegian case study. *Online Information Review*, *33*(5) :973-985.

[Cacaly et al., 2004] CACALY, S., LE COADIC, Y-F, POMART, P-D et SUTTER, E. (2004). *Dictionnaire de l'information*. Armand Colin, Paris.

[Cacaly, 2001] CACALY, S . (dir.) (2001). *Dictionnaire encyclopédique de l'information et de la documentation*. Nathan, Paris.

[Capurro et Hjørland, 2003] CAPURRO, R., et HJØRLAND, B. (2003). The concept of information. *Annual review of information science and technology*, *37*(1):343-411.

[Carmes et Noyer, 2013] CARMES, M. et NOYER J-M. (2013). *Les débats du numérique*. Presses de Mines, Paris.

[Chaudiron et Ihadjadène, 2010] CHAUDIRON, S., et IHADJADENE, M. (2010). De la recherche de l'information aux pratiques informationnelles. *Études de communication. langages, information, médiations*, (35) :13-30. http://journals.openedition.org/edc/2257[En ligne].

[Chaudiron, 2004] CHAUDIRON, S. (2004). *Évaluation des systèmes de traitement de l'information.* Lavoisier, Paris.

[Chaudiron, 2007] CHAUDIRON, S. (2007). Technologies linguistiques et modes de représentation de l'information textuelle. *Documentaliste-Sciences de l'information*, *44*(1): 30-39.

[Chisenga, 2003] CHISENGA, J (2003). Digital libraries and Virtual libraries : definitions, concepts and goals. http://repository.uneca.org/bitstream/handle/10855/1899/Bib-15173.pdf?sequence=1. [En ligne].

[Choi et Rasmussen, 2006] CHOI, Y., et RASMUSSEN, E. (2006). What is needed to educate future digital librarians: A study of current practice and staffing patterns in academic and research libraries. *D-lib magazine*, *12*(9):19.

[Claerr et Westeel, 2013] CLAERR, T. et WESTEEL, I. (dirs). (2013). *Manuel de constitution de bibliothèques numériques*. Electre, Paris.

[Cleveland, 1998] CLEVELAND, G. (1998). Digital libraries: definitions, issues and challenges. http://ifla.queenslibrary.org/vi/5/op/udtop8/udt-op8.pdf. [En ligne].

[Collins et Lotfus, 1975] COLLINS, A. M. et LOFTUS, E. F. (1975). A spreading-activation

theory of semantic processing. *Psychological review*, *82*(6):407.

[Collins et Quillian, 1969] COLLINS, A. M. et QUILLIAN, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2):240-247.

[Compiègne, 2011] COMPIÈGNE, I. (2011). *La société numérique en question(s)*. Sciences Humaines Éditions, Paris.

[Cothran, 2011] COTHRAN, T. (2011). Google Scholar acceptance and use among graduate students: A quantitative study. *Library & Information Science Research*, *33*(4):293-301.

[Cotte, 2004] COTTE, D. (2004). Le concept de «document numérique». *Communication et langages*, *140*(1):31-41.http://www.persee.fr/doc/colan_0336-1500_2004_num_140_1_3265 [En ligne]

[Cotte, 2013] COTTE, D. (2013). Qui (re) travaillera la donnée?. *Documentaliste-Sciences de l'Information*, *50*(3):24-25.

[Courbières, 2011] COURBIERES, C. (2011). Les langages documentaires : éléments dorganisation des connaissances. *In* GARDIES, C. (éd.) *Approche de l'information-documentation : concepts fondateurs,* p. 131-150. Cépaduès. Toulouse.

[Couzinet, 2001] COUZINET, V. (2001). *Jean Meyriat théoricien et praticien de l'information-documentation*. ADBS Editions, Paris.

[Couzinet, 2009] COUZINET, V. (2009). *Dispositifs info-communicationnels : questions de médiations documentaires*. Lavoisier, Paris.

[Couzinet, 2011] COUZINET, V. (2011). Les dispositifs : question documentaire. *In* GARDIES, C. (éd.) *Approche de l'information-documentation : concepts fondateurs,* pages 117-130. Cépaduès. Toulouse.

[Cox, 2012] COX, A. M. (2012). An exploration of the practice approach and its place in information science. *Journal of Information Science*, *38*(2):176-188.

[Creusot et Schopfel, 2005] CREUSOT, J., et SCHÖPFEL, J. (2005). Le Tao de la bibliothèque numérique–bibliothèque sans bibliothécaire? https://archivesic.ccsd.cnrs.fr/sic_00107231/document [En ligne].

[Cuche, 2010] CUCHE, D. (2010). *La notion de culture dans les sciences sociales.* La Découverte, Paris.

[Dabin et al., 2011] DALBIN, S., BERMÈS, E., ISAAC, A., WENZ, R., NICOLAS, Y.,

MERABTI, T., ... et VATANT, B. (2011). Approches documentaires: priorité aux contenus. *Documentaliste-Sciences de l'Information*, *48*(4):42-59.

[Davis et al., 1993] DAVIS, R., SHROBE, H., et SZOLOVITS, P. (1993). What is a knowledge representation?. *AI magazine*, *14*(1):17.

[Ding, 2010] Ding, Y. (2010). Semantic Web: Who is who in the field : a bibliometric analysis. *Journal of Information Science*, *36*(3):335-356.

[DiNucci, 1999] DiNucci, D. (1999). Fragmented future. http://darcyd.com/fragmented_future.pdf. [en ligne] .

[Favier et El-hadi, 2013] FAVIER, L.et EL-HADI, W.M (2013). L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité du savoir. In PAPY, F. : *Recherches ouvertes sur le numérique*, pages 23-38. Lavoisier, Paris.

[Fondin, 2001] FONDIN, H. (2001). La science de l'information : posture épistémologique et spécificité disciplinaire. *Documentaliste-Sciences de l'Information*, 38(2):112-122.

[Fox et Marchionini, 2001] FOX, E. A., et Marchionini, G. (2001). Digital libraries: introduction. *Communications of the ACM*, *44*(5):30-32.

[Fox, 1993] Fox, E. A. (1993). Sourcebook on Digital Libraries: Report for the National Science Foundation. (Technical Report TR-93-35). Blacksburg, VA: Dept. of Computer Science, Virginia Tech. [Online at]: fox.cs.vt.edu

[Gandon et al., 2012] GANDON, F., CORBY, O., et Faron-ZUCKER, C. (2012). *Le web sémantique: Comment lier les données et les schémas sur le web?*. Dunod, Paris.

[Gardiès et al., 2010] GARDIÈS, C., FABRE, I., et COUZINET, V. (2010). Re-questionner les pratiques informationnelles. *Études de communication. langages, information, médiations*, (35):121-132.

[Gardiès, 2011] GARDIÈS, C. (2011). *Approche de l'information-documentation: concepts fondateurs*. Cépaduès, Toulouse.

[Gardiès, 2012] GARDIÈS, C. (2012). *Dispositifs info-communicationnels de médiation des savoirs : cadre d'analyse pour l'informationdocumentation [HDR]*. Université de Toulouse.

[Gimazane et al., 2010] GIMAZANE, R., Smith, F., NGUYEN, C., BRULEY, C., BERTRAND, A., EPRON, B. et CHARTRON, G. (2010). Nouveaux documents, nouvelles compétences. *Documentaliste-Sciences de l'Information*, 47(2):56-67.

[Goddard et Byrne, 2010] GODDARD, L., et Byrne, G. (2010). The strongest link: Libraries and linked data. *D-Lib magazine*, 16(11/12).

[Godin, 2004] GODIN, C. (2004). *Dictionnaire de philosophie*. Fayard, Paris.

[Gonzales, 2014] GONZALES, B. M. (2014). Linking libraries to the web: linked data and the future of the bibliographic record. *Information Technology and Libraries (Online)*, *33*(4): 10.

[Greenberg, 2007] GREENBERG, J. (2007). Advancing the semantic web via library functions. *Cataloging & classification quarterly*, *43*(3-4):203-225.

[Gruber, 1993] GRUBER, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing?. *International journal of human-computer studies*, *43*(5-6):907-928.

[Guibert et Jumel, 1997] GUIBERT, J., JUMEL, G. (1997). *Méthodologie des Pratiques de Terrains en Sciences Humaines et Sociales*. Armand Colin, Paris.

[Guyot, 2000] GUYOT, B. (2002). Mettre en ordre les activités d'information, nouvelle forme de rationalisation organisationnelle. *Les Enjeux de l'information et de la communication*, (1): 49-64.

[Hallo etal., 2016] HALLO, M., LUJÁN-MORA, S., MATÉ, A., et TRUJILLO, J. (2016). Current state of Linked Data in digital libraries. *Journal of Information Science*, *42*(2):117-127.

[Harter, 1996] HARTER, S. P. (1996). What is a digital library? Definitions, content, and issues. In *International conference on digital libraries and information services for the 21st century, pages* 8-17.

[Heery, 2004] HEERY, R. (2004). Metadata Futures: Steps Toward Semantic Interoperability. In HILLMAN D. I. et WESTBROOKS, E. L., éditeurs : *Metadata in Practice*, pages 257–271. American Library Association, Chicago.

[Heiler, 1995] HEILER, S. (1995). Semantic interoperability. *ACM Computing Surveys (CSUR)*, *27*(2):271-273.

[Hjørland, 2009] HJØRLAND, B. (2009). Concept theory. *Journal of the American Society for Information Science and Technology*, *60*(8):1519-1536.

[Hudon et El-Hadi, 2010] HUDON, M., et EL HADI, W. M. (2010). Organisation des connaissances et des ressources documentaires. *Les Cahiers du numérique*, *6*(3):9-38.

[Hudon et El-Hadi, 2012] HUDON, M., et EL HADI, W. M. (2012). Organisation des connaissances : épistémologie, approches théoriques et méthodologiques. *Études de communication*, 39:9-14.

[Ibekwe-Sanjuan, 2012] Ibekwe-Sanjuan, F. (2012). *La science de l'information : origines, théories et paradigmes.* Lavoisier, Paris.

[Ihadjadene et Chaudiron, 2008] Ihadjadene, M. et Chaudiron, S., (2008). L'Étude des dispositifs d'accès à l'information : approches croisées. *In* PAPY, F. : *Problématiques émergentes dans les sciences de l'information*, pages 183-207. Hermès-Lavoisier, Paris.

[Ihadjadène et Favier, 2008] IHADJADENE, M., et FAVIER, L. (2008). Langages documentaires: vers une «crise de l'autorité»?. *Sciences de la Société*, (75):10-21.

[Illien et al., 2013] ILLIEN, G., HOLOGNE, O., POUYLLAU, S., ALFONSI, G., TROEIRA, J. P., DELAHOUSSE, J., ... et HUOT, C. (2013). Enjeux professionnels. *Documentaliste-Sciences de l'Information*, *50*(3):26-41.

[Javeau, 1992] JAVEAU, C. (1992). *L'enquête par questionnaire : Manuel à l'usage de praticien.* Éditions de l'Université de Bruxelles, Bruxelles.

[Jeanneret, 2000] JEANNERET, Yves. (2000). *Y a-t-il (vraiment) des technologies d'information?* Presses universitaires de Septentrion, Paris.

[Jeanneret, 2008] JEANNERET, Y. (2008). *Penser la trivialité : la vie triviale des êtres culturels*. Hermès, Paris.

[Jouet, 1993] JOUËT, J. (1993). Pratiques de communication et figures de la médiation. *Réseaux*, *11*(60):99-120.

[Lagoze et al., 2006] LAGOZE, C., KRAFFT, D. B., PAYETTE, S., et JESUROGA, S. (2006). Qu'est-ce qu'une bibliothèque numérique, au juste? Au-delà des fonctions recherche et accès dans la National Science Digital Library. *AMETIST: Appropriation, Mutialisation, Expérimentations des Technologies de l'IST.*https://archivesic.ccsd.cnrs.fr/sic_00123423/document [Enligne].

[Lamizet, 1995] Lamizet, B. (1995). *Les lieux de la communication.* Mardaga, Liège.

[Latour, 2006] LATOUR, B. (2006). *Changer de société Refaire de la sociologie.* La Découverte, Paris.

[Le Coadic, 2004] LE COADIC, Y-F. (2004). *Usages et usagers de l'information.* Armand Colin, Paris.

[Le Coadic, 2007] LE COADIC, Y-F (2007). *Le besoin d'information : formulation, négociation, diagnostic*. ADBS, Paris.

[Le Crosnier, 2007] LE CROSNIER, H. (2007). *Réseau, bibliothèques et documents numériques: architecture informatique et construction sociale* [Thèse, Université de Caen].

[Le Roy et Pierrette, 2012] LE ROY, J., PIERRETTE, M. (2012). *Petit manuel méthodologique du questionnaire de recherche : De la conception à l'analyse*. Enrick B. Editions.

[Lesk, 1997] LESK, M. (1997). *Practical digital libraries: Books, bytes, and bucks*. Morgan Kaufmann.

[Lyman, 1999] LYMAN, P. (1999). The social functions of digital libraries: Designing information resources for virtual communities. *Retreived March*, *5*, 2007 .

[Manes-Gallo et Paganelli, 2003] MANES-GALLO, M. C., et PAGANELLI, C. (2003). La recherche d'information assistée par ordinateur: quelle représentation des connaissances?. *Les Enjeux de l'information et de la communication*, *2003*(1):36-51.

[Maniez et El-Hadi, 1999] MANIEZ et Jacques, EL-HADI, W.M. (1999). *Organisation des connaissances en vue de leur intégration dans les systèmes de représentation et recherche d'information*, Actes des Premières Journées du Chapitre Français de l'ISKO, Actes des Premières Journées du chapitre français de l'ISKO, Edition du Conseil Scientifique de l'Université Charles De Gaulle Lille 3, Collection UL3: Travaux & Recherches.

[Maniez, 1987] MANIEZ, Jacques. (1987). *Les langages documentaires et classificatoires*. Ed. D'Organisation, Paris.

[Menon, 2003] MENON, B. (2003). Journée d'étude ADBS-IUT Paris-V. *Documentaliste-Sciences de l'Information*, *40*(6):387-391.

[Meyriat, 1985] MEYRIAT, J. (1985). Information versus communication ? In A-M, Laulin (dir). L'espace sociale de la communication : concepts et théories. Retz-CNRS, Paris.

[Meyriat, 1994] MEYRIAT, M. (1994).Y a-t-il une place pour une théorie de la documentation ? *In* COUZINET, V. Je*an Meyriat, théoricien et particien de l'informaiton-documentation*, pages 393-402. ADBS, Paris.

[Miège, 2005] MIÈGE, B. (2005). *La pensée communicationnelle*. PUG, Grenoble.

[Moirez et Stutzmann, 2013] MOIREZ, Pauline et STUTZMANN, D. (2013). *Manuel de constitution de bibliothèques numériques. Signaler les ressources numérisées: enrichissement, visibilité, dissémination*. Electre, Paris.

[Mucchielli, 1996] MUCCHIELLI, A. (1996). *Dictionnaire des méthodes qualitatives en sciences humaines et sociales*. Armand colin, Paris.

[Nevers, 2000] NEVERS, B. (2000). Implication Des Mécanismes D'activation, D'intégration et de Construction Sur La Nature D'une Trace Mnésique : Etude Des Effets de Fréquence Sur L'amorçage de Répétition. [Thèse], Université Lumière Lyon 2.

[Nonthacumjane, 2011] NONTHACUMJANE, P. (2011). Key skills and competencies of a new generation of LIS professionals. *IFLA journal*, *37*(4):280-288.

[Ory-Lavollée, 2002] ORY-LAVOLLÉE, B. (2002). La diffusion numérique du patrimoine, dimension de la politique culturelle, remue à Mme la Ministre de la culture et de la communication.http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/024000087.pdf [En ligne]

[Ory-Lavollée, 2009] ORY-LAVOLLÉE, B. (2009). Partager notre patrimoine culturel, remis à Monsieur le Ministre de la Culture et de la Communication.http://www.enssib.fr/bibliotheque-numerique/documents/48591-partager-notre-patrimoine-culturel.pdf [en ligne]

[Otlet, 1934] OTLET, Paul (1934). *Traité de documentation*. Editions Mundaneum, Bruxelles.

[Paganelli, 2012] PAGANELLI, C. (2012). *Une approche info-communicationnelle des activités informationnelles en contexte de travail: Acteurs, pratiques et logiques sociales* (Doctoral dissertation, Université de Grenoble).

[Paganelli, 2013] PAGANELLI, C. (2013). Les activités informationnelles en contexte de travail : questionnements en information-communication. *In* CLAVIER, V. et PAGANELLI, C. *L'information professionnelle,* pages 221-240. Lavoisier, Paris.

[Papy, 2005] Papy, F. (2005). *Les bibliothèques numériques*. Hermès science publications, Lavoisier, Paris.

[Pédauque, 2006] PÉDAUQUE, R.T. (2006). *Le document à la lumière du numérique.* C&F éditions, Caen.

[Pédauque, 2007] PÉDAUQUE, R. T. (2007). *La redocumentarisation du monde*. Cépaduès, Toulouse

[Pradel, 2013] PRADEL, C. (2013). *D'un langage de haut niveau à des requêtes graphes permettant d'interroger le web sémantique* [Thèse, Université de Toulouse, Université Toulouse III-Paul Sabatier].

[Rastier, 1991] RASTIER, F. (1991). S*émantique et recherches cognitive*. Presses Universitaires de France, Paris.

[Rey-Debove, 2004] REY-DEBOVE, J. (2004). *Le Robert Brio: analyse des mots et régularités du lexique* . Robert, Paris.

[Rey, 2005] REY, A. (2005). *Dictionnaire culturel en langue française*. Le Robert, Paris.

[Rhee et al., 2008] RHEE, S. K., LEE, J., et PARK, M. W. (2008). Semantic relevance measure between resources based on a graph structure. In *Computer Science and Information Technology,* pages 229-236. IEEE.

[Ristoski et Paulheim, 2016] RISTOSKI, P., et PAULHEIM, H. (2016). Semantic Web in data mining and knowledge discovery: a comprehensive survey. *Web semantics: science, services and agents on the World Wide Web*, *36:*1-22.

[Role, 2000] Role., F. (2000). Métadonnées et structuration des bibliothèques numériques. In *Bibliothèques numériques : cours INRIA*, pages 143–170. Editions de l'ADBS. La Bresse.

[Sabah, 1990] SABAH, G. (1990). *L'intelligence artificielle et le langage*. Hermès, Paris.

[Salaun et Arsenault, 2009] SALAUN, J-M. et ARSENAULT, C. (2009). *Introduction aux sciences de l'information*. Les Presses de l'Université de Montréal.

[Salaun, 2008] SALAÜN, J-M. (2008). Le défi du numérique : redonner sa place à la fonction documentaire. https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/2225/salaun-jm-defi-documentaliste-mktg.pdf?sequence=1&isAllowed=y [En ligne]

[Salaun, 2012] SALAÜN, J-M. (2012). *Vu, lu, su: les architectes de l'information face à l'oligopole du Web*. La Découverte, Paris.

[Saleh et Mkadmi, 2008] SALEH, I., MKADMI, A. (2008). *Bibliothèques numériques et recherche d'information*. Hermès Science Publications, Lavoisier, Paris.

[Saracevic, 1999] SARACEVIC, T.(1999). Information science. T*he Journal of the American Society for Information Science* , 50(12):1051–1063.

[Senié-Demeuresse et Fabre, 2011] SENIÉ-DEMEURESSE, J. et FABRE, I. (2011). Usages et pratiques de l'information. In Gardiès, C. : *Approche de l'information-documentation. Concepts fondateurs,* pages 231-252. Cépaduès, Toulouse.

[Shadbolt et Berners-Lee, 2006] SHADBOLt, N., BERNERS-LEE, T., et HALL, W. (2006). The semantic web revisited. *IEEE intelligent systems*, *21*(3):96-101.

[Sharon et Frank, 2000] SHARON, T., et FRANK, A. J. (2000). Digital Libraries on the Internet.https://files.eric.ed.gov/fulltext/ED450757.pdf [En ligne]

[Simonnot, 2012] SIMONNOT, B. (2012). *L'accès a l'information en ligne : moteurs, dispositifs et médiations* . Lavoisier, Paris.

[Singer, 2009] SINGER, R. (2009). Linked library data now!. *Journal of Electronic Resources Librarianship*, *21*(2):114-126.

[Souchier et al., 2003] SOUCHIER, E., JEANNERET, Y., LE MAREC, J. (2003). *Lire, écrire, récrire : objets, signes et pratiques des médias informatisés.* Bibliothèque publique d'Information, Paris.

[Sowa, 1984] SOWA, J.F. (1984). *Conceptual structures: information processing in mind and machine*.Addison-Wesley Longman Publishing, Boston, MA, USA.

[Sreenivasulu, 2000] SREENIVASULU, V. (2000). The role of a digital librarian in the management of digital information systems (DIS). *The electronic library*, *18*(1):12-20.

[Tennis, 2008] TENNIS, J.T. (2008). Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework. *Knowledge Organization,* 35(2/3):102-112.

[Tennis, 2012] TENNIS, J.T. (2012). Le poids du langage et de l'action dans l'organisation des connaissances : Position épistémologique, action méthodologique et perspective théorique. *Études de communication*, 39:15-40.

[Tennnant, 1999] TENNANT, R. (1999). Digital v. electronic v. virtual libraries. *Última atualização em* (1999). [En ligne].

[Tessier, 2010] TESSIER, M. (2010). Rapport sur la numérisation du patrimoine écrit, Paris, remis à M. le Ministre de la Culture et de la Communication, 12 janvier 2010, 64 p. http://www.lefigaro.fr/assets/pdf/rapport-numerisation.pdf [en ligne]

[Thiolon, 2003] THIOLON, C. (2003). L'image des métiers de l'I-D: Diagnostic et recommandations pour une communication active. *Documentaliste-Sciences de l'Information*, 40(6):396-400.

[Van Noorden, 2014] VAN NOORDEN, R. (2014). Scientists and the social network. *Nature*, *512*(7513):126.

[Vitali-Rosati, 2014] VITALI-ROSATI, M. (2014). Pour une définition du "numérique". In Sinatra, M.E. et Vitali-Rosati, M. (dir.), *Pratiques de l'édition numérique*, pages 63-75. Les Presses de l'Université de Montréal, Montréal.

[Wahlster et al., 2006] WAHLSTER, W., DENGEL, A., TELEKOM, D., DENGEL, W., DENGLER, C. D., HECKMANN, D., ... et Schwarzkopf, E. (2006). Web 3.0: Convergence of web 2.0 and the semantic web. In *In Technology Radar, Feature Paper, 2nd ed.; Deutsche Telekom Laboratories*.

[Wenger et al., 2002] WENGER, E., MCDERMOTT, R. A., et SNYDER, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business Press.

[Witten et al., 2009] WITTEN, I.H., BAINBRIDGE, D. and NICHOLS, D.M (2009). *How to build a digital library.* Elsevier, Morgan Kaufmann, Amsterdam.

[Zacklad, 2005] ZACKLAD, M. (2005, February). Vers le Web Socio Sémantique: introduction aux ontologies sémiotiques. In *Ingénierie des Connaissance*. https://archivesic.ccsd.cnrs.fr/sic_00001347/document [En ligne].

Communauté UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Sciences de l'Information et de la Communication**

Arrêté ministériel : 25 mai 2016

Présentée par

Hiba MELHEM

Thèse dirigée par **Laurence BALICCO**

préparée au sein du **Laboratoire GRESEC (Groupe de recherche sur les enjeux de la communication)**
dans **l'École Doctorale Langues, Littératures et Sciences Humaines**

Usages et applications du web sémantique en bibliothèques numériques

Volume 2 : Annexes

Thèse soutenue publiquement le **27 octobre 2017**,
devant le jury composé de :

Madame Viviane COUZINET
Professeure en Sciences de l'Information et de la Communication
Université Paul Sabatier, Toulouse3, Rapporteur

Monsieur Madjid IHADJADENE
Professeur en Sciences de l'Information et de la Communication
Université Paris 8, Président et Rapporteur

Madame Viviane CLAVIER
Maître de conférences HDR en Sciences de l'Information et de la Communication
Université Grenoble Alpes, Examinateur

Madame Laurence BALICCO
Professeure en Sciences de l'Information et de la Communication
Université Grenoble Alpes, Directrice de thèse

Liste des annexes

Annexe I Description des données de l'étude sur l'usage des termes liés au web sémantique dans les bases de données et les moteurs de recherche étudiés …….. 1

Annexe II Jeux de données publiés dans le cadre de groupe « bibliothèques et web de données » Library Linked Data …… 3

Annexe III Les vocabulaires d'autorité publiés dans le cadre de groupe « bibliothèques et web de données » Library Linked Data …… 6

Annexe IV Éléments de descriptions de métadonnées élaborés dans le cadre de groupe « bibliothèques et web de données » Library Linked Data …… 12

Annexe V Recensement de la bibliothèque nationale de France ……23

Annexe VI Recensement fait par l'URFIST de Lyon …… 24

Annexe VII Recensement fait par l'inventaire Multilingue du Patrimoine Culturel en Europe..27

Annexe VIII Les bibliothèques sélectionnées …….. 35

Annexe IX Le courrier électronique envoyé aux bibliothèques cibles pour demander leur participation au questionnaire …….. 42

Annexe X Le courrier électronique envoyé aux professionnels de l'information pour demander un entretien …… 43

Annexe XI Classement des bibliothèques numériques sélectionnées …….. 44

Annexe XII Les entretiens …… 47

Annexe XIII Questionnaire destiné aux bibliothécaires …… 102

Annexes

Annexe I : Description des données de l'étude sur l'usage des termes liés au web sémantique dans les bases de données et les moteurs de recherche étudiés

Nombre de documents correspondant à chaque période dans la base de données « Web of science »			
	1994-2000	**2001-2005**	**2005-2016**
Semantic web	10	2845	10204
Linked data	7	100	2518
Linked open data	0	0	691
Web 3.0	1	0	179
Web of data	0	3	337

Nombre de documents correspondant à chaque période dans la base de données « Science Direct »			
	1994-2000	**2001-2005**	**2005-2016**
Semantic web	11	545	5908
Linked data	11	812	4673
Linked open data	0	0	415
Web 3.0	0	1	330
Web of data	5	37	287

Nombre de documents correspondant à chaque période dans la base de données « Sage journals »			
	1994-2000	**2001-2005**	**2005-2016**
Semantic web	5	50	434
Linked data	25	42	435
Linked open data	0	0	33
Web 3.0	0	0	86
Web of data	2	0	26

Nombre de documents correspondant à chaque période dans la base de données « Cairn »			
	1994-2000	**2001-2005**	**2005-2016**
Web sémantique	0	40	196
Linked data	0	1	38
Linked open data	0	0	17
Web 3.0	0	0	72
Web de données	0	1	56
Web of data	0	0	14

Nombre de documents correspondant à chaque terme dans le moteur de recherche Google	
Semantic web	4 640 000
Linked data	15 100 000
Linked open data	378 000
Web 3.0	533 000
Web of data	2 000 000
Web sémantique	402 000
Web de données	329 000

Nombre de documents correspondant à chaque terme dans le moteur de recherche Google scholar	
Semantic web	315 000
Linked data	75 800
Linked open data	16400
Web 3.0	20900
Web of data	12700
Web sémantique	5740
Web de données	677

Annexe II : Jeux de données publiés dans le cadre de groupe « bibliothèques et web de données » Library Linked Data (Source :https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/)

Published Datasets

The present section lists all datasets (mostly bibliographic) that are available on the Library Linked Data group from the Data Hub at the time of the publication of the report. For more information, the reader is invited to follow the links for individual packages.

BibBase

BibBase.org facilitates the dissemination of scientific publications over the Internet.

British National Bibliography (BNB)

British National Bibliography (BNB) published as Linked Data, linked to external sources including VIAF, LCSH, Lexvo, GeoNames, MARC country, and language, Dewey.info, RDF Book Mashup....

Calames

Calames is the French academic union catalog of archives and manuscripts, maintained by ABES.

Chronicling America

Chronicling America provides access to information about historic newspapers and select digitized newspaper pages. It contains 140,000 newspapers and 3.2 million pages.

Cambridge University Library dataset #1

This data marks the first major output of the COMET project, a JISC-funded collaboration between Cambridge University Library and CARET.

data.bnf.fr - Bibliothèque nationale de France

data.bnf.fr gathers data from the different databases of the Bibliothèque nationale de France, so as to create Web pages about works and authors, together with a RDF view on the extracted data.

Scottish Mountaineering Council Journals Issues 1-36

Digital archive of 'Scottish Mountaineering Club Journal Issues 1 to 36, 1890-1901', which was created by Alan Dawson, University of Strathclyde, with funding from the Scottish Mountaineering Trust.

CrossRef DOI Resolver

Digital Object Idenfiers (DOI) are a persistent identifier strategy used by around 3,000 publishers to identify their documents, mostly scholarly publications.

Europeana Linked Open Data

The data.europeana.eu pilot is part of Europeana's effort of making its metadata available as Linked Open Data on the Web. It currently serves metadata on 3.5 million items.

Freebase

Freebase is an open database of the world's information. It is built by the community and for the community — free for anyone to query, contribute to, built applications on top, or integrate into

their websites.

Hungarian National Library (NSZL) catalog

OPAC and Digital Library and the corresponding authority data as Linked Open Data.

Linked Periodicals Database

The Linked Periodicals Database is a dataset from the Data Incubator which aggregates journal metadata provided by CrossRef, Highwire Press, and the National Library of Medicine.

lobid. Index of libraries and related organisations

lobid-organisations provides URIs for library organizations, based on the existing and well-established International Standard Identifier for Libraries and Related Organizations (ISIL).

lobid. Bibliographic Resources

lobid-resources is a service which offers access to metadata about bibliographic resources (books, articles, pdfs etc.). Currently there are more than 7 million records.

medline

RDF representation of the Medline catalog. Information about 19 million articles linked to http://dx.doi.org/ with article identifiers and http://crossref.org/ with journal identifiers.

NTNU special collections

The digitized historical manuscripts held in the special collections of the Norwegian University of Science and Technology (NTNU).

The Open Library

One Web page for every book ever published. Currently gathers over 20 million records from a variety of large catalogs as well as single contributions.

English Language Books listed in Printed Book Auction Catalogs from 17th Century Holland

The books are those listed in the English-language section of Dutch printed book auction catalogs of collections of scholars and clergy.

ePrints3 Institutional Archive Collection (RKBExplorer)

Linked Data version of a number of ePrints3 archives.

ECS Southampton EPrints

This is live data produced by the EPrints server, which is distinct from the service provided by RKB Explorer.

Sudoc bibliographic data

Sudoc is the French academic union catalog, maintained by ABES. It contains 10 million bibliographic records.

Open Library data mirror in the Talis Platform

Modeled using the JSON data dumps from the Open Library. Provides a SPARQL endpoint and OpenSearch interface (with RSS 1.0 output).

theses.fr

theses.fr is the french dissertations search engine, maintained by ABES.

Linked Data Service der Universitätsbibliothek Mannheim

Publishes RDF for a number of bibliographic resources: Bibliograhic data of the Südwestdeutscher Bibliotheksverbund, Bibliographic data of the Hessisches Bibliotheksinformationssystem, and others.

University of Sussex Reading Lists

Linked Data version of the resources available through the university's reading-list search engine.

20th Century Press Archives

More than 30 million documents, mostly press clippings about individual persons, companies and other corporate bodies, products and a wide variety of economics-related topics.

Annexe III : Les vocabulaires d'autorité publiés dans le cadre de groupe « bibliothèques et web de données » (Source : https://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/)

Published value vocabularies

This section describes value vocabularies that have been made available as Linked Data or are mentioned as being relevant by one of the Incubator Group's use cases.

Every entry features a brief introduction to the vocabulary, as well as links to their locations. Cases collected by the Incubator Group that refer to the value vocabulary are also listed under each entry.

Classification systems

Dewey Decimal Classification (DDC)

Dewey Summaries is a suitable data set containing the top classes of Dewey Decimal Classification (DDC) 22. It provides access to the top three levels of the DDC in eleven languages along with access to Abridged Edition 14 (assignable numbers and captions) in three languages.

- Usage examples: Browsing And Searching In Repositories With Different Thesauri, Pode, Europeana

Universal Decimal Classification (UDC)

The Universal Decimal Classification (UDC) is a multilingual classification scheme for all fields of knowledge. The UDC Summary represents a selection of around 2,000 classes extracted from the UDC scheme. [1]

- Usage example: Europeana

Subject headings/subject authority files

Library of Congress Subject Headings (LCSH)

LCSH is a comprehensive list of subject headings published in print and as Linked Data. Subject authority headings can be accessed through the Library of Congress Authoritiesand Vocabularies service.

- Usage examples: Component Vocabularies, Polymath Virtual Library, Recollection, Digital Text Repository, potentially to be crosswalked from Civil War Data 150, Vocabulary Merging

Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié (RAMEAU)

RAMEAU is a subject heading vocabulary used by the French National Library (BnF). It has been developed starting from the subject heading repository of the Quebec University, being derived itself from the Library of Congress Subject Headings (LCSH). RAMEAU has been published as Linked Data by the TELplus project.

- Usage examples: Polymath Virtual Library, Vocabulary Merging

Schlagwortnormdatei (SWD)

A controlled vocabulary system managed by the German National Library (DNB) in cooperation

with various library networks. The inclusion of keywords in the SWD is defined by "Rules for the Keyword Catalogue" (RSWK).

•Usage examples: Publishing 20th Century Press Archives, Browsing And Searching In Repositories With Different Thesauri, Vocabulary Merging

National Diet Library List of Subject Headings (NDLSH)

The National Diet Library List of Subject Headings (NDLSH) is a list of subject headings applied to the catalog of the National Diet Library (Japan), including mainly the topical headings and some proper name headings.

•Usage example: Vocabulary Merging

Name authority data

Virtual International Authority File (VIAF)

VIAF is a joint project of multiple national libraries in the world which virtually combine the name authority files of participating institutions into a single name authority service. As of the date of this report, there are 21 authority files of personal, corporate, and conference names from 18 organizations participating in VIAF.

•Usage examples: Polymath Virtual Library, Publishing 20th Century Press Archives, Digital Text Repository, Europeana, Virtual International Authority File (VIAF)

Getty Union List of Artist Names (ULAN)

ULAN is a structured vocabulary containing more than 225,000 names as well as biographical and bibliographic information about artists and architects, including a wealth of variant names, pseudonyms, and language variants.

Although ULAN is not yet published as Linked Data per se, it is included in the VIAF as the Getty Research Institute's contribution.

•Usage example: Europeana

Library of Congress Name Authority File (LC/NAF)

LC/NAF provides authoritative data for names of persons, organizations, events, places, and titles, with over 8 million descriptions created over multiple decades, according to different cataloging policies. LC Names is officially called the Name Authority Component (NACO) Authority File and is a cooperative effort in which participants follow a common set of standards and guidelines.

•Usage example: Recollection

GeoNames

The GeoNames geographical database contains over 10 million geographical names and consists of 7.5 million unique features whereof 2.8 million populated places and 5.5 million alternate names.

•Usage examples: Pode, Publishing 20th Century Press Archives, Europeana

Thesauri

STW Thesaurus for Economics

This thesaurus provides vocabulary on any economic subject. It also covers technical terms used in law, sociology, or politics, and geographic names.

•Usage example: W3C STW Thesaurus Case Study

AGROVOC

AGROVOC is a multilingual structured and controlled vocabulary published by the Food and Agriculture Organization of the United Nations (FAO). It is designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food, and related domains (e.g., environment).

•Usage example: AGROVOC Thesaurus

Eurovoc

Eurovoc is a multilingual, multidisciplinary thesaurus covering the activities of the European Union, the European Parliament in particular. It contains terms in 24 languages (as of the date of this report).

•Usage example: AGROVOC Thesaurus

Thesaurus for Graphic Materials (TGM)

The Library of Congress' Thesaurus for Graphic Materials includes more than 7,000 subject terms to index topic shown or reflected in pictures, and 650 genre/format terms to index types of photographs, prints, design drawings, ephemera, and other categories.

•Usage example: Recollection

Other controlled vocabularies

DCMI Type Vocabulary

A general, cross-domain list of Dublin Core Metadata Initiative (DCMI) approved terms that may be used as values for the resource type element to identify the genre of a resource.

•Usage example: Enhanced Publications

MARC Code List of Relators (also in element sets)

The MARC (MAchine-Readable Cataloging) Relators provide a list of properties for describing the relationship between a name and a bibliographic resource.

•Usage example: Component Vocabularies

PRONOM

PRONOM is the online registry of technical information about the file formats, software products, and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical, or business value.

•Usage example: Digital Preservation

Creative Commons (CC) License set

Creative Commons provides an infrastructure that consists of a set of copyright licenses and tools to create a balance inside the traditional “all rights reserved” setting that copyright law creates.

•Usage example: Civil War Data 150

Preservation vocabularies from LoC

Two main vocabularies are provided. Preservation Events is a concept scheme for the preservation events, i.e., actions performed on digital objects within a preservation repository. Preservation Level Role is a concept scheme for the preservation level roles, i.e., values that specify in what context a

set of preservation options is applicable.

•Usage example: Digital Preservation

Additional sources

Wordnet

WordNet is a lexical database of English in which nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (called "synsets"). Each synset expresses a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. Wordnet has been published as Linked Data by the Vrije Universiteit Amsterdam.

•Usage example: Language Technology

Freebase (also in datasets)

Freebase is an open, Creative Commons licensed collection of structured data, and a platform for accessing and manipulating that data via the Freebase API. Freebase imports data from a wide variety of open data sources, such as Wikipedia, MusicBrainz, and others. Note that Freebase is essentially a dataset, but its inclusion of many reference resources allows some parts of it to be used as value vocabularies in certain cases.

•Usage example: Freebase is mentioned as reference for alignment in Civil War Data 150

DBpedia

DBpedia extracts structured information from Wikipedia. The DBpedia data set features labels and abstracts for over three million things, with a half of them classified in an ontology, and contains millions of links to images, external Web pages, and external links to other RDF datasets . Similarly to Freebase, DBpedia can be seen as a general dataset, but some of the entities it describes — places, persons, "categories" — can be used as reference value vocabularies in some cases.

•Usage examples: Polymath Virtual Library, Language Technology, Digital Text Repository, Publishing 20th Century Press Archives, Europeana, potentially to be crosswalked from Civil War Data 150

Work in progress, or relevant for cases but not officially in progress

Aquatic Sciences and Fisheries Abstracts (ASFA) thesaurus

This thesaurus is used for the subject indexing of the Aquatic Sciences and Fisheries Abstracts (ASFA), an abstracting and indexing service that covers the world's literature on the science, technology, management, and conservation of marine, brackish water, and freshwater resources and environments, including their socio-economic and legal aspects.

•Usage example: AGROVOC Thesaurus

Fisheries Reference Metadata

The Fisheries Reference Metadata system stores all the classification systems (for species, countries, water areas, commodities, fishing vessels, fishing gears, etc.) used by FAO to describe fisheries observations such as time-series data on fisheries capture and production and species fact sheets.

•Usage example: AGROVOC Thesaurus

National Agricultural Library's Agriculture Thesaurus and Glossary

The Agricultural Thesaurus and Glossary are online vocabulary tools of agricultural terms in English and Spanish provided by the USDA National Agricultural Library (NAL). The subject scope of agriculture is broadly defined in the NAL Agricultural Thesaurus, and includes terminology in the supporting biological, physical, and social sciences. The definitions of terms in the thesaurus were separately published as the Glossary of Agricultural Terms.

•Usage example: AGROVOC Thesaurus

Getty Art and Architecture Thesaurus (AAT)

A multilingual controlled vocabulary for fine art, architecture, decorative arts, archival materials, and material culture for the purposes of indexing, cataloging, searching, as being a research tool.

•Usage example: Europeana

Medical Subject Headings (MeSH)

A comprehensive controlled vocabulary produced by the National Library of Medicine (NLM) for biomedical and health-related information and documents. Spanish and French MeSH are available as Semantic Web ontologies at BioPortal. The Norwegian translation of MeSH is published as Linked Data by the Norwegian University of Science and Technology. Another version of MeSH, in Simple Knowledge Organization System (SKOS), is available through OCLC Terminology Services.

•Usage example: Vocabulary Alignment Cluster

Iconclass

A classification system for describing and classifying the subject of images represented in various media such as paintings, drawings, and photographs.

•Usage example: Europeana

Getty Thesaurus of Geographic Names (TGN)

A structured, world-coverage vocabulary of over 1.3 million names, including vernacular and historical names, coordinates, place types, and descriptive notes, focusing on places important for the study of art and architecture.

•Usage example: Europeana

Other value vocabularies relevant to the Library Linked Data field, not mentioned in the cases

New York Times subject headings

The New York Times uses approximately 30,000 tags to power its Times Topics Pages. These tags (categorized into 'people', 'organization', 'place', and 'descriptor') as published as Linked Data and are mapped to Freebase, DBpedia, and GeoNames.

MARC List for Countries

The MARC List for Countries identifies current national entities, states of the United States, provinces and territories of Canada and Australia, divisions of the United Kingdom, and internationally recognized dependencies. The entries include references to their equivalent ISO 3166 codes.

MARC List for Languages

The MARC List for Languages provides three-character lowercase alphabetic strings that serve as

the identifiers of languages and language groups. It have been cross referenced with ISOs 639-1, 639-2, and 639-5, where appropriate.

MARC List for Geographic Areas

The MARC List for Geographic Areas identifies separate countries, first order political divisions of some countries, regions, geographic features, areas in outer space, and celestial bodies. The list contains over 550 different codes.

Annexe IV - Éléments de descriptions de métadonnées élaborés dans le cadre de groupe « bibliothèques et web de données » Library Linked Data

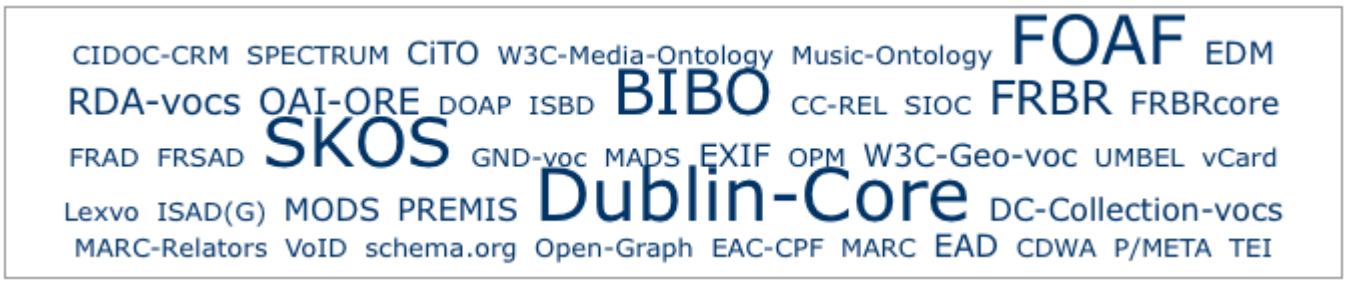
Metadata Element Sets

This section lists metadata element sets mentioned in the cases gathered by the Library Linked Data group in 2010-2011. These include some of the most relevant RDFvocabularies for practitioners who want to re-use available Semantic Web technology for creating or converting data from the library domain.

These RDF vocabularies are represented using the constructs offered by the RDF Schema (RDFS) and OWL Web Ontology Language modeling languages. In addition to the documentation provided by its maintainers, an ontology can also be viewed using generic ontology creation and visualization tools such as Protégé, the Manchester ontology browser, OWL Sight or the Live OWL Documentation Environment (LODE) (see for example the Description of a Project (DOAP) ontology rendered in LODE).

For each element set, we give a pointer to a human-readable website and indicate the corresponding RDF namespace, as well as a common prefix abbreviation used for it, using the XML namespace declaration syntax. We also provide or re-use a short description, focused on the main scope or usage domain for the element set. We have sometimes emphasized important design decisions that characterize the element set, including indications on whether the element set is connected to another one, and its relation to traditional library usages. Finally, cases collected by the Incubator Group are also listed under each entry as relevant usage examples.

For illustration purposes, we include a tag cloud rendition of the element sets presented in this section, adapting a site created by Paul Walk:



Note that this tag cloud is a context-specific snapshot of the usage of metadata element sets. In particular, the size of each tag is directly related to the number of individual cases that use it, as gathered by the Library Linked Data Incubator Group. Beyond this analysis based on the Incubator Group cases, Library Linked Data community members should consider helping maintain precise and up-to-date listings of datasets and value vocabularies, such as the Data Hub Library Linked Data group, so that the usage of element sets can be measured. A refined, domain-specific version of the usage statistics for the Linked Open Data Cloud would help the community to develop a clearer idea about which elements sets are widely used and which are less common.

Rendering links between metadata element sets would also be valuable for practitioners willing to re-use data across vocabularies, or to make their data better usable by a wide community. The Upper Mapping and Binding Exchange Layer (UMBEL) constellation has been the first to illustrate connections between classes from popular Linked Data vocabularies. The Linked Open Vocabulary effort generalizes and automatizes the gathering of such information. For a wide range of metadata element sets, e.g., Dublin Core, Linked Open Vocabulary offers a detailed view of the

relationships with other element sets, based on the available machine-readable definitions (ontologies).

Metadata element sets published as RDF vocabularies

This sub-section lists the relevant ontologies (OWL or RDFS) available at the time of writing this report. To help readers orient themselves in our selection, we first introduce metadata element sets that originate from the Libraries, Archives, Museums, and Information communities. We then present relevant element sets, which are rooted in other communities. This categorization is often arbitrary, as many vocabularies already result from cross-community work. We believe, however, that this shows the great potential for the Linked Data approach, where easily sharing, re-using or extending a diversity of element sets independently from their origin is the rule.

Originating from the Libraries, Archives, Museums, and Information communities

Dublin Core and DCMI Metadata Terms

The original properties of the Dublin Core Metadata Element Set — fifteen generic property elements for describing information resources — are identified using the namespace http://purl.org/dc/elements/1.1/. Declared as RDF properties in 2000 — before the finalization of RDFS in 2004 — these properties lack defined ranges (rdfs:range), allowing them to be used both with literal values or fully-fledged RDF resources.

A second namespace in the larger set of DCMI Metadata Terms — http://purl.org/dc/terms/ — includes fifteen properties in parallel to the "unrestricted" /elements/1.1/ properties, with added rdfs:range restrictions, plus several dozen additional properties. Interoperability of the "restricted" /terms/ properties with the "unrestricted" /element/1.1/ properties is preserved with sub-property relationships (rdfs:subPropertyOf).

- xmlns:dc="http://purl.org/dc/elements/1.1/"
- xmlns:dcterms="http://purl.org/dc/terms/"
- Usage examples: Enhanced Publications, Publishing 20th Century Press Archives, Data BNF, VIAF, Browsing And Searching In Repositories With Different Thesauri, Pode, Archipel, SEO, LOCAH, Digital Text Repository, Europeana, Migrating Library Legacy Data, Library Address Data, NLL Digitized Map Archive, Collecting material related to courses at The Open University

Open Archives Initiative - Object Reuse and Exchange (OAI-ORE)

The Open Archives Initiative Object Reuse and Exchange model define elements to describe aggregations of Web resources, which together form complex digital objects, such as a journal article and its different digital variations and accompanying material. It also proposes a "resource map" mechanism for indicating and describing provenance of metadata on these aggregations, as well as "proxies" to describe any given resource from the perspective of a specific aggregation, when resources are included in different aggregations.

- xmlns:ore="http://www.openarchives.org/ore/terms/"
- Usage examples: Enhanced Publications, 20th Century Press Archives, Digital Preservation, Europeana, National Digital Newspaper Program (NDNP)

Simple Knowledge Organization System (SKOS)

SKOS provides a model for expressing the basic structure and content of concept schemes such as

thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary [17]. SKOS deliberately avoids providing rdfs:domains with some of its properties (especially labelling and note properties), enabling one to re-use them for any kind of resource.

•xmlns:skos="http://www.w3.org/2004/02/skos/core#"

•Usage examples: Publishing 20th Century Press Archives, Data BNF, LOCAH, Browsing And Searching In Repositories With Different Thesauri, Component Vocabularies, Pode, Subject Search, Europeana, VIAF, AGROVOC Thesaurus, AGRIS, Vocabulary Merging (SKOS mapping), Migrating Library Legacy Data, NLL Digitized Map Archive, Collecting material related to courses at The Open University

SKOS eXtension for Labels (SKOS-XL)

SKOS-XL is a SKOS extension that provides support for describing lexical entities attached to concepts. It "reifies" the labels of skos:Concepts, treating them as fully-fledged RDFresources. This allows them to be annotated further, or support linking them using, say, a "isTranslationOf" property.

•xmlns:xl="http://www.w3.org/2008/05/skos-xl#"

•Usage examples: VIAF, AGROVOC Thesaurus

MARC Code List of Relators (also in value vocabularies)

The MARC Relators vocabulary provides a list of properties for describing the relationship between a name and a bibliographic resource.

•xmlns="http://id.loc.gov/vocabulary/relators"

•Usage example: Component Vocabularies

CIDOC Conceptual Reference Model (CRM)

The CIDOC object-oriented Conceptual Reference Model (CRM) is developed by the International Council of Museums (ICOM) to represent and make interoperable description of objects from the cultural sector. It makes intensive use of events to link objects, persons, places, and more conceptual notions together.

•xmlns:crm="http://www.cidoc-crm.org/rdfs/cidoc-crm" for the latest official release.

•Version 5.0.2 RDFS files: http://www.cidoc-crm.org/official_release_cidoc.html (each file is a partial definition of the CIDOC Conceptual Reference Model version 5.0.2 in RDF).

•Usage example: Europeana

Alternative OWL-Description Logic (OWL-DL) versions (OWL 1 and 2) are available at http://erlangen-crm.org (namespace: http://erlangen-crm.org/current/) and http://bloody-byte.net/rdf/cidoc-crm/ (namespace: http://purl.org/NET/cidoc-crm/core#).

Dublin Core Collection Description vocabularies

A Task Group of the DCMI Collection Description Community developed a Dublin Core collections application profile and several vocabularies. Its work was based on the Research Support Libraries Programme (RSLP) Collection Description Schema.

•Dublin Core Collection Description Terms

•xmlns:cld="http://purl.org/cld/terms/"

•Describes the properties, classes, vocabulary encoding schemes and syntax encoding schemes created for use by the application profile for collection-level description.

•Dublin Core Collection Description Type Vocabulary

•xmlns:cdtype="http://purl.org/cld/cdtype/"

•Provides a set of classes that can be used as values of the dc:type property in descriptions of collection descriptions.

•Dublin Core Collection Description Accrual Method Vocabulary

•xmlns:accmeth="http://purl.org/cld/accmeth/"

•Provides a set of terms that can be used as values of the dcterms:accrualPolicy property in descriptions of collections.

•Dublin Core Collection Description Frequency Vocabulary

•xmlns:freq="http://purl.org/cld/freq/"

•Provides a set of terms that can be used as values of the dcterms:accrualPeriodicity property in descriptions of collections.

•Dublin Core Collection Description Accrual Policy Vocabulary

•xmlns:accpol="http://purl.org/cld/accpol/"

•Provides a set of terms that can be used as values of the dcterms:accrualPolicy property in descriptions of collections.

•Usage examples: Collection-Level Description, Nearest physical collection

Functional Requirements for Bibliographic Records (FRBR) and related ontologies

FRBR is a conceptual reference model developed by the International Federation of Library Associations and Institutions (IFLA) "to provide a (...) framework for relating the data that are recorded in bibliographic records to the needs of users of those records" (FRBR Final Report, sec. 2.1) and for assessing their actual relevance. See more details here.

•Usage examples: Bibliographic Network, Europeana, FAO Authority Description Concept Scheme, LOCAH, NDNP (National Digital Newspaper Program), Open Library Data, VIAF, Pode

The IFLA "FRBR family" consists of three conceptual models each covering an aspect of the data recorded in bibliographic and authority records. The entities, attributes, and relationships defined by each of the models are included in the Open Metadata Registry:

•IFLA's FRBRer ("entity-relationship")

•xmlns:frbrer="http://iflastandards.info/ns/fr/frbr/frbrer/"

•Functional Requirements for Authority Data (FRAD)

•xmlns:frad="http://iflastandards.info/ns/fr/frad/"

•Functional Requirements for Subject Authority Data (FRSAD)

•xmlns:frsad="http://iflastandards.info/ns/fr/frsad/"

The FRBR Final Report describes an entity-relationship model that has been the source of a number of other ontology implementations:

•Ian Davis' FRBRcore - Expression of Core FRBR Concepts in RDF.

•xmlns:frbr="http://purl.org/vocab/frbr/core"

•RDF: http://vocab.org/frbr/core.rdf

•Usage examples: Pode, Bibliographic Network

•RDA: Resource Description and Access (RDA) - RDA is a cataloging standard aimed at replacing the AACR2 guides (for more details see here). It is influenced by the FRBR model. Work is ongoing to make available RDF vocabularies for RDA via the Open Metadata Registry, e.g., for the RDA Group 1 elements. Note that RDA comes with value vocabularies to be used in combination with the element set below, such as the Forms of Musical Notation. These are also being published as Linked Data.

•xmlns:rdaEnt="http://rdvocab.info/uri/schema/FRBRentitiesRDA/"

•xmlns:rdaGr1="http://rdvocab.info/Elements/"

•xmlns:rdaGr2="http://rdvocab.info/ElementsGr2/"

•xmlns:rdaGr3="http://rdvocab.info/ElementsGr3/"

•Usage examples: Data BNF, Publishing 20th Century Press Archives, VIAF, FAO Authority Description Concept Scheme, Regional Catalog

•FaBiO- the FRBR-aligned Bibliographic Ontology

•xmlns:fabio="http://purl.org/spar/fabio/"

•OWL file: http://sempublishing.svn.sourceforge.net/svnroot/sempublishing/FaBiO/fabio.owl

International Standard Bibliographic Description (ISBD)

This is a preliminary registration of classes and properties from the International Standard Bibliographic Description (ISBD) consolidated edition. The ISBD (see more explanations here) is useful and applicable for descriptions of bibliographic resources in any type of catalog.

•Usage example: Migrating Library Legacy Data

Metadata Authority Description Schema in RDF (MADS/RDF)

MADS/RDF is designed for use with controlled values for names (personal, corporate, geographic, etc.), thesauri, taxonomies, subject heading systems, and other controlled value lists. The MADS/RDF ontology is mapped to SKOS.

•xmlns:mads="http://www.loc.gov/mads/rdf/v1#"

•Usage example: Component Vocabularies

Gemeinsame NormDatei (GND) vocabulary

For its Linked Data services, the German National Library has created a namespace that is devoted to detailed description of authority resources (Gemeinsame NormDatei, GND). This set of classes and properties especially refines the possibilities offered by SKOS and the RDA vocabularies.

•xmlns:gnd="http://d-nb.info/gnd/"

•Usage example: Linked Data Service of the German National Library

Originated from other communities

Friend of a Friend (FOAF)

FOAF is a widely used ontology for describing persons and their relationships to other persons and to Web resources.

•xmlns:foaf="http://xmlns.com/foaf/0.1/"

•Usage examples: Enhanced Publications, Data BNF, Pode, SEO, Digital Text Repository, FAO Authority Description Concept Scheme, VIAF, LOCAH, Library Address Data, Open Library Data, Authority Data Enrichment, International Registry for Authors, AGRIS, Collecting material related to courses at The Open University, Talis Prism 3

Vocabulary of Interlinked Datasets (VoID)

VoID is an RDF-based schema for describing linked datasets. With VoID the discovery and usage of linked datasets can be performed both effectively and efficiently. A VoID dataset is a collection of data, published and maintained by a single provider, available as RDF, and accessible, for example, through dereferenceable HTTP URIs or a SPARQLendpoint.

•xmlns:void="http://rdfs.org/ns/void#"

•Usage example: Publishing 20th Century Press Archives

Bibliographic Ontology (BIBO)

BIBO can be used as a citation ontology or document classification ontology, or a way to describe any kind of bibliographic things in RDF.

•xmlns:bibo="http://purl.org/ontology/bibo/"

•Usage examples: Bibliographic Network, LOCAH, Community Information Service, Regional Catalog, Pode, Migrating Library Legacy Data, Talis Prism 3, FAO Authority Description Concept Scheme, AGRISInternational Registry for Authors, NDNP (National Digital Newspaper Program), Collecting material related to courses at The Open University, Collection-Level Description

Upper Mapping and Binding Exchange Layer (UMBEL) Vocabulary

The Upper Mapping and Binding Exchange Layer (UMBEL) Reference Concepts dataset is derived from the OpenCyc ontology. It includes thousands of coherently structured and linked concepts, and is broadly applicable to provide orienting nodes to any knowledge domain. The UMBEL vocabulary provides classes and properties to describe this conceptual knowledge. It also intended to function as the basis for constructing domain ontologies. It re-uses external vocabularies whenever possible.

•xmlns:umbel="http://umbel.org/umbel#"

•Usage example: VIAF

vCard

The vCard ontology enables representing business card profiles defined by vCard (RFC2426).

•xmlns:vcard="http://www.w3.org/2006/vcard/ns#"

•Usage example: Library Address Data

Lexvo.org ontology

The name Lexvo is derived from the Ancient Greek λεξικόν (lexicon) and the Latin vocabularium (vocabulary). The ontology provides a vocabulary for defining global URIs for languages, words, characters, and other human language-related objects.

•xmlns:lvont="http://lexvo.org/ontology#"

•Usage example: Pode

Exchangeable Image File Format (EXIF)

This is a RDF Schema for EXIF — a standard for images that supports mainly technical metadata, usually embedded in an image file (e.g., JPEG file), where each key of the EXIF specification has been directly mapped to a corresponding property. In order to preserve the groupings of metadata keys that is provided in the original EXIF specification (e.g., pixel composition and geo location), other efforts have been reported, such an EXIF OWL ontology .

•xmlns:exif="http://www.w3.org/2003/12/exif/ns"

•Usage examples: Publishing 20th Century Press Archives, NDNP

Open Provenance Model (OPM)

The Open Provenance Model is a generic model to express and share provenance information. It consists of a lightweight Open Provenance Model Vocabulary which enables basic representation of provenance data, and a more expressive Open Provenance Model OWL Specification geared towards inference.

•xmlns:opmv="http://purl.org/net/opmv/ns#"

•xmlns:opmo="http://openprovenance.org/model/opmo#"

•Usage example: Citation of Scientific Datasets

Music Ontology

"The Music Ontology Specification provides main concepts and properties for describing music (i.e., artists, albums, and tracks) on the Semantic Web". It applies the FRBRdistinctions to the music domain.

•xmlns:mo="http://purl.org/ontology/mo/"

•Usage example: Talis Prism 3

Creative Commons Rights Expression Language (CC REL)

CC REL enables describing copyright licenses in RDF.

•xmlns:cc="http://creativecommons.org/ns#"

•Usage example: SEO

A Citation Type Ontology (CiTO)

CiTO, one of the SPAR ontologies, is a minimal ontology for describing reference citations in research articles.

•xmlns:cito="http://purl.org/net/cito/"

•OWL file: http://imageweb.zoo.ox.ac.uk/pub/2009/citobase/cito-20090311/cito-content/owldoc/

•Usage examples: Community Information Service, Citation of Scientific Datasets

Description Of A Project (DOAP)

Description of a Project (DOAP) is a vocabulary for describing software projects, especially open-source projects.

•xmlns:doap="http://usefulinc.com/ns/doap#"

•OWL file: "http://usefulinc.com/ns/doap"

•Usage example: Digital Preservation

W3C Geo vocabulary

This small ontology is aimed at representing geo-positioning (latitude, longitude, and altitude) for spatial objects, according to the WGS84 standard.

•xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"

•Usage examples: Library Address Data, NLL Digitized Map Archive

Semantically-Interlinked Online Communities (SIOC)

The SIOC core ontology can be used to describe online communities and their activities (e.g., message boards, wikis, weblogs, etc.).

•xmlns:sioc="http://rdfs.org/sioc/ns#"

•Usage example: Social Uses Cluster

Schema.org vocabulary

Schema.org is a set of constructs that allow website designers to include structured metadata in their Web pages, to be consumed by the major search engines Bing, Google, and Yahoo! Schema.org is designed to represent resources from a great diversity of domains. It thus duplicates many elements from other element sets, and fails to capture the richness of library data. However, it can be used to exchange simple information about libraries and the resources they own, as demonstrated in a blog post by Eric Hellman.

•xmlns:schema="http://schema.org/"

•Usage example: SEO

Open Graph

Facebook's Open Graph "protocol" enables the description of resources (movies, books, etc.) that can be of interest to social network members. Its main purpose is to allow websites to include RDFa markup, which is used in combination with the "Like" button to communicate to the Facebook service data about the objects mentioned on web pages.

•xmlns:og="http://ogp.me/ns#"

•Usage example: SEO

W3C Ontology for Media Resources

The Ontology for Media Resources defines a core set of metadata properties for media resources, along with their mappings to elements from a set of existing metadata formats. It is mainly targeted towards media resources available on the Web, as opposed to resources that are only accessible in local archives or museums.

•xmlns:ma="http://www.w3.org/ns/ma-ont#"

•Usage example: Collecting material related to courses at The Open University

Work in progress to make RDF vocabularies available

General International Standard Archival Description (ISAD(G))

The General International Standard Archival Description defines the elements that should be included in an archival finding aid.

- Usage example: LOCAH

Europeana Data Model (EDM)

The Europeana Data Model is a vocabulary aimed at representing metadata for cultural objects and giving access to digital representations of these objects. It is positioned in a data aggregation context, where objects can be complex, and several data providers may entertain different views on them. EDM re-uses, extends or has been inspired by other element sets, notably OAI-ORE, Dublin Core, SKOS, and CIDOC CRM.

- xmlns:edm="http://www.europeana.eu/schemas/edm/"
- Usage examples: Europeana, Polymath Virtual Library

Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF)

EAC-CPF is aimed at representing authoritative information about the context of archival materials, including "the identification and characteristics of the persons, organizations, and families (agents) who have been the creators, users, or subjects of records, as well as the relationships amongst them". It is a parallel effort to the Encoded Archival Description (EAD) standard for representation of archival finding aids.

A core concept in EAC-CPF is the distinction between agents and identities: a same agent can have different identities, and one identity can correspond to several agents.

- Resources from Istituto per i beni artistici culturali e naturali of the Regione Emilia-Romagna (IBC) and Regesta.com: documentation, OWL file
- Usage example: Archive and heterogeneous data Cluster. See a note by Ed Summers.

MARC21

MARC (MAchine-Readable Cataloging) has played a crucial role in the creation and exchange of library metadata. An RDF version of the complete MARC21 elements has been published via the Open Metadata Registry as a "basic loss-less way to transition MARC 21 data into RDF." Prior to this, the MarcOnt initiative had created an OWL ontology that includes a small sub-set of MARC elements, related to other ontologies.

- xmlns="http://marc21rdf.info/elements/"
- Usage example: Archipel

Preservation Metadata: Implementation Strategies (PREMIS)

PREMIS defines a core set of preservation metadata elements, with a supporting data dictionary, applicable to a broad range of digital preservation activities.

- Resources from the Archipel project: documentation, OWL file
- Usage examples: Digital Preservation, Archipel

Encoded Archival Description (EAD) and other archive-oriented element sets

EAD is a standard for encoding archival finding aids using Extensible Markup Language (XML).

•Usage example: Archives and heterogeneous data Cluster

•Work relevant for EAD in RDF has been done in LOCAH (see website and documentation) and EuropeanaConnect (see schema)

Note that the LOCAH element set only handles a part of EAD, and introduces other elements that the LOCAH participants found useful to publish archival collection data as Linked Data. Readers may also be interested in the lightweight Archival vocabulary maintained by Aaron Rubinstein for describing archives and the named entities associated with them.

Metadata element sets from cases for which no RDF vocabulary is available

Categories for the Description of Works of Art (CDWA)

Categories for the Description of Works of Art (CDWA) includes 532 categories and subcategories for describing describing and accessing information about works of art, architecture, other material culture, groups and collections of works, and related images. A simpler subset of these elements, CDWA Lite, has been created.

•Usage example: Archipel

EBU P/Meta Semantic Metadata Schema (P/META)

A standard vocabulary for information relating to program information in the professional broadcasting industry.

•Usage example: Archipel

SPECTRUM

SPECTRUM is a UK-originated standard for managing museum collections, from descriptive metadata for objects to loan information.

•Usage example: Archipel

Metadata Object Description Schema (MODS)

MODS includes a subset of MARC fields and uses language-based tags rather than numeric ones, in some cases regrouping elements from the MARC 21 bibliographic format. MODS is expressed using XML.

•Usage examples: Archives and heterogeneous data Cluster, FAO Authority Description Concept Scheme

Text Encoding Initiative (TEI) Guidelines

The "Guidelines for Electronic Text Encoding and Interchange" is a standard for representing all kinds of literary and linguistic texts for online research and teaching.

•Usage example: Digital objects Cluster

Other metadata element sets (no RDF vocabulary) relevant to the library field, not mentioned in the cases

Visual Resources Association Core Categories (VRA Core)

Visual Resources Association Core Categories (VRA Core) specifies a set of core categories for creating records to describe works of visual culture as well as the images that document them.

•An OWL ontology for VRA Core 3.0 has been created by Mark van Assem for the W3C

Semantic Web Best Practices and Deployment Working Group.

Public Broadcasting Metadata Dictionary (PBCore)

PBCore is a metadata standard designed to describe media, both digital and analog. The PBCore XML Schema Definition (XSD) defines the structure and content of PBCore. The element set and related value vocabularies are available at the Open Metadata Registry.

Annexe V : Recensement de la bibliothèque nationale de France

Les bibliothèques recensées par la BNF		Type de bibliothèque	Bibliothèques répondant à nos critères
1	Université de Toulouse-Réseau des bibliothèques (Tolosana)	Bibliothèque universitaire	Oui
2	Bibliothèque valenciennes	Bibliothèque publique	Oui
3	ABU	Bibliothèque spécialisée	Oui
4	Bibliothèque numérique de l'Institut national d'histoire de l'art INHA	Bibliothèque spécialisée	Oui
5	Conseil des musées de Poitou-Charentes (Alienor)	Bibliothèque spécialisée	Oui
6	ELEC : Éditions en ligne de l'École des chartes	Bibliothèque spécialisée	Oui
7	ENSSIB : école nationale supérieure des sciences de l'information et des bibliothèque	Bibliothèque spécialisée	Oui
8	Fonds virtuels	Bibliothèque spécialisée	Non (site inaccessible)
9	Gallica	Bibliothèque nationale	Oui
10	La Bibliothèque André Desguine	Bibliothèque publique	Oui
11	La Bibliothèque électronique de Lisieux	Bibliothèque publique	Oui
12	La bibliothèque numérique DIGIMOM	Bibliothèque spécialisée	Oui
13	Les Bibliothèques virtuelles humanistes	Bibliothèque spécialisée	Oui
14	Numelyo	Bibliothèque publique	Oui
15	Numes	Bibliothèque universitaire	Non (site inaccessible)
16	Numistral	Bibliothèque nationale	Oui
17	Patrimoine numérisé	Bibliothèque publique	Non (site inaccessible)
18	Pôlib, Bibliothèque patrimoniale virtuelle	Bibliothèque universitaire	Oui
19	SICD Universités de Strasbourg - Patrimoine numérisé	Bibliothèque universitaire	Non (site inaccessible)
20	Textes rares	Bibliothèque spécialisée	Oui

ANNEXE VI : Recensement fait par l'URFIST de Lyon

Bibliothèques recensées	Type de bibliothèque	Bibliothèque répondant à nos critères
1. Bibliothèque de Cerf	Non accessible	
2. Marcel Proust (Cujasnum)	Bibliothèque universitaire	Oui
3. Bibliothèque électronique de Lisieux	Bibliothèque publique	Oui (double BNF)
4. Bibliothèque municipale de Grenoble	Bibliothèque publique	Oui
5. Bibliothèque municipale de Lyon	Bibliothèque publique	Oui (double BNF)
6. Bibliothèque municipale de Lille	Bibliothèque publique	Oui
7. bn-r	Bibliothèque publique	Oui
8. Cartolis	Bibliothèque publique	Oui
9. Chroniques de Port-Royal, Les	Non (ne répond pas à nos critères de sélection)	
10. Digimom	Bibliothèque spécialisée	Oui (double BNF)
11. Gallica	Bibliothèque nationale	Oui (double BNF)
12. Médiathèque de l'agglomération troyenne	Non accessible	
13. Pole International de la Préhistoire	Non (ne répond pas à nos critères de sélection)	
14. Revues online	Non (ne répond pas à nos critères de sélection)	
15. Cefael	Bibliothèque spécialisée	Oui
16. Cnum – Conservatoire numérique des Arts & Métiers	Bibliothèque spécialisée	Oui
17. Ecole centrale de Lyon	Bibliothèque universitaire	Oui
18. Ecole française de Rome	Non (ne répond pas à nos critères de sélection)	
19. ELEC – Editions en ligne de l'Ecole des chartres	Bibliothèque spécialisée	Oui
20. Oscar Wilde Collection, The	Non (ne répond pas à nos critères de sélection)	
21. IRIS – Histoire des sciences	Bibliothèque universitaire	Oui
22. Jubliothèque, Bibliothèque numérique scientifique de l'UPMC	Bibliothèque universitaire	Oui
23. PoliB	Bibliothèque universitaire	Oui (double BNF)
24. Premiers socialismes, Les	Bibliothèque universitaire	Oui
25. Presses Universitaires de Lyon	Non (ne répond pas à nos critères de sélection)	
26. SICD de Toulouse	Bibliothèque universitaire	Oui (double BNF)
27. Académie royale des sciences de France	Non (ne répond pas à nos critères de sélection)	
28. Archives de l'assemblée Nationale française	Non (ne répond pas à nos critères de sélection)	
29. Bibliothèque numérique mondiale (BNM)	Bibliothèque internationale	Non
30. COLISciences	Non (ne répond pas à nos critères de sélection)	
31. Europeana	Bibliothèque internationale	Non

32. Mémoires de hommes	Non (ne répond pas à nos critères de sélection)	
33. Napoleonica	Bibliothèque spécialisée	Oui
34. Nouvelle France	Non (ne répond pas à nos critères de sélection)	
35. Pleins Textes	Non (ne répond pas aux critères de sélection)	
36. @ampère	Bibliothèque spécialisée	Oui
37. Archives Bourbaki	Non accessible	
38. Buffon et l'histoire naturelle	Bibliothèque spécialisée	Oui
39. Charcot – Fonds numérisé	Non accessible	
40. Claude Bernard (Bibliothèque virtuelle)	Bibliothèque spécialisée	Oui
41. D'Alembert	Bibliothèque spécialisée	Oui
42. Darwin correspondence Project	Non (zone géographique hors France)	
43. Einstein Archives Online	Non (zone géographique hors France)	
44. Euler Leonhard	Non accessible	
45. Lamarck Jean-Baptiste	Bibliothèque spécialisée	Oui
46. Lavoisier	Bibliothèque spécialisée	Oui
47. Newton Project	Non (zone géographique hors France)	
48. Niels Bohr Library & Archives	Non accessible	
49. Problem of the planets	Non (zone géographique hors France)	
50. The complete work of charles Darwin online	Non (zone géographique hors France)	
51. Balzac : La comédie humaine	Non (ne répond pas à nos critères de sélection)	
52. Gustave Flaubert – Madame Bovary	Non accessible	
53. Manuscrits de Flaubert	Bibliothèque universitaire	Oui
54. Manuscrits de vie de Henry Brulard	Non accessible	
55. Marcel Proust	Non (ne répond pas à nos critères de sélection)	
56. Mark Twain project online	Non (zone géographique hors France)	
57. Oscar wild collection	Non (zone géographique hors France)	
58. Proscenium	Bibliothèque spécialisée	Oui
59. Mémoires de Saint-Simon	Non (ne répond pas à nos critères de sélection)	
60. Vincent Van Gogh – The letters	Non (zone géographique hors France)	
61. XMLIttré	Non accessible	
62. Bibnum	Bibliothèque spécialisée	Oui
63. Biodiversity Heritage Library	Bibliothèque internationale	Non
64. Medic@ - BIUL de Paris	Bibliothèque universitaire	Oui
65. Scientifica	Non accessible	
66. Bibliothèques virtuelles humanistes	Bibliothèque spécialisée	Oui (double BNF)
67. Criminocorpus	Bibliothèque spécialisée	Oui
68. E-codices	Non (zone géographique hors France)	
69. E-corpus	Non (ne répond pas à nos critères de sélection)	
70. ETANA	Non (zone géographique hors France)	

71. Normannia	Bibliothèque spécialisée	Oui
72. ABU- Association des bibliothèques universels	Bibliothèque spécialisée	Oui (double BNF)
73. CTLF – Corpus des textes linguistiques fondamentaux	Non (ne répond pas à nos critères des sélection)	
74. Project Gutenberg	Bibliothèque internationale	Non
75. Persée	Non (ne répond pas à nos critères de sélection)	
76. Classiques des sciences sociales, Les	Non (zone géographique hors France)	
77. PSYCHOLOGIA	Non (ne répond pas à nos critères de sélection)	

Annexe VII : Recensement fait par l'inventaire Multilingue du Patrimoine Culturel en Europe MICHAEL

Bibliothèques recensées par MICHAEL		
1. Service commun de la documentation - Université de Poitiers (Coutumiers du centre-ouest/ Biblioteca virtual cordel / Biblioteca virtual cibeles / Bibliothèque virtuel des premiers socialismes) **[4 bibliothèques]**	BU	Oui
2. École nationale supérieure d'art de Nancy	BU	Non
3. Service Commun de la Documentation de l'Université de Bourgogne (Queneau)	BU	Oui
4. Bibliothèque de documentation internationale contemporaine (Argonnaute)	BU	Oui
5. Photothèque – Bibliothèque de Fels	BU	Non
6. INHA	BS	Oui
7. Service Commun de la Documentation de Lille 3 (Polib et Nordnum) **[2 bibliothèques]**	BU	Oui
8. Service Commun de la Documentation - Université Droit et Santé Lille 2	BU	Non
9. Service commun de la documentation de l'Université Paris 8 Vincennes-Saint-Denis	BU	Oui
10. Service Commun de la Documentation de l'Université d'Aix-Marseille, BU droit-sciences économiques	BU	Non
11. Bibliothèque interuniversitaire de médecine et d'odontologie - Paris	BU	Non
12. Bibliothèque interuniversitaire de Montpellier - Section médecine	BU	Oui
13. Service Commun de la Documentation - Université des Sciences et Technologies de Lille (Iris)	BU	Oui
14. Bibliothèque Diderot de Lyon	BU	Oui
15. Bibliothèque de l'Ecole Polytechnique	BU	Non
16. Centre culturel Rémy Nainsouta	BS	Non
17. Bureau des bibliothèques et de la lecture	BP	Non
18. Médiathèque Nadia Boulanger	BP	Non
19. Bibliothèque du DEFAP - Service protestant de mission	BS	Non
20. Conservatoire national des arts et métiers – Bibliothèque	BU	Oui

21. Bibliothèque de l'Ecole Centrale de Lyon	BU	Oui
22. La Châtre Sainte-Sévère	BP	Non
23. Centre inter-régional de développement de l'occitan (CIRDOC)	BP	Oui
24. Médiathèque Michel Crépeau - La Rochelle	BP	Non
25. Centre régional des Lettres de Basse-Normandie	BS	Oui
26. Bibliothèque du Conseil d'Etat	BS	Non
27. Ecole nationale supérieure des Beaux-arts - médiathèque	BU	Non
28. Centre culturel irlandais	BS	Non
29. Cité de la musique	BS	Non
30. Bibliothèque du Château de Chantilly	BP	Non
31. Bibliothèque André-Desguine	BP	Oui
32. Centre d'Etudes Supérieures de la Renaissance	BU	Non
33. Médiathèque municipale de Tourcoing	BP	Non
34. Cité internationale de la bande dessinée et de l'image	BS	Non
35. Association des bibliophiles universels	BS	Oui
36. Pôle Universitaire Européen Lille-Nord-Pas-de-Calais	BU	Non
37. Bibliothèque Mazarine	BP	Oui
38. Bibliothèque de l'Assemblée Nationale	BS	Non
39. Bnf	BN	Oui
40. Bibliothèque de Rennes Métropole - Les Champs libres	BP	Oui
41. Cité des sciences et de l'industrie (Universcience) - Médiathèque d'histoire des sciences	BS	Non
42. Bibliothèque des Arts décoratifs (Les Arts décoratifs)	BS	Non
43. Maison de la culture yiddish - Bibliothèque Medem	BS	Non
44. Bibliothèque communautaire et interuniversitaire de Clermont-Ferrand	BU	Oui
45. Comédie-Française	BS	Non
46. COBB Agence de coopération des bibliothèques et centres de documentation en Bretagne	BS	Non

47. Conservatoire National Supérieur de Musique et de Danse de Paris(CNSMDP) - Médiathèque Hector Berlioz	BS	Oui
48. Languedoc-Roussillon livre et lecture	BS	Non
49. BPI – Centre pompidou	BN	Oui
50. Bibliothèque de l'Ecole des Mines de Paris	BU	Oui
51. Médiathèque municipale de Bayeux	BP	Non
52. Bibliothèque municipale de Riom	BP	Non
53. Bibliothèque municipale de Charleville-Mézières	BP	Non
54. Médiathèque d'Epernay	BP	Non
55. Bibliothèque municipale d'Aurillac	BP	Non
56. Bibliothèque Carnegie	BP	Oui
57. Médiathèque de l'Agence de développement de la culture kanak	BS	Oui
58. Bibliothèque municipale de Toulon	BP	Non
59. Bibliothèque municipale de Saint-Claude	BP	Non
60. Bibliothèque municipale d'Orléans (Aurelia)	BP	Oui
61. Bibliothèque municipale d'Arles	BP	Non
62. Bn-r	BP	Oui
63. Archives-Médiathèque du Pays de Flers	BP	Non
64. Bibliothèque Armand Salacrou	BP	Non
65. Bibliothèque municipale de Baud -Conservatoire régional de la carte postale (Cartolis)	BP	Oui
66. Bibliothèque municipale de Chambéry	BP	Oui
67. Médiathèque Elisabeth et Roger Vailland	BP	Non
68. Bibliothèque municipale de Dole	BP	Non
69. Médiathèque de Perpignan	BP	Oui
70. Bibliothèque municipale de Saint-Quentin	BP	Non
71. Bibliothèque municipale de Joigny	BP	Non
72. Bureau des bibliothèques et de la lecture	BP	Non

73. L'Astrolabe : Médiathèque et Archives de Melun (Numel)	BP	Oui
74. Bibliothèque du Tourisme et des Voyages	BP	Non
75. Médiathèque Louis Aragon	BP	Oui
76. Bibliothèque municipale de Nancy	BP	Non
77. Bibliothèque municipale de Sens	BP	Non
78. Bibliothèque municipale de Valenciennes	BP	Oui
79. Bibliothèque municipale de Saint-Mihiel	BP	Non
80. Bibliothèque municipale de Cambrai	BP	Non
81. Médiathèque Alphonse Daudet d'Alès	BP	Non
82. Bibliothèque municipale d'Avignon	BP	Non
83. Mémorabilia,	BP	Oui
84. Bibliothèque du Patrimoine "Tommaso Prelà"	BP	Non
85. Médiathèque départementale Puy-De-Dome	BP	Oui
86. Bibliothèque municipale de Chartres	BP	Non
87. Bibliothèque municipale de Vierzon	BP	Non
88. Médiathèque du Grand Troyes	BP	Oui
89. Réseau des bibliothèques municipales de Saint-Brieuc	BP	Non
90. Médiathèque Equinoxe	BP	Non
91. Bibliothèque municipale de Montargis	BP	Non
92. Bibliothèque municipale de Tonnerre	BP	Non
93. Médiathèque de la Narbonnaise	BP	Non
94. Bibliothèque municipale de Besançon	BP	Oui
95. Bibliothèque municipale de Lons-le-Saunier	BP	Non
96. Centre iconographique de la Flandre	BP	Non
97. Bibliothèque municipale d'Ajaccio	BP	Non
98. Bibliothèque municipale de Tours	BP	Non
99. Bibliothèque municipale de Cherbourg (Coriallo)	BP	Oui

100.	Bibliothèque municipale de Douai	BP	Non
101.	BMVR de Toulouse (Rosalis)	BP	Oui
102.	Bibliothèque municipale de Saint-Étienne	BP	Non
103.	Bibliothèque-médiathèque de Metz	BP	Non
104.	Bibliothèque municipale de Grenoble	BP	Oui
105.	Bibliothèque municipale de Vendôme	BP	Non
106.	Bibliothèque municipale de Montbéliard	BP	Non
107.	Bibliothèque municipale de Tournus	BP	Non
108.	Médiathèque intercommunale des Trois Vallées (Gassendi)	BP	Oui
109.	Bibliothèque municipale de Senlis	BP	Non
110.	Bibliothèque municipale de Lyon	BP	Oui
111.	Bibliothèque municipale de Bordeaux (Séléné)	BP	Oui
112.	Bibliothèque municipale du Puy-en-Velay	BP	Non
113.	Bibliothèque municipale de Carpentras	BP	Non
114.	Médiathèque municipale de Tourcoing	BP	Non
115.	Bibliothèque municipale de Soissons	BP	Non
116.	Bibliothèque de l'Hôtel de Ville de Paris	BP	Non
117.	Bibliothèque municipale d'Abbeville	BP	Oui
118.	Bibliothèque municipale de Châteaudun	BP	Non
119.	Bibliothèque municipale Marcel Arland	BP	Non
120.	Médiathèque François Mitterrand	BP	Oui
121.	Bibliothèques d'Amiens Métropole	BP	Non
122.	Médiathèque de Lorient	BP	Non
123.	Médiathèque communautaire de Draguignan	BP	Non
124.	Médiathèque municipale de Saint-Dié	BP	Non
125.	Bibliothèque des littératures policières	BP	Non
126.	Médiathèque André Malraux de Lisieux	BP	Oui

127.	Bibliothèque numérique du Limousin	BP	Oui
128.	Bibliothèque Forney	BP	Non
129.	Réseau des bibliothèques de Bourges – Bibliothèque des Quatre Piliers	BP	Oui
130.	Bibliothèque municipale d'Autun	BP	Non
131.	Réseau des médiathèques de la communauté d'agglomération du Beauvaisis	BP	Non
132.	Bibliothèque Marguerite Durand	BP	Non
133.	Médiathèque de la communauté urbaine d'Alençon	BP	Oui
134.	Bibliothèque municipale de Foix	BP	Non
135.	Médiathèque numérique Valence Romans	BP	Oui
136.	Bibliothèque municipale de Grasse	BP	Oui
137.	Bibliothèque municipale de Nîmes	BP	Oui
138.	Médiathèque municipale Pierre-Amalric d'Albi	BP	Non
139.	Bibliothèque municipale de Brest	BP	Non
140.	Cité du livre – Bibliothèque Méjanes	BP	Oui
141.	Bibliothèque municipale de Gray	BP	Non
142.	Bibliothèque historique de la ville de Paris	BP	Non
143.	BMVR Alcazar	BP	Non
144.	Bibliothèque municipale de Provins	BP	Non
145.	Bibliothèque municipale d'Evreux	BP	Non
146.	Bibliothèque municipale de Nantes	BP	Oui
147.	Bibliothèque municipale d'Auxerre	BP	Non
148.	Bibliothèque municipale de Hyères	BP	Non
149.	Rouen Nouvelles Bibliothèques	BP	Non
150.	Bibliothèque municipale de Verdun	BP	Non
151.	Bibliothèque municipale d'Auch	BP	Non
152.	Bibliothèque municipale de la Roche-sur-Yon / Bibliothèque Vendée	BP	Oui

153.	Médiathèque de Chaumont	BP	Oui
154.	Bibliothèque municipale de Romorantin-Lanthenay	BP	Non
155.	Bibliothèque du cinéma François Truffaut	BP	Non
156.	Bibliothèque municipale de Saint-Bonnet-le-Château	BP	Non
157.	Bibliothèque municipale de Vesoul	BP	Non
158.	Bibliothèque Humaniste de Sélestat	BP	Oui
159.	Médiathèque musicale de Paris	BP	Non
160.	Bibliothèque municipale de Laon - Bibliothèque Suzanne Martinet	BP	Oui
161.	Médiathèque Villa-Marie de Fréjus	BP	Non
162.	Bibliothèque municipale de Loches	BP	Non
163.	Bibliothèque intercommunale Pau-Pyrénées	BP	Oui
164.	Bibliothèque municipale de Pontarlier	BP	Non
165.	Bibliothèque municipale de Chalon-sur-Saône	BP	Non
166.	Bibliothèque municipale de Châlons-en-Champagne	BP	Oui
167.	Médiathèque municipale de Sedan	BP	Oui
168.	Bibliothèque municipale de Versailles	BP	Non
169.	Médiathèque municipale de Meaux	BP	Non
170.	Bibliothèque patrimoniale et d'étude Romain Gary	BP	Oui
171.	Bibliothèque municipale de Conches-en-Ouche	BP	Non
172.	Bibliothèque intercommunale Épinal-Golbey	BP	Oui
173.	Bibliothèque municipale de Blois	BP	Non
174.	Bibliothèque municipale de Beaune	BP	Non
175.	Bibliothèque municipale de Lille	BP	Oui
176.	Bibliothèque municipale de Mâcon	BP	Non
177.	Bibliothèque de l'Heure Joyeuse	BP	Non
178.	Bibliothèque municipale de Nevers	BP	Non
179.	Bibliothèque municipale de Semur-en-Auxois	BP	Non

180.	Bibliothèque municipale d'Avallon	BP	Non
181.	Bibliothèque municipale de Salins-les-Bains	BP	Oui
182.	Bibliothèque municipale de Vire	BP	Non
183.	Bibliothèque municipale d'Angers	BP	Non
184.	Bibliothèque municipale de Caen	BP	Non
185.	Bibliothèque municipale de Dijon	BP	Oui
186.	Médiathèque intercommunale du Père Castor	BP	Non
187.	Bibliothèque municipale de Moulins	BP	Oui
188.	Bibliothèque municipale d'Avranches	BP	Non
189.	Médiathèque centrale d'Agglomération Emile Zola	BP	Oui
190.	Centre de recherche bretonne et celtique	BU	Oui
191.	NUMDAM	BS	Oui
192.	Bibliothèque universitaire – Nice	BU	Oui
193.	Médiathèque numérique Valence Romans	BP	Oui

BU : Bibliothèque universitaire / **BN** : Bibliothèque nationale /**BS** : Bibliothèque spécialisée /**BP** : Bibliothèque publique

Oui : la bibliothèque répond à nos critères de sélection / **Non** : la bibliothèque ne répond pas à nos critères de sélection

Annexe VIII : Les bibliothèques sélectionnées

Bibliothèque	Type	Site web	Établissement
1. Biblioteca virtuel cordel	BU	http://cordel.edel.univ-poitiers.fr/	Université de Poitiers
2. Les premiers socialismes	BU	http://premierssocialismes.edel.univ-poitiers.fr/	Université de Poitiers
3. CIBeLES	BU	http://cibeles.edel.univ-poitiers.fr/collections/show/1	Université de Poitiers
4. Les coutumiers du Centre-Ouest	BU	http://coutumiers.edel.univ-poitiers.fr/index.php	Université de Poitiers
5. Fonds Queneau	BU	http://www.queneau.fr/	Université de Bourgogne
6. Argonnaute (BDIC)	BU	http://argonnaute.u-paris10.fr/	Université Paris 10
7. INHA	BS	http://bibliotheque-numerique.inha.fr/	Institut national d'histoire de l'art
8. Polib	BU	http://polib.univ-lille3.fr/	Université Lille 3
9. Nordnum	BU	http://nordnum.univ-lille3.fr/	Université Lille 3
10. Bibliothèque numérique du paris 8	BU	http://www.bibliotheque-numerique-paris8.fr/	Université Paris 8
11. la bibliothèque numérique Foli@	BU	http://www.biu-montpellier.fr/florabium/servlet/LoginServlet	Université Montpellier
12. Iris	BU	http://iris.univ-lille1.fr/	Université Lille 1
13. la médiathèque numérique occitane	BP	http://occitanica.eu/index.php/fr/	Le Centre inter-régional de développement de l'occitan
14. Normannia	BS	http://www.normannia.info/	Centre régional des Lettres de Basse-Normandie
15. Bibliothèque virtuelle André Desguine	BP	http://bibliotheque-desguine.hauts-de-seine.net/desguine/Bibliotheque-virtuelle?lang=fr	Département des Hauts-de-Seine / Nanterre
16. ABU : la Bibliothèque Universelle	BS	http://abu.cnam.fr/	Association des Bibliophiles Universels
17. Mazarinum	BP	http://mazarinum.bibliotheque-mazarine.fr/	Institut de France / Paris
18. BCU	BU	http://bibliotheque-virtuelle.clermont-universite.fr/	Université Clermont

19. Médiathèque Hector Berlioz/ la bibliothèque numérique	BS	http://mediatheque.cnsmdp.fr/opacwebaloes/opac/index.aspx?IdPage=117	Conservatoire national supérieur de musique et de danse de Paris
20. Bibliothèque patrimoniale numérique	BU	https://patrimoine.mines-paristech.fr/	École des mines de Paris
21. Bibliothèque de Reims / Bibliothèque Carnegie	BP	http://www.bm-reims.fr/medias/medias.aspx?INSTANCE=exploitation&PORTAL_ID=WBCT_WBCTDOC_131.xml&SYNCMENU=PAT_LIGNE#	Bibliothèque municipale de Reims
22. Médiatique centre culturel Tijibaou	BS	http://mediatheque.adck.nc/mediath/default.htm	l'Agence de développement de la culture kanak (ADCK)
23. Aurelia	BP	http://aurelia.orleans.fr/	Bibliothèque municipale d'Orléans
24. Cartolis	BP	http://www.cartolis.org/presentation_fond.php	la municipalité de Baud
25. BM Chambéry	BP	http://bibliotheque-numerique.chambery.fr/	Bibliothèque municipale de Chambéry
26. BM Perpignan	BP	http://www.numilog.com/bibliotheque/BM-Perpignan/default.asp	Bibliothèque municipale de Perpignan
27. Numel	BP	http://www.numel.fr/	Médiathèque de Melun
28. Médiathèque Louis Aragon	BP	http://80.82.239.59/ClientBookLine/toolkit/p_requests/formulaire.asp?GRILLE=MANUSCRITSSIMPLE_0	Médiathèque Louis Aragon /Le Mans
29. La bibliothèque numérique de Valenciennes	BP	http://patrimoine-numerique.ville-valenciennes.fr/in/collections/survol	Bibliothèque municipale de Valencienne
30. Gallica	BN	http://gallica.bnf.fr/	BNF
31. Tolosana	BU	http://tolosana.univ-toulouse.fr/	SICD de l'Université de Toulouse
32. Alienor	BS	http://www.alienor.org/	Conseil des musées de Poitou-Charentes
33. Elec	BS	http://elec.enc.sorbonne.fr/	École nationale des chartes (Paris)
34. Enssib	BS	http://www.enssib.fr/bibliotheque-numerique/	École nationale

			supérieure des sciences de l'information et des bibliothèques
35. La bibliothèque électronique de Lisieux	BP	http://www.bmlisieux.com/	Bibliothèque Municipale (Lisieux
36. DIGIMOM	BS	http://www.mom.fr/digimom/	La bibliothèque de la Maison de l'Orient et de la Méditerranée – Lyon 2
37. Les bibliothèques virtuelles humanistes	BS	http://www.bvh.univ-tours.fr/	Centre d'études supérieures de la Renaissance (Tours) Institut de recherche et d'histoire des textes (France)
38. Numelyo	BP	http://numelyo.bm-lyon.fr/index.php	Bibliothèque municipale de Lyon
39. Numistral	BN	http://www.numistral.fr/	Bibliothèque nationale et universitaire de Strasbourg
40. Textes rares	BS	http://pages.textesrares.com/	
41. Pagella	BP	http://pagella.bm-grenoble.fr/	Bibliothèque municipale de Grenoble
42. B N de Roubaix	BP	http://www.bn-r.fr/	Ville de Roubaix
43. Cefael	BS	http://cefael.efa.gr/site.php?ce=fu5g75upll0ntuqmm7i4cr082dbp8ju9	École française d'Athènes
44. Cnum	BS	http://cnum.cnam.fr/	Conservatoire numérique des Arts & Métiers (cnam)
45. ECL	BU	http://histoire.ec-lyon.fr/index.php?id=790	École Centrale de Lyon
46. Jubilothèque	BU	http://jubilotheque.upmc.fr/index.html	Bibliothèque numérique scientifique de l'UPMC /Paris
47. Napoleonica	BS	http://fondationnapoleon.org/activites-et-services/recherche/la-bibliotheque-numerique/	Fondation Napoléon.
48. Ampère	BS	http://www.ampere.cnrs.fr/	Cnrs
49. Buffon et l'histoire naturelle	BS	http://www.buffon.cnrs.fr/	Cnrs

50. La bibliothèque virtuelle « Claude Bernard »	BS	http://claudebernard.in2p3.fr/index.php?name=paratexte&page=presentation	Cnrs
51. Les Œuvres complètes de D'Alembert (1717-1783)	BS	http://dalembert.academie-sciences.fr/	Cnrs
52. Lamarck Jean-Baptiste	BS	http://www.lamarck.cnrs.fr/	Cnrs
53. Lavoisier	BS	http://www.lavoisier.cnrs.fr/	Cnrs
54. Manuscrits de Flaubert	BU	http://flaubert.univ-rouen.fr/	Université de Rouen
55. Proscenium	BS	http://www.leproscenium.com/	Lusi sensu société
56. Bibnum	BS	https://www.bibnum.education.fr/	Fondation Maison des sciences de l'homme
57. Criminocorpus	BS	https://criminocorpus.org/fr/	Association Criminocorpus
58. Mémorabilia	BP	http://www.mediatheque-montauban.com/mediatheque-montauban.com/index/index?id_profil=26	Bibliothèque municipale Antonin Perbosc
59. Bibliothèque virtuelle de Clairvaux	BP	https://www.bibliotheque-virtuelle-clairvaux.com/	Médiathèque du Grand Troyes
60. Bibliothèque Cujasnum	BU	http://cujasweb.univ-paris1.fr/	Université Paris 1
61. Bibliothèque municipale de Besançon	BP	http://www.bm-besancon.fr/index.php?p=1857	Bibliothèque municipale de Besançon
62. Coriallo	BP	https://phototheque.cherbourg-octeville.fr/#	Bibliothèque Municipale Jacques Prévert /Bibliothèque municipale de Cherbourg
63. Rosalis	BP	http://rosalis.bibliotheque.toulouse.fr/	Bibliothèque municipale de Toulouse
64. Gassendi	BP	https://www.gassendi.fr/	Médiathèque intercommunale des Trois Vallées
65. Séléné	BP	http://bibliotheque.bordeaux.fr/le-patrimoine/bibliotheque-numerique.dot	Bibliothèque municipale de Bordeaux

66. Bibliothèque municipale d'Abbeville	BP	http://www.abbeville.fr/2-non-categorise/110-les-collections-numerisees-de-la-bibliotheque.html	Bibliothèque municipale d'Abbeville
67. Médiathèque François Mitterrand	BP	http://www.bm-poitiers.fr/EXPLOITATION/patrimoine/patrimoine-numerise.aspx	Bibliothèque municipale de Poitiers
68. Réseau des bibliothèques de Bourges – Bibliothèque des Quatre Piliers	BP	http://mediatheque.ville-bourges.fr/EXPLOITATION/PATRIMOINE/bibliotheque-patrimoine.aspx	Réseau des bibliothèques de Bourges – Bibliothèque des Quatre Piliers
69. Médiathèque de la communauté urbaine d'Alençon	BP	http://my.yoolib.com/bmalencon/	Médiathèque de la communauté urbaine d'Alençon
70. Bibliothèque municipale de Grasse	BP	http://www.bibliotheques.ville-grasse.fr/EXPLOITATION/bibliotheque-numerique.aspx	Bibliothèque municipale de Grasse
71. Cité du livre – Bibliothèque Méjanes	BP	http://bibliotheque-numerique.citedulivre-aix.com/	Ville d'Aix-en-Provence.
72. BM Nantes	BP	https://catalogue-bm.nantes.fr/in/faces/browse.xhtml?facetClause=%2BTypeOfDocumentFacet%3ABDMDocument%3B&pageNo=1&query=*&pageSize=20&sort=author_sort_ASC&view=grid&debug=false&mz=%D0%B6	Bibliothèque municipale de Nantes
73. Médiathèque de Chaumont	BP	http://silos.ville-chaumont.fr/flora/jsp/banque_images.jsp	Médiathèque de Chaumont
74. Bibliothèque Humaniste de Sélestat	BP	http://bhnumerique.ville-selestat.fr/client/fr_FR/bh/	Bibliothèque Humaniste de Sélestat
75. Bibliothèque municipale de Laon - Bibliothèque Suzanne Martinet	BP	http://bibliotheque-numerique.ville-laon.fr/collection/?esa=resetall&navigation=0	Bibliothèque municipale de Laon - Bibliothèque Suzanne Martinet
76. Bibliothèque intercommunale Pau-Pyrénées	BP	http://mediatheques.agglo-pau.fr/ClientBookline/toolkit/p_requests/formulaire.asp?GRILLE=GED&PORTAL_ID=erm_portal_medias_pat.xml&INSTANCE=exploitation&OUTPUT=PORTAL&SYNCMENU=GED	Bibliothèque intercommunale Pau-Pyrénées
77. Bibliothèque municipale de Châlons-en-Champagne	BP	http://bmvr.chalons-en-champagne.net/in/faces/browse.xhtml?facetClause=%2BTypeOfDocumentFacet	Bibliothèque municipale de Châlons-en-Champagne

		%3AOAIDocument%3B	
78. Médiathèque municipale de Sedan	BP	http://www.bm-sedan.fr/medias/medias.aspx?INSTANCE=exploitation&PORTAL_ID=portal_model_instance__Recherche_dans_les_collections_numerisees.xml&SYNCMENU=RECHERCHESNUM	Médiathèque municipale de Sedan
79. Bibliothèque patrimoniale et d'étude Romain Gary	BP	http://manuscrits.nice.fr/_app/index.php	Bibliothèque municipale de Nice
80. Bibliothèque intercommunale Épinal-Golbey	BP	http://www.bmi.agglo-epinal.fr:8084/base_patrimoine/Francais/index.php	Bibliothèque intercommunale Épinal-Golbey
81. BM Lille	BP	http://www.bm-lille.fr/la-bibliotheque-numerique/	Bibliothèque municipale de Lille
82. Bibliothèque municipale de Salins-les-Bains	BP	http://bib-salins.dnsalias.net/public/	Bibliothèque municipale de Salins-les-Bains
83. BM Dijon	BP	http://patrimoine.bm-dijon.fr/pleade/ead.html?id=FR212316101_presentation#!{%22content%22:[%22FR212316101_presentation_e0000007%22,true,%22%22]}	Bibliothèque municipale de Dijon
84. Bibliothèque municipale de Moulins	BP	http://mediatheques.agglo-moulins.fr/agglo-moulins.fr/index/index/id_profil/161#/page/1	Bibliothèque municipale de Moulins
85. Mémonum	BP	https://mediatheques.montpellier3m.fr/MEMONUM/accueil-memonum.aspx	Médiathèque centrale Émile Zola Montpellier
86. Bibliothèque Jean-François Séguier	BP	http://bibliotheque-numerique.nimes.fr/fre/virtualcollections/seguier/index.html	la Ville de Nîmes-Carré d'art bibliothèque.
87. Medi@	BU	http://www.biusante.parisdescartes.fr/histoire/medica/	BIU santé, Paris
88. Bibliothèque Diderot de Lyon	BU	http://numerisation.bibliotheque-diderot.fr/R/2GT69YE3MADIX3AXXXKN8CGKK7Q1838JF4KJDSVNA2PXMY5KMK-03426?RN=776800808&pds_handle=GUEST	BIU Lyon
89. Bibliothèque numérique du Limousin	BP	http://www.bn-limousin.fr/	Bibliothèque francophone multimédia de Limoges

90. BPI- Centre Pompidou	BN	http://www.bpi.fr/home.html	Bibliothèque publique d'information
91. Bibliothèque municipale de la Roche-sur-Yon / Bibliothèque Vendée	BP	http://bibliotheque.vendee.fr/	La Direction des Bibliothèques de la Vendée
92. Médiathèque numérique Puy-de-Dôme	BP	http://mediatheque-numerique.puy-de-dome.fr/	Médiathèque de Puy-de-Dôme
93. Centre de recherche bretonne et celtique	BU	https://www.univ-brest.fr/crbc/#	Université de Brest
94. Médiathèque numérique Valence Romans	BP	http://mediatheques.valenceromansagglo.fr/bibliotheque-numerique	Médiathèque de Valence Romans
95. Bibliothèque de Rennes Métropole - Les Champs libres	BP	http://www.bibliotheque.leschampslibres.fr/	Bibliothèque de Rennes métropole
96. Bibliothèques de l'université Nice Sophia Antipolis – Collections remarquables	BU	http://bibliotheque.unice.fr/ressources/presentation-des-ressources/les-collections-remarquables	Université Nice Sophia Antipolis
97. Bibliothèque numérique – Université Saint Étienne	BU	https://scd.univ-st-etienne.fr/fr/bibliotheque-numerique.html	Université Jean Monet
98. NUMDAM	BS	http://www.numdam.org/about/	Mathdoc – Université Grenoble Alpes

Annexe IX - Le courrier électronique envoyé aux bibliothèques cibles pour demander leur participation au questionnaire

Madame, Monsieur,

Doctorante en sciences de l'information et de la communication, je réalise une thèse sur les applications du web sémantique en bibliothèques numériques, sous la direction de Laurence Balicco.
L'objectif de notre travail est de faire l'état des lieux des applications du web sémantique en bibliothèques numériques en France. Le but est de s'interroger sur les pratiques documentaires des bibliothécaires qui travaillent dans la gestion des collections numériques.
A ce titre, nous vous serions reconnaissantes de bien vouloir prendre quelques instants pour répondre à notre questionnaire. Cela vous prendra environ cinq minutes.
Vous pouvez remplir le questionnaire via le lien internet
suivant :https://goo.gl/forms/F4ABS1k12yS0KHjz2

En vous remerciant par avance du temps que vous voudrez bien consacrer à cette enquête

--
Hiba MELHEM
Enseignante à l'IUT2 / UPMF - Grenoble
Doctorante au Groupe de Recherche sur les Enjeux de la Communication (Gresec)
Université Stendhal - Grenoble 3

Annexe X - Le courrier électronique envoyé aux professionnels de l'information pour demander un entretien

Bonjour,

Je tiens à vous remercier d'avoir pris le temps de répondre au questionnaire que je vous ai adressé récemment. (voir questionnaire ci-dessous)
Dans le cadre de ma thèse, ce questionnaire doit être complété par une étude qualitative sous forme d'entretiens semi-directifs. L'objectif de cette seconde partie est de s'interroger sur les pratiques documentaires suite à l'introduction du web sémantique en bibliothèque numérique.

Pour cette raison, je me permets de vous solliciter par un entretien. Ce dernier sera organisé en quatre parties :
1. Appréhender le profil du bibliothécaire /professionnel d'information
2. Comprendre la gestion des collections numériques et les métadonnées
3. Discuter sur les enjeux du web sémantique en bibliothèques numériques
4. Interroger les bibliothécaires sur leurs pratiques documentaires

Merci de me contacter si vous êtes intéressé. Je suis disponible pour vous rencontrer au moment qui vous conviendra le mieux.
Nous pouvons également réalisons cet entretien par skype ou par téléphone.

Je reste à votre disposition pour toute question complémentaire.

En vous remerciant pour l'investissement que vous portez à ce projet, je vous souhaite une bonne journée.

--
Hiba MELHEM
Enseignante à l'IUT2 / UPMF - Grenoble
Doctorante au Groupe de Recherche sur les Enjeux de la Communication (Gresec)
Université Stendhal - Grenoble 3

Annexe XI : Classement des bibliothèques numériques sélectionnées

Bibliothèques numériques nationales	**Bibliothèques numériques universitaires**	**Bibliothèques numériques de lecture publique**	**Bibliothèques numériques spécialisées**
Gallica	Biblioteca virtuel cordel	La médiathèque numérique occitane	Bibnum
BPI	Les premiers socialismes	Mazarinum	ABU : bibliothèque universelle
Numistral	CIBeLES	La Bibliothèque virtuelle André Desguine	Normannia
	Les coutumiers du Centre-Ouest	La bibliothèque électronique de Lisieux	Textes rares
	Fonds Queneau	Bibliothèque municipale de Lille	Proscenium
	Argonnaute (BDIC)	La bibliothèque numérique de Valenciennes	Napoleonica
	Bibliothèque numérique – Université Saint Étienne	Bibliothèque de Reims / Bibliothèque Carnegie	Criminocorpus
	Polib	Bibliothèque municipale de Salins-les-Bains	Alienor
	Nordum	Aurelia	INHA
	Bibliothèque numérique du paris 8	Bibliothèque de Rennes Métropole - Les Champs libres	Médiathèque Hector Berlioz/ la bibliothèque numérique
	La bibliothèque numérique Foli@	Cartolis	Médiatique centre culturel Tijibaou
	Iris	Bibliothèque municipale de Chambéry	DIGIMOM
	BCU	Bibliothèque municipale de Perpignan	Les bibliothèques virtuelles humanistes
	Bibliothèque patrimoniale numérique	Médiathèque numérique Valence Romans	Enssib
	Tolosana	Numel	Elec

	Jubilothèque	Médiathèque Louis Aragon	Cnum
	ECL	Numelyo	Cefael
	Manuscrits de Flaubert	Pagella	Ampère
	Bibliothèque Cujasnum	Bibliothèque numérique de Roubaix	Buffon et l'histoire naturelle
	Medic@	Bibliothèque numérique du Limousin	La bibliothèque virtuelle « Claude Bernard »
	Bibliothèque Diderot de Lyon	Bibliothèque municipale de la Roche-sur-Yon / Bibliothèque Vendée	Les Œuvres complètes de D'Alembert (1717-1783)
	Centre de recherche bretonne et celtique	Médiathèque numérique Puy-de-Dôme	Lamarck Jean-Baptiste
	Bibliothèque Nice Sophia Antipolis	Bibliothèque intercommunale Épinal-Golbey	Lavoisier
		Bibliothèque municipale de Dijon	NUMDAM
		Mémorabilia	
		Bibliothèque virtuelle de Clairvaux /Grand Troyes	
		Mémonum	
		Bibliothèque Jean-François Séguier	
		Bibliothèque municipale de Besançon	
		Coriallo	
		Rosalis	
		Gassendi	
		Séléné	
		Bibliothèque municipale	

		d'Abbeville	
		Médiathèque François Mitterrand	
		Réseau des bibliothèques de Bourges – Bibliothèque des Quatre Piliers	
		Médiathèque de la communauté urbaine d'Alençon	
		Bibliothèque municipale de Grasse	
		Cité du livre – Bibliothèque Méjanes	
		Bibliothèque municipale de Moulins	
		Bibliothèque municipale de Nantes	
		Médiathèque de Chaumont	
		Bibliothèque Humaniste de Sélestat	
		Bibliothèque municipale de Laon - Bibliothèque Suzanne Martinet	
		Bibliothèque intercommunale Pau-Pyrénées	
		Bibliothèque municipale de Châlons-en-Champagne	
		Médiathèque municipale de Sedan	
		Bibliothèque patrimoniale et d'étude Romain Gary	

Annexe XII : Les entretiens

Entretien 1

Réalisé le 26 septembre 2016

Durée 1h10

L'entretien est organisé en quatre parties : La première partie porte sur le profil de bibliothécaire, la deuxième sur la gestion des collections numériques, la troisième sur les web sémantique et la dernière sur les pratiques professionnelles.

Je veux vous demander de décrire votre cursus universitaire et votre parcours professionnel

Je suis bibliothécaire depuis 2003, avant j'ai fait des études de lettres, en lettres moderne, et je me suis reconverti vers la documentation, surtout vers l'informatique documentaire en faisant des cours du soir au CNAM, essentiellement pour pouvoir, au début, par curiosité et ensuite là je me suis rendu compte que ça sera utile dans mon futur métier bibliothécaire. Savoir tout ce qui touche les bases des données, les langages type html xml et puis les réseaux, avoir des bases sur le fonctionnement des réseaux, donc j'ai suis des cours de soir au CNAM pour me réorienter pour passer le concours des bibliothèques, que j'ai passé entre 2000 et 2003, et j'ai commencé à travailler en bibliothèque en 2003.

Est-ce que vous avez suivi des formations depuis que vous travaillez en bibliothèque ?

Régulièrement oui, à peu près une ou 2 formations par an, et comme je suis dans un domaine assez technique, il n'y a pas forcement de formation qui me convienne, j'ai souvent donné des formations aussi […] je pense que depuis que j'ai commencé à travailler j'ai suivi 2 ou 3 formations techniques intéressantes.

Quelle est votre fonction ou votre titre à la bibliothèque ?

Responsable du service informatique, parce que cette bibliothèque a ses propres services dédiés aux ressources informatiques et numériques, ça fait presque cinq personnes et on est intégré à la bibliothèque.

Quelles sont vos missions ?

Les missions sont en train d'évoluer.Au départ, c'était essentiellement faire fonctionner tous les moyens informatiques, alors ça concerne le logiciel de gestion, ça concerne les postes accès internet qui sont proposés au public dans toutes les bibliothèques, et de plus en plus ça évolue vers les bibliothèques numériques, donc tout ce qui est gestion des ressources, c'est pour ça qu'on va passer sur un portail dans 15 jours. Et aussi tout ce qui est nouveau support numérique. Tablettes lisseuses smartphone, les missions du service ont un peu évolué pour accompagner les usagers, on organise des ateliers …

Je passe à la deuxième partie sur la gestion des collections numériques. Dans ma thèse je m'intéresse beaucoup à la notion de bibliothèque numérique, qu'est-ce que c'est une bibliothèque numérique ? comment la définir etc., et je m'intéresse aussi à comment vous les professionnels, vous définissez une bibliothèque numérique, qu'est-ce qu'une bibliothèque numérique pour vous ?

Je pense le cadre de nos bibliothèques est très intéressant, parce qu'il reflète vraiment l'évolution historique. Au départ, la bibliothèque a été lancée en même temps que les autres, une bibliothèque numérique patrimoniale avec des images numérisées très classiques, une interface de recherche, feuilletage en ligne etc., uniquement sur des documents libres de droits dans le domaine public. Il y avait des photos des estampes des objets pris en photos, ça été considéré comme la bibliothèque

numérique pendant longtemps par opposition à la bibliothèque physique patrimoniale. On a répondu à plusieurs appels à projet du ministère sur des projets innovants liés aux ressources numériques. En 2010 la BM a travaillé sur une offre de vidéo à la demande [..] qui couvrait des films sous droits, donc c'était la première fois ici que la bibliothèque numérique a commencé à englober des ouvrages et des documents sous droits qui n'étaient pas libres. En parallèle même depuis longtemps en 2005 la bibliothèque était aussi abonnée à des livres numériques chez un fournisseur Numilog [..] qui a eu un certain succès, ça veut dire petit à petit entre 2000 et 2012 plutôt entre 2005 et 2012 la notion de la bibliothèque numérique a commencé à s'accroître pour englober des documents qui n'étaient pas forcement libres, pas forcement patrimoniaux et pas forcément des livres. Donc c'est sur cette base qu'on a travaillé avec les collègues pour réaliser un portail, c'est-à-dire la solution informatique qui permet de naviguer dans tous les contenus numériques, on écarte tous les contenus physiques mais on essaye d'agréger le maximum de contenu numérique pour permettre aux usagers d'aller jusqu'à la consultation des documents même s'ils ne sont pas dans la bibliothèque, et même s'ils ne sont pas libres. Bien sûr que l'accès est réservé aux usagers inscrits, ceux qui ont payé leur abonnement. Nous achetons pour eux des ressources sous droits, donc ces ressources sous droits sont maintenant comprises dans les bibliothèques numériques, et ce qui est très bien accepté fin je ne sais pas si c'est bien accepté mais ça va finir par être accepté par l'ensemble des collègues.

Et donc ça sera une recherche fédérée ?

Non ce n'est pas de la recherche fédérée, justement ça on a essayé de l'éviter, c'est ce qu'on appelle la recherche unifiée, c'est-à-dire que c'est le portail qui moissonne les métadonnées et celui qui fait la présentation. J'ai travaillé sur la recherche fédérée pendant plusieurs années, ça ne marche pas parce que ce n'est pas suffisamment fluide pour l'usager, là c'est ce qu'on a vraiment cherché à éviter.

Quels sont les langages documentaires utilisés pour gérer les collections numériques ?

C'est surtout du marc unimarc, on a un peu de xml dans la bibliothèque numérique patrimoniale puisqu'elle repose sur une solution qui s'appelle Plead notice Plead qui produit du xml. Et à l'intérieur on a de OAI, du dublin core c'est surtout du dublin core, ensuite sur le livre numérique est apparue depuis deux ans un projet nationale qui s'appelle PNB prêt numérique en bibliothèques, il y a un catalogue des livres numérique sous droits géré par la société Digicom et qui nous propose des notices en format xml, xml d'éditeur qu'on utilise parfois.

L'un des éléments essentiels dans une bibliothèque numérique est la production des métadonnées. Pouvez-vous me parler des métadonnées produites dans votre bibliothèque ?

En fait, on en produit pas, non j'exagère, sur le livres numériques on en produit pas mais le seule domaine qu'on en produit c'est le patrimoine […] je n'ai pas de ressources humaines nécessaires pour produire davantage des métadonnées sur toutes les autres ressources, les métadonnées sont fournies dans le cadre de l'abonnement ou dans le cadre de l'achat. Elles ne sont pas forcément corrigées et enrichies par la suite, on les utilise telles qu'elles sont.

Est-ce qu'il y a des pistes stratégiques pour la production des métadonnées dans l'avenir?

Non pas de piste.

Pourquoi ?

Parce que ça prend beaucoup de temps, je pense que vous avez dû constater si vous avez interrogé d'autres bibliothèques

Non vous êtes la première

Il y a une intervention intéressante d'un bibliothécaire qui s'appelle Nicolas Morin il y a quelques années, que vous pouvez le trouver sur le site de l'ADBU association des directeurs des bibliothèques universitaires, et en fait il disait que depuis 10 ans les bibliothécaires ont privatisé leurs métadonnées en les externalisant, en fait en bibliothèques de lecture publique c'est totalement vrai, ici le seul domaine où on en produit c'est le patrimoine, ce sont les documents uniques qui ne sont pas forcément référencés, [..] le temps que ça prend est devenu exorbitant par rapport aux exigences qu'on a et aux personnels qu'on a à mettre sur cette tâche, [..] Donc là il y a 90 000

documents à inventorier, et on est à 4500 et pour le reste les métadonnées sont importées au moment où on fait les achats. Si c'est un document physique on reçoit le document et un lien pour télécharger la notice, et si c'est un document numérique c'est variable, de plus en plus de côté des éditeurs il y a eu un progrès, mais des fois ils nous fournissaient juste une feuille Excel, après il faut les re-cataloguer, ils ont bien compris de leur côté que c'était plus possible chez nous [..] je pense à un de touch qui est là, c'est une ressource où il y a 80 milles albums, l'éditeur a fait l'effort de mettre à disposition un entrepôt des métadonnées pour le 3000 album et nous allons juste récupérer ça et on l'intègre dans le portail.

Je pense à la troisième partie sur le web sémantique : Quels sont les enjeux de l'utilisation du web sémantique en bibliothèques numériques?

Je pense il y a deux aspects le premier c'est effectivement la visibilité via les moteurs de recherche, parce que actuellement tous les catalogues des bibliothèques sont dans le web profond on va dire. [...] c'est le premier aspect le 2eme aspect c'est gagner de la qualité sans mobiliser des ressources humaines, toujours pareil, et c'est là où ça devient extrêmement complexe, parce qu'il faut arriver à passer du marc, des outils fermés et propriétaires etc. à une logique complètement ouverte complètement orientée web.

Vous avez répondu que la bibliothèque n'utilise pas les standards du web sémantique pourquoi pour quelles raisons?

Il y a 150 personnes dans ce réseau les compétences sont très largement distribuées, je ne sais combien il y a des catalogueurs mais je pense qu'il y a une centaine, tous ces gens ont été formés par la filière habituelle du métier du livre dont laquelle le web sémantique était à peine abordé [...] quand je dis manque de compétences ça veut qu'il n'y a pas sur un équipe assez important suffisamment de personnes sur lesquels s'appuyer, en plus ça se voit pas forcement mais c'est un réseau de 12 bibliothèques maintenant complètement décentralisées, chacun bosse dans son coin, si je voulais que le web sémantique se développe, il faut que je forme 12 personnes, et j'en ai une dans chaque bibliothèque, et qu'elle soit capable de former ses collègues, d'autant que le seul élément commun à tous ces collègues c'est le logiciel, tant que le logiciel n'est pas mûr pour le web sémantique ça ne sert à rien. On a un logiciel très marc très métier qui est costaud, mais qui n'est pas RDA , qui ne passera pas la transition bibliographique, donc c'est un des problèmes qu'on aura à gérer d'ici quelques années, et ensuite il est incapable de fournir des triplets rdf, il est incapable d'intégrer des notices xml c'est compliqué donc voilà il faudra évoluer.

C'est plus simple d'évoluer le logiciel ?

Oui c'est plus simple, mais ça ne marche pas seul, il faut former l'équipe derrière là on vient de changer la version de logiciel ça nous a pris quand même 6 mois parce c'est un réseau très complexe [...] j'ai déjà envisagé de préparer tout le monde à RDA […] aujourd'hui je ne vois pas trop comment. Il faut voir aussi que dans la bibliothèque de lecture publique, peut-être ce n'est pas la même chose en BU, mais dans la bibliothèque de lecture publique, le catalogage, la description des documents, c'est en train de devenir le parent pauvre, même si c'est toujours revendiqué comme le cœur du métier ce sont des tâches minimalistes, donc c'est pour ça qu'on importe les métadonnées. [...] les tâches des bibliothécaires ont totalement évolué vers le service au public vers la médiation.

Si la bibliothèque venait à évoluer et à introduire l'utilisation du web sémantique, quels seraient selon vous les principaux défis ?

En bibliothèque numérique je pense qu'on peut y arriver plus facilement, parce qu'on récupère des métadonnées qui sont déjà plus faciles à manipuler, qui sont déjà en XML, le défi va toujours être les identifiants. A savoir que sur le portail on intègre aussi des métadonnées qui viennent de 8 sources, on a aucun identifiant commun, le point commun c'est le portail qui les reconstruit [..] les métadonnées aujourd'hui elles n'ont pas des uris, donc ça ne va pas marcher, il y a des sources très hétérogènes, derrière il y a une qualité de production qui est aussi hétérogène, je pense que les meilleurs métadonnées que nous fournissons, ça doit être le projet PNB, on a des identifiants internes qui sont liés au projet, mais on a pas du tout de arc, d'identifiant arc ou on n'a pas

d'identifiants universels. On a beaucoup travaillé sur ce portail sur un artiste transversale qui s'appelle Ibrahim Maalouf avec des surprises, c'est-à-dire on a le Touch qui nous fournit de la musique en ligne, on a cinévod qui nous fournit des films, et dans certains cas Ibrahim Maalouf a fait les bandes originales des films qu'on propose via cinévod, et en fait la seule interface qui s'en est rendu compte est le portail, parce qu'il trouvait le mot Maalouf en commun. En livre numérique on a aussi un livre de Amin Maalouf et là c'est bien on a toute la famille ce n'est pas pertinent ! En même temps les usagers s'habituent à ça, ils savent quand ils tapent Maalouf ils ont de la musique de la vidéo et du livre de deux personnes ou trois différentes […] ce n'est pas au point parce que les éditeurs ne font pas le travail, nous on prend les métadonnées telles quelles sont, qu'est-ce qu'on fait avec ça, est-ce qu'on les réécrit, est-ce qu'on les enrichit, on va ajouter les identifiants qui manquent ça va prendre des semaines, et ça sert à rien, ça va les écraser à chaque mise à jour, c'est sans fin [..] on va faire la pression auprès des éditeurs pour qu'ils fassent correctement leurs métiers, c'est-à-dire qu'un éditeur qui publie Amin Maalouf, c'est lui qui doit s'arranger pour qui a une autorité et ensuite c'est celle-ci qu'on va utiliser [...] pourquoi ça marche à la bnf, moi vu de l'extérieur, je dis que ça marche parce que tout d'abord il y a les compétences, les moyens informatiques qu'on n'a pas, et ils ont la masse de données suffisante, le gros problème qu'on rencontre nous sur la bibliothèque numérique, qu'on manque de volume, c'est-à-dire on a rarement deux objets du même auteur, des objets avec des points commun ! Là j'ai parlé d'Ibrhaim Maalouf c'est une exception […] c'est à nous de construire le sens, [...] on a beaucoup réfléchi avec les collègues, comment est-ce qu'on va rapprocher des ressources numériques qui n'ont aucun point commun […] et là on attend pas mal du web sémantique, c'est-à-dire qu'il fasse lui-même ces liens entre toutes les ressources.

Est-ce que le web sémantique va être capable tout seul de produire du sens ?

Le web sémantique n'arrive pas tout seul, il arrive dans un monde documentaire où on a vu à un moment le FRBR. Le FRBR n'est pas tout à fait compris et accepté par tout le monde, en revanche le fonctionnement notice bibliographique notice autorité notice d'exemplaire, la distinction entre 3 objets est à peu près comprise par tout le monde, donc les fiches autorités ça relève complètement de ce que pourrait faire le web sémantique dans un premier temps. Les autorités posent néanmoins des problèmes, puisqu'elles ne sont pas tout à fait normalisées [..] si vous regardez bien des fois il y a presque autant des notices d'autorités que des notices bibliographiques, ce qui n'est pas un bon signe par ce qu'il y a des doublons. En lecture publique c'est un peu mieux parce que il n'y a pas des réservoirs d'autorités, on est obligé de recréer des autorités en interne. Là je me rends compte avec la bibliothèque numérique que mes collègues écrivent les noms d'auteurs avec ou non ses dates de naissances et de morts. Quand je déverse ça sur dbpedia qui est quand même un réservoir intéressant plutôt bien, ça ne marche pas puisque dbpedia ne prend pas les dates. Donc une fois sur deux ça marche parfois non. Ça c'est le premier exemple c'est un exemple vraiment typique pourtant ça à l'air simple, [..] c'est très normalisé sur dbpedia, c'est super, mais pour faire des requêtes de type web sémantique qui marche bien, malheureusement non s'il y a pas la date de vie, la date de naissance ou la date de mort ça ne marche plus. Ça c'est le premier cas. 2eme cas où on aimerait avoir beaucoup du web sémantique c'est tout ce qui est autorité sujet, et là on tombe sur un problème historique des bibliothèques françaises qui est rameau ! Parce que rameau c'est des vedettes construites, c'est extrêmement riche en sens et aucun système informatique est capable de comprendre les vedettes matières rameau. […] je trouve que c'est informatiquement ingérable c'est quelque chose qui a été pensé avant l'informatique, par contre documentairement c'est extrêmement riche, c'est passionnant mais parce que juste en une ligne vous savez exactement ce qui est dans le bouquin, c'est formidable mais malheureusement chaque ligne est unique chaque ligne est propre à chaque document, là où ça devient plus intéressant si on arrive à déconstruire la chaîne et à relier entre elles les chaines qui sont proches, mais ça c'est pareil ça demande des montagnes de données pour y arriver, donc la bnf peut travailler là-dessus, mais une bibliothèque lambda n'arrivera jamais, on aura très peu de livres qui ont suffisamment des éléments en commun pour que ça soit pertinent, on aura souvent la racine commune France 20 siècle là c'est bon on a la moitié du fonds.

Je passe à la dernière partie sur les pratiques professionnelles. Que pensez-vous de l'influence de la technologie sur vos pratiques professionnelles ?
L'omniprésence de Google le moteur de recherche aurait tendance à renvoyer les catalogues dans la préhistoire, mais nos utilisateurs quand ils font l'effort de venir en bibliothèques, Google ne les intéresse plus, parce qu'ils sont dans les murs [...] C'est pour cela je pense que les catalogues ont encore de l'avenir au moins dans la bibliothèque, de l'extérieur les pratiques documentaires sont balayées par les moteurs de recherche, et là est-ce qu'on fait l'effort de s'adapter comme je disais, nous on a un effort de visibilité à faire, il faut que les ressources numériques soient visibles, après les ressources physiques c'est un peu différent, parce que les usagers n'ont pas encore l'habitude d'aller les chercher de façon transversale, [..] Il y a une évolution des pratiques mais elle est plus délicate à gérer sur les ressources numériques, là il faut vraiment s'adapter parce que sinon on vod on va être écrabouillé par Netflix, en musique on va être écrabouillé par Deezer etc., […] Là il faut qu'on évolue sinon nos pratiques documentaires seront obsolètes.
Quels types de compétences mobiliser en termes de pratiques documentaires en vue d'accompagner l'utilisation du web sémantique en bibliothèques numériques?
Ce sont des compétences de fonds, ce sont des compétences documentaires classiques de type description indexation catalogage, [..] il y a une communication là-dessus pendant une journée à l'Enssib l'an dernier, je peux la retrouver parce que j'ai été intervenu juste avant ou après, c'était un professeur d'iut qui recensait toutes les formations liées au numérique dans les maquettes des iut métier de livres à travers la France, parce qu'ils ont un point commun ensuite chaque iut peut s'adapter .. C'est intéressant pour votre question parce qu'il dit que c'était extrêmement pauvre [...]. les compétences fondamentales ne sont pas différentes, après les logiciels évoluent, la pensée de travail liée aux logiciels va forcément évoluer […] aucun logiciel aujourd'hui n'est capable de, quand vous tapez Victor Hugo d'aller vous sortir la fiche data.bnf.fr, alors que théoriquement c'est possible, alors je dis aucun logiciel c'est faux, il y a la nouvelle génération des logiciels documentaires qui est en train d'arriver [..] ce sont des logiciels qui ont commencé déjà à travailler avec une logique du web sémantique, et qui sont directement branchés sur data.bnf et qui permettent de cataloguer un document physique ou un document numérique, aujourd'hui la plupart des logiciels qui sont présents dans les bibliothèques, les SIGB ne font pas ça ! Tant que les outils n'évoluent pas, ce n'est pas la peine de faire évoluer les compétences.
Dernier point qui n'est pas négligeable, si on dit web sémantique on dit web, et ça pour beaucoup de collègues c'est quand même la révolution, c'est-à-dire tout ce que vous produisez doit être disponible et utilisable sur le web par tout le monde. Pa contre c'est une autre pratique parce que si on demande à des collègues de décrire pour le web, là je peux dire que 10 % de mes collègues seront favorables à ça, les autres ne savent pas faire ou ils n'ont pas envie. Ils sont habitués à choisir les documents, en parler devant des lecteurs, faire des tables de présentation, ça leur va. Mais rédiger un article qui va être visible sur le web, c'est une autre dimension et pour l'instant ce n'est pas acquis, ça fait pas partie de la profession, ça s'apprend, il y en a qui ont été formés, il y en a qui refusent et puis il y en a beaucoup qui sont entre les deux.
Pensez-vous qu'il soit nécessaire de suivre une formation sur le web sémantique avant de l'intégrer en bibliothèque? Pourquoi ?
ça dépend si ça reste uniquement de l'infrastructure ce n'est pas nécessaire, mais par contre si on veut que les bibliothécaires travaillent dans le web sémantique, collaborent et produisent des données là oui bien sûr, je pense que c'est quelque chose qui ne se fera pas en bibliothèque publique, parce que ça arrive trop tard par rapport à un métier qui a tellement évolué, que la métadonnée est considérée comme acquise […] je pense qu'aujourd'hui c'est un peu ça c'est pas forcément le même discours dans un BU ou bien une bibliothèque de recherche qui sera très patrimoniale, parce que là on est fier de son exemplaire il faut le mettre en valeur c'est différent.
Comment voyez-vous l'avenir de votre profession ?
Alors moi je vais donner l'avenir de la profession de quelqu'un qui travaille dans la bibliothèque

numérique, la profession n'a pas mal évolué, déjà vers des choses qu'on aura pas imaginé, qui sont de l'ordre de la négociation de contrat,[..] mais dans la bibliothèque numérique il y a d'autres problématiques techniques, l'accès aux métadonnées, la qualité des métadonnées et la diffusion, et puis il y a tous les aspects statistiques, on en n'a pas parlé mais c'était déjà assez présent sur les documents physiques, mais sur les documents numériques le statistique est extrêmement important, sur les livres numériques on a des statistiques dans le temps réel mois par mois, c'est devenu très important pour nous, ce sont des choses qui ne sont pas du tout dans la profession, ils sont dans la profession mais catégorie A, il y a pas mal des cours et des formations sur les statistiques la production et l'utilisation des statistiques, mais pas dans les autres catégories B et C [..] alors pourquoi statistique, parce qu'il y a un aspect marketing qui est en train d'envahir la profession, il y a du bon et il y a du mauvais, ça veut dire qu'on doit être capable de gérer une campagne de communication, capable d'aller chercher le public, capable de mener des enquêtes de satisfaction, mais aussi des enquêtes sociologiques avec des personnels compétents, c'est pas nous que nous allons les mener et il faut être capable de faire appel au bon laboratoire [..]
On est un service culturel qui doit rendre ses comptes [..] c'est un service culturel qui est en concurrence avec d'autres avec le théâtre avec le mc2, alors que 10 ou 15 ans avant c'était acquis, il y a eu une période de forte construction des bibliothèques, les maires savaient que c'est bien pour la culture pour l'éducation etc., aujourd'hui ce n'est pas tout à fait comme ça, parce qu'il y a le numérique. Ah oui vous avez une super bibliothèque numérique, pourquoi vous avez encore autant de bâtiments ? C'est compliqué à expliquer c'est très compliqué.
Voilà on a parlé marketing, qu'est-ce qu'il y a encore, voilà en termes du métier la métadonnée est revendiquée comme le cœur du métier, je pense qu'il faut continuer un certain temps à cataloguer au moins la catégorie B, mais les outils du catalogage vont évoluer, les méthodes vont évoluer, c'est- à-dire qu'avant on récupérait des notices sur electre, peut être dans quelques années on les récupérait directement de databnf, qui eux-mêmes vont les récupérer de oclc, ça sans doute va évoluer, mais le problème va rester à savoir qu'il faudra être toujours capable de décrire correctement un document et de le rattacher à la bonne notice.
Ce qui évolue par contre et ça je pense c'est beaucoup plus flagrant en lecture publique qu'en BU, c'est la part des animations, la médiation, le service au public […] c'est une évolution qui se fait par le biais de toutes les animations qui sont organisées par les bibliothécaires, [..] ils font la médiation comme aucune BU l'en fait, c'est-à-dire ils sont tout le temps en train de proposer des choses, ils n'arrêtent pas, j'ai l'impression qu'ils réfugient derrière leur banque de prêt pour souffler, ils font beaucoup d'animations, beaucoup d'activités, la bibliothèque en train de devenir dans certains endroits vraiment un troisième lieu [..]

Entretien 2

Réalisé le 29 /09/ 2016

Durée 1h05

Cet entretien est organisé en 4 parties : première partie sur le profil de bibliothécaire, deuxième partie sur la gestion des collections numériques, troisième partie sur le web sémantique et la dernière sur les pratiques professionnelles

Pouvez-vous décrire votre cursus universitaire ?

Avant de faire le concours de conservateur de bibliothèque et de passer par l 'Enssib à Lyon, auparavant j'ai commencé par faire des études scientifiques et ensuite j'ai fait des études de philosophie, et donc voilà et après les études de philosophie j'ai passé le concours, et donc j'ai fait un cursus de conservateur à l'Enssib à Lyon, après ce cursus j'ai eu un poste à la bibliothèque nationale de France, ensuite j'ai enchaîné par un poste au ministère de la culture et maintenant je suis au service informatique de la bibliothèque interuniversitaire de santé.

Quelle est votre fonction ou bien titre à la bibliothèque ?

Je suis adjointe responsable du service informatique […] et dans la perspective de succéder quelqu'un qui part en retraite [..] j'ai été recruté parce que les bases de données que gèrent la bibliothèque sont des bases de données maison, et donc il fallait former quelqu'un pour pouvoir manipuler ces bases et de prendre succession de mon chef de service.

Quelles sont vos missions ?

Alors les missions sont de plusieurs ordres, c'est aider à la coordination de service informatique, que ce soit le suivi des commandes des matériels ou des logiciels, la validation des factures et puis c'est l'intervention au niveau de hardware, c'est aussi en lien avec le service informatique de l'université et aussi prendre contact avec les prestataires, que ce soit des prestataires du secteur informatique ou développement des bases de données et notre éditeur de SIGB voilà.

Est-ce que vous avez suivi des formations depuis que vous travaillez en bibliothèque ?

Oui j'ai suivi un certain nombre organisé par plusieurs établissements de l'Enssib, de l'ABES et dans le passé j'ai donné des cours de catalogage et puis de manipulation des métadonnées à des étudiants, soit en préparation de concours ou de DEUST métiers des bibliothèques et de la documentation, pour les apprendre les bases Marc et des autres formats dont on utilise pour le catalogage aujourd'hui Dublin core, EAD.

Qu'est-ce que c'est une bibliothèque numérique pour vous?

Très bonne question, au départ il y a les collections numérisées et les collections numériques natives et aussi il y a tous les services associés à ces collections numériques, c'est d'abord l'outil de mise en ligne, il y a notamment les outils de EAD, c'est vraiment la chose brute pour pouvoir présenter les documents sur le web, mais il y a aussi tout le site internet et il y a aussi tous les services autour de la dissémination des collections, comment on peut faire disséminer les contenus de la bibliothèque numérique, comment on peut les partager selon les fonctionnalités du web collaboratif.

Quels sont les langages documentaires utilisés dans votre bibliothèque pour décrire les collections numériques ?

Alors pour les collections numériques, il y a au départ un catalogue qui décrit les documents en Marc, le langage classique de bibliothèque, on a aussi OAI-PMH associé à notre bibliothèque

numérique et le format c'est Dublin Core, le Dublin Core bon c'est très souple il y a vraiment une possibilité de correspondance avec la possibilité de se faire moissonner très facilement, et on est le premier fournisseur de données de Gallica, on est le premier entrepôt moissonné par Gallica.

L'un des éléments essentiels dans une bibliothèque numérique est la production des métadonnées. Pouvez-vous me parler des métadonnées produites dans votre bibliothèque.

Alors nous produisons des métadonnées auprès de nos catalogues et puis nous les convertissons en Dublin Core pour les rendre visibles dans notre entrepôt OIA-PMH et notre bibliothèque numérique, ensuite au niveau de la production des métadonnées bibliographiques au sein des logiciels de données bibliographiques Sudoc voilà, on a aussi des métadonnées concernant une base de nos étudiants […] pour les exposer en RDF c'est dans un état de test en ce moment.

On passe à la troisième partie sur les applications du web sémantique en bibliothèque numérique. Quels sont les enjeux de l'utilisation du web sémantique en bibliothèques numériques? Qu'est-ce que le web sémantique apporte aux bibliothèques ?

Le web sémantique apporte avant tout une visibilité supplémentaire aux données des bibliothèques numériques, avant les bibliothèques ne parlaient pas les langages du web, il y avait déjà un décalage, les bibliothèques ont continué à parler en Marc alors que le web continue à parler des langages d'abord le html, et au niveau des protocoles d'échange on était toujours avec z39-50, alors que le web c'est http et fttp,[..] on a négocié le virage de XML un peu plus tard que tout le monde, c'est-à-dire il y a eu des initiatives pour transporter le langage Marc en langage XML, il y a eu le Marc XML et puis pour le protocole de communication d'échange des données bibliographiques, on est passé à quelque chose qui s'appelle SRU et SRW qui utilisent les technologies XML […] les bibliothèques numériques ont permis l'introduction de nouveaux formats de métadonnées dans les bibliothèques, Dublin core notamment, ensuite il y a les archives qui ont apporté beaucoup aux modes de description des bibliothèques avec donc la DTD EAD, donc vraiment le monde des archives a apporté beaucoup aux bibliothèques de ce point de vue-là, parce que les bibliothèques se sont mises à décrire leurs archives et leurs manuscrits avec ce schéma XML EAD, et il y a eu d'autres schémas de métadonnées et de DTD qui ont permis vraiment la multiplication de formats des métadonnées présentes dans les bibliothèques numériques, notamment les bibliothèques numériques ont posé la question de l'utilisation de leurs documents, de la conservation pérenne de l'information et donc il y a eu l'apparition du format ALTO pour créer des métadonnées pérennes et le schéma METS, c'est très important parce que ça permet de décrire intégralement un document numérique non seulement les métadonnées descriptives mais aussi les métadonnées administratives, c'est vraiment METS est un schéma de métadonnées pour les bibliothèques numériques, les bibliothèques numériques sont intéressées par ce schéma là parce que il y avait la question de l'archivage pérenne des données qui se posait.

Je dérive pour tout ce qui est du web sémantique, justement les bibliothèques ont compris très vite l'intérêt qui avait à ne pas louper, il s'agit cette fois ci de rendre visible leur données sur le web, pour les bibliothèques universitaires c'est vraiment un enjeu crucial [..] parce que si on n'expose pas nos données, les gens ne savent pas qu'elles existent, un fonds qui n'est pas signalé c'est un fonds qui est mort, il s'agit vraiment d'exposer nos données et de fournir des services aux étudiants et aux chercheurs.

Vous avez répondu dans le questionnaire que votre bibliothèque utilise les standards du web sémantique

Oui on est vraiment en train de les tester, on est dans une configuration un peu plus classique avec le Sudoc d'une part pour l'ensemble de collections, et pour la bibliothèque numérique les données sont en Dublin Core, mais on est en train de réfléchir à la potentialité d'exposition des données en Rdf, on est en train de tester ça sur un corpus réduit, un corpus iconographique

Est-ce que vous utilisez des identifiants des URI ?

Alors justement sur cette base on est en train de tester avec le numéro ISNI, c'est le chantier sur lequel on travaille, ISNI est maintenant le standard normalisé pour identifier d'une manière univoque les personnes comme les collections, et puis donc on a vraiment un identifiant unique pour chaque identité, à partir de ça on prend l'uri associé et ensuite à partir de là on peut associer des prédicats

La deuxième étape est de lier les données entre elles en utilisant le triplet RDF et vous avez commencé à lier les données

Tout à fait

Et qui fait ce travail

Donc on est en collaboration avec un autre agent du service informatique, il travaille à mi-temps donc pour monter cette exposition des données géographiques en RDF

C'est quoi l'étape prochaine

C'est comment on va pouvoir donc associer cette base de données biographique avec les documents, voir quel graphe mettre en place pour justement développer ça, et comment on va pouvoir associer les contenus entre eux en prenant cette base biographique comme point de départ. On va d'abord essayer de tester le plus des choses possibles avec ce que permet l'association de l'ISNi par rapport à notre base biographique, et puis ensuite on va voir comment établir les correspondances grâce notamment aux autorités auteur [..] il y a tout un problème d'arbitrage de sujet et on s'est dit est ce qu'on se fonde sur la manière dont fonctionne les référentiels de la bnf, est ce qu'on utilise d'autres référentiels, chez nous par exemple une bibliothèque spécialisée dans le domaine de la santé on a d'autres vocabulaires contrôlés on a le thesaurus MeSH qu'on utilise .

Vous pensez qu'il faut forcement se baser sur un vocabulaire contrôlé un thesaurus

Je ne dis pas forcement mais c'est quelque chose qui permet néanmoins de faire avancer certain nombre de choses, effectivement on s'est dit que avec l'arrivée de la recherche par texte c'est la fin de vocabulaire contrôlé, et même c'est la fin des catalogues, quand on fait une recherche dans un catalogue on va d'abord regarder la notice et ensuite on va aller vers le texte intégral, quand on fait une recherche par texte avec un moteur de recherche on tombe d'abord sur le texte intégral et puis on remonte vers la notice qui décrit le contenu intellectuel du document, donc c'est la démarche inverse [...]alors que pour la simple raison que la recherche par texte ne donne pas d'indices sur la manière dont les connaissances sont organisées et liées entre elles.

C'est un enjeu pour les bibliothèques de proposer des métadonnées de qualité justement, les données des bibliothèques sont des données qui ont été établies selon un consensus confiant rationnel, qui ont été vraiment sélectionnées et validées au niveau de leur qualité, et c'est là vraiment où les bibliothèques ont une carte à jouer, et d'autant plus l'intérêt de rendre visible leur métadonnées et là on voit l'efficacité du facteur humain face aux algorithmes.

Je passe à la dernière partie. C'est une question qui se pose tout le temps c'est l'influence de la technologie sur les pratiques professionnelles qu'est-ce que vous en pensez ?

Si on prend le métier de bibliothécaire sans l'informatique et bibliothécaire avec informatique ça rien à voir, c'est quand même arrivé relativement tôt les premiers langages, marc c'est un format qui date il y a 50 ans maintenant, après il y a eu la logique de réseau avec l'arrivée d'internet, parce qu'avant les gens faisaient leur notices bibliographiques dans leur coin et échanger assez peu leurs métadonnées bibliographiques. […] Après il s'agit toujours de décrire le document d'une manière normée avec du vocabulaire contrôlé

Quels types de compétences mobiliser en termes de pratiques documentaires en vue

d’accompagner l’utilisation du web sémantique en bibliothèques numériques?

Je pense que ce sont des compétences informatiques notamment au niveau du RDF, les technologies XML sont vraiment ce qu'il y a de plus indiqué pour pouvoir accompagner le web sémantique en bibliothèque.

Pensez-vous qu’il soit nécessaire de suivre une formation sur le web sémantique avant de l’intégrer en bibliothèque?

Ce que je vois assez souvent, les gens se forment beaucoup au niveau de la manipulation d'une part sur XML pour pouvoir cataloguer en Dublin Core et puis ensuite ils vont s'intéresser effectivement à RDF à d'autre chose par exemple aux ontologies ils vont s'intéresser à SKOS

Comment voyez-vous l'avenir de votre professionnel ?

On sera toujours aussi multi casquettes, on sera toujours autant orienté donnée qu'orienté service aux lecteurs, ça sera toujours une question de gestion de réservoir des métadonnées et d’exposition des données sur le web que de rendre service à des lecteurs, en tout cas effectivement les profils concernant le web sémantique, les profils techniques seront de plus en plus recherchés, parce que c'est vraiment quelque chose en quoi les bibliothèques sont profondément engagées, donc je pense que ça sera de moins en moins d'établissements traditionnels, de faire des données chacun dans son coin etc., on va avoir un développement du contrôle de la qualité des données, [..] le but des bibliothécaires sera de produire des données fiables, des données de confiance, qu'elles soient interopérables aussi et puis et qu'elles puissent être bien exposées, bien visibles et qu'elles puissent être aussi facilement réutilisables.

Nos données sont interopérables, elles sont réutilisables, elles sont stables, elles sont disponibles, elles sont visibles, et il y a aussi avec ça l'enjeu de statut juridique, il y a l'enjeu de la liberté de la réutilisation notamment les données bibliographiques.

Entretien 3

Réalisé le 06/10/2016 avec deux personnes

Durée 2h15

C'est entretien est organisé en 4 parties : première partie porte sur le profil du bibliothécaire ; 2eme partie sur la gestion des collections numériques, 3eme partie sur l'utilisation du web sémantique en bibliothèque numérique et la dernière partie sur les pratiques documentaires.
Je veux vous demander de décrire votre profil professionnel et votre cursus universitaire
E1) j'ai fait l'école de Chartes, j'ai fait 3 ans de prépa après un bac scientifique, à l'école de chartes j'ai pas eu trop l'occasion de faire du numérique,[..] j'ai consacré mon mémoire de recherche aux éditions françaises de Marx, après j'ai passé le concours de conservateur le concours de l'enssib […] quand je l'ai fait il y avait pas d'option pour le numérique, c'est pour cela j'ai fait un mémoire de conservateur consacré aux ouvrages politiques en bibliothèques publiques, et je suis arrivée à la bnf au premier juillet 2016. c'est assez récent j'ai passé des entretiens, c'était un poste qui m'intéressait le plus dans la liste qu'on nous a proposé, parce que je trouve que data.bnf est un très bel outil, c'est quelque chose que j'ai un peu utilisé, et je trouve que c'est vraiment génial, [..] je trouve que c'est vraiment quelque chose de très important pour toutes les raisons qu'on connaît avec le web, il faut référencer les choses, c'est d'autant plus important que contrairement à ce qu'on pense ce n'est pas parce qu'il y a le web qu'on a accès à tout, c'est grâce au référencement qu'on peut retrouver dans la masse ce qu'on cherche, et donc ici dans ce service dans ce département on est avec des gens qui sont des catalogueurs confirmés, des gens qui font du catalogage depuis 30 ans et c'est vraiment du travail du bibliothécaire très classique, et je trouve que c'est super de travailler sur des questions très techniques avec des gens qui ont une expérience de catalogage, qui est d'abord très diverse parce qu'ils ont connu le catalogue sur fiche, et après ils ont connu la conversion rétrospective du catalogue et maintenant ils voient ce qu'on fait avec data.bnf, ils ne participent pas forcement mais ils connaissent tout ce qui est automatisation des chantiers de correction, c'est-à-dire tout ce qui faisait à la main maintenant, il faut construire des outils pour le faire d'une façon plus automatisée et je trouve ça super voilà .
Et vous ?
E2) moi j'ai un parcours très similaire à celui de E1 sauf que j'ai un peu louvoyé, c'est-à-dire que déjà j'ai fait trois années de prépa de plus des années de fac, je suis passée aussi par la fac avant d'intégrer l'école de chartes. À l'école de chartes j'ai travaillé plus spécifiquement sur le mariage et le divorce au 17eme siècle pour lequel j'ai utilisé des outils numériques, j'ai surtout utilisé des bases de données pour faire des croisements statistiques, ce n'était pas la grosse partie de travail mais j'ai vraiment tenu à utiliser un peu des outils de traitement des données, j'ai pas eu l'enssib tout de suite, et ça m'a laissé le temps de faire un stage à la bibliothèque publique d'information centre Georges Pompidou, où là c'était super parce qu'ils m'ont mis un gigantesque entrepôt de données entre les mains, on m'a dit : tiens c'est du SQL, et donc j'ai appris un peu de SQL, et c'était absolument génial, donc c'est comme ça que j'ai commencé à travailler avec les données, et c'est comme ça que j'ai pu avoir aussi une idée du potentiel des données des bibliothèques, des données bibliographiques notamment. Une fois intégrer à l'enssib, j'ai donc décidé de faire un mémoire là-dessus, donc j'ai fait un mémoire qui spécifiquement porte sur les données des bibliothèques, qu'est-ce que ça apporte aux bibliothèques et dans quelles conditions elles peuvent être vraiment utiles pour le métier de bibliothécaire, voilà je suis passée par un super stage aux états unies à Boston où là en fait ils ont mis en place une formation de data scientist pour les bibliothécaires, donc là-bas ils forment les bibliothécaires au traitement de données parce que de plus en plus dans des

bibliothèques scientifiques ils sont amenés à faire ce qu'ils appellent data curation, où leur rôle est de mettre à disposition de chercheurs des jeux de données qui sont issus de la recherche, donc ils accompagnent les chercheurs pour déposer sur les plateformes, surtout pour créer des métadonnées autour des données, pour pouvoir créer de contenu, donc voilà ça était une super expérience aussi et en sortant de l 'enssib, il y a eu une miracle il y avait un poste qui correspondait complètement à ce que j'avais fait et c'était data.bnf. J'ai commencé en juillet 2015 à la bnf, on est toutes les deux très récentes quand même globalement dans le domaine et le web sémantique à proprement parler moi je le découvert surtout ici, ce qui m'intéressait surtout c'était le traitement de données à proprement parler et sans forcément la publication et la diffusion sur le web.

Vous avez suivi des formations depuis que vous travaillez en bibliothèque à la bnf ?

E2): sur le web sémantique il y a des formations qui sont faites sur le tas, en fait quand on arrive en poste on a des gens qui sont là depuis longtemps, qui nous prennent une à deux heures et qui nous expliquent voilà le modèle de données voilà le rdf ils nous amènent dans les formations qu'ils font à l'extérieur, aussi il y a eu pas mal de formation à l'extérieur, donc dans ces formations-là souvent à la fois on participe et à la fois c'est l'occasion aussi pour nous d'approfondir tout ce qu'on a pas encore su, et ensuite il y a vraiment l'apprentissage par soi-même, c'est-à-dire c'est comme le cas de SQL, on tape quelques exercices à faire, puis voilà au fur et à mesure on bricole, c'est vraiment les moments où on a approfondi le plus c'est souvent quand on doit rendre des services à des gens ...

Quel est votre titre ou fonction à la bnf ?

E2) Je suis chef de projet, chef responsable de projet plus exactement data.bnf et E1 est mon adjointe

Quelles sont vos missions ?

E2) ça relève beaucoup plus de la gestion de projet à proprement parler, c'est-à-dire que mon rôle c'est un peu d'être un chef d'orchestre, c'est-à-dire que je coordonne les activités des développeurs, les activités des personnes qui intègrent les développements dans l'application data.bnf et surtout je parle à nos responsables hiérarchiques sur le développement de projet, aux personnes qui sont actuellement à la bnf qui vont avoir l'œil aussi sur le projet, et j'ai aussi une main ça tient à ma personnalité, mais j'ai la main aussi un peu sur le développement de l'interface à proprement parler, sur tout ce qui est graphisme et ergonomie, sachant que pour l'instant c'est dans les cartons mais il y a un suivi presque personnel je dirais là-dessus, parce que voilà pour des raisons de pratique de développement en fait c'est plus simple si je me colle directement.

E1) Alors moi j'ai une mission particulière de médiation, de réflexion comment on va organiser une médiation autour des données qui sont disponibles dans data.bnf

E2) oui puis il y a toute la question de la communication, communication à l'extérieur avec des partenaires possibles

Je passe à la 2eme partie sur la gestion des collections numériques dans ma thèse je m'intéresse à la notion de bibliothèque numérique, qu'est-ce que c'est une bibliothèque numérique comment la définir ? Qu'est-ce que c'est pour vous une bibliothèque numérique ?

E2) Bonne question

E1) moi je ne sais pas ce que c'est une bibliothèque numérique. Moi je regrouperais sous ce terme toutes les traces de la présence des bibliothèques de façon numérique, mais c'est une question que je trouve difficile, oui ce n'est pas assez facile que ça. Après quand on s'intéresse à faire la médiation, c'est vrai qu'il faut s'intéresser à la façon dont les gens se représentent la bibliothèque numérique. Je pense que dans l'esprit des usagers, c'est vraiment Gallica c'est-à-dire avoir accès aux documents numérisés. Je pense que dans l'esprit commun c'est ça après je pense qu'ils ne se rendent pas compte de l'ingénierie qui a derrière pour accéder à ces documents numérisés, et c'est finalement beaucoup plus large que ça.

Oui une bibliothèque numérique ce n'est pas juste mettre des documents numérisés en ligne il y a un travail documentaire derrière

E1) oui je pense que les usagers ne font pas la différence. Je pense même qu'il y a plusieurs façon

de mettre en ligne un document ça peut être un simple pdf ça peut être un ocr c'est précis, je pense que les gens voient les services qu'on les rend, mais je ne suis pas sûre qu'ils se rendent compte techniquement de ce que ça peut représenter et de la différence qu'il y a entre un document numérisé en interne à la bnf et un ebook disponible par l'intermédiaire du prêt de telle bibliothèque[...].
E2) pour moi la bibliothèque numérique ça recouvre beaucoup la notion d'accès à distance, [..] pour moi c'est paradoxale c'est-à-dire qu'il y a d'une part la bibliothèque numérique et il y a d'autre part la documentation électronique, pour autant ce sont des objets qui ne devraient pas être séparés parce qu'ils ont un format qui se regroupe énormément, même si finalement ce qui caractérisera la documentation électronique que c'est un document électronique natif contrairement à la bibliothèque numérique où le document était numérisé, mais en même temps est ce que la numérisation du document par un autre document, est ce que de ce point de vue là on peut pas dire que c'est un document électronique natif ? la question se pose de fait qu'il y a un problème de signalement qui ne se pose pas de la même manière entre la documentation électronique et les bibliothèques numériques, dans le sens où quand on numérise, on a pas forcement en tête les métadonnées, quand on produit un document nativement numérique au contraire souvent il est structuré, donc il y a une possibilité de produire des métadonnées qui est quand même beaucoup plus facile, et tout l'enjeu aujourd'hui c'est d'arriver à transformer un gigantesque dépôt de version numérisés de documents en quelque chose qui soit assez accessible que la documentation électronique, ça c'est le grand défi qui se pose à nous, et le web sémantique peut être une solution pour ça, et je pense que c'est très lié mais ça fait pas tout il y a aussi toute la part de la médiation, ici à la bnf je crois qu'il y a une équipe de trois personnes qui est entièrement consacrée à la médiation autour de documents de Gallica sur le web et sur les réseaux sociaux. […] Ce qu'il y a aussi derrière c'est la notion de conservation c'est-à-dire par exemple tout ce qui est périodique de début de 20e siècle sont en train de tomber dans le domaine public, aujourd'hui c'est vrai que les bibliothèques numériques permettent finalement la consultation de ces documents, encore il faut les trouver c'est toujours le même problème. Pour autant ce sont des documents formatés et structurés autant que la documentation électronique, sauf que quand ils ont été numérisés ce n'est pas forcement ce qu'on a eu en tête, je pense au départ c'est souvent une numérisation au km, qui est plus une numérisation de prévention, qui doit permettre de fournir un autre support qui n'est pas pensé dans la communication et il est pensé du côté de la conservation, alors que à la base moi ce que j'ai en tête, une bibliothèque numérique c'est quand même l'accès distant c'est la communication mais en fait on s'aperçoit que c'est un domaine qui est complètement déchiré entre cet aspect de conservation et cet aspect de communication
En fait dans les bibliothèques municipales j'ai l'impression que le rôle principal des bibliothécaires c'est la médiation que ce soit à travers les animations les activités etc.
E2) Alors le catalogage n'est pas pensé comme une médiation. alors que pour nous en fait on est de plus en plus emmené à le penser comme une médiation, surtout dans un contexte des bibliothèques numériques, [..] nous par exemple notre expérience c'est qu'on diffuse les données qui ont été faites par des bibliothécaires pour des bibliothèques, donc quand on diffuse sur le web disons qu'on donne aux données une portée qu'elles n'avaient pas auparavant, une portée notamment d'ouverture qui n'est pas inhérente à la dimension d'initiation propre au monde des bibliothèques, il faut passer par l'initiation, il y a les codes professionnels à intégrer etc., sauf que pour pouvoir comprendre les données il faut avoir ces codes, et c'est un peu la raison pour laquelle E1 est là, il y a une médiation à faire autour des données, il y a une médiation à faire avec les gens qui réutilisent les données, c'est-à-dire pour pouvoir les réutiliser correctement il faut avoir conscience que ces données ont un caractère archivistique, quasiment celles qui ont été produites dans le cadre d'une activité, elles ne prennent sens que du point de vue de cette activité [..]
Sur la question de la bibliothèque numérique vous voulez ajouter quelque chose
E2) on pose beaucoup de questions par rapport à comment exploiter le potentiel des textes

numérisés pour enrichir les données, et comment enrichir les textes numérisés avec nos données aussi, [..] comment faire communiquer ces deux dimensions qui auraient dû communiquer dès le début, c'est un défi qui se pose dans notre organisation interne, et de fait que le web sémantique est quelque chose qui permet de dessiloter les bases de données qui sont distantes, et nous on s'aperçoit qu'il y a aussi un travail de dessilotage qui doit pouvoir se faire au sein même de l'institution, par exemple dans une bibliothèque territoriale, la bdp ne va pas avoir forcement son service informatique dans l'organisation, donc il faut qu'elle puisse se communiquer avec ce service informatique [...]

On va parler des langages documentaires et des formats utilisés pour décrire les collections numériques

E1 : à la bnf on utilise comme format de description documentaire l'intermarc voilà, ça c'est traité dans les applications du catalogage classique, mais le problème ça ne permet pas de décrire tous les documents qui sont conservés à la bnf, parce qu'il y a quand même une variété de documents conservée ici, donc pour décrire les archives et les manuscrits on utilise le EAD, ça a des impacts sur l'organisation des bases de données ici, puisque ce sont des choses qui existent dans deux bases de données différentes, après il y a une base de données pour les médailles antiques, ce sont plutôt les objets et les monnaie qui sont décrites en intermarc [..] quand c'est une grande bibliothèque comme la bnf et avec un fonds assez varié, [..] on ce qui nous concerne il y a une sorte de mapping qui existe entre le format intermarc et le modèle rdf.

E2 : Même à la bnf on ne voit pas de la même manière rdf et le web sémantique, [..] moi par exemple j'observe ça plus d'un point de vue sciences humaines, c'est-à-dire que je le regarde vraiment du côté des humanités numériques [..]

Un des éléments les plus essentiels dans les bibliothèques numériques est la production des métadonnées, vous pouvez me parler des métadonnées produites à la bnf

E2:[..] gallica agrège les documents des bibliothèques partenaires avec le protocole OAI PMH, qui est un peu ce qu'on utilisait avant d'utiliser rdf et d'agréger autour des entités, donc OAI PMH permettait d'agréger certaines métadonnées qui parlaient déjà entre elles [..]

les métadonnées de gallica, c'est surtout du dublin core, c'est un format qui est assez pauvre par rapport à intermarc, c'est un grand souci aussi, c'est que pour l'instant les métadonnées de gallica sont dans un format assez pauvre, et maintenant ça pose ici la question de bibliothèque numérique à un niveau supérieur qui est Europeana, Europeana agrège, mais pas de la même manière, Europeana agrège des données et les documents de bibliothèques partenaires avec un modèle qui s'appelle EDM donc qui est European Data Model […] il y a plein de processus différents et il y a une évolution, il y a plusieurs modèles d'agrégation [...] la question pour nous c'est comment utiliser le texte même de gallica c'est-à-dire l'OCR, comment est-ce qu'on peut arriver en faire des métadonnées, sachant qu'il y a déjà des techniques qui sont déjà utilisées par les moteurs de recherche, il y a des circuits qui sont rajoutés aux moteurs de recherche pour les rendre performant, qui utilisent du sémantique etc., comment est-ce que maintenant on peut utiliser cette richesse pour enrichir nos propres données, c'est une grande question aussi

Vous avez des pistes sur ce point là

E2 : on est vraiment en cours de réflexion, pour l'instant ce n'est pas du tout abouti

Je passe à la 3ème partie sur le web sémantique vous préférez utiliser le terme web sémantique ou web de données

E2: j'emploie beaucoup le web sémantique bizarrement, mais c'est justement parce que je le regarde du côté sciences humaines

E1 : quand je suis arrivée en juillet je devais traduire un texte pour l'Ifla, donc c'est un texte qui était écrit entre autre par Emmanuelle Bermès en français, c'était un très grand problème, comment je traduisais web de données et web sémantique, j'ai eu vraiment des soucis […] surtout que je ne comprenais pas la différence, j'ai un peu cherché à la comprendre, [..] je pense que c'est quelque chose au fond désigne la même chose, voilà moi je connais pas la différence

E2: en fait ça fait peur comme expression, c'est-à-dire dans un article sur le web sémantique j'emploie le web sémantique comme expression, mais dans une formation j'emploie plutôt web de données ou linked open data parce que sémantique c'est trompeur, c'est-à-dire on a l'impression qu'on fait l'intelligence artificielle et que c'est du linguistique, pour moi il y a un fondement linguistique, un fondement historique, et c'est pour ça quand j'écris un article sur le web sémantique j'emploie cette expression, parce qu'elle regroupe à mon avis une tradition scientifique qui est sémantique, qui est une tradition du signe scientifique, et de ce point de vue là je trouve que le web sémantique est une bonne expression, mais pour les formations pour expliquer qu'est e que c'est je préfère le terme de linked open data, parce que c'est moins trompeur et dans linked open data on a la notion d'ouverture, donc de vocabulaire contrôlé, et le linked, c'est-à-dire identifiant web donc identifiant url donc c'est plus simple à mon avis de partir de là pour expliquer aujourd'hui il n'y a pas de différence sémantique, mais il y a une différence vraiment de point de vue du traitement de l'objet.

Quels sont les enjeux de l'utilisation du web sémantique en bibliothèque numérique en général et à la bnf en particulier ?

E2: il y a un enjeu interne pour employer le web sémantique, c'est qu'on voulait dessiloter les applications, on voulait avoir un modèle qui soit organisé autour d'entités, et qui puisse faire communiquer les bases de données entre elles, et quant à faire puisque ces bases de données devaient communiquer entre elles, ça était bien qu'elles puissent communiquer avec l'extérieur, donc le format linked open data était particulièrement indiqué, parce que ça permet de faire communiquer les bases des données en interne et les ouvrir aussi à l'extérieur […] donc pour cela le web sémantique était particulièrement indiqué, même si ce n'a pas été forcement la première chose à laquelle on aurait pensé, un collègue me dit souvent on aurait pu prendre une autre technologie on aurait pu faire autrement, on a fait comme ça pour des raisons historiques, aussi parce que il y avait tout le mouvement de open data en France, qui commençait à démarrer, on a la naissance de datalab en 2014, et donc voilà l'application remplissait plusieurs objectifs différents qui étaient des objectifs internes et externes, c'est surtout pour ça que le web sémantique a été employé, [..] on s'est aperçu que les gens qui étaient intéressés par les données en rdf n'étaient pas forcement des gens du monde du livre et des bibliothèques, c'était plutôt des gens de laboratoires extérieurs comme l'institut national de géographie, comme des gens qui s'intéressaient à la culture de façon amateur, on a un site qui réutilise nos données qui est « libre théâtre », c'est un site de publication de texte de théâtre amateur dans le domaine public, on a eu éventuellement le moteur de recherche Isidore. c 'est vraiment des laboratoires scientifiques qui sont particulièrement intéressés par ce format, ce qui ne m'étonne pas, parce que comme je disais à la base c'est un format qui vient vraiment à mon avis du scientifique, c'est-à-dire que Tim Berners Lee, c'est quelqu'un qui était au Cern et à mon avis qu'il l'a pensé vraiment de point de vue de scientifique, et finalement assez peu de développeurs qui s'y sont intéressés [...] et paradoxalement ce n'est pas forcement quelque chose qui parle beaucoup à l'informatique et aux développeurs, ce qui n'était pas forcement ce que j'aurais imaginé au départ, je pense que les développeurs sont plus intéressés par des applications très classiques, ils ne sont pas forcément intéressés par ce format [..]

Il n'y a pas de bibliothèques qui réutilisent vos données

E2: les bibliothèques qui sont intéressées par nos données sont surtout intéressées par le fait qu'elles sont frbrisées, c'est-à-dire qu'elles sont organisées autour d'entités, mais le format en tant que tel il rebute un peu, c'est compréhensible parce qu'il faut quand même se coltiner un langage, se farcir le sparql, se farcir un peu la syntaxe rdf etc., donc ça ne parle pas forcément aux professionnels, mais les professionnels sont intéressés par l'évolution de catalogue, ça fait quand même partie de l'évolution de catalogue, c'est très lié parce que souvent la modélisation rdf se fait autour d'entités, et de fait l'évolution de catalogue va vraiment vers ce modèle d'entités frbrisation surtout qui est un modèle d'entités

Certaines bibliothèques comptent sur vous sur la bnf pour récupérer des données des

identifiants du rdf
E2: ça ne sera pas forcement du web sémantique, ça ne sera pas forcement du rdf, au contraire la tradition de diffusion des données à la bnf ne date pas du web de données, et globalement ce qui intéressant c'est le format qui soit vraiment exhaustif, d'une certaine manière et surtout le problème c'est que nos données sont vraiment marquées historiquement par des strates de catalogage, des manières de cataloguer, par exemple nos données leur grand défaut qu'elles ne sont pas homogènes, [...] il y a un problème de fraîcheur, c'est-à-dire les données qu'on diffuse ne sont pas mises à jour d'une manière régulière, rien que le catalogue il faut une semaine pour le mettre à jour, et une semaine d'un point de vue économique c'est beaucoup, [..] en tout cas le web de données n' intéresse pas immédiatement le réutilisateurs coté bibliothèques et livres, mais la frbrisation des données et la diffusion traditionnelle par fttp ça c'est très utilisé [..]
Vous avez parlé des enjeux du web sémantique à la bnf et si on va parler des enjeux en générale
E1 : il y a quand même un enjeu de signalement, nous en tant que responsables de databnf, on se demande quel est l'intérêt du rdf, qu'est-ce qu'il fait en plus par rapport aux autres, par rapport justement au fait que finalement c'est pas le rdf qui parle le plus, il y a d'autres choses plus traditionnelles qui continuent de fonctionner, qui vont être mises en place, je pense que le signalement, le fait que ça remonte dans les moteurs de recherche, [..] ça répond à un grand enjeu de visibilité des données des documents, ça c'est le premier enjeu, après on peut faire aussi de la visualisation des données, [..] ce sont des choses qu'on va explorer, il y a des questions que se posent, qu'est-ce qu'on peut faire avec le rdf qu'on ne peut pas faire avec d'autres choses, il y a un moment qu'on pose ces questions, je pense que le fait que databnf remonte dans les résultats Google, rien que ça c'est un enjeu important
Pas tout le temps j'ai un peu fait quelque recherche notamment pour voir si les résultats databnf vont apparaître, ce n'est pas toujours le cas
E1 : c'est mieux que zéro avant on était à zéro on arrive rarement sur un contenu de la bnf
Est-ce que c'est la seule solution ?
E2 : pour parler aux moteurs de recherche, ce n'est pas la seule solution, [..] grâce au fait qu'on est organisé autour d'entités et aujourd'hui la recherche dans les moteurs de recherche se fait quand même beaucoup par entité, [..] les moteurs de recherche ne vont pas voir le rdf, ils vont regarder le html et l'organisation de la page surtout ce qui fait qu'on remonte c'est surtout qu'on a du contenu rare, c'est-à-dire qu'on a de contenu même Wikipédia n'a pas, donc on a cette exhaustivité qui vient au fait qu'on est le dépôt légal, [..] c'est vraiment ça qui permet de faire remonter, c'est-à-dire que le moteur de recherche n'a pas à interroger un autre moteur interne, il va interroger des pages publiées et donc ça c'est très différent, mais le rdf en tant que tel, ce n'est pas forcement ce que va voir le moteur, en revanche le fait d'avoir organisé et modélisé les données selon un modèle d'entités qui est plutôt un modèle rdf de graphe, oui cette organisation-là parle au moteur de recherche, du moins elle parle aux usagers sur internet parce que les usagers ont tendance à chercher par titre par auteur etc, donc de ce point vue-là la modélisation rdf aide beaucoup à faire remonter, même si ce n'est pas forcement la seule solution, il y a aussi d'autres formats qui utilisent le modèle d'entité. Je crois que le véritable avantage de mon point de vue c'est que le web sémantique permet d'ouvrir la communauté des bibliothèques sur d'autres communautés, c'est-à-dire faire parler des communautés entre elles, qui ne parlaient pas auparavant, [..] c'est vraiment ça qui est super c'est-à-dire on est amené à rencontrer des gens qui sont complètement en dehors de notre domaine.
Selon les résultats de mon questionnaire sur l'utilisation du web sémantique en France, il y a peu de bibliothèques qui l'utilisent, et c'est que à la bnf que c'est bien développé, pourquoi ?
E1 : oui c'est le fait que la bnf est si grande et il y a quand même des ressources, il y a le question de budget, des moyens humains, de stratégie, parce que c'est un établissement qui se repilote sur des projets innovants, et ici on peut essayer de choses [..]
Est ce qu'il y a aussi une question géoculturelle derrière

E2 : je pense qu'il y a une question historique et culturelle derrière absolument, à mon avis il faut bien voir une chose, c'est qu'un de premier catalogue en France c'était un catalogue qui a été fait par le bibliothécaire Clément au 17 siècle, ce bibliothécaire il a institué ce qu'on appelle la cotation Clément ,or qu'est-ce que c'est que la cotation Clément, c'est un petit code qui permet à la fois d'identifier et de localiser dans un ensemble scientifique de description de collection un document, et c'est spécifiquement ce que se fait un uri aujourd'hui sur le web, je pense qu'il y a quand même deux domaines qui se parlent complètement cette tradition de catalogage qu'on a historiquement et le langage qui utilise sur le web, qui est un langage finalement scientifique et de toute façon le bibliothécaire est un scientifique, c'est un scientifique de collection mais qui est aussi un bibliothécaire en tant que tel dans cette volonté de description et de décrire et de localiser

Le rôle du bibliothécaire est toujours d'identifier de décrire de localiser un document mais ce qui change c'est la technique utilisée pour faire ça

E2 : exactement la technique et la portée. oui absolument le web en fait, il ressuscite les pratiques de la profession en tout cas la profession telle qu'elle a été pensée au début de l'époque moderne, à mon avis le rdf en tout cas le web sémantique, c'est un langage qui sort tout droit de la tradition scientifique de 17eme siècle [..]

Est-ce que le web sémantique pour vous est une technique une évolution c'est une révolution c'est quoi exactement

E2 : c'est une révolution au sens étymologique du terme, c'est-à-dire qu'on revient à quelque chose, on revient avec du neuf, on revient à ce qu'on faisait au départ

[..]

Vous avez répondu que la bnf utilise le web sémantique, vous utilisez quoi comme identifiant ?

E2: on utilise Arc, Arc est un identifiant qui a été produit par une bibliothèque californienne américaine, et donc ce sont des serveurs géants sur lesquels les bibliothèques sont amenées à se créer un compte, en quelque sorte à se créer un identifiant, donc il y a cet identifiant ARC et il y a l'identifiant de la bnf qui en ce qui nous concerne c'est 1248, et à partir de là il y a l'identifiant des différents documents qui sont dans les différentes applications de la bnf, ça se situe vraiment par couches successives.

Il faut créer un compte

E1 : ce n'est pas un compte c'est un enregistrement

E2 : on enregistre et on obtient un identifiant

Est-ce que les autres bibliothèques en France seraient obligées à utiliser ses identifiants ?

E2 : on peut faire du rdf sans utiliser arc, par contre c'est ce qu'il faut pour pouvoir faire du rdf d'avoir un identifiant pérenne

Il faut aussi que cet identifiant soit plus en moins homogène et unique ?

E2 : ça c'est le travail de viaf, on a pas parlé de viaf encore, alors de toute façon il faut bien voir que les entités similaires d'une institution à une autre ont toujours des identifiants différents, de toute façon l'identifiant de Victor Hugo ici n'a pas le même identifiant à la bibliothèque nationale d'Espagne, utiliser arc n'est pas une obligation, il y a des gens qui prennent DOI il y a plein d'autres systèmes d'identification, ce qu'il faut c'est qu'il y a une centralisation d'identifiants, nous on a voulu créer vraiment une gestion centralisée des identifiants, tout simplement pour assurer la pérennité des identifiants.

E1: oui

E2 : ce qui est intéressant avec le rdf, ça permet de reconstituer l'identifiant, [..] parce que justement c'est un format de lien, c'est un format qui permet d'exprimer ce lien d'identité entre des identifiants qui sont différents, c'est « same as » qui permet de définir ça. ce qui est intéressant ici c'est qu'on peut introduire des nuances, [..] par exemple exprimer « psychologie d'adolescent» en français, la manière dont on voit l'objet n'est pas pareil que les allemands, eux ont une tradition autour de la psychologie d'adolescent qui n'est pas la même, donc du coup ils ont plusieurs termes

pour définir ça, et alors que nous on a qu'un, et voilà c'est une casse-tête, mais ce sont des choses très intéressantes, du côté métier c'est vraiment génial, plus que jamais ces formats-là font appel au métier des bibliothécaires aux catalogueurs, parce qu'il faut quelqu'un qui a cette vision de métier pour dire bon c'est exactement la même chose, c'est pas tout à fait la même chose, c'est un peu la même chose mais pas complètement, et donc ça il faut pouvoir l'exprimer dans le format, il y a une souplesse du format, c'est pour ça que le rdf est quand même bien pour ça

Vous avez dit qu'on peut faire du rdf sans identifiant

E2 : il faut un identifiant pérenne, mais ce n'est pas forcément le même type d'identifiant [..] c'est ce que Foucault appelle la dérivation, c'est-à-dire le fait qu'un nom ne va pas forcement designer toujours la même chose, viaf en fait c'est une plate-forme gérée par le oclc, qui permet justement de réconcilier les identités des auteurs d'un pays à un autre, par exemple Israël va appeler Jean Jaurès ou un truc comme ça, alors que la bibliothèque la dnb va l'appeler Jean Auguste Joseph Jaurès, quelle version est la bonne ? est-ce qu'un pays qui a l'exhaustivité du domaine français a plus d'autorité sur le domaine français, j'avais posé la question, un collègue m'a dit que non pas forcement, il y a des pays qui sont plus connaisseurs que nous sur certains auteurs, c'est intéressant mais du coup viaf ça permet de centraliser tous les identifiants et de les réconcilier entre eux.

Et pour la gestion centralisée des identifiants

E2 : alors dans les cartons il y a un projet de fichiers nationales d'autorités, qui effectivement vise à gérer des autorités produites dans un même pays par deux agences bibliographiques différentes, l'idée est d'arriver à produire d'une manière unifiée ces autorités et ces identifiants, je sais pas si le projet devra servir à la gestion d'identifiants sans doute ça sera une partie de projet [..]

Je pense que le défi c'est vraiment d'avoir des identifiants pérennes et uniques

E2: oui ce n'est pas insoluble, c'est-à-dire qu'il y a des solutions qui existent, et puis on a choisi arc pour des raisons aussi très institutionnelles, je pense que les bibliothèques municipales et les bibliothèques universitaires ont l'habitude de travailler avec des identifiants, les bu travaillent avec orchid par exemple, elles travaillent avec DOI, elles ont quand même cette tradition d'identifiant, elles ont aussi une ouverture sur le web sémantique par leur caractère scientifique, c'est-à-dire qu'elles sont souvent amenées à utiliser des vocabulaires dans les différents domaines qu'elles documentent [..] ce qui est sûr, c'est qu'il faut arriver à penser numérique d'un point de vue pérennité, et donc penser numérique sur le web ça implique de penser identifiant ça c'est sûr

Pour lier les données entre elles vous vous appuyez sur quoi ?

E2 : rien que sur la modélisation, rien que le mapping entre les formats ça demande un travail intellectuel, ça c'est sûr, là on a deux équipes ici qui ont participé à cette modélisation, en fait deux modélisateurs des gens qui ont fourni le travail intellectuel de modélisation et de mapping dans le rdf ça c'est sûr

Il y a un projet pour mettre rameau en RDF

E2 : rameau est déjà en SKOS tout rameau est publié en skos. c'est un vocabulaire contrôlé et en même temps joue le rôle d'ontologie, il définit des domaines. Il est publié dans databnf en skos, mais il contenue d'être produit en marc en intermarc, le format de base de production de rameau reste l'intermarc, et à priori s'est appelé à rester comme ça, et de fait on a pas dans l'idée qu'il faut reproduire nativement du rdf, c'est-à-dire on continuera à produire dans des formats d'autre que rdf, et il y aura toujours ce mapping qui sera fait entre ces formats de productions et les formats de diffusion, ce qui est sûr c'est que la production et la diffusion d'un point de vue format, elle sera toujours distincte en tout cas dans notre politique .

Comment vous créez les liens ?

E2 : c'est un travail intellectuel, c'est-à-dire que c'est le catalogueur qui va faire le lien avec un auteur, qui va faire le lien avec une vedette rameau etc., le lien est fait dans la production, et même de plus en plus on commence à se demander si on ne pourrait pas leur demander de lier à des urls, et non pas lier à un mot qui existe dans un référentiel ailleurs, d'aller chercher l'url qui n'est pas

forcément évident d'un point de vue de pratique de catalogage, c'est plus contraignant, [..] par ailleurs il faut savoir que nous quand on a commencé à frbriser le catalogue, c'est- à-dire à lier des auteurs à des œuvres ou des œuvres à des éditions, il y avait quand même une tradition de lier les identifiants dans leur production, c'est-à-dire que dans marc il y avait une zone qui permettait d'insérer l'identifiant de l'auteur lors du catalogage, [..] je crois aussi qu'au départ il y avait une zone dans le format marc qui permettait aussi de lier une autorité de titre, donc une œuvre à des éditions [..]A l'origine il y avait cette possibilité d'insérer des identifiants dans les zones du marc et de lier les entités entre elles.

Pouvez-vous parler de l'interopérabilité suite à l'utilisation du web sémantique

E1 : c'est un enjeu interne important, c'est un format qui permet l'interopérabilité de nos propres bases produites ici à la bnf, et ça permet d'articuler avec des bases documentaires dans d'autres institutions, et ça permet de se greffer sur le web, c'est une notion essentielle, c'est à la base du projet, [..] , je pense que c'est un enjeu emportant du web, l'interopérabilité c'est au fondement du web, c'est-à-dire c'était une question très importante dès le départ, c'est pour ça que c'est une des recommandations de w3c

E2 : il y a quelque chose d'intéressant avec ce modèle d'entités,[..] par exemple pour décrire les lieux on a deux logiques différentes, on a la logique carte et plan que le département des cartes et plans utilise, et là il y a une vision plus scientifique, plus science exacte, et on a la manière rameau de décrire le lieu, rameau décrit plutôt d'un point de vue sciences humaines, [..] il y a une couche historique politique dans rameau, elle n'est pas dans geo, [..] etc ce sont deux manières de décrire pendant des années n'ont jamais pu parler à la bnf,[..] c'est quand même ridicule d'avoir deux manières différentes de décrire les choses, et pas pouvoir les réconcilier, et en fait databnf ça permet de réconcilier au sens où ils commencent à se parler, mais c'est-à-dire que l'entité va permettre d'agréger sans forcément fusionner les deux, le fait de les voir comme ça agréger autour des entités sans avoir alors demander de fusionner ces deux approches, c'est-à-dire d'essayer de faire un compromis entre deux manières de voir les choses différentes, [..]

Quels sont les défis à utiliser le web sémantique ?

E2 : c'est parler aux réutilisateurs, je trouve c'est un grand défi, ce qui n'est pas évident avec le web sémantique c'est qu'on sait qu'on a des réutilisateurs, mais on ne sait pas qui, parce qu'on n'a pas de retour par rapport à ça. on a des retours que quand ça ne marche pas, alors du coup on est content que ça ne marche pas, généralement l'idée c'est qu'un utilisateur de moteur de recherche devrait pouvoir arriver aux ressources de la bnf sans connaître la bnf, de même il faut que le réutilisteur puisse récupérer des données sans avoir jamais parler à quelqu'un d'une manière fluide, mais du coup on sait pas vraiment si nos données ont convaincu ou pas, [..] la dessus on essaye de construire des espaces de dialogue

Je pense que là il y a une médiation à faire pour la réutilisation des données

E1 : oui tout à fait, mais le problème c'est qu'on a un public très large

E2 : très large et très spécifique en même temps

E1 : c'est-à-dire on va avoir des gens qui vont arriver sur databnf juste parce qu'ils sont passés par Google, et donc qui n'ont pas conscience qui sont sur le site de la bnf, et on va avoir de gens qui vont être intéressés par les données, [..] on pourrait être plus didactique sur le fonctionnement de sparql, pour expliquer comment faire des requêtes simples et trouver des choses assez précises, ça pourrait être intéressant pour des projets de recherches en master, on pourrait montrer comment utiliser des outils de visualisation, pour une visualisation simple des lieux, il y a des choses à trouver de ce point de vue-là, on s'adresse vraiment à des publics très différents, c'est pour ça qu'il faut mettre en place la médiation, mais il y a plusieurs médiations à mettre en place, on s'adresse en même temps à tout le monde même si on cible plutôt certain public

Est-ce que vous avez des statistiques sur les usages de databnf, est ce que le fait de mettre les documents ici a changé ou pas l'usage des documents de la bnf ?

E2 : déjà le fait d'avoir l'exhaustivité de contenu de catalogue ça augmentait considérablement le

trafic sur databnf, parce que justement on va publier du contenu rare, du coup on arrive à capter l'internaute en fait grâce à ces contenus rares, databnf a pris progressivement une part de consultation par rapport au total des applications de la bnf, maintenant je crois que c'est le quart de la consultation des applications de la bnf, c'est quand même gigantesque pour un petit projet alors que le catalogue est là depuis un certain temps, on a bien grandi de ce point de vue là on a en gros 12 à 13 milles visiteurs par jour en moyenne sur une année , c'est quand même pas mal

Ce sont des utilisateurs qui utilisent un moteur de recherche et tombent sur databnf

E2: exactement, c'est-à-dire que souvent on a des gens qui ne connaissaient pas de tout la BNF, [..] finalement databnf remplie bien son rôle dans le sens où l'utilisateur qui connaît Google n'a pas besoin de connaître autre chose que Google pour arriver à l'information, je pense à la base l'application a été pensée surtout comme ça, c'est-à-dire on capte le trafic des internautes sur le web et on les redistribue dans les différentes applications de la BNF, pour qu'ils puissent accéder à l'information, et même maintenant on est quasiment à se dire il faudrait qu'ils puissent consulter directement sur une seule interface tout ce qu'ils veulent, avoir une interface unifiée où ils n'auraient pas besoin de cliquer trop jusqu'à arriver au document, de ce point de vue là je trouve qu'on joue très bien ce rôle-là de parler à un public distant, à un public nouveau qui aurait pas forcement pensé à la bnf [..] il y a vraiment une évolution, aussi on donne une visibilité à des données qui en avaient pas avant, donc on a des auteurs qui tout d'un coup voient leurs informations diffusées sur internet, alors qu'ils ont pas même créé de fichier Wikipédia par exemple, et donc de ce point de vue là ça pose problème aussi, il y a des questions de droit de l'oubli, des gens qui ont participé à des films et qui aimeraient bien qu'on oublie leur participation, des choses comme ça, on a vraiment des questions éthiques et morales qui se posent par rapport à cette nouvelle visibilité sur le web [..]

E2: le catalogue parle aux bibliothécaires, il ne parle pas à des gens en dehors des bibliothèques, [..] le catalogage c'est des codes et donc ça demande d'initiation et ça ne parle pas à tout le monde, il faut faire de sorte que ça puisse parler à tout le monde et ce n'est pas évident

Je passe à la dernière partie sur les pratiques documentaires, que pensez-vous de l'influence de la technologie sur vos pratiques documentaires ?

E1 nous on a grandi avec la technologie, donc elle n'a pas d'influence, ça a toujours existé dans nos pratiques à nous, donc la technologie elle est présente tout le temps et on n'imagine pas comment les gens faisaient pour travailler quand il n' y avait pas de technologie, quand il n' y avait pas de mail, quand il n' y avait pas tout

Et dans un contexte professionnel

E1 : oui je ne sais pas comment

Je pose la question par rapport à l'arrivée du web sémantique comme étant une nouvelle technique pour la présentation des données en bibliothèques

E1: oui je comprends la question mais comme c'est mon premier poste, ça ne change pas grand-chose à mes pratiques professionnelles

E2 : alors il y a tout cet aspect de traitement de données qui change beaucoup les pratiques professionnelles, [..] l'ouverture des données sur le web, c'est que finalement le bibliothécaire n'a plus le monopole de ses données, donc forcément il n'est pas amener à produire, il est amené à examiner, à vérifier et à croiser les références, il devient plus expert qu'initiateur de quelque chose, je pense que c'est ça qui change beaucoup, et comment dire l'aspect collaboratif et participatif du web, c'est que chacun peut être amené à produire des données, [..] je trouve que c'est vraiment fascinant, ça l'oblige des citoyens lambda, si le citoyen lambda existe, à se poser des questions sur la manière dont l'accès à l'information est fait, c'est-à-dire que ça l'oblige à se poser des questions à la manière dont on décrit l'information et dont on présente l'information, et donc ça pose la question des médias et cette question des médias aujourd'hui elle est cruciale dans l'idée de démocratie et de citoyenneté, elle est fondamentale et donc le bibliothécaire avait un rôle traditionnel il avait quasiment un monopole institutionnel sur l'accès à l'information, et le fait d'ouvrir l'accès à l'information, ça oblige je pense les gens vraiment à avoir des réflexions sur ces cadres de

production et d'accès à l'information, aujourd'hui on pose la question de comment le citoyen peut arriver à produire des données comme nous, est ce que demain on ne posera pas la question de comment un citoyen peut arriver à définir un algorithme de moteur de recherche, comment est-ce qu'on peut associer le citoyen à la production d'un moteur de recherche, qu'il sera beaucoup plus transparent et ouvert, [..]

Quels types de compétences mobiliser en termes de pratiques documentaires pour l'utilisation du web sémantique en bibliothèques numériques

E2 : alors là on est dans l'évolution des métiers, de l'apprentissage comme je disais aux états unies ils font des formations data scientist pour les bibliothécaires, donc ce qui est sûr dans l'apprentissage du métier à mon avis il y a une part d'informatique qui doit, si ce n'est pas déjà le cas, prendre une part plus conséquente, ça me paraît en fait inévitable. On peut très bien continuer à travailler comme on fait aujourd'hui, on peut très bien avoir des gens coté métier qui sont très connaisseurs de leur données etc., qui sont extrêmement experts, et puis qui parlent avec des informaticiens qui eux vont développer pour eux, on peut continuer comme ça, mais je pense qu'un bibliothécaire doit pouvoir avoir la main sur la technologie dans le sens où il doit pouvoir comprendre la technologie et l'infuser, c'est-à-dire en être à l'origine

[..]

Pensez-vous qu'il soit nécessaire de suivre une formation sur le web sémantique avant de l'intégrer en bibliothèque

E1 : oui quand même je dirais oui

E2 : oui par ce que c'est un prolongement du métier. C'est même quand on fait du catalogage c'est super d'avoir en tête le web sémantique, parce que du coup on voit le cycle entier des données de la production jusqu'à leur diffusion

E1: un jour ça va changer le travail de tout le monde, c'est quand même important de préparer le terrain [..]

E2 : Ce n'est évident en formation d'arriver à faire comprendre les bibliothécaires que ça peut les intéresser, ils n'y pensent pas tout de suite parce que ce n'est pas forcément dans leur problématique quotidienne, mais à un moment donné ils vont poser des questions et éventuellement le web sémantique pourrait y répondre, c'est le plus dure à faire passer en formation

E2: de toute façon c'est l'apprentissage, si on s'est investi dans l'apprentissage on peut ne pas arriver à s'intéresser

Les gens vont lier le web sémantique à l'informatique

E2 : pour autant ce n'est pas forcément un truc qui intéresse les informaticiens

On peut appliquer les standards du web sémantique en bibliothèques parce que les bibliothèques sont habituées à travailler avec les données les métadonnées leur identification leur localisation c'est dans leur pratiques professionnelles

E2 : oui ce n'est pas faux, mais en même temps on ne peut pas parler du web sémantique en général, mais dans des communautés très spécifiques mais qui sont très larges et très diversifiées

La dernière question comment voyez-vous l'avenir de votre profession

E1 : je ne suis pas inquiète, je pense qu'ils auront besoin toujours de nos compétences, ce que je disais au début, dans notre équipe il y a des gens qui font du catalogage depuis 30 ans, et il reste de catalogage à faire au moins pour toutes les années à venir, donc je pense comme tout le monde que l'informatique va prendre de plus en plus de place, ce sont des compétences qu'il va falloir les acquérir, qu'on va les demander de plus en plus aux jeunes, et d'une certaine manière dans nos pratiques de tous les jours on utilise de plus en plus le web, la technologie a une place très importante, [..] il y a quand même la culture de faire les choses tout seul quand même, il y a des outils qui existent, [..] je pense que la profession va évoluer et que la médiation va prendre beaucoup d'importance

Quand vous parlez de la médiation c'est plutôt la médiation autour de la réutilisation des données

E1 non je pense de la médiation pour tout

Oui mais ce que vous faites à databnf c'est la médiation autour de la réutilisation des données

E1 : on s'aperçoit qu'il faut qu'on réfléchisse sur la médiation, qu'est-ce qu'on fait de ces données, comment on peut être en contact avec un réalisateur potentiel, comment on discute avec ces gens

E2 : comme je disais sparql n'est pas évident, et surtout que les données des bibliothèques ont été faites dans un contexte particulier, pour un public particulier, [..] les gens qui arrivent sur notre sparql, ils disent oui chouette, il y a toutes les données, en plus ce sont des données exactes et elles sont exhaustives. Nos données dépendent de la politique de l'institution, ça veut dire il faut une communication autour de l'histoire de l'institution, c'est-à-dire [..] nos données sont produites à partir des codes qui correspondent à une institution particulière, ces codes il faut les comprendre, il faut arriver à les communiquer et c'est vraiment là qu'on a besoin de communiquer, [..] en fait les données n'ont pas du sens si elles ne sont pas accompagnées de quelque chose, et quand vous visualisez des données, ce que j'appelle visualisation des données c'est un petit graphique qui vous oblige à un temps d'adaptation, c'est-à-dire un temps d'appropriation des codes de l'outil graphique, c'est pareil pour le web de données il faut un temps d'adaptation des codes, et pour ça on a vraiment besoin d'un métier nouveau qui est vraiment la médiation autour de cet objet documentaire particulier qui est la donnée, c'est pour ça aussi qu'il y a le thème data librarian [..]

Et comment vous faites pour faire cette médiation

E1 : on ne sait pas encore c'est quelque chose de nouveau

E2 : c'est en cours de réflexion, et il y a quelque chose qui est intéressant sur les pratiques professionnelles, c'est que databnf à l'origine était juste un outil, c'était un projet à part dans le département des métadonnées, maintenant c'est devenu carrément un service c'est le service d'ingénierie des métadonnées, c'est-à-dire on a été intégré à part entière dans le département, on est très polyvalent on parle à tout le monde, on parle à des techniciens, on parle à des politiques, on parle à des bibliothécaires, [..] moi je pense que le métier évolue vers ça, il évolue vers le bibliothécaire, c'est quelqu'un qui sera un polytotechnicien c'est-à-dire quelqu'un qui est complètement ancré dans la synthèse politique entre ce que veulent les usagers entre ce que veut lui et ce qu'il perçoit de la profession et la technologie, c'est à dire comment est-ce qu'on infuse tout ça dans la technologie, et on l'adapte à l'environnement technologique, d'aujourd'hui c'est vraiment infuser l'aspect politique documentaire de métier dans l'environnement technologique d'aujourd'hui, et ce n'est pas forcément une vision évolutive progressive, c'est-à-dire on va plus vers le progrès etc. je pense que chaque époque produit ses propres conditions qui sont liés à ses propres contraintes ses propres manières de penser, et ce n'est pas du tout une évolution linéaire

Entretien 4

Réalisé le 24/10/ 2016

Durée 55 min

Pouvez-vous décrire votre cursus universitaire et votre parcours professionnel

Je suis recherchiste, j'ai fait l'école de chartes et j'ai travaillé sur quelque chose qui n'a rien à voir ou presque avec qu'est-ce que je fais maintenant, et qui est donc un sujet de thèse sur l'édition de texte d'un poète latin du 12 siècle, donc voilà et ensuite j'ai fait l'enssib, et je suis du coup conservateur depuis 2008 à la BNF où j'occupais trois postes successifs, et j'ai fait aussi pardon un mémoire à l'Enssib sur l'accès aux collections de masse, et ensuite je suis donc arrivé à la BNF sur un poste d'expert de préservation, c'est un poste d'expertise sur les métadonnées de notre magasin numérique donc SPAR, qui en fait je dis magasin numérique d'autres disent systèmes de préservation, en fait c'était le dispositif technique en cours d'élaboration avec un prestataire au moment que je suis arrivé, qui consistait à préserver et sauvegarder les collections patrimoniales de la BNF qui présentaient sous forme numérique, que ce soit des documents numérisés ou bien des documents nés numériques, qui arrivaient soit par voie de dépôt légal comme le dépôt légal du web par exemple soit par voie d'acquisition ou autre, voilà j'ai travaillé dans l'équipe de projet SPAR et j'avais une spécialisation sur tout ce qui tournait autour des métadonnées de documents numériques, notamment sur les métadonnées de préservation qui sont en gros les métadonnées essentiellement techniques et qui permettent la structuration du document, qui permettent de le préserver sur le long terme, ça c'était mon premier poste, et ensuite j'ai eu un deuxième poste, en fait le premier je l'ai occupé 5 ans jusqu'à 2013. En 2014/2015 donc pendant 2 ans je travaillais dans l'équipe data.bnf et j'avais une spécialisation qui portait sur l'articulation entre databnf et le catalogue général de la BNF, [..] et depuis juin je suis chef du service ou se trouve notamment databnf, qui s'appelle ingénierie des métadonnées, c'est un service dans le département des métadonnées qui s'occupe de tout un tas de problématiques qui coordonnent un certain nombre de questions liées au catalogue, comment on va récupérer des données éventuellement du l'extérieur ce qui est le cas par exemple pour les données qu'on récupère depuis les éditeurs, comment est-ce qu'on fait évoluer les règles de catalogage alors, c'est ici qu'il y a une coordination qui se passe et c'est ici aussi qu'on gère les formats, c'est notamment notre format marc et le format Ead de catalogue, il y a une coordination autour des dispositifs, autour de ça il y a des personnes qui travaillent sur la normalisation documentaire, voilà il y a un aspect administration des données lié à la fois à l'équipe data.bnf et l'équipe de correcteurs qui travaille à la mise en niveau des données qui ont été chargées au fil de temps dans notre catalogue, et qui sont de qualité très hétérogène, [..] voilà et un truc que je n'ai pas dit et qui est encore actuellement une casquette que j'ai commencé sur mon premier poste et que je continue à avoir tout au long de mes postes à la BNF et qui en fait que j'assure la coordination métier sur nos identifiants ARC, en fait ce sont les identifiants pérennes qu'on utilise pour faire des liens dans Gallica dans databnf dans notre catalogue notamment.

Dans ma thèse il y a deux axes principaux, il y a les bibliothèques numériques et le web sémantique, donc je m'intéresse à la notion de la bibliothèque numérique comment vous définissez une bibliothèque numérique ? qu'est-ce que c'est pour vous une bibliothèque numérique ?

Alors il y a plein de manière de la définir, ce qui vient tout de suite à l'esprit c'est grosso modo ce qu'on a derrière gallica par exemple, c'est-à-dire c'est un dispositif d'accès aux collections numériques, je le vois vraiment sous l'aspect accès, [..] si j'étais rigoureux je dirais que c'est l'ensemble de dimensions de la bibliothèque, [..] en tout cas moi je pense systématiquement à l'accès, aux fonctions liées à l'accès et à la diffusion sur le web, qui correspond grosso modo à ce

qu'on a dans gallica, voilà mais ce n'est pas que la numérisation c'est pour ça que j'ai dit que ça rassemble à des problématiques à lesquelles Gallica répond, mais qui peuvent s'élargir à tout optique de collections, en particulier des collections numériques comme les acquisitions des publications scientifiques qui se trouvent dans les archives ouvertes, pour moi aussi ça correspond à une bibliothèque numérique, il y a des contenus amenés à faire partie de bibliothèque numérique.

Vous pouvez me parler des langages documentaires utilisés dans gallica pour décrire les documents et aussi des métadonnées produites dans gallica

Ce que vous voudriez savoir c'est par rapport à l'écosystème des différents formats de métadonnées qu'on utilise à la BNF où est ce que je place les langages liés utilisés dans le cadre du web sémantique

Oui

Alors pour moi pour le moment rdf c'est avant tout un langage dont la valeur ajoutée ça se discute mais la valeur ajoutée est de permettre un partage efficace, et une réutilisation efficace par des partenaires qui ne sont pas nécessairement des bibliothèques. [..] le rdf pour le moment est un format de diffusion et non pas un format de production, d'ailleurs c'est actuellement le cas je pense que ce n'est pas forcement en tout cas pour le moment la peine de changer le format de production pour passer à quelque chose en rdf, pour plusieurs raisons en fait d'une pour des problématiques techniques et pour des problématiques culturelles .Pour des problématiques techniques, parce que le rdf c'est compliqué à sourcer, c'est compliqué à gérer en mode mise à jours, [..] En fait pour moi c'est plus une question d'architecture qu'une question de format, et de fait on peut très bien faire de linked data avec des données en marc dans l'écosystème de production interne, parce que depuis 87 pour être précis les catalogueurs font déjà des liens entre les notices bibliographiques et les notices d'autorités, ce que j'appelle en fait de vintage linked data, parce que les notices ont des identifiants mais pas des uris, mais des identifiants qui sont bien gérés, derrière il y a des données en marc, mais on a déjà des entités qui sont liées entre elles, et finalement de ce point de vue là on a déjà une logique de production qui pense à linker en entités, on va lier les unes aux autres avec du coup un travail qui depuis la fin les années 80 est un travail de catalogueur, et qui est non seulement de la production des données mais l'administration des liens, et là que je vous rejoint c'est que probablement que ma vision des choses est très connotée, parce que je tiens ce discours là parce que justement le discours qu'on a pas besoin de rdf pour la production, [..] l'enjeu c'est de faire évoluer nos pratiques pour avoir un peu plus d'entités, un peu moins de descriptifs de publications, par exemple d'accentuer et de systématiser ce qu'on fait déjà, plutôt que de se focaliser sur le fait de les avoir dans un format différent, puisque déjà on peut continuer à le faire, on peut déjà le faire en utilisant un modèle de données qui se base sur marc, c'est pour cela je disais le deuxième aspect n'est pas technique il est culturel, le fait que les experts des données et des catalogues connaissent par cœur le format marc, ils connaissent très bien les strates de notre catalogue, [..]

ET quand vous dites il fallait prendre une autre technologie ou il fallait faire autrement vous pensez à quoi exactement quelle autre technique ?

[..] au niveau métier il faut un modèle qui peut rester du marc à mon avis, c'est ce que les gens ont l'habitude de faire, c'est sur ces formats là que s'opèrent toute la coordination bibliographique, en long terme on peut s'appuyer la dessus et faire évoluer, derrière le modèle physique ça peut être une base de données ça peut être une base de données relationnelle du sql, ensuite on peut dire peu importe le flacon pourvu qu'on ait l'ivresse, après on a finalement un modèle qui est plus conceptuel qui pour le coup actuellement est du marc, et à mon avis comment dire on a plus de 300 personnes à la bnf qui travaillent dans ce format-là, et qui le maîtrise et plutôt que de le changer d'une manière très destructive, le mieux est de le faire évoluer pour accorder de nouveaux besoins puis ensuite ça devient un format de diffusion parmi d'autres, [..] on a une culture professionnelle à la bnf où les catalogueurs n'utilisent pas de masque de saisie, ils codent directement en marc et de coup c'est pour ça que la question de format est aussi connotée culturellement, si on change le format ça a un

impact énorme sur toutes les personnes, [..] en plus on a tout un tas de circuits qui fonctionne bien pour faire évoluer le catalogue à travers le format marc qui est une de pierres angulaires de ce système, et on a l'avantage aussi d'avoir un format l'intermarc qui est notre propre format, du coup concentrons-nous plutôt sur la préparation des données, sur le fait d'avoir des nouvelles entités, je ne sais si vous voyez un peu FRBR c'est un truc très bibliothèque, [..] la question c'est comment est-ce qu'on va passer de ce contexte, de cet état actuel de catalogue à un état futur où on aurait potentiellement plusieurs millions des notices d'œuvres qui se régénèrent automatiquement. aussi une problématique un autre truc culturel que je n'ai pas dit, c'est la manière dont on gère les alignements, de plus en plus il y a des alignements effectués, certains peuvent être automatiques d'autres sont à faire directement par le catalogueur ou le gestionnaire des interférentiels, par exemple là nos gestionnaires du centre rameau qui est notre langage d'indexation, quand ils produisent une notice rameau, ils vont vérifier s'il y a une équivalence dans la contrepartie anglaise qui est lcsh *Library of Congress Subject Headings,* l'objectif est de continuer ça, d'avoir une politique d'alignement, et de rajouter éventuellement pour parler très concret une sous zone, les sous zones qui vont aller plus davantage vers le web sémantique, ils rajoutent les URI plutôt que de passer effectivement en rdf, je ne vous dis pas que dans 10, 15, 20 ans qu'il ne soit pas pertinent, c'est plus une question de priorité en termes d'évolution à un moment où en plus il y a quand même un certain nombre d'inconnus vis à vis de la capacité de rdf [..] je ne serai pas forcement convaincu qu'il faut passer au rdf, par contre c'est indispensable qu'on a un format pour exporter nos données [..]

Je passe à la question des identifiants à la bnf vous utilisez les identifiants uri Arc

Les uris contiennent des Arc plus exactement, on utilise un standard qui s'appelle arc qui correspond à un type d'identifiant, qui prend en compte de manière on va dire prioritaire la question de la stabilité de cet identifiant, et qui prévoit la manière dont on peut l'inclure dans une url

Est ce qu'on peut lier les données sans avoir des identifiants pérennes ? Est ce qu'il faut que ces identifiants soient universels ?

C'est une vaste question, on peut toujours publier des données en linked data sans avoir des identifiants pérennes, mais ça va être trop compliqué je dirais, [..] il y a une réponse graduée, il y a des risques de stabilité d'une part de l'identification et des liens à d'autres uri, [..] je vais arrêter de dire pérennes d'ailleurs, c'est aussi le problème ça fait peur aux gens quand on dit uri pérenne, on a l'impression qu'on a un truc qui ne doit jamais bouger et qui est pour l'éternité, et en fait c'est simplement qu'il faut des uris de la même manière des bibliothèques. [..] Il faut choisir un nom de domaine qui soit le plus stable possible, [..] en fait ARC prend la question au niveau de l'identifiant il va au-delà du web il prend la question de l'identifiant qui doit être globalement unique, et qui doit tenir la route et bien voyager, [..] après entre Arc et le numéro, le fait qui a un arc ce n'est pas du tout une obligation, ce qui est important ce qu'il y a une politique de stabilité pour ces identifiants pour ces uris, [..] il faut la prévoir dans le cahier des charges [..] en gros ce que je disais par rapport à arc vs les identifiants classiques, c'est en effet arc fournit un ensemble de bonnes pratiques, il oblige à poser ces questions-là mais on peut très bien les poser à travers des uris bien gérés, du coup la question est plutôt qu'il faut être conscient de problème et de bien le gérer [..]

Qu'est-ce que le web sémantique apporte à la bnf ?

D'une part c'est une technologie qui parle déjà la manière dont les données sont déjà organisées chez nous, et d'autre part qui est un style d'interopérabilité qui permet d'ouvrir beaucoup plus largement le champs et avoir des nouveaux types de partenariats autour de données, [..] pour construire des partenariats avec des nouveaux acteurs, et c'est comme ça qu'on avance parce que on a un langage qui nous permet de causer nos données en fait pour moi c'est ça la valeur ajoutée du web de données.

Vous préférez utiliser le terme web de données ou web sémantique?

Pour mai clairement web de données

Pourquoi
[..] le problème c'est que pour plein de gens sémantique ne veut pas dire données structurées, et du coup ça aboutit à se poser tout de suite des questions autour de traitement automatique de la langue autour des moteurs de recherche, [..] pour moi le fait de parler du web de données me paraît plus simple, je ne sais pas comment dire, ce truc, j'ai du mal à trouver un mot qui convient, ce paradigme se traduit en terme d'architecture, je pense c'est mieux de parler d'architecture des données, de parler dont la manière que ça se présente, plutôt que d'insister sur le fait que les données sont structurées, le web déjà c'est de la donnée structurée, mais ça se discute largement, je vous donne juste mon opinion, après pareil on va poser des questions de l'intelligence artificielle, c'est vrai la technologie du web sémantique permet de répondre en partie à certaines problématiques, je n'oserais pas dire de l'intelligence artificielle parce que quand même c'est très sophistiqué, en tout cas j'ai envie de dire c'est le degré zéro de l'intelligence artificielle c'est d'alimenter des règles dans les ontologies en fait qui structurent des données. [..] d'ailleurs c'est un ensemble de standards et le fait de dire sémantique on va insister qu'il y a des données structurées, et quand on dit web de données on insiste sur le fait qu'on ne relie plus des documents mais on relie des données, et pour moi c'est ça l'intérêt de la nomination web de données, que ça informe de manière que je trouve assez clair sur qu'est-ce que c'est l'enjeu par rapport au web traditionnel, ça permet de lire d'une manière claire l'évolution que ça constitue par rapport au web traditionnel, et ce qui est important il y a le web dedans, web de données ou web sémantique le mot clé le plus important est le web, parce que le graphe existait déjà sauf que c'était un graphe de documents et non pas un graphe de données, et c'était juste un changement de paradigme qui consistait à passer au web de données et appliquer en grande partie les mêmes technologies, puisque finalement rdf est une surcouche et d'autre part les uris existaient déjà par exemple, et par rapport à ça oui on va rajouter une surcouche sur les technologies qui existaient pour aller plus loin, de la même manière que le web, il a constitué un changement hyper important par rapport à internet et aux réseaux pour moi c'est l'étape d'après
J'ai plus de question je vous remercie beaucoup
Une chose que je n'ai pas dit que j'ai aussi fait beaucoup de formations, notamment sur ces choses-là, je trouve important de signaler en tout cas ça contextualise mon discours.

Entretien 5

Réalisé le 14/10/ 2016

Durée 50 min

Vous pouvez décrire votre cursus universitaire et votre parcours professionnel
J'ai fait une maîtrise à l'époque donc M1 de lettres moderne à l'université de Lyon 2, deux après j'ai commencé un DEA recherche que je n'ai pas fini, après j'ai commencé un DESS en informatique documentaire à l'enssib que je n'ai pas fini non plus, parce que j'ai eu le concours conservateur des bibliothèques donc j'ai commencé ma scolarité de conservatrice à l'enssib en ce moment-là. Du coup après l 'enssib, j'ai d'abord eu le concours de conservateur territorial, après l'enssib j'ai été recrutée pour un CDD à l'agence de développement de livres [..] et ensuite j'ai été recrutée comme une conservatrice des bibliothèques au conseil général [..] pour être directrice adjointe de la bibliothèque départementale de presse, et ensuite j'ai demandé une mise à disposition dans la fonction publique d'état pour venir travailler au service commun de documentation à l'université de Poitiers où je suis depuis 2012 sur mon poste actuel/
Est-ce que vous avez suivi des formations depuis que vous travaillez en bibliothèques
Oui tout le temps tout le temps, plusieurs fois par an
Quel est votre fonction ou votre titre ?
Je suis responsable du service des ressources informatiques documentaires
Quelles sont vos missions
c'est encadrer le service en question c'est un service d'environ une dizaine de personnes [..] qui touche tout ce qui concerne l'informatique documentaire dans la BU, que soit des matériels de logiciels que des ressources numériques, la mise en ligne des thèses et des mémoires, le signalement, le catalogue, le site web, le sgib, donc voilà encadrer ce service là et moi j'ai des missions plus spécifiques sur le site web et le catalogue et l'outil de découverte actuellement
Qu'est-ce que c'est pour vous une bibliothèque numérique ?
ça va être la mise à disposition de tout ce qui est document numérique, en occurrence donc ça va brasser aussi bien les documents qu'on produit et qui sont mis à disposition via un autre site ou bien via nos différents sites par exemple, ainsi que les documents qu'on s'abonne ou qu'on achète, les ebook et les bases de données et ,les documents numérisés qui sont issus de la numérisation des collections patrimoniales, notamment qui sont mis en valeur soit par exposition soit par la bibliothèque virtuelle voilà c'est tout ça la bibliothèque numérique.

Quels sont les langages documentaires utilisés dans votre bibliothèque pour décrire les collections numériques ?
Pour décrire les collections numériques, on utilise le langage unimarc pour décrire un certain nombre des collections numériques auxquelles on s'abonne ou qu'on achète, [..] au niveau de la mise en ligne des thèses et des mémoires on a utilisé le langage XML [..] et puis pour les documents numérisés ça va être du dublin core
Moi d'accord l'un des éléments essentiels dans une bibliothèque numérique est la production des métadonnées. Pouvez-vous me parler des métadonnées produites dans votre bibliothèque ?
[..] je prends un exemple pour les thèses et les mémoires, on produit effectivement des métadonnées au format xml et en unimarc ,on va aussi cataloguer ces documents dans sudoc donc là ce sont des bibliothécaires qui s'en occupent, des catalogueurs de catégorie B qui vont cataloguer ces documents, donc ils savent bien comment cataloguer respecter les normes etc., voilà pour les bibliothèques numériques alors du coup c'est peut-être un peu plus compliqué, parce que parfois il y

a des projets qui ne sont pas faits par des bibliothécaires, ils ne sont pas portés par des gens du métiers qui savent décrire un document selon des normes, par exemple on va aider des laboratoires à porter ces projets là, dans ce cas-là une partie de rôle c'est de les aider justement à produire des métadonnées qui soient de qualité, alors que ce sont des personnes qui ne sont pas dans le métier, […] voilà c'est une partie des missions de service en fait de vérifier que ces métadonnées-là respectent les normes, du coup on s'engage à les respecter, on accompagne les gens qui travaillent sur les différents projets pour qu'ils les respectent. [..] On a beaucoup de bibliothécaires qui travaillent dans le même catalogue et du coup ça fait aussi partie des missions du service de vérifier que tous les catalogueurs produisent des métadonnées de qualité, donc qu'elles soient conformes aux règles de sudoc par exemple, et du coup si ce n'est pas dans sudoc, si c'est pour des bibliothèques numériques c'est le même travail au final, ça fait partie des missions du service.

Est-ce que ce sont des métadonnées ouvertes réutilisables ?

Alors du coup celles qu'on produit sur Sudoc, oui puisque de toute façon on travaille selon les règles du sudoc, et donc maintenant le sudoc a ouvert ces métadonnées il y a deux ou trois ans, donc du coup les métadonnées sont ouvertes et réutilisables. Pour les thèses et les mémoires, on n'a pas mis en place des possibilités de moissonner, alors pour les bibliothèques virtuelles, oui il y a deux qui sont moissonnées dans gallica, donc du coup elles sont ouvertes, les thèses et les mémoires ce n'est pas encore fait du coup mais par contre on a des entrepôts spécifiques pour les thèses de l'université de Poitiers par exemple. Oui on est plus pour l'ouverture, après si ce n'est pas mis en place c'est plus une question de temps

Est-ce qu'il y a des pistes stratégiques pour la production et la diffusion des métadonnées dans l'avenir ?

oui de toute façon, quand on porte un projet de bibliothèques numériques, donc ça concerne les collections numérisés ou d'autres, on va intégrer la notion de diffusion et d'ouverture des métadonnées dedans, en fait ça va être intégrer dans le projet, par exemple là on a un projet qui est à l'étape de prototype on va dire, d'un portail qui centraliserait toutes nos bibliothèques et nos expositions virtuelles qui sont consacrées au patrimoine, en fait pour l'instant on a plusieurs sites web séparés et on voudrait les rassembler dans un même portail donc du coup dans ce projet de portail il y a la dimension d'ouverture des métadonnées.

Vous avez répondu dans le questionnaire que vous utilisez un peu le web sémantique

oui j'ai mis dans le cadre de sudoc, parce que moi c'était la chose que je voyais, c'était comment dans le processus de transition bibliographique que l'abes met en place pour faire travailler les catalogueurs de sudoc, du coup dans une optique de transition justement bibliographique via ces notions qui sont intégrées par l'abes, et du coup des règles qui arrivent au fur et à mesure de l 'abes pour gérer cette production bibliographique, elle intègre cette dimension du web sémantique, donc voilà moi la seule chose en fait que je voie concrètement qui impacte notre travail sur le web sémantique .

Quels sont les enjeux d'utilisation le web sémantique en bibliothèques ?

Alors si j'ai bien compris qu'est-ce que c'est le web sémantique, les enjeux du coup peuvent être justement de la réutilisation facilitée des métadonnées et de pouvoir réadapter les métadonnées qui sont fournies par d'autres pour reproduire de l'information différente, et pouvoir mettre en valeur certaines informations qui sont disponibles parce qu'elles sont conformes aux standards du web sémantique, [..] c'est une façon donc de pouvoir réutiliser les métadonnées et remettre en valeur certaines choses

Dans votre bibliothèque pour le moment on peut dire que vous n'utilisez pas les standards du web sémantique ?

Juste on suit les règles de l'abes pour le catalogage, puis on suit des normes de description des documents

Vous n'utilisez pas par exemple les uri pour identifier les données ou rdf pour lier les données

Non

Est ce qu'il y a un projet pour introduire des uris pérennes pour identifier les données
Je suis en train de réfléchir à ce que vous dites, de réfléchir à une application concrète vous pouvez me donner un exemple. Par exemple j'ai une notice-là qui est décrite en dublin core qu'est-ce que je dois faire concrètement.
Moi là concrètement il y a deux choses, identifier les données et après les relier entre elles, et pour les identifiants par exemple la bnf utilise ARC pour les identifiants et après relier ces données identifiées en utilisant des triplets rdf. Ma question portait sur ces identifiants-là est ce que vous avez des projets pour l'avenir ?
Non je ne peux pas dire que ce soit en projet mais après je n'ai pas l'impression si on décide de s'y mettre que ça poserait vraiment un problème, outre que des problèmes de temps parce que j'ai l'impression que nos notices sont déjà bien telles quelles sont, elles sont déjà bien conformées donc après oui ça demandera plus de temps

Le problème pour vous pour la mise en place du web sémantique c'est le temps
Ah oui il faut du temps pour s'y mettre
Par rapport aussi aux personnels est ce qu'il faut par exemple faire des formations là-dessus
Oui sans doute oui tout à fait, oui ce sont des notions qu'il faut les réapprendre, [..] après pour des gens qui sont catalogueurs ce sont des notions qui leur parlent, qu'ils connaissent très bien
Si la bibliothèque venait à évoluer et à introduire le web sémantique quels seront selon vous les principaux défis ?
Formation, temps de travail, normalisation voilà des choses classiques
Sur l'interopérabilité est ce que vous avez eu une discussion autour de ce sujet ?
Récemment on est moissonné par gallica sur des bibliothèques virtuelles et récemment elles ont été intégrées à Europeana. Il me semble que récemment on a eu des allers-retours entre la BNF et nous sur l'interopérabilité des données
Que pensez-vous de l'influence de la technologie sur vos pratiques professionnelles ?
Je n'ai pas d'opinion, je ne sais pas mon poste est la dessus, je ne suis ni pour ni contre voilà
Quel type de compétences mobilisées en termes de pratiques documentaires pour utiliser le web sémantique en bibliothèques numériques
La rigueur, la capacité de s'adapter à des nouveaux langages donc voilà ça me parait plutôt le principal
Pensez-vous qu'il soit nécessaire de suivre une formation avant l'utilisation de web sémantique en bibliothèque numérique ?
Oui je pense oui même [..] il faudra se former d'une manière ou d'autre que ce soit sur le tas ou en stage
Comment voyez-vous le rôle des catalogueurs suite à l'introduction du web sémantique en bibliothèque ?
Leur rôle est essentiel [..] pour moi ça ne change pas, ils continuent à faire leur travail à appliquer les règles ey à faire attention qu'elles soient respectés [..] c'est eux qu'ils ont la compétence technique au sens respect de la règle, respect de la norme etc.
Comment voyez cous l'avenir de votre profession votre métier ?
Alors je suis nulle pour ça, je n'arrive absolument pas à avoir une vision prospective des trucs, je ne sais pas où on sera dans 5 ou 10 ans voilà j'en ai absolument aucune idée

Entretien 6

Réalisé le 15/12/ 2016

Durée 45 min

Vous pouvez décrire votre cursus universitaire et votre parcours professionnel
J'ai fait une licence d'anglais et une année spéciale en métier de documentation, ensuite j'ai travaillé à l'Insa pour un certain temps dans une école d'ingénieur, après j'ai passé le concours, [..] j'ai travaillé à la bibliothèque nationale où je faisais à la fois, je suis à la fois sur les métadonnées parce que j'ai fait du catalogage depuis 15 ans 20 ans, catalogage indexation signalement et j'étais correspondante informatique, déjà à la bnf j'ai censé faire de l'informatique documentaire et puis après je travaille à Toulouse 3 depuis 2004, depuis 2004 je suis bibliothécaire j'ai passé un autre concours et je travaille à l'université Toulouse 3
Quel est votre fonction à la bibliothèque ?
Je suis responsable du service de catalogage centralisé qui s'occupe des ouvrages de l'ensemble de SICD, je suis responsable des métadonnées et je m'occupe également de la mise en ligne des thèses
Vous avez suivi des formations depuis que vous travaillez en bibliothèques ?
Oui j'ai fait des formations classiques en bibliothéconomie, oui j'ai fait à l'enssib, puis après j'ai fait des formations sur XML, j'ai suis après des formations plus sur le côté informatique, sur tout ce qui est html php etc., je me suis occupée pendant quelques années du site web et des bases de données locales
Comment définissez-vous une bibliothèque numérique ?
C'est un terme que je n'utilise pas du tout
Quel terme vous utilisez alors ?
On parle de ressources électroniques, on ne parle pas de bibliothèque numérique parce que bibliothèque numérique on va penser à Gallica, on va penser à quelque chose qui a été numérisée etc., c'est un peu ça qu'on a en tête, ou alors on va penser uniquement aux ressources électroniques mais comme une bibliothèque, on va proposer les deux, on a pas une bibliothèque numérique en tout cas, nous on va proposer un catalogue qui continuera des ressources électroniques et des ressources papiers, c'est un accès uniquement à distance ça ne regroupe pas totalement les pratiques de nos utilisateurs. Il y a des collections numérisées c'est Toulousana gérée par un service interuniversitaire le SCID qui gère à la fois le SIGB, et ils gèrent aussi les livres anciens, la numérisation des livres anciens donc voilà il y a une plateforme qui s'appelle Toulousana qui est une bibliothèque virtuelle des fonds anciens.

Quels sont les langages documentaires utilisés dans votre bibliothèque ?
Pour tout ce qui est dans le Sudoc on utilise unimarc et FRBR[..] on intègre le RDA progressivement [..] et moi je fais partie des formateurs qui forment les collègues à l'évolution RDA et le RDA qui sera notre nouveau code de catalogage qui utilise le FRBR globalement
Vous pouvez me parler des métadonnées produites dans votre bibliothèque ?
Alors tout ce qui est Sudoc, on utilise ces formats, on indexe en rameau, donc ça c'est plutôt du classique, [..] par ailleurs on a les revues les périodiques qui ne sont pas dans Sudoc, ils sont accessibles dans les outils de découverte et dans le réservoir des liens, donc on utilise des métadonnées forcément moins propres de ce qu'on trouve dans le Sudoc mais on essaye d'utiliser des choses qui sont labellisés, il y a un travail sur une base de données nationale qui s'appelle Bacon, ils utilisent des métadonnées au format KBART ou label KBART, c'est un label qui dit que ces métadonnées sont de bonne qualité, elles remplissent des critères qui ont été établis, donc c'est

un travail pour les éditeurs pour avoir des métadonnées de bonne qualité, et l'ABES travaille la dedans, il récupère l'ensemble de métadonnées, il les retravaille et les exporte, voilà nous on les récupère dans les outils de découverte ou les réservoirs des liens, soit on les récupère chez les éditeurs et on les intègre dans nos outils, soit quand on les coche on voit label KBART donc elles sont de bonne qualité et réutilisables etc.

Est-ce que la bibliothèque utilise les standards du web sémantique ?

Oui et non, en ce qui concerne les standards du web sémantique on ne les utilise pas, on utilise des URI parce qu'on met en ligne les thèses on a aussi une plateforme locale de mise en ligne de thèses sur un outil open source, et là on donne accès à nos thèses on s'engage à donner une adresse qui sera fixe et par ailleurs l'Abes attribue un URI à chaque thèse, c'est le numéro national de la thèse [..] après toutes les notices dans Sudoc ont des liens pérennes donc ont des URI,[..]

Et pour les autorités auteurs ?

Oui justement il y a l'ISNI le numéro international, c'est attribué pour l'Europe par British library et la BNF, ce sont des numéros qui correspondent aux auteurs [..] par ailleurs il y a un équivalent dans la recherche, pas dans les outils mais du côté HaL c'est Orchid [..] Tout ça ce n'est pas nous, c'est l'abes et la bnf qui les mettent en place, mais on travaille dedans puisque que comme on est dans le réseau national on va récupérer ces numéros là et on va faire en sorte, le vrai travail de catalogueur aujourd'hui est de faire en sorte qu'on peut retracer la source, on est sûr de ce numéro qui identifie d'une manière certaine [..] le FRBR se découpe en triplet très simple ça sera très compatible avec RDF mais on y est pas encore …

Vous préférez utiliser le terme web sémantique ou web de données?

Moi quand je donne des formations j'utilise le terme web de données, [..] historiquement on a appelé ça web sémantique dans la fin des années 90, [..] le web sémantique est un peu confus parce qu'on va penser à la traduction de langage naturel qui n'est pas le cas, en fait le web sémantique n'est pas ça, en gros historiquement au début on a appelé ça web sémantique, et comme ce n'est pas très clair, le terme web de données a été utilisé, je trouve que la différence n'est pas trop clair, en gros le web sémantique c'est plus les standards et les formats et le web de données c'est un peu le résultat de ce qu'on veut en faire, je trouve que le coté chronologique est plus intéressant de dire on est parti de ça et aujourd'hui on n'a pas de web sémantique dans le sens où on n'a pas de traduction automatique de langues, mais on normalise et on met on place des choses très normalisées, pour que les machines puissent les utiliser pour qu'on puisse faire une réutilisation intéressante et intelligente derrière [..] C'est vrai que le web de données c'est beaucoup plus parlant.

Quels sont les enjeux de l'utilisation du web sémantique en bibliothèque ?

La visibilité je pense que c'est le 1er aspect, [..] une visibilité mais une visibilité intéressante à l'exploitation de nos données et à la réutilisation des données d'autres établissements, c'est bien l'idée data.bnf, pour moi ce que je présente aux gens en général c'est Google [..] on peut faire la même chose avec la bibliothèque mais d'une manière beaucoup plus fiable, donc voilà à la fois qu'on soit visible et que ça soit intéressant à l'utilisateur final, après il y a des enjeux qui sont mis en avant de côté de l'Abes, je ne suis pas bien convaincue sur le côté économie, après ça permet de faire communiquer des communautés qui ne parlent pas parce que les formats ne parlent pas entre eux comme les musées archives bibliothèques, voilà ce genre de choses c'est encore à discuter ça, c'est l'interopérabilité effectivement

Que pensez de l'influence de la technologie sur vos pratiques documentaires ?

Plus précisément ce qui impacte notre métier moi et mes collègues au niveau de signalement, c'est que le cœur du métier qui permet la communication ne change pas, par contre sur les pratiques effectivement, alors avant la description était plus importante aujourd'hui c'est le lien , comment dire on passe moins de temps à décrire et on passe beaucoup de temps à vérifier et à lier , à lier un

auteur à un livre ; à lier un auteur à un autre auteur etc. à s'assurer de la fiabilité de ce lien, aujourd'hui c'est plus de la création d'autorité que la création de notice bibliographique et la fiabilité de ça, avant on crée aujourd'hui on récupère, on passe de catalogueurs au catalinkeurs

Quel type de compétences mobilisées en termes de pratiques documentaires pour utiliser le web sémantique en bibliothèques numériques?

J'imagine qu'il faut une personne qui maîtrise un petit peu les enjeux [..] quelqu'un qui a de connaissances en informatique documentaire [..] après les compétences directes c'est classique c'est beaucoup de rigueur voilà maîtrise un peu les outils type xml etc...

Comment voyez-vous l'avenir de votre métier ?

Beaucoup moins de travail sur l'acquisition, moins de travail sur la création de notices et beaucoup de travail sur la mise en valeur, ça c'est du côté catalogueurs au niveau des [..] c'est la sélection plus que l'achat […] la médiation en bibliothèque universitaire est devenue la moitié du temps du travail les formations 1er année 2eme année 4eme année etc. ça prend énormément de temps [..]

Entretien 7

Réalisé le 23/11/ 2016

Durée 1h10min

Pouvez-vous me décrire votre cursus universitaire et votre parcours professionnel

D'accord alors après le bac j'ai passé à l'époque c'était un DEUG de médiation culturelle, ensuite je me suis réorienté vers le français langue étrangère jusqu'au DEA, et après j'étais prof de français langue étrangère au Japon pendant deux ans, et je suis revenu en France faire un Dut année spéciale à Dijon métier de livres avec ensuite deux contrats et puis le concours bibliothécaire adjoint spécialisé à l'époque voilà. Alors au niveau du parcours professionnel c'est pour l'instant j'ai connu qu'un seul établissement donc université de Nice service commun de documentation

Et vous travaillez là-bas depuis combien de temps

alors ici depuis septembre 2007 ça va faire presque 10 ans, et alors au départ c'était un poste orienté classique, avec édition catalogage accueil formation et quelques tâches sur la documentation électronique, et petit à petit voilà c'est un poste qui réorientait pour devenir aujourd'hui un poste plus centré sur les ressources électroniques, sur aussi la responsabilité à l'accueil et sur la gestion d'un fonds spécialisé qui est en fait au cœur du projet de bibliothèque numérique, une partie de ce fonds spécialisé est une photothèque de 6000 photographies anciennes donc sur l'Asie et principalement sur ce corpus là qu'on a travaillé sur la mise en ligne selon le principe du web sémantique

Est-ce que vous avez suivi des formations depuis que vous travaillez à la bibliothèque

Alors oui des formations sûrement liées à beaucoup de choses : catalogage accueil etc. mais alors pour tout ce qui est web sémantique j'ai suivi une formation de formateur en juillet 2013 à la bibliothèque nationale de France, qui était assurée par Sébastien Peyrard et Ignés Simon qui sont l'équipe data.bnf, voilà qui était assurée dont le but de créer un réseau national de formateurs pour les aider un peu sur les formations du web sémantique

Quelle est votre fonction ou votre titre maintenant à la bibliothèque

Maintenant je suis coordinateur des services au public pour le pôle lettres arts sciences humaines et sociales, donc il y a trois bibliothèques dans ce pôle, ensuite je suis responsable de la collection Assemi qui est le fonds dont je vous ai parlé où il y a à peu près 20000 documents, ensuite je suis plusieurs secteurs d'acquisition et je m'occupe aussi donc du traitement bibliographique, parce que je suis correspondant national du groupe transition bibliographique groupe de formateurs transition bibliographique, donc je m'occupe de former les collègues ici après ça, j'ai donc aussi la responsabilité de la documentation électronique uniquement pour la bibliothèque de lettres arts sciences humaines, avec aussi le côté formation aux ressources en ligne des usagers et promotion de ressources et promotion open access également

Qu'est-ce que c'est une bibliothèque numérique pour vous ?

Il y a deux côtés à voir, il y a le côté stratégique de la bibliothèque numérique, et il y a le côté technique, au niveau stratégique je pense qu'une bibliothèque numérique c'est l'action de rendre visible sur le web selon des technologies et des standards, c'est l'action de rendre sur le web des contenus des savoirs qui sont conservés physiquement dans la bibliothèque ou qui sont nativement

numériques, voilà donc ça c'est au niveau stratégique, ça peut prendre le nom d'une bibliothèque numérique, mais ça peut dépasser simplement le mot bibliothèque qui est quand même très connoté livre dans l'esprit du grand public, donc plus à mon avis c'est une action de mettre à disposition de tout le monde tous les internautes des ressources multiformes multi supports, et voilà techniquement quand on parle de bibliothèques numériques dans le métier on a souvent l'idée d'un portail derrière, qui soit en fait une adresse unique dans laquelle on va mettre tout le contenu qui sera accessible par une page d'accueil, une interface de recherche etc., et techniquement une bibliothèque numérique je trouve que sa définition elle est là physiquement sur le web sur un infrastructure web, il faut avoir un endroit avec un site web, qui soit le site qui permet à la fois de stocker de consulter, mais ce qui est souvent n'est pas pris en compte dans les projets des bibliothèques numériques, c'est l'exposition des données qui sont dans ce site web vers les moteurs de recherches.

Est-ce que pour vous la notion de l'accès libre est une des critères d'une bibliothèque numérique

Oui je pense que ça fait partie des critères, c'est l'accès libre effectivement au contenu, je pense que c'est incontournable, même si rien empêche dans une bibliothèque numérique de proposer un accès complètement libre et un accès à des ressources sous forme d'abonnements qui sont accessibles uniquement sur identification pour telle ou telle catégorie d'usager

Vous pouvez me parler de la bibliothèque numérique

Le projet est né en 2013 suite effectivement à la formation que j'ai pu avoir avec l'équipe data.bnf, on a ici un département qui s'appelle le sidoc qui gère tous les aspects techniques communs à tous les services documentaires de l'université, donc c'est à la fois l'administration du SIGB et aussi les questions qui tournent autour de la bibliothèque numérique, c'est eux qui viennent faire l'assistance, ils sont associés au projet pour les outils techniques choisis pour la bibliothèque numérique donc le projet a commencé à l'époque en 2013 avec un collègue que je vais nommer qui s'appelle Étienne Cavalier, qui était responsable de ce département, alors il se trouve que Etienne Cavalier est un conservateur qui aujourd'hui travaille à la bnf, il a eu sa mutation il y a deux ans, il travaille dans l'équipe data.bnf, c'est quelqu'un qui était très au point sur l'informatique documentaire et sur tout ce qui est métadonnées et standards du web sémantique, donc moi j'étais plutôt sur le volet stratégique et relation avec les chercheurs à la fois à l'université et ailleurs qui pourraient être intéressés par la collection dont je suis en charge, en fait on a commencé notre projet uniquement en sélectionnant le corpus de photographies anciennes donc sur l'Asie selon deux arguments, c'est-à-dire c'est un corpus qui est unique au monde, donc ça fait partie des choses à mon avis prioritaire, une bibliothèque numérique c'est d'abord de diffuser de donner de la visibilité aux contenus qu'on est les seuls à conserver, donc on a commencé la dessus et on a commencé, lui a commencé de faire le travail technique et moi j'ai fait le travail de fédération de tous les acteurs qui peuvent être concernés ,et donc on a réussi quand même à avoir un groupe de chercheurs jusqu'en dehors de notre université, on avait des collègues notamment à Marseille parce qu'il y a un laboratoire qui travaille sur l'Asie et on a fait plusieurs réunions, certaines réunions à distance parce qu'il y avait des collègues qui étaient en Asie dans des universités partenaires à la nôtre, et on a pu recueillir les besoins des chercheurs qui voulaient effectivement, qui ont ciblé comme prioritaire cette collection photographique et qui voulaient avoir une interface collaborative dans laquelle ils pourraient venir à la fois consulter mais aussi enrichir les métadonnées, et nous ça nous a intéressés aussi parce qu'on a pas forcement les ressources humaines pour, et on est pas non plus des chercheurs donc on a pas forcement cette possibilité d'aller ajouter des données scientifiques, après il y a eu du travail via mon collègue pour rechercher les outils [..] de réaliser ça et puis il a fini par porter son choix sur Omeka [..] et donc on a commencé à faire des expérimentations sur un serveur que l'université nous a prêtés en local ,voilà et ce projet-là a eu un coup d'arrêt au moment où il y a eu un changement d'organisation et un changement de directeur au niveau du pôle, le nouveau directeur est arrivé et il a voulu que ce projet soit porté par lui, il a voulu revoir effectivement tout ce qui a été fait pour

avoir un projet vraiment, je pense que dans son idée c'est d'avoir plus quelque chose qui soit au niveau d'un portail dans lequel il n' n'y a pas forcement que les photos, il voulait ouvrir avec beaucoup plus de choses à proposer et surtout avec une identité forte que soit marquée l'université Nice Sophia Antipolis, un portail maison en local qui soit avant tout une vitrine pour l'établissement et ça passe avant malheureusement l'intérêt promis qu'on avait choisis, qui était avant tout la diffusion à un plus grand nombre d'une manière collaborative avec les standards du web sémantique, voilà mais tout le travail qu'on a fait, il va servir, il est là, il est présent, il a fait un grand travail technique, [..] il a fait un grand travail, on a numérisé en couleur et en définition les photos, les métadonnées sont, on n' a pas les 6000 photos, mais on a déjà bien avancé sur ça avec des chercheurs, on a un corpus Usad, un milliers de photos pour lesquelles les métadonnées sont enrichies et traduites en anglais et dans des langues asiatiques

Vous utilisez quoi comme langages documentaires pour décrire les collections numériques

Aujourd'hui on est dépendant de notre SIGB pour tout ce qui est signalement de ressources numériques, donc c'est Alef, et on est dépendant du Marc format marc complètement, on a quand même grâce au travail du Sidoc, ce sont des choses qui permettent de signaler dans notre catalogue sachant qu'on n'a pas un catalogue de dernière génération, donc c'est assez compliqué mais on arrive à signaler tout ce qui est dans calames, on arrive à signaler aussi tout ce qui est dans HAL, voilà et pour le reste on s'appuie complètement sur Marc, voilà sachant qu'après pour la bibliothèque numérique le travail qu'on a commencé à faire c'était, lui il maîtrisait vraiment bien tout ce qui est Rdf SKOS FOAF etc. donc à la fois on a sélectionné du rdf c'était une évidence, après c'était tout le travail autour du vocabulaire avec une sélection qui est une sélection, on voulait dans un premier temps vraiment suivre les recommandations w3c avec les grands vocabulaires mis en place, donc Geonames pour les lieux, SKOS pour les sujets, FOAF pour les relations des personnes et dublin core pour la description, on a envisagé à un moment TEI mais on est pas allé beaucoup plus loin, TEI c'est très utilisé par les chercheurs en fait et à l'étranger notamment, mais on est pas allé plus loin et on s'appuyait essentiellement sur ce que je viens de vous dire là

Est-ce que vous avez commencé à utiliser ces vocabulaires

On avait fait effectivement des essais en interne, et le choix pour lequel on était un peu bloqué c'était au niveau des URI les identifiants, mais au niveau du vocabulaire l'architecture Rdf on avait effectivement commencé à faire des graphes, Étienne Cavalier a commencé de faire des graphes qui étaient intéressants et qui permettaient effectivement de mettre en valeur toutes les métadonnées et toutes les données qu'on avait déjà pu constituer avec des exports en Json, pour moi ça a été quelque chose de vraiment décevant pour Étienne Cavalier, aussi parce qu'il a fini du coup par partir, il y a ça et il y a eu d'autres choses, mais il est parti à la Bnf et maintenant il profite bien de l'expérience data.bnf, ici on a arrêté du coup le jour au lendemain suite à la nomination de cette personne, et aussi on a eu des soucis avec l'INRIA, parce qu'on a un laboratoire INRIA qui est installé aussi à l'université, et qui considère en fait qu'ils ont le monopole de tout ce qui est web sémantique, [..] , ils nous ont pris de haut, [..] donc il n'y a pas eu de possibilité parce qu'au départ on est parti sur une idée de coopérer […] en fait c'est simplement pour pouvoir justifier et déposer des projets à leur nom et ça leur permet d'avoir du budget, voilà au lieu de coopérer on a senti qu'il y avait un blocage, c'est-à-dire ils voulaient piloter le projet, et nous on sera simplement des associés, [..] qu'on fasse le travail du fourmi, c'est vraiment une posture qu'ils ont ici à l'INRIA pour le web sémantique, du coup on a senti qu'il n'y a pas de possibilité de coopérer, par contre c'est mieux passé avec la DSI la direction de service informatique, parce qu'ils nous ont mis à disposition un espace sur le serveur pour qu'on puisse, ils nous ont accompagné pour l'installation d'Omeka et tout ce qu'on a besoin en termes de stockage etc.

Vous avez parlé des métadonnées, vous avez dit que c'est les chercheurs qui enrichissent les données

Alors notre idée était d'avoir une interface collaborative un peu sur le mode de ce qui fait le musée à New York, ils ouvraient effectivement la possibilité aux internautes de venir enrichir sur le profil Wikipédia aussi, on avait l'idée effectivement de faire quelque chose qui soit accessible avec login et mot de passe, mais ça on a jamais pu le mettre en place, ça reste quelque chose que je vais essayer de défendre par rapport au nouveau projet, […] mais je pense que c'est essentiel de faire ça, l'idée était d'avoir un login et un mot de passe pour les chercheurs qu'on a déjà identifié, et d'autres qui voudraient venir passer sur l'interface une fois qu'ils découvrent ce qu'il y a dedans, de toute façon pour tout ce qui est spécialiste de photos, parce qu'on a commencé avec les photos qui sont plutôt à l'étranger, qui sont plutôt des anglo-saxons c'est pour cela qu'on a besoin de multilinguisme, on a tout de suite fait en anglais puis dans d'autres langues, parce que effectivement il y a des chercheurs de ces pays-là qui sont amenés d'apporter une valeur très importante.

Je passe à la troisième partie sur le web sémantique, vous préférez utiliser le terme web sémantique ou web de données

Je préfère utiliser le terme web sémantique parce qu'il est plus ambitieux, sachant qu'on est loin, c'est une belle utopie en général, je trouve que les bibliothèques de toute façon depuis les antiquités ont fonctionné sur des utopies, la bibliothèque d'Alexandrie a fonctionné sur l'utopie de donner accès à tous les savants dans un espace unique, ça été aussi le cas pour Otlet et la fontaine de l'institut bibliographique universel, c'est un peu vers quoi le bibliothécaire essayait d'avoir un jour, un endroit où les savoirs humains soient accessibles à tous, voilà Internet c'est un outil qui pourrait permettre ça, ça ne sera jamais réalisable mais ça pourrait permettre de s'approcher en tout cas, pour tout ce qui est du web libre après pour le reste, ça reste beaucoup plus compliqué.

Quels sont selon vous les enjeux de l'utilisation du web sémantique en bibliothèques ?

Je pense que c'est tout simplement si on rate le virage on va avoir de moins en moins du poids dans la société, c'est l'occasion de redonner aux bibliothèques et aux établissements culturels en générale une visibilité importante sur le web, beaucoup plus de diffusion du savoir des contenus culturels, enfin je pense il y a une place plus importante à jouer dans la société, c'est effectivement de faire en sorte que les valeurs véhiculées des bibliothèques en termes d'accès aux savoirs pour tous, en termes de tolérance etc. ce sont des choses importantes que l'Unesco et l'IFLA ont souligné ces dernières années. On a un rôle aussi de devenir des acteurs du web sémantique, parce que c'est essentiel ce n'est pas seulement être présent sur le web sémantique et être visible, à mon avis être acteur du web sémantique, essayer de faire entendre notre voix auprès des instances qui font l'architecture du web auprès w3c, essayer de peser sur les choix stratégiques pour essayer d'améliorer la société tout simplement ça fait partie aussi du rôle et des missions des bibliothèques publiques, je pense au contraire si on rate effectivement, l'enjeu si on n'arrive pas à relever le défi on risque de devenir des établissements de seconde zone, de moins en moins fréquentés, on ne sera plus un reflex ce qui est le cas un peu aujourd'hui, on est plus un reflex pour la recherche d'information ou la recherche de savoir, les gens se tournent vers les moteurs de recherche, et effectivement aujourd'hui on passe au second plan pour beaucoup de personnes [..]

Il faut les mettre en valeur

Oui c'est ça, c'est certain que quelqu'un quelque part dans le monde est intéressé par la collection qu'on a, même si on ne trouve pas quelqu'un sur place, c'est un peu le problème qu'on a eu ici parce qu'on avait le laboratoire spécialisé sur l'Asie, les chercheurs sont partis à la retraite, les jeunes chercheurs de ce laboratoire ont décidé de s'installer à Marseille, donc on n'a plus un laboratoire spécialisé sur l'Asie, alors que ce fonds avait été constitué par le laboratoire en question, et donc on a aujourd'hui un enjeu local de faire travailler des jeunes doctorants et des chercheurs sur nos collections, ça on a réussi à le faire, mais on a surtout un enjeu international qui tout simplement de

faire connaître notre collection, de faire en sorte que les chercheurs du domaine ou toute personne puisse accéder à ce qu'on a ici, voilà ça c'était aussi le discours qu'avait l'équipe data.bnf, c'est-à-dire il faut d'abord commencer par une bibliothèque numérique en mettant les corpus pour lesquels on a une vraie valeur à rajouter, on va venir enrichir le web parce qu'on est les seuls à avoir ces contenus-là.

Qu'est-ce que c'est le web sémantique pour vous ? Une technique, une évolution une révolution

C'est une évolution du web qui est naturelle, parce qu'elle est liée à tout le travail technologique des acteurs du web, mais ce n'est pas une révolution c'est une opportunité pour les bibliothèques, c'est comme toutes les évolutions qu'on a pu avoir depuis que les bibliothèques existent, c'est simplement se saisir de cette opportunité et puis de faire en sorte qu'elle nous permette de mieux réaliser nos missions, je pense avant d 'être quelque chose de technologique le web sémantique c'est d'abord quelque chose de stratégique, fondamentalement l'idée de se dire on a la possibilité aujourd'hui de passer à une dimension supérieure au niveau de la diffusion de signalement de nos ressources avec celles du web et puis de participer à tous ces mouvements d'ouverture des données publiques aussi

Le web sémantique tel qu'il est présenté par Tim Berners Lee est une évolution du web actuel, on ne peut pas transformer tout ce qui est sur le web aujourd'hui au web sémantique. Ce n'est pas réalisable, moi je pense qu'on peut appliquer les standards du web sémantique dans des milieux organisés, les bibliothèques par exemple parce qu'ils ont l'habitude d'identifier les données de le décrire etc.

Non je pense qu'on est d'accord que le web sémantique ne concerne que les données publiques et les ressources qui sont en licence libre dans les établissements culturels, on a les archives qui ont fait une grosse expérience avec des projets intéressants, peut être que dans les musées il y a aussi tout un travail à faire, qui est fait notamment par la bpi c'est intéressant, puis il y a les bibliothèques, on a quelque chose qui est fait nul part ailleurs, c'est-à-dire on a l'habitude de cataloguer on a l'habitude de créer des fichiers d'autorités on a l'habitude de faire ça à un niveau très poussé surtout en France, et data.bnf a vraiment montré tout ce qui pouvait apporter le travail du bibliothécaire, simplement utiliser tout ce qui a été fait c'est pour ça que ce n'est pas une révolution, c'est vraiment utilisé tout ce qui a été fait d'une manière très précis dans les bibliothèques et notamment les agences bibliographiques nationales, pour faire en sorte qu'exposer ça sur le web sémantique, pour faire en sorte que l'utilisateur trouve plus facilement et découvre les choses, mais ça concerne que le web public sous licence libre, [..] le web sémantique nous donne l'occasion d'exister justement par rapport aux grands géants du web qui sont du domaine privé, et nous permet d'exister un peu plus, data.bnf encore une fois ont beaucoup plus de visibilité dans les moteurs de recherche, leur statistique montre qu'il y a beaucoup de visites qui arrivent depuis Google depuis Bing beaucoup plus qu'avant .. Ce qu'ils ont fait doit servir d'exemple à tous les projets des bibliothèques numériques, malheureusement ce n'est pas encore le cas, alors ils offrent un accompagnement très intéressant par le biais de Gallica, en tout cas ils ont montré la voie, pour les bibliothèques universitaires il y a quand même quelque chose qui est aussi différent qui est liée vraiment à l'établissement de l'université, il y a la politique de l'université, il y a toutes ces questions d'affichage aussi qui sont importantes, mais il y a quelque chose que je trouve pour lesquelles les universités ont beaucoup de retard, c'est la création d'interface de type data.gouv.fr ou data.bnf, c'est-à-dire une interface qui mêle à la fois les ressources qu'on peut trouver dans la bibliothèque, dans le service de documentation et les données qu'on peut trouver dans les laboratoires, et là je pense que les services commun de documentation doivent être pilotes de ces projets-là, et devraient initier des projets d'interfaces data.univenice.fr data.grenoble.fr je ne sais pas mais vraiment qu'on soit porteur, mais malheureusement ce n'est pas la cas, soit parce que les bibliothèques se heurtent à l'organisation administrative comme ça était le cas ici, à partir de moment où on a évoqué ça on s'est pris la réaction tout de suite c'est pas à vous de faire ça [..]

Vous avez parlé des identifiants et vous avez dit que vous avez eu quelques problèmes concernant ces identifiants

Oui ça été un problème parce qu'on ne voulait surtout pas créer des URI maison, on a beaucoup travaillé avec Sébastien pour ARC et on aussi analysé PURL, mais bon PURL leur évolution n'est pas du tout bonne, parce qu'apparemment les derniers échos qu'ils ont du mal à maintenir, en fait on a quelque chose qui nous a paru aussi évident c'est les identifiants ISNI, sachant que l'Abes travaille déjà en partenariat avec ISNI et finalement aujourd'hui les identifiants ISNI sont les plus riches pour les bibliothèques, [..] mais si vous voulez on n'est pas accroché ? [..] ARC on avait la possibilité de le faire mais c'était parce qu'on avait un collègue qui était capable techniquement de le faire et qui travaillait en collaboration avec l'équipe data.bnf, donc aujourd'hui moi je vais continuer à proposer ARC, ça me semble la solution la plus intéressante, mais ça dépendra aussi de l'outil, le projet il a été remis un petit peu à plat [..] j'ai peur que le projet se transforme en un portail qui soit une coquille qui brille qui montre le logo de l'université partout , mais derrière ne soit pas du web sémantique ça sera simplement un portail avec une interface de recherche qui ne soit pas bien moissonnée par les moteurs, c'est de ça que j'ai peur.

L'enjeu le plus important pour vous c'est la visibilité sur le web, c'est-à-dire qu'un internaute puisse accéder à des contenus de la bibliothèque en utilisant un moteur de recherche sans forcément passer par le site ou le portail de la bibliothèque

Oui c'est ça on ne peut pas demander aux internautes de connaitre tous les portails, et on a l'exemple d'un portail extraordinairement riche qui est Europeana qui est complétement sous utilisé, ils font des efforts en ce moment-là pour [..]

Que pensez-vous de l'influence de la technologie sur vos pratiques professionnelles ?

Concernant les technologies du web sémantique aujourd'hui il y a peu d'impact concret sur le catalogage courant, pour l'Abes de toute façon actuellement on a juste des consignes de FRBRisation, pour l'instant on est vraiment dans le catalogage qui reste MARC avec des consignes en plus au niveau de FRBR, maintenant pour la majorité des collègues ça passe par le programme de transition bibliographique, on leur dit que le web sémantique a été pris en compte par RDA par FRBR etc et que ça va arriver sauf que effectivement il n'arrive pas forcément dans l'environnement du travail quotidien de la plupart des collègues, par contre les collègues qui travaillent sur les autorités, eux ça touche déjà un peu plus, et ceux qui travaillent au niveau des thèses theses.fr ils ont déjà intégré FRBR, [..] mais d'une manière générale il y a très peu de collègues qui sont impactés aujourd'hui par les technologies du web sémantique au quotidien.

Oui de toute façon il y a peu de professionnels qui maitrisent les standards du web sémantique en France, donc on ne peut pas passer maintenant au web sémantique sans former les gens sans avoir un temps d'adaptation ça va être compliqué

Oui c'est pour ça d'abord je pense que c'est quelque chose qui passe par la sensibilisation au niveau stratégique, c'est-à-dire malheureusement il peut y avoir parfois l'impression pour les collègues, ceux qui ont suivi des formations sur le web sémantique, certains peuvent avoir l'impression qu'on leur présente des choses techniques, d'avoir quelque chose d'abstrait et ça donne un reflex protectionniste, on dit non c'est pour moi etc. et on a pas besoin, mais je pense à partir du moment où on a un discours que c'est simplement un moyen technologique de mieux réaliser la mission d'une bibliothèque pour diffuser les savoirs etc. ça passe beaucoup mieux, après il y a beaucoup de collègues qui sont convaincus par ce discours et qui sont en attente d'une solution logicielle, et malheureusement il y a une situation d'attente c'est l'Abes et la BNF qui vont nous proposer, [..] je pense qu'ils ont fait leur travail maintenant ils ne peuvent pas imposer à des établissement de commencer à travailler sur les corpus uniques, il faut que les directeurs de bibliothèque et les responsables de fonds patrimoniaux de prendre la mesure [..] Il faut qu'il y a une prise de

conscience au niveau des décideurs, et je pense que c'est urgent que dans la formation des conservateurs qu'il y a quelque chose de solide qui soit proposée à l'enssib en niveau du web sémantique, [..] ils n'insistent pas du tout sur le côté enjeu, ils n'insistent pas non plus sur le fait que le web sémantique n'est pas quelque chose d'automatisable si on a pas d'abord les données qui sont prêtes,[..]

C'est dommage parce qu'en France il y a des fonds patrimoniaux très intéressants très riches je pense que la Bnf ils ont bien compris ça et qu'il y a derrière le web sémantique un enjeu géopolitique

Oui effectivement il y a une dimension géopolitique, on est déjà dans une sous zone pour la recherche d'information, oui il y a effectivement des géants du web qui sont des réflexes pour beaucoup, ils ont un avancé technologique par rapport à nous, par contre il y a un enjeu très important pour la France et c'est vrai pour beaucoup de culture, les cultures des pays qui sont situés en Asie, qui sont situés en Afrique, qui ont des choses très importantes et très intéressantes pour qu'il y ait une meilleure compréhension entre les peuples, quand on voit tout ce qui est capable de se propager en termes d'information sur les réseaux sociaux ou Google etc. Les bibliothèques se servent de Google pour la visibilité, elles se servent de Facebook pour la popularité, mais finalement on a jamais une réflexion, on a jamais une position par rapport à ces outils-là, et pourtant au congrès de l'IFLA l'année dernière en 2015 il y a quand même une déclaration par rapport aux données utilisées dans les bibliothèques, on s'engage à ne pas donner et de collecter des données privées, et on s'engage à des outils qui ne collectent pas des données privées, à priori on a rien à faire sur Facebook les bibliothèques n'ont rien n'à faire sur Facebook, et d'un autre coté en 2016 on a eu une déclaration d'IFLA sur la neutralité du web, [..] Google est en train d'acheter des kilomètres de Datacenter un peu partout dans le monde, ils ont rachetés des câbles sous-marins, on voit aussi qu'ils sont en train d'essayer d'avoir leur propre autonomie énergétique pour justement avoir l'alimentation suffisante de Datacenter, [..] au niveau politique nous on a une grande influence c'est au gouvernement et à l'union européenne de transformer toutes les déclarations qui peuvent avoir autour de la neutralité du web, la nouvelle loi sur le numérique sur les données personnelles etc. [..]

Oui en fait les bibliothèques ne peuvent pas maintenant juste attendre l'usager qui va venir à la bibliothèque ou bien sur leur site il faut aller le chercher où il est

Oui tout à fait mais il y a effectivement ce qui s'est passé à la BNF avec data.bnf qui est très intéressant, parce qu'il y a cette visibilité auprès des moteurs de recherche, mais je pense qu'il y a des bibliothèques qui ouvrent un compte Facebook ou un compte twitter sans aucune réflexion sur ce qu'ils mettent d'information là-dessus, et je trouve que c'est dommage d'avoir dans les ressources humaines et des efforts consacrés à l'alimentation du compte Facebook, alors qu'ils y a des collections qui ne sont pas encore signalées, je trouve quand même que c'est une situation qui n'est pas normal, mais aussi il y a le côté idéologique on peut se positionner comme des partenaires très fort du w3c et suivre les positions du w3c sur tout ce qui est données personnelles etc., mais donc voilà chaque collègue dans sa bibliothèque il se sent impuissant, il y a effectivement une forme d'impuissance, parce je pense qu'on rate quelque chose en Europe et non seulement en France, le web sémantique à mon avis c'est un virage à ne pas rater, pour nous, pour l'opendata, pour les services publics en générale, tous les établissements qui ont des ressources à proposer. Quand on voit l'interface open.gouv, elle est vide, il y a rien, quelques chiffres sur les lycées, quand on la compare avec celle de Royaume Unies, déjà ils ont beaucoup de choses plus intéressantes

Comment voyez-vous l'avenir de votre profession

L'avenir de la profession je le vois encore une fois un métier très ancien la bibliothèque mais qui a toujours réussi à relever le défi des évolutions techniques ou technologiques, on a une place dans la société que personne ne la remise en cause, donc je pense l'avenir de notre métier dépend de notre

capacité à relever le défi du web sémantique, pour notre existence sur le web, pour garder une place importante dans la diffusion des savoirs, pour être des établissements de références pour la recherche d'informations et la recherche des connaissances, voilà je pense que vraiment c'est un moment crucial, après si on rate on va payer à des prestataires qui vont faire le travail à notre place, mais peut-être on va gagner on gagnera en visibilité, mais malheureusement sur le stratégique on aura perdu on aura raté quelque chose qui pourrait nous positionner comme des acteurs pas majeurs mais importants [..] Sinon les bibliothèques en tant que lieux resteront de toute façon, l'idée qu'elles restent, mais qu'elles soient fréquentées et pour ça je pense que le web sémantique c'est le levier fondamental

Entretien 8

Réalisé le 16/12/ 2016

Durée 1h10min

Pouvez-vous décrire votre cursus universitaire et votre parcours professionnel

J'ai choisi de faire une reconversion professionnelle parce qu'à l'origine j'étais dans un domaine complètement différent, et donc j'ai passé un DUT de documentation en formation initialement c'était prévu pour 3 ans, et donc à la suite de ce DUT j'ai eu plusieurs petits contrats sur le campus notamment à la bibliothèque de l'IMAG, à l'époque c'était une fédération IMAg qui existait, c'est une institution qui fédérait 8 laboratoires de recherche dans le domaine de l'informatique et de mathématiques appliqués, et donc je suis rentrée comme contractuelle et après j'ai fini par passer un concours qui était ouvert en assistant ingénieur au CNRS, et comme ça que j'ai mis les pieds au CNRS voilà donc à ce moment-là je m'occupais simplement du catalogage basique avec un format maison un truc qui n'existe plus de nos jours, et comme j'aimais beaucoup l'informatique documentaire petit à petit en fait l'informaticien qui était sur place m'a ouvert un peu plus les portes et les accès et j'ai fini par gérer vraiment le logiciel de catalogue en ligne, j'étais un peu chargé de faire les contrôles des saisies de mettre en place des procédures de voir mes collègues etc., donc ça évoluait dans le temps, après je me suis occupée plus de la diffusion des thèses sous format électronique, on a eu pas mal des projets avec la cellule Metdoc qui sont juste derrière nous là et j'ai mis plus les pieds du côté informatique documentaire sur les projets plus large, on mettait du Dublin core dans des fichiers, du html, il y avait un robot qui tournait qui récupérait les données qui créait, une base qui permettait de faire de recherche de thèses dans les différents laboratoires de mathématique et informatique en France, et voilà j'ai continué dans ce sens et petit à petit après on a ré-informatisé aussi la bibliothèque on est passé à Alexandrie à l'époque, et puis on a rentré dans un processus où les bibliothèques associées ont été impliquées dans le processus de la ré-informatisation, et on a vraiment travaillé avec eux et exprimé nos besoins et faire valoir nos besoins pour qu'ils soient intégrés dans le paramétrage et [..] et en fait à l'institut Fourier la démarche a été similaire et on a travaillé ensemble voilà et donc je me suis aussi occupée des archives ouvertes, j'ai été un peu chargée de tous ces aspects-là, et ensuite j'ai travaillé beaucoup avec l'informaticien et j'avais envie de savoir un petit plus au niveau informatique et au niveau de développement, donc j'ai décidé de préparer DESS de compétences en informatique […] j'ai eu un congés individuel de formation pour préparer ce DESS. Et puis après je suis retournée à l'IMAG mais ça ne passait pas très bien, il y a eu des changements au niveau de la direction et moi je ne me retrouvais plus la dedans, et il y a eu l'opportunité qui s 'est présentée, la collègue qui était responsable à l'institut Fourrier partait et elle m'a proposé, j'ai longtemps hésité parce que j'avais l'impression que j'allais perdre tout ce que je venais d'acquérir parce que je ne vais pas mettre en œuvre la partie développement informatique sur lequel je venais me former, ce qui était une réalité mais voilà mais j'ai quand même pris le choix de basculer, donc j'ai pris la direction de cette bibliothèque depuis septembre 2006 et bon je ne regrette pas du tout mon choix, il y a quand même beaucoup de choses au cours de ma formation qui m'ont servi notamment sur tout ce qui tourne autour de la gestion du projet, ça était vraiment un plus mais bon malgré tout le background informatique [..] et c'est un vrai problème dans le profil de notre métier, on en enchaîne par la suite, mais il faut absolument que les documentalistes ont une solide formation en informatique […] juste pour faire une précision dans le domaine des mathématiques on a un grand réseau qui est national qui s'appelle le réseau national des bibliothèques de mathématique, qui fédèrent une cinquantaine de bibliothèques de math au niveau recherche dans toute la France, ce sont des bibliothèques diverses

et variées, je le mentionne parce que c'est très important, tout à l'heure j'ai précisé dans le fonctionnement c'est vraiment la gestion de bibliothèque au sens basique, par le biais de ce réseau on travaille beaucoup en collaboration avec les mathématiciens, et ça c'est vrai même au niveau de la bibliothèque on a toujours un binôme bibliothécaire responsable scientifique, c'est vrai aussi au niveau de ce réseau et on travaille aussi avec la cellule Matdoc qui est un service d'ingénierie documentaire, il y a beaucoup d'informaticiens [..] du coup on travaille vraiment en partenariat tous ensemble, [..] on mène ensemble des collaborations sur des projets visés nationaux, il y a tous les aspects sur l'acquisition de documentation électronique en particulier, là on a ouvert un portail d'accès à des ressources documentaires qui fait qu'on touche un petit peu à la technique, justement auquel on n'a pas l'occasion quand on est à la bibliothèque, et du coup ça permets aussi de voir un petit peu les besoins, les manques et les différences qui peuvent y avoir dans les profils dans notre métier, et à travers ce réseau il y a vraiment la volonté d'essayer de faire monter en compétences les collègues, donc on fait des réunions très régulières, [..] on avait fait intervenir Emmanuelle Bermès à l'occasion d'une de ces journées.

Vous avez suivi des formations depuis que vous avez travaillé en bibliothèque

Beaucoup

Les plus récentes

La dernière à laquelle j'insistais c'était la transition bibliographique de Sudoc avec le passage RDA et la FRBRisation, et puis après on a fait plein de formation un peu d'initiation qu'est-ce que c'est le web de données, j'ai suivi ce cours là c'est vrai que ça tombe beaucoup autour de ces thématiques là

Quel est votre titre ici à la bibliothèque

Je suis responsable de la bibliothèque, et mon statut je suis ingénieur de recherche à la CNRS

Quelles sont vos missions

Je gère la bibliothèque, les moyens humains financiers

Comment vous définissez une bibliothèque numérique ? Qu'est-ce que c'est une bibliothèque numérique pour vous ?

C'est une bibliothèque avant tout, c'est un ensemble de documents alors en mathématique on est très sensible à ces questions-là, les usages des chercheurs sont très proches des usages SHS que des usages j'allais dire en sciences dur, et j'ai vraiment vu la différence on est sur un mode de fonctionnement où les gens veulent accéder vraiment à des documents, parfois très anciens c'est une documentation qui ne périme pas dans le temps, mais au contraire souvent ils font appel à des documents très anciens parce qu'ils vont travailler sur un théorème pour essayer de prouver et de remonter dans le temps, ça c'est une grande spécificité du domaine [..] il y a une grande volonté d'assurer une pérennité de cette documentation et d'assurer cette pérennité au sens premier du terme dans le temps mais aussi sur les modalités d 'accès avec des outils qui ne soient pas propriétaires, ce qui fait que depuis longtemps la cellule Matdoc a été mise en place dans les années 90 et on travaille beaucoup avec eux sur des différents projets, il y a des projets de numérisation donc on les prête parfois des collections, il y a tout un travail qui est fait de numérisation mais aussi d'analyse et de récupération des références bibliographiques avec un enrichissement des liens vers d'autres documents et de XMLisation de tout ça pour les conserver sous forme textuel, il y a une grande sensibilité des aspects là en mathématiques et du coup la bibliothèque numérique c'est quelque chose vers lequel on va c'est évident, qu'on prépare assez longtemps dans le domaine notamment la numérisation, maintenant Matdoc travaille sur la source et la création de revues et mettre en place des plateformes open access, il y a une sensibilité à ces questions-là, et du coup nous on travaille à gérer cette bibliothèque numérique comme une bibliothèque, c'est-à-dire en créant de collections et en faisant en sorte d'indexer d'identifier, mais c'est vraiment créer une collection, on n'est pas dans les mêmes objectifs que les SCD qui avaient comme objectif de créer des grosses bases, c'est génial on va vous proposer 3000 revues mais on s'en fiche si on en utilise que 100, nous on veut que 100 c'est donc pour nous c'est aspect-là c'est vraiment créer une collection qui a un sens et non pas mettre tout pour le plaisir de le mettre, et on travaille actuellement, on est loin du côté bibliothèque

numérique, mais ça va ensemble parce que la pérennisation de l'électronique c'est un vrai problème quand on a tous ces abonnements, les SCD sont revenus sur le tout est électronique dans les premiers accords qui ont eu avec Elsevier, ils sont partis à fond et à un moment donné ils ont dit qu'est que on est en train de faire, les articles sont là chez les éditeurs ils ne sont pas chez nous. [..] du coup je dirais qu'ils sont venus assez tard à ces problématiques, je trouve que c'est très bien cette prise de conscience, ils sont en train de revenir en arrière avec tout ça et au niveau du réseau des bibliothèques de mathématique depuis 2009 on a commencé à travaillé sur un plan de conservation partagé des périodiques de mathématique, on n'est pas dans la bibliothèque numérique mais justement l'idée de s'assurer qui est dans l'idéal au moins deux collections de revues qui soient identifiées validées scientifiquement par une comité sur laquelle on s'appuie.
Je pense que la grande différence, ce qui est numérisé c'est clairement identifié, il peut avoir des travaux d'indexation pour permettre de retrouver ce document avec de l'indexation avec éventuellement de la création des métadonnées sur les documents et les ressources électroniques, ça rejoint la problématique justement de ces énormes packages où on a pas le temps de les identifier indexer etc. Donc on a externalisé tout l'aspect de recherche documentaire qui est le cœur de notre métier, en fait il est externalisé auprès des éditeurs, et quand on prend des ressources électroniques, ce qu'on paye ce sont les accès, on ne crée pas une interface qui permet de faire la recherche et de créer des collections, parce qu'on paye des accès, et c'est un prestataire extérieur qui s'occupe de tout ce qui est métadonnées et interface de recherche etc.[..].

Quels sont langages documentaires et formats utilisés pour décrire les collections numériques ou numérisés

Au niveau de la bibliothèque numérique je pense qu'à l'époque il y a avait Dublin core mais là je ne sais pas ce qu'ils utilisent. Disons que là on est en train de travailler sur le plan de conservation partagée, on est en train de travailler sur des exécrations avec l'abes et on récupère du marc, mais on a affaire avec des informaticiens qui balancent ça dans un format Json.

Est-ce que vous produisez des métadonnées

On n'utilise pas à notre niveau à la bibliothèque, il n'y a pas d'enrichissement au niveau du portail, on n'est pas encore là, on a tous conscience qu'il faut travailler la dessus, pour l'instant on n'a pas travaillé sur ces questions là

Et les personnes qui font du catalogage

Je n'ai pas précisé qu'on est déployé dans Sudoc en fait, on catalogue dans le Sudoc ensuite les notices descendent, il y a un chargement qui se fait automatiquement, on utilise le SIGB du SCD qui est Aleph, on complète on gère nos exemplaires du coup dans le sudoc, oui on catalogue en Unimarc

[..]

Vous préférez utiliser le terme web sémantique ou web de données

Je n'ai pas une préférence particulière pour moi c'est la même chose

Qu'est-ce que c'est le web sémantique pour vous c'est une évolution une révolution une technique une architecture etc.

Pour moi c'est un peu une révolution, c'est une approche différente et ça peut permettre de revoir le positionnement de nos métiers, ça peut nous donner une opportunité de valoriser nos compétences [..] on a tout un travail à faire sur les contenus, la mise en place de la gestion des référentiels et des connaissances, travailler justement sur les référentiels auteurs par exemple il y aura tout un travail super intéressant à faire. [..] Ce n'est pas une révolution qui va se faire de jour au lendemain, parce que c'est compliqué, l'Abes commence à travailler très sérieusement sur ces questions-là, sur un discours qui est de dire on a un travail, notre travail va consister à un travail plus intellectuel, on va travailler sur le contenu [..] on peut déjà commencer sur ces aspects-là d'enrichissement intellectuel, [..] ça nécessite vraiment une prise de conscience des professionnels, ça nécessite de monter en compétences que beaucoup n'en pas aujourd'hui, et en amant de tout ça, ça nécessite des formations qui seront proposées qui aillent dans ce sens, on reste sur un schéma assez classique finalement.

Ici à la bibliothèque vous n'utilisez pas les standards du web sémantique
Non
Si la bibliothèque venait à évoluer et utiliser le web sémantique quels seraient les défis ?
Ce que je disais là à l'instant ce sont les compétences des collègues, pour moi c'est vraiment le gros défi, c'est vraiment de prendre conscience que le métier évolue, après ce n'est pas nécessairement le souhait d'un certain nombre de collègues, la difficulté qu'on peut avoir dans notre domaine c'est qu'on gère une bibliothèque et que les personnes qui s'occupent de la bibliothèque, les bibliothécaires sont très souvent des bibliothécaires, donc ils n'ont pas de formation conservateur [..] j'ai envie de dire que les bibliothécaires sont souvent en manque de confiance en elles, et tous ces aspects l'informatique ça les fait vraiment peur ; aujourd'hui le web de données c'est quelque chose qui leur fait peur. Aujourd'hui il y a déjà ce problème technique, après on leur dit on va avoir un travail à faire plus intéressant sur l'enrichissement et l'indexation, et un travail sur les référentiels etc. (moi je connais pas je ne suis pas mathématicienne) … nous à notre niveau c'est encore plus compliqué d'imaginer qu'on va enrichir au niveau du contenu … c'est quasiment une double compétence qu'aujourd'hui personne ne l'a dans nos domaines, aujourd'hui c'est comme ça qu'on fonctionne dans toutes nos bibliothèques, il y a une direction double une bibliothécaire un mathématicien … et c'est un binôme qu'on trouve dans les bibliothèques et on travaille avec le service Matdoc qui s'occupe plutôt de l'informatique c'est le trio gagnant.
Quels sont les enjeux de l'utilisation du web sémantique en bibliothèque ? Qu'est-ce qu'il apporte de plus aux bibliothèques ?
Dans l'état actuel pas vraiment parce que il n'est pas vraiment mis en œuvre, même s'il y a des choses qui commencent à se faire via le Sudoc, on va bien que ça va faciliter la recherche d'information. A termes les enjeux pour moi c'est tout ce défi d'enrichissement de faciliter la recherche de l'information, et à notre niveau notre enjeu ça va être là qu'est qu'on va pouvoir apporter pour permettre cette enrichissement en termes de moyens et en termes de données voilà
Que pensez-vous de l'influence de la technologie sur les pratiques professionnelles
Elle est importante, elle est très importante, elle implique qu'on fonctionne différemment, je dirais simplement déjà le passage au niveau alimenter le Sudoc ça change beaucoup les choses, les bibliothécaires ne se retrouvent plus seules isolés dans leurs coins qui était souvent le cas dans notre domaine … La technologie dans ce cas-là, quand on commence un catalogue national, elle va avoir une grosse influence sur nos pratiques, parce qu'on s'ouvre à l'extérieur parce que du coup on va être plus visible, donc il va falloir répondre aussi à des demandes, on va travailler en collaboration avec un autre collègue, ça change beaucoup la pratique documentaire, ça ouvre tout un monde de collaboration par le biais de la technologie, on change fondamentalement nos pratiques professionnelles mais aussi notre façon d'être vis à vis les autres
Quel type de compétences mobilisées en termes de pratiques documentaires pour utiliser le web sémantique en bibliothèques numériques?

Pour moi toute la partie indexation, c'est clairement sur ces compétences là qu'on va apporter, [..] le cœur de notre métier qui va basculer je dirais, il va être même mis en valeur, j'espère que dans le web sémantique que cette partie d'enrichissement par le biais de l'indexation et de la gestion des référentiels qui va être mis en valeur
Pensez-vous qu'il soit nécessaire de suivre une formation sur le web sémantique avant de l'intégrer en bibliothèque
Je pense que c'est important et nécessaire de suivre ce qui se passe, [..] si on a un niveau technique un peu avancé on peut imaginer à quoi ça peut servir les bibliothécaires [..] pour moi la difficulté est la mise en pratique, pour l'instant ça reste un peu théorique pour les bibliothécaires
Comment voyez-vous l'avenir de votre professionnel
C'est la continuité de tout ce qui s'est passé depuis toutes ces années, aujourd'hui il y a un vrai problème de positionnement, en même temps le métier de bibliothécaire le métier de livre j'ai envie de dire au sens plus général c'est un métier qui a toujours évolué, pour moi c'est un métier

extrêmement passionnant parce qu'on est en perpétuelle remise en cause, on est en perpétuelle évolution, ça a un côté un peu angoissant parce que du coup on est toujours en train de chercher où on est où on se situe, alors ce n'est pas évident vis à vis de nos lecteurs, quand on travaille comme ça dans un domaine hyper pointu, qu'est-ce qu'on les apporte effectivement, est ce qu'ils se rendent compte du travail que ça suppose que ça nécessite,[..] Il y a un vrai méconnaissance de tout le travail qui a à coté, parce tout ce travail n'est pas visible, et ce n'est pas possible de l'expliquer, on peut essayer de l'expliquer ils ne vont pas comprendre, après je sais que le directeur de l'IMAG la fédération, il disait il faut arrêter de faire des notices dans vos coins de toute façon, elles existent dans le Sudoc, ok mais le catalogage c'est une mutualisation, il comprenait pas du tout ces aspects là et avec la gestion de la documentation électronique c'est encore plus difficile je trouve, parce qu'il y a beaucoup de chercheurs qui disent que les bibliothèques on va plus en avoir besoin tout est en ligne, on essaye de les faire comprendre que nous notre rôle ça va être justement de faire le ménage dans tout ce truc là les aider à chercher etc., mais est-ce qu'on a besoin d'autant de bibliothécaires, on centralise tout ça au niveau national, ils embauchent les bibliothécaires ils font le boulot pour tout le monde, on a plus besoin de tout ça donc voilà je trouve que c'est la grosse difficulté de notre métier aujourd'hui, et c'est vers quoi ça va dépend un petit peu des compétences de la bibliothécaire, et de sa volonté de suivre les évolutions pour ça que je dis pour moi c'est très important même si on ne comprend pas tout, il faut suivre ce qui se passe pour pouvoir s'y préparer, parce que le jour que nous va tomber dessus, soit on va nous dire vous faites ça vous appliquez dans le meilleur des cas au moins on servira à quelque chose, soit on va nous dire écoutez on a plus besoin de vous au revoir, donc voilà je trouve qu'on est dans un période un petit peu compliqué, voir vers quoi le métier va aller en tout cas je suis persuadée que si on est pas actif la dedans on va passer à côté de quelque chose, c'est comme les bibliothèques qui gèrent avec des fiches papiers, aujourd'hui il en existe encore quelques-unes mais ce n'est plus possible, je pense qu'il y a un travail énorme à faire de coté de la formation et je pense que les gens qui font les formations suivent toutes ces évolutions, mais aujourd'hui il y a un tel décalage entre ce qui se fait en vrai et vers quoi on va et puis la formation il y aura un vrai travail à faire à ce niveau-là [..] j'aimerais plus de relation entre les gens qui font les formations et les gens qui sont sur le terrain pour mieux se rendre compte de tout ce qui se passe.

Entretien 9

Réalisé le 08/02/2017

Durée 55 min

Pouvez-vous me parler de votre cursus universitaire et votre parcours professionnel
Donc je suis titulaire d'un bac littéraire et après j'ai passé trois ans de préparatoire à l'école de Chartres, et puis j'ai enchaîné en parallèle sur une licence d'Histoire et une maîtrise à l'université paris 4, ensuite j'ai préparé le concours en bibliothèque que j'ai eu en 2006. Et ensuite j'ai suivi la formation de conservateur de bibliothèques à l'Enssib, et quand je suis sortie j'ai pris mon poste à la bibliothèque publique d'information en juillet 2008, c'était le poste d'adjointe de service de la coordination bibliographique, je suis restée trois ans et demi sur ce poste là et ensuite mon chef est parti de la bibliothèque et j'ai pris la suite, donc je suis le chef de ce service qui a changé de nom à la suite de la réorganisation de l'organigramme en 2012 qui s'appelle donc maintenant « Données et Accès » voilà. Ce service a des missions qui sont les mêmes depuis très longtemps, il s'occupe en fait à la fois de l'harmonisation des données bibliographiques et des autorités produites par la bibliothèque et de l'administration de base et des interfaces de recherche, et donc c'est un service qui est assez complet sur la chaîne du livre et du document, [..] donc il se concentre vraiment sur le catalogue du coup et l'administration de la base de SIGB et l'interface de recherche, donc voilà sur les autorités on a un poste sur la classification, un poste sur les autres types d'autorités [..] deux personnes qui s'occupent de l'administration de l'interface de recherche et du moteur qui est derrière et une personne qui est chargé du projet modélisation […] La bibliothèque a un site web qui comprend des sous parties qui correspondent à la bibliothèque numérique, avant on avait quelque chose qui s'appelle web SG et web radio qui est une base de données de tous les enregistrements numériques, des débats en fait qui sont restés à la bibliothèque, c'est surtout de l'audio, et il y a les une base qui comprend les données d'accès aux documents numériques de la bibliothèque [..]
Est-ce que vous avez suivi des formations depuis que vous travaillez à la bibliothèque

Oui on a régulièrement des formations qu'on demande qu'on obtient, j'ai des formations relatives aux interfaces sur le référencement, j'ai eu de formations sur les langages XML j'ai fait de formations d'informatique documentaire à la fois sur les données et sur les interfaces, j'ai fait une formation de formateur aussi [..]
Quelle est votre fonction ou votre titre maintenant à la bibliothèque

Je suis chef de service Données et Accès et à la fois responsable du catalogue public
Qu'est-ce que c'est une bibliothèque numérique pour vous?

C'est une bibliothèque qui donne accès à des documents numériques, accès à une base bibliographique. Une bibliothèque traditionnellement est un ensemble organisé de collections, une bibliothèque numérique doit avoir une interface qui ressemble à celui d'un site web, du coup pour moi voilà ça dépend beaucoup de la nature du site, le site de la bpi ne constituera pas forcément de bibliothèque numérique, il est organisé comme une interface prioritairement pour donner accès à des documents, pour moi il faut qu'il soit organisé.
Pouvez-vous me parler des langages documentaires utilisés dans la bibliothèque ?
Là je me suis dit que je ne suis pas la meilleure interlocutrice pour vous, parce que les autres bases je ne les connais pas forcément. Voilà si je prends le web radio, le schéma d'indexation doit rassembler à un mélange Dublin Core et de l'indexation maison et les données sont issues du catalogage […] franchement pour répondre à vos questions peut être ma collègue Nathalie pourrait

être plus utile. Je suis désolée je ne pense pas que je peux répondre à cette question parce que je n'ai pas envie de dire des bêtises sur le schéma d'origine.

Un des éléments les plus essentiels dans les bibliothèques numériques est la production des métadonnées, vous pouvez me parler des métadonnées produites à la bnf

Dans ma service le catalogue n'est pas constitué uniquement des données qui viennent du SIGB en UniMARC, sachant qu'on récupère beaucoup de la BNF et on récupère également de Electre on a trois sources de récupération BNF, Electre et (). On récupère également de la BNF depuis quelques années l'ensemble des notices d'autorités, [..] on a eu un grand projet entre 2013 et 2016 où on a aligner l'ensemble d'autorités avec celles de la BNF, et il y avait bien sûr la question de la liaison avec l'identifiant et tout ce qu'on peut faire avec l'identifiant dans le cadre du web sémantique, donc ça c'est le catalogue. [..] Si vous allez sur catalogue.bpi.fr vous trouverez la liste de ces bases. Ça permet d'offrir à nos lecteurs quelques chose enrichie et du coup on utilise les bibliothèques numérique qui ont été produites en interne ou par d'autres.

Quand vous dites en interne ce sont des documents numérisés ?

Il y a les documents numérisés et il y a les documents qui sont produits en interne mais avec des sources externes, parce qu'on est en partenariat avec ces sources externes comme Open Edition, on est en partenariat avec Open Edition qui fait qu'on a accès à la ressource parce que on est producteurs des données dans la base, en fait on récupère de la base externe des choses qu'on produit chez nous, de même pour les documents qui sont dans HAL. Sur le système de recherche, donc il y a un schéma de données qui est basé sur le Dublin Core simple auquel on ajoute un certain nombre de champs pour coller un petit peu à nos besoins puisque évidement le Dublin core ne contient pas toutes les richesses des métadonnées qui sont produites en Unimarc (). Sinon pour les notices d'autorités on a décidé de conserver le format Unimarc [..] On va exposer les notices d'autorités sur l'interface on va en faire des points de rentrée pour la recherche. En l'occurrence pour les autorités on continue quand même d'en créer à la bpi parce que malheureusement la BNF ne recouvre pas l'intégralité des besoins, des publications étrangères c'est genre de chose on estime encore qu'on produit 10 % d'autorités

Vous préférez utiliser le terme Web sémantique ou web de données ?

Je pense que je préfère utiliser le terme de « web de données » [..] Le projet qu'on a mené à la bpi autour du web de données et le web sémantique [..] il y a le web magazine et le site professionnel et donc l'idée était de travailler donc mon service et le responsable du projet web magazine de poser les bases, c'est-à-dire une correspondance en fait de vocabulaires communs entre le catalogue et le site de la bpi, on a choisi le Dublin core pour les sites et puis on a décidé d'utiliser l'ontologie FOAF pour tout ce qui est données d'utilisateurs, on a aussi décidé à cette occasion-là d'utiliser les identifiants Arc pour les articles produits dans le cadre des sites, puis on a au niveau de l'interface on a décidé d'utiliser des schéma de schéma.org ou de faire du balisage en html en utilisant donc le web de données, [..] Je pense que je n'exprime pas bien en fait je suis fatiguée

Quels sont les enjeux de l'utilisation du web sémantique en bibliothèques numériques ?

L'enjeu d'utiliser le web de données en bibliothèques en générale c'est de pouvoir lier les données entre elles et de permettre d'enrichir l'expérience de l'utilisateur, ça permet d'enrichir nos entrepôts même si ce n'est pas forcément l'objectif principale avec des données tiers, et éventuellement de se procéder à des alignements sur la base de ces données tiers. Et en particulier pour la bpi pour l'instant on a juste posé les bases, on a pas créé les liens qu'on aimerait créer avec d'autres applications, ça fait partie du projet initial et qu'on a pas les outils pour, mais on aimerait beaucoup à l'avenir exploiter les identifiants qu'on a mis en place du côté du site et du catalogue, pour l'instant on a l'idée d'en faire quelque chose à l'avenir pour enrichir les interfaces de recherche, ça peut aussi nous permettre par l'analogie des algorithmes et le FRBIsation de créer des données qui n'existent pas dans l'entrepôt à partir des données extraites d'ailleurs .

Et l'interopérabilité ?

Tout à fait pour les raisons que j'ai déjà évoqué c'est quand même l'intérêt principale, c'est déjà pour

l'instant c'est un défi en interne pour essayer nous rendre interopérables entre nous, c'est-à-dire entre les différents silos mais c'est une dimension assez importante du projet, ça permet aussi de réinterroger les pratiques et de structurer les choses à l'intérieur de l'établissement, pour l'instant je trouve que ce n'est pas mal [..] l'idée c'est qu'on puisse un jour proposer à nos usagers des services enrichis sur la base des liens internes qui ne reposent pas forcément sur le mécanisme du web de données parce qu'on est pas forcement obligé d'utiliser ce type de mécanisme, bon jusqu'à là ça nous a servi à adopter les bonnes pratiques voilà et la deuxième chose effectivement qu'on puisse un jour publier nos données dans le web de données et que d'autres puissent se lier à nous et que nous puissions récupérer les données des autres.

Les défis à l'utilisation du web sémantique ?

Je ne suis pas sûre comment répondre à cette question-là, au niveau des difficultés je pense qu'il faut acquérir des compétences sur le sujet, les compétences en interne ne suffiront pas […] la difficulté ça consiste déjà de pouvoir mobiliser du temps de travail pour se pencher sur ces problématiques, pour essayer d'en tirer des leçons en interne, deuxièmement joindre des projets ou des groupes de travail qui travaillent sur ces problématiques, il faut avoir du temps et des moyens pour ça et puis troisièmement il faut avoir certains moyens financiers de lancer dans ces genres de chose, parce que dans le cadre d'une réflexion sur les autorités du catalogue on avait réfléchi, qu'est-ce qu'on pouvait faire avec nos autorités et on a contacté des sociétés, et effectivement les sociétés qui proposent des outils puissants d'alignements basés sur les technologies du web sémantique les vendent très cher, il faut avoir un budget conséquent et après il faut qu'on puisse maintenir le projet sur la durée. Pour l'instant à la bpi on est dans une phase de stabilisation dans certain nombre de choses déjà basiques on s'est pas encore lancé dedans mais on sait que le jour on voudra le faire on aura déjà au moins préparé le terrain on peut dire.

Que pensez-vous de l'influence de la technologie sur vos pratiques professionnelles ?

Pour un service que le mien, elle est assez essentielle [..] pour moi c'est tellement centrale c'est le cœur de mon métier si vous voulez.

Quels types de compétences mobiliser en termes de pratiques documentaires pour l'utilisation du web sémantique en bibliothèques numériques

Ça nous touche parce que c'est lié à des projets des agents bibliographiques nationales et de l'évolution des normes, [..] il faut de la rigueur et du suivi derrière la pédagogie je ne sais est ce que c'est une réponse qui vous parle. Après effectivement chacun son niveau c'est-à-dire la bpi c'est une bibliothèque qui ne fait pas partie d'un réseau, donc nous si vous voulez notre optique on regarde ce que fait la BNF on regarde ce que fait l'Abes on est très content de voir que les optiques de ces agences bibliographiques nationales convergent de plus en plus, et on essaye de récupérer de part et d'autre, nous on a été très heureux de voir l'ouverture d'un service comme data.bnf.fr, on a été très heureux de voir l'ouverture de Idref pour récupérer des données, et pour ne pas à créer ce que est déjà été créé ailleurs, et effectivement c'est une question intéressante parce que ça suppose une certaine évolution du métier, du catalogage qu'on peut faire en local, à la personne qui va surtout chercher des liens, chercher à établir les bonnes connexions entre les données. Mais je suis tout à fait d'accord avec vous effectivement dans mon service c'est l'optique qu'on a choisi, quand je vous ai parlé de fait qu'on a arrêté de produire nos autorités en interne pour décider de les récupérer de la bnf, de regarder qu'est ce qui se passe de coté de l'Abes et après éventuellement d'enrichir les choses, et en tout cas en se basant sur des référentiels et du coup en adoptant les normes et en introduisant le plus des identifiants Arc etc. [..] Je pense d'ailleurs même les acteurs du web sémantique eux-mêmes disent que ce n'est pas forcement la réponse à tout, en fait c'est un très bel outil mais il a quand même quelques défauts sur les calculs de SPARQL.

[..]

Comment voyez-vous l'avenir de votre profession ?

Je pense que les bibliothèques peuvent aller dans deux directions entre les collections et les usagers. Du point de vue usager il y a tout un mouvement qui va de la bibliothèque comme troisième lieu à

la proposition des ateliers etc. qui je pense devrait se développer si on veut que la bibliothèque continue à prendre une place dans l'espace publique ça c'est du coté d'usager.
Et côté collection, la bibliothèque dispose des collections patrimoniales importantes sur lesquelles on peut continuer à avoir une certaine expertise, et il est aussi extrêmement important l'aspect qu'on puisse valoriser l'expérience très importante qu'on a dans la description des collections, c'est-à-dire que les bibliothèques ont des années d'expertises en description, on a élaboré des normes et des formats extrêmement riches et l'enjeu est d'arriver à conserver cette richesse et à l'améliorer en partenariat avec d'autres filières professionnelles avec aussi le monde privé.
Je pense que les bibliothécaires ont aussi une place et ils continueront d'avoir une place, en tant qu'expert en métier de livre il faut qu'on travaille sur les forces qu'on a et pas forcément sur les collections en tant que telles en fait.

Entretien 10
Réalisé le 01/03/2017
Durée 1h

Pouvez-vous me décrire votre cursus universitaire ainsi que votre parcours professionnel ?
J'ai un cursus universitaire de mathématiques avant de découvrir les métiers des bibliothèques par hasard à la faveur de vacations étudiantes. J'ai donc passé ensuite le concours des bibliothèques. Et je suis aujourd'hui bibliothécaire. J'ai toujours travaillé en bibliothèque publique, donc dans la fonction publique territoriale. Donc j'ai commencé en bibliothèque publique tout d'abord en région parisienne puis à la bibliothèque de Lyon. Et depuis 5 ans je suis chargé des projets numériques à la médiathèque départementale du Puy de Dôme. Et plus particulièrement depuis un an je suis chargé du projet de réinformatisation et d'une plate-forme de service numérique à destination des bibliothèques du département [..]. Et je suis pilote du groupe de travail sur l'évolution des systèmes d'information et des formats de données pour le programme national « transition bibliographique ».
Quel est votre titre à la bibliothèque ?
Chef de projet système d'information.
Et quelles sont vos missions ?
Actuellement nous menons un projet qui va se concrétiser prochainement par une consultation pour la mise en place d'une plate-forme de service pour les professionnels des bibliothèques. Donc avec un SIGD mutualisé, un hub de données et puis différentes fonctionnalités professionnelles. Donc je suis chargé de suivre ce projet mais par ailleurs je suis aussi en charge de l'analyse des processus de traitements de données à la médiathèque et de leur optimisation, y compris dans les logiciels actuels.
Qu'est-ce que c'est pour vous une bibliothèque numérique ?
En ce que me concerne, j'ai toujours milité pour l'insertion du numérique comme un support, un nouveau média parmi d'autres. Après la diversification des bibliothèques et desmédiathèques (CD, DVD), il m'a semblé que la migration vers la bibliothèque en ligne n'est pour moi qu'une poursuite de la diversification des modalités d'accès à l'information en bibliothèque. Donc je ne mettrai pas forcément la bibliothèque numérique comme un espace autonome complètement scindé de la bibliothèque dans ces différents services et ces différentes composantes.
Vous trouvez que c'est un service qui est devenu nécessaire ou obligatoire ?
Oui tout à fait. Je pense que c'est complètement nécessaire. Il me semble que ça doit s'inscrire tout d'abord dans le grand mouvement de l'e-administration. Le fait de pouvoir rendre des services à distance est aujourd'hui indispensable pour les administrations. Et dans la mesure où les bibliothèques donnent accès à des collections, une partie de leurs collections doit pouvoir être accessible en ligne. Et d'autre part c'est en phase avec l'évolution des pratiques culturelles puisqu'on constate, dans le domaine de la musique évidemment mais aussi dans celui de la vidéo et dans d'autres également, que l'usage de collections numérisées est déjà bien répandu donc c'est aussi une manière pour la bibliothèque de prendre acte de l'évolution des pratiques culturelles tout simplement.
La question suivante concerne les langages documentaires utilisés dans les bibliothèques pour décrire les collections numériques.
En ce qui nous concerne, je penche quand même pour des alignements en ce qui concerne les langages documentaires utilisés entre ce que nous avons développé pour les collections physiques et ce qui serait utilisé pour les collections numériques. De toute façon, l'enjeu du bibliothécaire de faire collection, c'est-à-dire de construire une collection, mais aussi de la décrire et de constituer des points d'accès avec les outils numériques. Simplement, nous sommes beaucoup moins maîtres

des métadonnées qui nous sont fournies parce que nous nous abonnons le plus souvent à des bouquets ou à des catalogues entiers, là où auparavant, quand on acquiert titre par titre, avec le document en main on était capable de refaire une description plus précise. Mais je pense que d'ailleurs quand on regarde ce qui nous est proposé par les éditeurs, la notion de métadonnées, les métadonnées avec lesquelles ils travaillent sont loin d'être aussi précises et efficaces que celles qui ont été développées par les documentalistes, les archivistes ou les bibliothécaires. Donc à mon sens, c'est à nous de nous emparer de ces ressources. C'est ce qu'on fait actuellement de notre côté. Par exemple, on est abonné à des bouquets, très présent en bibliothèque, comme ceux d'Arte VOD par exemple. Quand on voit les informations qui nous sont transmises, évidemment on est loin de la gestion finie des points d'accès, des fichiers d'autorité ou de l'indexation matière qu'on peut développer en bibliothèque, donc c'est à nous de partir de ces éléments et de construire des alignements. En ce qui nous concerne, on utilise le langage RAMEAU, donc à nous de voir comment les Tags proposés par l'équipe d'Arte VOD, qui sont dans une logique de Tags grand public et qui surtout ne constituent pas une liste contrôlée, peuvent être rapprochés d'un langage plus scientifique comme le nôtre. Après il y a évidemment des langages, des spécificités, des objets numériques qui nécessitent de développer certains thésaurus ou certaines métadonnées propres aux objets numériques, que ce soit la gestion de droit ou d'autre chose. Mais à part ces quelques spécificités, pour le reste, il me semble qu'il faut plutôt aller vers des alignements et une convergence des langages documentaires plutôt que de construire des outils spécifiques.

Vous avez parlé des métadonnées. Or l'un des éléments essentiels d'une bibliothèque numérique est la production de métadonnées, leur collecte et leur partage. Est-ce que votre bibliothèque produit des métadonnées et est-ce que vous avez une stratégie pour la production des métadonnées dans l'avenir ?

Oui puisqu'on s'est positionné à la bibliothèque départementale, en tant que bibliothèque support au service des autres bibliothèques publiques du département, comme un pivot de constitution et de mise à disposition de métadonnées documentaires. Effectivement, il y a celles qui sont récupérées, soit auprès des éditeurs eux-mêmes, dans ce cas-là avec peut être un niveau d'exigence et de qualification un peu moindre que celui qu'on peut avoir en bibliothèque. Il y a celles qui peuvent être récupérées auprès des agences bibliographiques. Et il y a celles que nous avons décidé de produire. Il y a deux solutions, soit on récupère les métadonnées des éditeurs et on fait un travail de qualification et de fiabilité en les rapprochant de référentiels que nous utilisons nous, en considérant que leur indexation est bonne mais en les rapprochant d'un référentiel plus générique et que nous utilisons. Ou bien elles sont totalement absentes comme par exemple dans certaines collections qui peuvent être fournies par des agrégateurs comme « Tout apprendre », qui fournit des collections liées à des didacticiels, à un ensemble de modules d'auto formation qui sont de qualité. Là par contre les métadonnées sont extrêmement pauvres, donc là il va falloir les créer de toutes pièces. C'est un chantier qu'on a prévu d'engager dans le cadre du futur logiciel qu'on doit se doter à partir de début 2018. Dans ce cadre, ce qu'on a commencé à faire c'est aussi de former les bibliothécaires à la nouvelle modélisation FRBR et évidement aussi aux nouvelles règles de catalogage RDA puisque leur but est aussi de prendre en compte ces supports numériques pour lesquels les anciennes ou actuelles règles de catalogage sont de moins en moins pertinentes. Évidemment, quand on se pose des questions aussi basiques que le lieu ou la date de publication, etc., ce sont des choses qui sont beaucoup fluides ou labiles dans l'univers numérique et qui sont difficiles à déterminer. Après, il y a aussi des choses plus essentielles pour la recherche des utilisateurs qui sont plutôt l'indexation sujet.

Ce qui est très important c'est la formation des bibliothécaires. Lorsque j'ai interrogé des bibliothèques, principalement des bibliothèques municipales, c'était un des problèmes. Le personnel n'est pas vraiment formé à utiliser que ce soit les standard du web sémantique ou autres. Il y a une peur, une méfiance à changer de standard, d'habitude. D'où l'importance de la formation au sein des bibliothèques ainsi qu'au sein des établissements d'enseignements

supérieurs dans lesquels il devient possible d'apprendre ces nouveaux langages et standards.
Tout à fait. Il y a eu tout d'abord la nécessaire formation du personnel à l'usage et aux modalités de ces ressources, car télécharger un ebook pour un bibliothécaire qui ne l'a jamais fait ça posait évidemment des soucis et si les bibliothécaires ne sont pas capables de le faire eux même c'est illusoire de penser qu'ils vont pouvoir le proposer au public. Ensuite une fois cette formation faite, et encore elle n'est pas terminée, il reste une partie de la profession qui n'est pas à l'aise avec l'usage de ces ressources électroniques. Une fois cette première étape réalisée il est évident qu'il faut avancer vers une nouvelle structuration de l'information bibliographique et qu'il faut former les bibliothécaires, alors non pas à ces nouveaux langages, comme vous l'avez mentionné, qui ne sont pas une profonde révolution de leur paradigme de catalogage puisqu'au contraire dès qu'on touche du doigt la question des identifiants, des URI etc. on se rend compte qu'on retrouve des préoccupations de bibliothécaire, simplement il faut les réinterpréter dans un nouveau modèle.
Vous préférez utiliser le terme web de données ou web sémantique ?
En ce qui me concerne, je privilégie plutôt les termes web de données et linked data, où dans certain cas linked open data, parce ce que je le trouve effectivement plus précis et à la fois plus circonscrit même si peut être que ça donne l'impression d'être moins ambitieux dans un premier temps mais du coup ça me semble plus réaliste et ça me semble être plus en phase avec justement ce que vous évoquiez, c'est-à-dire les missions et les activités des bibliothécaires. Mais j'ai rien contre l'usage du terme de web sémantique et je présente aux collègues le web sémantique comme étant la boîte à outils qu'on va mobiliser pour aller vers le linked data et donc l'objectif je le présente plutôt comme étant le linked data. Aussi comme vous le disiez parce que le terme de web sémantique a donné lieu à des interprétations différentes. Selon les experts du web, on place derrière le web sémantique des enjeux qui peuvent être différents. En revanche, par rapport aux bibliothécaires dont la mission est finalement la modélisation, l'organisation et la mise à disposition des savoirs sous toutes leurs formes, je trouve que la question de présenter ça sous l'angle de linked open data me semble être un pas plus facile à franchir et plus opérant parce qu'on remet au cœur la question de la métadonnée et des relations qu'on construit entre ces métadonnées. Et ce sont ces relations qui font sens, et donc sans aller chercher l'adjectif sémantique, en parlant de relation on rejoint tout à fait la modélisation FRBR etc. et en ajoutant le terme « open » ça nous remet dans une posture d'agent public constituant une connaissance au service de tous. Donc c'est aussi pour derrière ces valeurs là que je privilégie le terme de « linked open data ».
Quels sont les enjeux du web sémantique en bibliothèques ?
Tout d'abord je dirai qu'il faut quand même s'intéresser au web sémantique mais aussi à d'autres technologies du web comme éventuellement à termes les technologies du big data ou des bases de données non relationnelles qui peuvent offrir des manières différentes de gérer l'information. Le web sémantique est ainsi une boîte à outils parmi d'autres qui participent à l'évolution de notre métier. Cela étant pour moi, l'enjeu qui est derrière c'est de mettre d'une part nos collections sur les parcours des internautes parce qu'actuellement on est bien trop souvent dans le web profond dans certains cas ou alors très mal référencés dans d'autres. L'enjeu est aussi de passer d'un paradigme où on était centré sur le document à un paradigme où on est centré sur l'information et sur la donnée. Aujourd'hui nos catalogues ont été construits comme le reflet des collections physiques, donc pour trouver un document, or le point de départ d'un internaute n'est pas forcément d'accéder à un document. Il recherche une information et à partir de cette information qui peut être contenue dans un document, il va mobiliser des documents éventuellement ou bien s'arrêter là ou rebondir sur autre chose etc. Le deuxième enjeu est donc de passer du catalogue centré sur le document au catalogue centré sur l'information. Et puis enfin, le troisième enjeu c'est de décloisonner nos collections parce qu'aujourd'hui on n'a été soit redondant, on saisit des informations peut être autrement mais qui peuvent être déjà produites par d'autres, les éditeurs ou des communautés d'internautes par exemple, ou bien il y a celles qu'on saisit et qui peuvent avoir un grand intérêt comme l'indexation matière ou la gestion de fichier d'autorité extrêmement fiable, mais qu'on est

pas capable de partager parce qu'on a un langage ou des bases de données qui ont du mal à communiquer avec d'autres. Donc cette mise en relation de nos contenus à la fois pour nous enrichir nous même, pour profiter de ce qui est fait dans Wikipédia par exemple, mais aussi pour redistribuer ce qu'on produit, me semble essentielle. De ce point de vu là, je rejoins les points de vue sur la bibliothèque qui ont déjà été exprimés de longue date bien avant la prédominance du numérique, à savoir que la fréquentation d'une bibliothèque n'est pas un indicateur de réussite, car si la seule fréquentation était indicateur de la réussite peut être qu'on pourrait mobiliser dans la bibliothèque tout un tas de choses qui font venir la population et qui n'ont rien à faire dans une bibliothèque mais qui marchent beaucoup mieux que les livres. L'indicateur de la réussite c'est ce que les citoyens qui fréquentent les bibliothèques vont faire de cette richesse qui est mise à leur disposition dans le domaine de leur loisir, culturel, de la connaissance, de leur participation citoyenne etc. Et bien c'est pareil, les informations qu'on construit pour un catalogue on ne les construit pas uniquement pour faire venir les gens sur nos catalogues. Si ces informations qu'on construit sont disséminées et qu'elles servent à enrichir d'autres pratiques ailleurs alors l'objectif est atteint. C'est vraiment pour avoir une vision de la bibliothèque qui n'est pas centrée uniquement sur la fréquentation de son équipement ou de son catalogue mais sur le fait de construire et de produire de l'information et de la richesse au service des citoyens quel que soit leur pratique et leur assiduité à nous fréquenter.
[..].

Dans votre bibliothèque vous utilisez des standards du web sémantique ?

Pas encore mais c'est prévu. Nous avons toujours prôné l'utilisation des identifiants produits par les agences bibliographiques en l'occurrence la BNF, donc de ce point de vue-là à partir du moment où déjà dans nos données on conserve un certain nombre d'information qui aujourd'hui sont insérées dans le web de données, je pense évidemment aux autorités de la BNF, au langage d'indexation RAMEAU etc., donc sans utiliser encore aujourd'hui les technologie du web sémantique on a quand même stocker dans nos données un certain nombre d'outils qui nous serviront à migrer vers le web de données. En tout cas, il est prévu dans nos prochains projets d'aller vers la modélisation FRBR le code de catalogage RDA et d'autre part d'exposer nos données en RDA ça sera pour 2018.

Si la bibliothèque venait à introduire l'utilisation du web sémantique, quels seraient les principaux défis ?

Il y en a un qui concerne la formation des collègues pour essayer de repérer quels peuvent être les enrichissements sémantiques et les alignements qui peuvent être utiles de faire, c'est-à-dire qu'il faut apprendre aux bibliothécaires à devenir non plus seulement des catalogueurs mais aussi des personnes capables de mettre en relation des données qui font sens mais qui sont issues de silos différents et qui ne sont pas produits au sein de notre communauté métier, donc c'est une nouvelle habitude à prendre, comment les juger fiables et comment ensuite les mettre en relation. Et puis l'autre aspect en termes de défi c'est qu'il ne faut pas demander à chaque bibliothèque de refaire ce qui a déjà été fait soit à un niveau national ou bien départemental. Et que ce qu'on n'a pas encore vraiment réussi à faire, c'est-à-dire de mieux partager nos données, il faut le faire, par exemple la seule exposition des données dans data.bnf, si on utilise leur notice ça permet de ne pas avoir à refaire ce travail-là. Mais ça on a encore peu de moyens de travailler avec une partie des données qui pourrait être dans les nuages, comme la notion d'œuvre par exemple. Là il y a encore des défis technologiques qui au-delà des défis technologiques sont aussi des défis institutionnels parce qu'on n'a pas d'organisme centralisé, [..]. Il y a un vrai défi institutionnel de mutualisation de l'exposition des données, je pense à celles qui ont vocation à être mutualisées donc par les données locales mais justement la question de la notion d'œuvre etc.

Que pensez-vous de l'influence de la technologie sur vos pratiques professionnelles ?

Je la trouve positive, c'est-à-dire que si je compare 20 ans auparavant, aujourd'hui on a des technologies qui nous permettent d'être un peu à la hauteur de nos ambitions. C'était quand même plus délicat de travailler sur de l'échange de données il y a 20 ans alors qu'internet venait tout juste

de naître et qu'on avait des capacités d'échange et de traitement de données très limités, donc je trouve que c'est très positif tout ça.

Quel types de compétences mobilisez-vous en termes de pratiques documentaires pour accompagner l'utilisation du web sémantique en bibliothèque ?

D'abord les compétences de bases sur la question du traitement des documents, le bagage technique du documentaliste et du bibliothécaire est indispensable et je pense qu'il était un peu perdu parce que les bibliothécaire aujourd'hui occupent des fonctions très diverses et il y en a qui sont plutôt du côté de la médiation, de l'action culturelle, du développement documentaire mais finalement qui maîtrisent les techniques documentaires, donc je dirai que déjà ce bagage il ne faut pas le négliger et parfois il faut même en remettre une couche pour mettre à jour les collègues. Et puis d'autre part, l'autre compétence c'est évidemment une initiation aux bases des autres langages mobilisés dans le web (html, xtml, etc.) parce que finalement on s'aperçoit que ce sont des langages auxquels la majorité des bibliothécaires ne se sont pas beaucoup intéressés mais on peut en comprendre les bases et ensuite comme ils sont très souvent mobilisés ils offrent un terrain, une couche très souvent utilisée dans le cadre du web sémantique qui me semble important de maîtriser, au moins les bases. On a vu ces langages là uniquement du côté de la publication, la mise en page et la réalisation de sites web alors qu'il faut comprendre que c'est une manière de structurer de l'information aussi et que le comprendre peut nous permettre de gagner du temps ensuite pour abandonner le format Marc et aller vers autres choses ou comprendre pourquoi on peut présenter du Marc dans du Xml etc. c'est finalement pas hors de portée des bibliothécaires, ce n'est pas beaucoup plus compliqué que Marc. Il faut désacraliser tout ça et donner les bases.

Pensez-vous qu'il soit nécessaire de suivre une formation sur le web sémantique avant de l'intégrer en bibliothèque ?

Oui, comme tout. Je suis une pro-formation.[..] Je pense qu'il faut toujours commencer par une formation aux enjeux, au contexte etc. On se rend compte ensuite après quand il y a des collègues qui s'y intéressent cela démultiplie l'effort qu'on peut y consacrer. On est même toujours agréablement surpris de l'intérêt que cela peut susciter. J'ai vu chez des collègues renaître un intérêt pour les formations qu'on avait mis en place.

Comment voyez- vous l'avenir de votre profession ?

Il me semble qu'on va aller, mais c'était déjà le cas, vers une plus grande spécialisation entre les bibliothèques d'un côté qui seront tournées vers la médiation et l'accueil du public et d'autre part celles qui seront centrées sur la donnée et sur la gestion documentaire…

Est-ce qu'on peut considérer le traitement documentaire, l'indexation etc. une médiation documentaire ?

Oui pourquoi pas, de toute façon ceux qui s'occupent de la médiation auprès du public doivent mobiliser aussi la médiation documentaire. Et le fait de faire de la description et de l'indexation est une forme de médiation puisqu'on le fait en direction d'un public. Je pense qu'il y a des grandes familles de compétences et des conditions d'exercice qui font que ça va être difficile de se consacrer pleinement aux deux. Donc je pense qu'on va aller vers une plus grande séparation des tâches. Je peux me tromper, c'est toujours difficile de faire des prédictions en ce qui concerne le futur.

Avez-vous des questions ou des remarques à ajouter ?

Non, je pense que c'est très complet. Sur la médiation, je n'oublie pas que c'est vraiment le cœur des bibliothèques publiques. Simplement, c'est que je pense qu'il va y avoir dans la médiation, d'autant plus dans les domaines socio-culturels, la participation des usagers, avec une visée de changement social. C'est pour ça qu'il ne faut pas réduire la médiation à la mise en avant des collections ou même leur description ou leur indexation. Il y a de plus en plus un rôle de la bibliothèque comme lieu de participation parce qu'aujourd'hui la culture comme objet de consommation, on en touche les limites pour la bibliothèque, justement parce qu'on peut consommer de la culture dans bien d'autres endroit qu'une bibliothèque. Et je pense que les bibliothécaire-médiateurs vont de plus en plus dans le faire et le faire-avec les usagers. Mais là

encore comme vous le dites, la question de l'indexation peut être avec ce prisme, c'est-à-dire comment on fait participer les usagers. Et quand on mobilise Wikipédia au service de l'enrichissement des catalogues des bibliothèques on ne fait rien d'autre que de profiter de l'intelligence collective pour améliorer le travail des bibliothécaires et réciproquement.

Annexe XIII - Questionnaire destiné aux bibliothécaires

Ce questionnaire est réalisé dans le cadre d'une thèse doctorale en sciences de l'information et de la communication qui porte sur les applications du web sémantique en bibliothèques numériques. Ce questionnaire anonyme et confidentiel ne vous prendra que cinq minutes.

Je vous remercie par avance du temps que vous voudrez bien consacrer à cette enquête.

Hiba MELHEM

Doctorante en sciences de l'information et de la communication Laboratoire GRESEC, UGA Grenoble melhem.hiba@gmail.com

* Required

Partie A : Coordonnées de la bibliothèque / Votre profil

1. Nom de l'établissement *

2. Site web : *

3. Courrier électronique : *

4. Le statut de la bibliothèque *

Mark only one oval.

Bibliothèque nationale

Bibliothèque publique

Bibliothèque universitaire

Bibliothèque spécialisée

Other:

Votre profil

5. Quelle est votre profession ? *

Mark only one oval.

Bibliothécaire

Conservateur(trice)

Bibliothécaire assistant spécialisé de classe supérieure

Bibliothécaire assistant spécialisé de classe normale

Assistant(e) de conservation principal du patrimoine et des bibliothèques

Assistant(e) de conservation du patrimoine et des bibliothèques

Other:

Partie B : Les applications du web sémantique en bibliothèque

6. Comprenez-vous la notion exprimée par l'expression « web sémantique » ? *

Mark only one oval.

1	2	3	4
Pas du tout			Tout à fait

7. Comprenez-vous la notion exprimée par l'expression « données liées » ou « linked data » ? *

Mark only one oval.

1	2	3	4
Pas du tout			Tout à fait

8. Comprenez-vous la notion exprimée par l'expression « Resource Description Framework » (RDF) ? *

Mark only one oval.

1	2	3	4
Pas du tout			Tout à fait

9. Comprenez-vous la notion exprimée par l'expression « Web Ontology Language » (OWL) ? *

Mark only one oval.

1	2	3	4
Pas du tout			Tout à fait

10. Comprenez-vous la notion exprimée par l'expression « Simple Knowledge Organization System » (SKOS) ? *

Mark only one oval.

1	2	3	4
Pas du tout			Tout à fait

11. Est-ce que votre bibliothèque utilise les standards, normes ou langages du web sémantique ? *

Mark only one oval.

Oui

Non

12. Si oui, pourquoi?

13. Si non, pourquoi? (Passez ensuite à la question n°18)

14. Quelles sont les standards du web sémantique utilisés dans votre bibliothèque ? (Plusieurs réponses possibles)

Check all that apply.

URI ☐

RDF ☐

OWL ☐

SKOS ☐

JSONLD ☐

SPARQL ☐

Other: ☐

15. Est-ce que l'utilisation des standards du web sémantique en bibliothèque a changé vos pratiques documentaires ?

Mark only one oval.

Oui

Non

16. Si oui, comment?

17. Si non, pourquoi?

18. L'utilisation du web sémantique en bibliothèque numérique permet de : (Plusieurs réponses possibles) *

Check all that apply.

Faciliter l'échange des notices bibliographiques entre plusieurs institutions

Faciliter l'interopérabilité entre les données

Faciliter la publication en ligne

Valoriser le fonds numérique

Améliorer la visibilité des collections numériques

Faciliter la recherche d'information

Trouver des résultats plus pertinents

Je ne sais pas

Other:

19. Quels sont les obstacles à l'utilisation du web sémantique en bibliothèque ? (Plusieurs réponses possibles) *

Check all that apply.

Sa complexité technique

La diversité des standards

Le manque d'expertise des professionnels d'information

Les contraintes juridiques

Je ne sais pas

Other:

20. Pensez-vous que le web sémantique est important pour le futur des bibliothèques ? *

Mark only one oval.

Oui

Non

21. Si oui, pourquoi ?

22. Si non, pourquoi ?

23. Souhaiteriez-vous suivre une formation vous permettant de maîtriser les standards du web sémantique ? *

Mark only one oval.

Oui

Non