

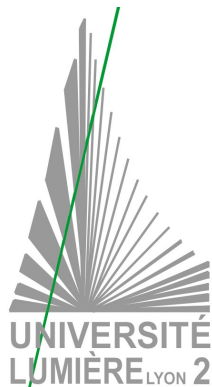


*Mémoire de recherche*

# Panorama de la bibliothèque 3.0

Par Florence TISSERAND  
Master 1 Information – Communication

Dirigé par Vincent MABILLOT



Année universitaire  
2010 – 2011

À la mémoire d'Isabelle Lux,  
*Professeur à l'IUT Infocom de Lyon 3*

*"Et de même que l'écriture n'est pas la même chez tous les hommes, les mots parlés ne sont pas non plus les mêmes, bien que les états de l'âme dont ces expressions sont les signes immédiats soient identiques chez tous, comme sont identiques aussi les choses dont ces états sont les images." Aristote, De l'interprétation.*

# Remerciements

Tout d'abord, je souhaiterais remercier mon directeur de mémoire, Vincent Mabillot, pour ses conseils, son intérêt pour le monde de la documentation et son soutien pendant cette année universitaire.

Je remercie également Jean-Claude Soulages pour son écoute et ses conseils durant les cours d'aide à la rédaction du mémoire.

Merci à Christian Cote et Mabrouka El Hachani, Maîtres de conférences en sciences de l'information et de la documentation à l'Université Lyon 3, pour m'avoir reçue et avoir répondu à mes nombreuses interrogations concernant mon sujet.

Je remercie les étudiants en information-documentation et les bibliothécaires de Lyon 2 avec lesquelles je me suis brièvement entretenue lors de mes recherches et qui ont apporté leur pierre à l'édifice.

Enfin, je remercie ma famille et mes amis pour m'avoir soutenue et avoir été présents durant la rédaction de ce mémoire. Merci, tout particulièrement, à Delphine, Coralie, Mireille et Jean-Damien de m'avoir proposé si souvent leur aide que j'ai si souvent refusée. Un grand merci à JH pour sa présence, son aide et ses relectures.

## Sommaire

Remerciements.....	3
Introduction.....	5
Partie 1 : De la préhistoire à l'histoire de la sémantique.....	10
1. Les utopies et travaux sur le langage.....	11
2. Les projets précurseurs en bibliothèque.....	18
3. Bibliothèque virtuelle : du web 1.0 au web 3.0.....	26
Partie 2 : Application du web de données en bibliothèque.....	33
4. Les formats du web de données.....	34
5. Réalisations pour l'information-documentation.....	41
6. La prise de décision en bibliothèque.....	50
Partie 3 : La bibliothèque hybride, une amorce à la bibliothèque du futur.....	58
7. Impacts socio-organisationnels.....	59
8. La politique documentaire.....	66
9. La place du livre.....	72
Conclusion.....	79
Bibliographie.....	84
Annexes.....	95

# Introduction

En entendant le titre de bibliothécaire, n'avez-vous jamais imaginé une mégère d'un âge plus qu'avancé, chignon serré et lunettes carrées entrain de vous observer du coin de l'œil et vous lançant de temps en temps un regard désapprobateur suivi d'un « chut » significatif ? Cette bibliothécaire peu avenante dans l'imaginaire social est paradoxalement la gardienne d'un objet symbolisant l'évasion et l'ouverture sur le monde dans ce même imaginaire : le livre. Aujourd'hui, d'autres moyens d'accéder au monde et aux connaissances sont apparus. Cependant, les stéréotypes persistent sur ces professionnels malgré leurs efforts pour évoluer avec la société et diffuser au mieux l'information.

Mon intérêt pour le milieu des bibliothèques et les nouvelles technologies qu'elles intègrent vient principalement de mon cursus universitaire qui a été profondément marqué par les problématiques liées à la gestion des connaissances et à leur diffusion électronique. J'ai ainsi choisi de traiter d'un sujet englobant le monde des bibliothèques et le web 3.0. En effet, titulaire d'un DUT en information-documentation et d'une licence en communication et logiciels libres, je désirais traiter de ces deux domaines, documentation et communication électronique, dans ce mémoire. De plus, l'opportunité offerte par le Master 1 Information-Communication de Lyon 2 de choisir librement un sujet d'étude m'a permis d'appuyer mes candidatures en Master 2 Information-Documentation. Le sujet que je traite dans ce mémoire est d'actualité dans ces formations et il est souvent étudié en Master 1 Information-Documentation. Grâce à ce mémoire, mes candidatures ont certainement été légitimées et ont reçu des avis favorables. Ainsi, ce mémoire est pour moi l'occasion de comprendre les stratégies et le contexte qui régissent la naissance d'un projet en bibliothèque. Il m'a permis également de me rapprocher un peu plus du monde professionnel où je souhaiterais exercer en tant que documentaliste-webmestre. Le choix de ce sujet n'est donc pas anodin.

Depuis le début de mon travail, je me suis doutée que ma méconnaissance du web 3.0 serait une source de difficultés. Pour moi, il était évident qu'en si peu de temps je ne pourrais pas réaliser une étude exhaustive et approfondie de cette technologie malgré mon désir de la comprendre. J'ai alors décidé de faire un panorama plutôt généraliste de ce sujet afin de le comprendre dans sa globalité. Je

me suis attachée à comprendre son histoire, son fonctionnement et ses enjeux au sein des bibliothèques et en particulier dans les bibliothèques académiques. Par bibliothèques académiques, j'entends les bibliothèques universitaires et les bibliothèques de grandes écoles. Toutefois, j'ai éprouvé une certaine difficulté à rester dans ce cadre lors de l'effervescence de la recherche et de l'écriture.

J'ai délibérément choisi de traiter d'un sujet d'actualité dans l'évolution du monde de la gestion des connaissances : le web 3.0. Toutefois, j'ai pu constater lors de mes recherches que cette expression de web 3.0 est floue et inappropriée pour désigner cette technologie. En effet, elle est souvent utilisée dans un objectif marketing ou pour mentionner ce qui n'existe pas encore, le futur du web. Il a donc fallu que je la redéfinisse clairement avant de débiter toute recherche approfondie. J'ai alors découvert l'expression de web sémantique qui restait tout de même très abstraite. Au fil de mes recherches, je suis enfin parvenue à découvrir l'expression que j'utiliserais tout au long de ce mémoire et qui m'a permis d'avancer dans mes recherches : le web de données ou *linked data* en anglais. Le fait que cette technologie soit d'actualité m'a également poussée à étudier son histoire pour savoir en quoi elle répond aux attentes des gestionnaires de l'information. Ce mémoire démontrera donc que les objectifs du web de données sont récurrents depuis beaucoup plus longtemps qu'on ne le pense.

Bien qu'étant encore d'actualité, ce sujet dispose tout de même d'une littérature datant de plus de dix ans pour les plus anciennes références concernant le web de données. Le désir d'établir une technologie similaire au web de données dans les bibliothèques remonte toutefois aux années 1990 avec les premières recherches sur les systèmes d'organisation des connaissances. Le milieu des bibliothèques est spécifique mais il est également un des plus touchés par les problématiques d'organisations et de diffusions des connaissances. Il est alors devenu en moins de dix ans l'un des plus spécialisés et avancés en matière de web de données. Cette technologie est donc considérée comme étant d'actualité au sein de ce milieu.

Suite à cette définition et à ce cadrage de mon sujet, j'ai pu me concentrer sur la véritable question à laquelle je souhaitais répondre dans ce mémoire. Je me suis alors demandé pourquoi les technologies du web de données suscitent autant d'engouement dans le monde de l'information-documentation, notamment en bibliothèque. La mise en place d'une telle technologie dans les bibliothèques ne

peut pas être anodine et dépend de nombreux facteurs. Parmi ces facteurs, j'ai pu en distinguer quatre qui sont devenus mes hypothèses de départ :

- l'application de cette technologie pourrait dépendre des usagers de la bibliothèque qui, habitués à l'interopérabilité des formats du web collaboratif, souhaiteraient élargir cette interopérabilité aux informations disponibles sur le web et pouvoir les repérer facilement grâce à des systèmes de classement.
- Les bibliothèques seraient aujourd'hui dans une logique de partage entre elles mais aussi avec l'extérieur (public, entreprises, associations, etc.). Le web de données et ses formats ouverts seraient alors un moyen de diffuser et d'échanger largement les connaissances.
- Le web de données pourrait avoir des avantages économiques en mettant à disposition des usagers et des autres professionnels des informations déjà traitées et organisées. Pour le public, mais également pour les bibliothèques, la récupération d'informations grâce aux formats ouverts pourrait être ainsi facilitée.
- L'application du web de données en bibliothèque pourrait également dépendre d'un simple effet de mode insufflé par le public, qui suit souvent les évolutions du marché, et/ou par des organismes influents.

Pour répondre à ces interrogations, j'ai choisi d'étudier des textes généraux concernant les bibliothèques, l'informatique documentaire ou le web pour ensuite m'orienter vers des documents plus approfondis et plus critiques sur le web de données et les bibliothèques numériques. Pour moi, il était également important de diversifier les types de documents que j'allais utiliser. Les ouvrages m'ont permis de survoler mon sujet et de le voir dans sa globalité. Les articles de revues spécialisées, telles que Documentaliste – Sciences de l'information et Archimag, m'ont permis d'avoir des informations plus précises et des avis sur des éléments spécifiques de mon mémoire. Quand aux articles de chercheurs, ils contiennent des analyses assez fiables sur les différents éléments de mon sujet de recherche. Enfin, les sites d'institutions engagées dans le web de données, comme la Bibliothèque Nationale de France ou le World Wide Web Consortium, m'ont été utiles pour obtenir des explications détaillées et parfois des points de vues subjectifs sur mon sujet. Durant cette étape, j'ai surtout rencontré des difficultés à trouver des ouvrages sur le web de données. En effet, la littérature sur ce sujet se compose

essentiellement d'articles de revues, de pages de sites Internet, de blogs de professionnels reconnus ou encore de comptes-rendus de colloques et de conférences. Les bibliothèques universitaires de Lyon 2 et de Lyon 3 ont été de véritables alliées pour trouver des références papiers et électroniques (articles de Cairn, Universalis, etc.). De plus, j'ai beaucoup utilisé les bases de données spécialisées telles que ArchiveSic et Hal qui m'ont permis d'accéder à des articles de chercheurs.

Pour compléter ces recherches textuelles et m'aider à maîtriser ce sujet, il me semblait nécessaire de rencontrer et de questionner des professionnels impliqués dans mon sujet. Étant donné que mes recherches étaient principalement axées vers les bibliothèques universitaires, je souhaitais m'entretenir avec les bibliothécaires et les conservateurs des bibliothèques universitaires de Lyon 2, de Lyon 3 et de Saint-Étienne. Je n'ai malheureusement pas pu le faire par manque de temps. En effet, ayant fait un cursus plus technologique que théorique plutôt axé sur la documentation et la communication électronique, l'obtention du premier semestre du Master 1 Information-Communication m'a demandé beaucoup plus de travail personnel que si j'avais eu un parcours principalement basé sur les théories de la communication. De plus, j'ai également choisi de réaliser, lors du second semestre du Master 1, un stage de quatre mois afin d'obtenir une mission correspond à mes attentes. Ces deux semestres et leur déroulement ont donc été pour moi assez énergivores et chronophages. Cependant, bien que je n'ai pas pu réaliser d'entretiens et diffuser de questionnaires auprès de professionnels, j'ai tout de même pu rencontrer diverses personnes mais aussi assister à deux journées de conférences qui m'ont aidées à concrétiser mon sujet. J'ai ainsi pris contact et rencontré au mois de février deux Maîtres de conférences en sciences de l'information et de la documentation de l'Université Lyon 3, Christian Cote et Mabrouka El Hachani, qui m'ont orientés dans ma réflexion. De plus, j'ai également pu m'entretenir avec des étudiants issus du Master 1 Information-Documentation de Lyon 3. Fin mai, j'ai pu aussi obtenir des informations concernant mon sujet en assistant à deux événements : le Symposium Koha organisé par Lyon 2 et la journée d'étude Afnor/BnF « Référentiels et données d'autorité à l'heure du Web sémantique » accueillie au sein de la Bibliothèque Nationale de France. Enfin, j'ai également pu échanger brièvement avec le personnel de la bibliothèque de Lyon 2 grâce à mes passages à la banque de prêt et durant les cours de recherche documentaire du premier semestre.



Grâce à ces recherches et à la synthèse que j'en ai fait, j'ai pu répondre à ma problématique en trois temps. Tout d'abord, afin de comprendre les raisons pour lesquelles le web de données commence à s'imposer dans les bibliothèques, il a fallu que je revienne à la naissance de l'écriture et aux premiers systèmes d'organisation des connaissances. En effet, bien que les techniques et les outils évoluent assez rapidement, j'ai pu constater que le désir de mettre en réseau les connaissances est né quasiment en même temps que l'écriture. Ainsi, les problématiques concernant la gestion des connaissances, auxquelles répond le web de données, sont apparues près de trois millénaires avant notre ère et ont déjà connus de nombreuses tentatives de solutions.

Ensuite, je me suis efforcée d'analyser comment et pourquoi le web de données a pu s'appliquer au monde des bibliothèques. On peut constater que les bibliothécaires ne sont pas les seuls à vouloir organiser les connaissances. Ils ne sont donc pas seuls dans l'avancée des recherches et des réalisations du web de données. De plus, on observe également que les bibliothèques sont très sensibles à l'environnement extérieur en ce qui concerne leurs prises de décisions.

Enfin, on peut observer que le web de données est étroitement lié à la documentation électronique. En effet, avec les questions liées à la numérisation des documents, le web de données est devenu essentiel au sein des bibliothèques. Il sert à organiser les documents en renseignant leurs métadonnées. Il se situe donc au cœur des documents numériques. Les enjeux du web de données correspondent donc à ceux des documents électroniques et des bibliothèques numériques. Sans les mutations nécessaires à la création et à la gestion d'une bibliothèque numérique, le web de données n'existerait pas en bibliothèque.

# **Partie 1 : De la préhistoire à l'histoire de la sémantique**

# 1. Les utopies et travaux sur le langage

De nombreux hommes ont cru en la force d'un langage commun, universel qui leur permettrait d'échanger et d'utiliser de façon optimale la totalité des connaissances. Des mythes religieux, tels que celui de la tour de Babel, aux utopies liées à la naissance des réseaux en passant par les idéologies de grands penseurs de la philosophie et de la linguistique, cette notion de langage universel a parcouru plus de cinq millénaires d'histoire de l'organisation des connaissances<sup>1</sup>.

## 1.1. Babel et les sumériens<sup>2</sup>

À l'époque sumérienne, Babylone était l'une des grandes capitales de la Mésopotamie par son emplacement géographique, sa richesse, son pouvoir mais également par les savoirs qu'elle détenait et les monuments colossaux qu'elle a réalisés. Ainsi, on retrouve dans divers récits de différentes cultures des références aux jardins suspendus de Babylone et, surtout, à la célèbre tour de Babel. En effet, des grecs et des juifs ayant séjourné à Babylone décrivent dans leurs textes la fameuse tour comme étant une construction impressionnante par sa grandeur. Toutefois, pour les juifs, la tour de Babel symbolisait l'orgueil de l'homme qui défiait la puissance de Dieu en voulant atteindre le ciel. Ainsi, selon l'Ancien Testament de la Bible, ces réalisations ont été possible uniquement grâce à la présence d'une langue universelle : « Voici que tous font un seul peuple et parlent une seule langue, et tel est le début de leurs entreprises ! Maintenant, aucun dessein ne sera irréalisable pour eux. »<sup>3</sup>. Toujours selon l'Ancien Testament, Dieu aurait décidé de contrer ce projet qu'il jugeait très orgueilleux en multipliant les langues pour que les hommes ne se comprennent plus entre eux. La confusion des langues, décidée par Dieu, aurait donc entraîné la confusion des savoirs et l'incompréhension entre les hommes.

---

1 MENON B. « Les systèmes d'organisation des connaissances, de Babel aux linked data » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_intro\\_menon.pdf](http://www.bnf.fr/documents/afnor2011_intro_menon.pdf) (Consulté le 3/07/2011)

2 CARDASCIA G, LAFFORGUE G. « Babylone ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/babylone/> (Consulté le 3/07/2011)

3 DE VAUX R, BENOIT P et al (dir). « Genèse 11.1-9 ». *La Bible de Jérusalem*. Paris : Éditions du Cerf, 1973. p. 29

Néanmoins, il existe aussi une explication scientifique de ce mythe, basée sur de possibles faits historiques, prouvant que la confusion des langues aurait bien débuté, entre autres, à Babylone. En effet, les rois babyloniens pourraient être à l'origine de la création des langages secrets dans le but d'asseoir leur pouvoir en confondant les esprits par l'utilisation des jargons et des termes à double sens. Tout comme la confusion des langues dans la Bible, les langages secrets auraient alors contribué à la multiplication des langues et ainsi à la difficulté des hommes à communiquer entre eux.

Parmi les mythes et suppositions, il existe tout de même un certain nombre de faits historiques démontrant l'intérêt de la langue et de l'écriture dans l'organisation des connaissances chez les sumériens. En effet, les premiers écrits datent de cette époque, c'est-à-dire de plus de trois millénaires avant notre ère. Parmi ces écrits, de nombreuses tablettes ayant servi de lexiques aux apprentis scribes ont été découvertes. Ces tablettes lexicales pouvaient donner différents types de classement des termes. Ainsi, les tablettes découvertes pouvaient être : de simples listes de vocabulaire, des listes thématiques avec parfois des définitions et des synonymes, des listes encyclopédiques et enfin des listes de traduction entre les différentes langues de Mésopotamie.

De nombreux autres documents, repris et traduits par des copistes en sumérien et en akkadien, ont été également découverts dans les ruines des bibliothèques et des archives de l'époque. Ces découvertes sont la preuve de l'importance que l'Homme accordait déjà aux écrits et donc à la transmission du savoir<sup>4</sup>.

Dès le IV<sup>e</sup> millénaire, à Babylone, et plus largement en Mésopotamie, des langues vernaculaires sont apparues, telles que l'akkadien et l'hébreu, tandis que certains religieux, savants et écrivains ont persisté pendant longtemps à utiliser le sumérien. Bien que le bilinguisme ait été très important durant ces millénaires, cette multiplication des langues et leur usage parfois élitiste reprend l'idée d'une confusion et d'une incompréhension entre les hommes.

Ainsi, on peut penser que, depuis la naissance de l'écriture, l'homme a toujours cherché à organiser, classer, rassembler et relier ses savoirs afin de les exploiter de la meilleure façon. À partir de cette époque, on observe donc déjà un effort de l'homme pour créer un sens commun aux mots, pour normaliser l'écriture et enfin pour communiquer de façon sûre avec une même langue.

---

4 BOTTÉRO J. *Babylone : À l'aube de notre culture*. Évreux : Gallimard, 1994.

Depuis l'Antiquité, on retrouve des recherches allant dans ce sens chez les philosophes. Ces recherches débutèrent par de simples analyses de la construction de la langue jusqu'à de réelles tentatives d'amélioration de la langue.

## 1.2. La philosophie du langage et la critique de la langue

Les tous premiers questionnements sur le langage sont apparus avec les philosophes antiques Socrate, Platon et Aristote. En effet, « Socrate cherche les « définitions », c'est-à-dire le sens permanent de nos mots et de nos phrases ; Platon, dans le *Cratyle*, s'interroge sur la « justesse » des mots [...]. Dans son traité *Sur l'interprétation*, Aristote établit que toutes les lois de la logique s'édifient sur les caractères de la proposition, laquelle consiste à affirmer ou nier « quelque chose au sujet de quelque chose »<sup>5</sup>. Ces philosophes sont donc à l'origine des différentes philosophies du langage de leurs successeurs. Toutefois, c'est bel et bien Platon qui amorça les réflexions autour de la langue en se demandant si les mots sont directement liés aux choses ou s'ils sont de simples créations de l'homme.

À partir du XVII<sup>e</sup> siècle, certains philosophes recherchèrent une langue universelle, appelée aussi « langue philosophique a priori »<sup>6</sup>, basée sur le langage adamique ou sur une nouvelle langue. Ce langage serait le moyen d'éviter les contre-sens. Cette idée est développée en particulier chez Bacon, Hobbes, Leibniz et Locke.

Ainsi, dès le début du XVII<sup>e</sup> siècle, le philosophe anglais Francis Bacon, a établi un schéma de langue universelle. Toutefois, ce schéma englobe uniquement le langage scientifique et non la langue dans sa totalité. En effet, le but de Bacon était surtout de créer une science universelle. Lors de ses recherches, il a notamment instauré une nouvelle classification des sciences et a prôné une science empirique, c'est-à-dire basée sur l'expérience. Ses idées seront notamment reprises par la Royal Society of London for the Improvement of Natural Knowledge, et en particulier par l'un de ses fondateurs, John Wilkins. En effet, en 1668, Wilkins compléta l'oeuvre de Bacon et créa une langue philosophique ou langue universelle.

---

5 RICOEUR P, SEBESTIK J. « Philosophies du langage ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/philosophies-du-langage/> (Consulté le 3/07/2011)

6 ECO U. *La recherche de la langue parfaite dans la culture européenne*. Paris : Éditions du Seuil, 1994.

Par la suite, en 1641, Hobbes a mis en doute la pensée cartésienne, et notamment le fameux « cogito ergo sum ») en déclarant que les mots ne sont pas l'image même des choses<sup>7</sup>. Ainsi, les mots choisis par Descartes seraient trompeurs et son argumentation serait alors irrecevable. Hobbes prend donc part à la critique du langage en dénonçant le côté conventionnel de la langue.

En 1679, Leibniz déclara dans sa *Lettre à Jean-Frédéric* : « J'ai le projet d'une langue ou écriture universelle qui (...) outre l'usage du commerce et la communication des peuples divers (ce qui la pourrait même rendre plausible au vulgaire), aurait des avantages incomparablement plus grands : car elle donnerait moyen de raisonner sur les matières capables de raisonnement par une espèce de calcul infaillible pourvu qu'on y apportât la même exactitude qu'à chiffrer, et les erreurs ne seraient que des erreurs de calcul »<sup>8</sup>. Pour Leibniz, les avancées scientifiques pourraient être plus grandes grâce à ce langage car il permettrait de supprimer les incompréhensions et erreurs dues au langage. Les scientifiques pourraient alors se focaliser uniquement sur la science malgré leurs nationalités différentes.

Enfin, John Locke, en 1690, déclara « les hommes croient que leurs paroles signifient la réalité des choses alors qu'elles n'expriment que leurs idées »<sup>9</sup>. Il critique donc lui aussi l'importance qu'on accorde à des mots qui sont des « signes » (au sens de Locke) arbitraires de la pensée. Dans son essai sur l'entendement humain, il reprend ainsi l'idée selon laquelle le côté conventionnel et la diversité des langues mèneraient à la confusion entre les hommes.

Ces différents philosophes sont des précurseurs de la philosophie du langage et de l'époque des Lumières. Ils ont tous pour point commun une certaine critique du langage. Toutefois, ils sont également très différents dans leurs méthodes de travail et donc dans leur argumentation. C'est certainement grâce à cette diversité que d'autres, tels que Diderot ou Voltaire suivront leur voie et affineront encore les recherches de leurs prédécesseurs.

Ainsi, la philosophie du langage a également permis de mettre en évidence les failles du langage tel qu'il existe aujourd'hui et l'intérêt d'un langage commun plus proche de la pensée réelle. À ce jour, le projet de langue universelle le plus abouti

---

7 HOBBS T. *Troisièmes Objections aux Méditations sur la philosophie première de Descartes*. 1640

8 GONDRAN M. « Bilan et perspective de dix ans d'Intelligence Artificielle en entreprise ». *Quaderni*, printemps 1995, n°25, pp. 99-123

9 LOCKE J. *Essai sur l'entendement humain*. Paris : Ellipses, 2003

reste encore l'Espéranto créé en 1887 par Ludwik Lejzer Zamenhof. Cette langue à vocation universelle a certainement réussi à voir le jour grâce à l'apparition de la linguistique au début du XVII<sup>e</sup> siècle.

### 1.3. La linguistique

Le souhait d'organiser et de diffuser les connaissances est donc l'une des préoccupations des hommes depuis au moins cinq millénaires. Toutefois, les premières recherches visant à comprendre le fonctionnement de la langue comme outil capable de diffuser et d'organiser le savoir sont très récentes.

En effet, on peut considérer l'ouvrage d'Antoine Arnauld et de Claude Lancelot, *Grammaire générale et raisonnée* (1660), comme l'un des précurseurs dans ce domaine. Dans ce livre, les auteurs mènent une analyse de la langue en l'étudiant grâce à la logique et à la philosophie<sup>10</sup>. De plus, bien que leur travail soit principalement axé sur la langue française, ils cherchent à distinguer un certain nombre d'éléments communs entre les langues.

Par la suite, la linguistique a été reconnue comme étant une discipline comportant plusieurs axes d'étude et regroupant de nombreux domaines<sup>11</sup>. Parmi les axes d'étude possibles, on retrouve celui de la linguistique contextuelle qui a, entre autres, permis de développer les analyses critiques du discours et a ainsi contribué à la philosophie du langage traitée précédemment. Bien entendu, la recherche fondamentale est essentielle pour la linguistique, comme pour toutes les autres sciences puisqu'elle permet de mener à bien des découvertes plus généralistes que la recherche appliquée. Toutefois, la recherche appliquée en linguistique, ou linguistique appliquée, constitue un axe de recherche qui a contribué à développer de nouvelles techniques en informatique et en intelligence artificielle.

En effet, la linguistique appliquée peut toucher de nombreux domaines. De plus, en se basant sur les recherches de linguistique moderne, c'est-à-dire sur les recherches concernant les éléments universels du langage, elle permet de contribuer à améliorer les systèmes d'organisation des connaissances. Ainsi, grâce à la linguistique appliquée, et plus particulièrement à la recherche

---

10 FUCHS C. « Linguistique ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/linguistique/#> (Consulté le 3/07/2011)

11 CHISS J-L., PUECH C. *Fondations de la linguistique : études d'histoire et d'épistémologie*. Louvain-la-Neuve : Duculot, 1997

interdisciplinaire de la linguistique informatique, des technologies telles que le Tal (Traitement Automatique des Langues) ou encore le Sampa (Speech Assessment Methods Phonetic Alphabet) sont apparues. Ce dernier est un moyen de transcrire toutes les langues sur informatique<sup>12</sup>. Il est basé sur une autre découverte linguistique : l'Api (Alphabet Phonétique International).

En ce qui concerne le Tal, « il s'agit majoritairement de traiter de grandes quantités de données textuelles, sur support électronique »<sup>13</sup>. Le but du Tal est notamment d'améliorer les outils déjà existants, d'extraire des données pertinentes mais aussi de créer des terminologies en se basant sur des corpus. Cette préoccupation concerne donc plus particulièrement le milieu de la documentation qui se base sur des terminologies spécifiques (généralistes, médicales, internes à une entreprise, etc.) pour développer les différents outils de GED (Gestion Électronique de Documents). La GED est un des systèmes de gestion des connaissances dans les organisations : elle formalise, synthétise et organise les données des organismes.

Ainsi, la terminologie et, plus largement, la linguistique permettent d'entretenir une logique et un vocabulaire communs au sein d'un organisme ou d'une communauté particulière. Toutefois, le fait que chacune de ces terminologies soient différentes les unes des autres, voire parfois incompatibles, constitue une des principales failles de ce système. En effet, le but premier des terminologies est d'être universelles et donc objectives. Or, il existe de multiples terminologies qui concernent parfois le même domaine. Pour l'instant, ces terminologies restent donc subjectives car, pour un même domaine, elles pourront être basées sur des interprétations et des corpus différents.

En ce qui concerne le langage informatique, son hérité est indéniablement linguistique. En effet, il est basé sur des normes internationales, que ce soit au niveau du langage binaire ou de l'Unicode. Bien entendu, la linguistique fait partie des nombreux éléments qui ont aidé au développement des normes informatiques. Sans une base commune, c'est-à-dire l'Api ou encore le Sampa, ces langages auraient eu beaucoup plus de difficultés à s'imposer en tant que normes dans le milieu informatique.

---

12 UNIVERSITY COLLEGE LONDON. « SAMPA - computer readable phonetic alphabet ». UCL - London's Global University [En ligne]. 3 juillet 2011. Disponible sur : <http://www.phon.ucl.ac.uk/home/sampa/> (Consulté le 3/07/2011)

13 CONDAMINES A. « Linguistique de corpus et terminologie ». *Langages*, 2005, n°157, pp. 36-47



Ainsi, la linguistique avec notamment le Tal et le Sampa, a largement contribué aux avancées de l'informatique documentaire. Dans la quête de l'intelligence artificielle, cette science dispose d'une place primordiale puisqu'elle permet de comprendre et de traiter du seul outil essentiel à la diffusion de la connaissance : le langage.

#### **1.4. L'intelligence artificielle**

Avant d'étudier la place de l'intelligence artificielle dans la naissance du web de données, il est important de définir ce terme abstrait.

L'intelligence artificielle est considérée pour beaucoup comme une utopie. En effet, sa définition première décrit une machine pouvant reproduire l'intelligence humaine, c'est-à-dire l'apprentissage et le raisonnement humain. Cette définition est très critiquée car elle ne prend pas en compte le caractère subjectif du raisonnement humain qui ne suit pas obligatoirement un cheminement logique pour autrui. De plus, le fonctionnement de la machine est basé sur un système de calculs programmé et corrigé continuellement par l'homme. Il est donc presque certain que la machine ne pourra jamais atteindre les capacités langagières de l'homme sans son intervention<sup>14</sup>.

Toutefois, face à cette définition utopiste, des technologies sont considérées par certains comme faisant partie des technologies de l'intelligence artificielle. Ce statut leur a été accordé suite à une redéfinition par certains scientifiques de l'intelligence artificielle.

Ainsi, le Tal, vu précédemment, peut être assimilé à une sorte d'intelligence artificielle puisqu'il permet d'automatiser certaines capacités langagières de l'homme. Par exemple, il a permis de développer des correcteurs orthographiques, des logiciels de traduction, etc. Toutefois, le Tal trouve beaucoup de ses applications dans le milieu de la documentation car il permet la modélisation des connaissances à l'aide de l'étude des textes et des comportements humains. Ces travaux en documentation ont pour principaux objectifs d'automatiser l'indexation et la classification des documents grâce notamment à des listes de mots-clés prédéfinis (thésaurus, etc.) ou encore au calcul des occurrences dans les textes. Un des autres intérêts de la documentation pour le Tal, et plus largement l'intelligence artificielle, vient également du fait que cette dernière tente de dépasser les

---

14 ABEILLÉ A. « Traitement automatique des langues » [En ligne]. *Encyclopaedia Universalis – Education*. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/traitement-automatique-des-langues/> (Consulté le 3/07/2011)

barrières de la langue. Pour les documentalistes, la diversité des langues est un frein incontestable à la gestion des connaissances. Les logiciels de traduction mais surtout les nouveaux formats et langages du web de données se rapprochent de ce qu'on pourrait nommer l'intelligence artificielle. Il sont donc d'une grande aide pour les documentalistes aussi bien au niveau de la recherche des connaissances que de leur compréhension et de leur traitement.

Cette définition plus large de l'intelligence artificielle engloberait alors les systèmes experts, c'est-à-dire des logiciels pouvant reproduire les tâches et possédant les mêmes connaissances qu'un spécialiste du domaine concerné<sup>15</sup>. Toutefois, c'est encore l'imperfection de ces systèmes et l'insatisfaction des chercheurs, informaticiens et professionnels sur leur travail qui contribuent à alimenter le débat autour de l'existence d'une réelle intelligence artificielle.

En bibliothèque, ces systèmes utilisent donc des bases déjà existantes en informatique, en approche cognitive, en linguistique mais surtout en méthodes de traitement de l'information. En effet, les recherches traitant de l'intelligence artificielle ont une place centrale dans le monde de la documentation car l'automatisation, la gestion et l'exploitation des connaissances font partie des rôles principaux des documentalistes.

Ainsi, les documentalistes ont aidé à l'élaboration de ces nouvelles technologies avec notamment différents systèmes et types de classification : des simples lexiques aux ontologies, en passant par les thésaurus, les taxonomies ou encore les classifications comme la classification décimale de Dewey.

## 2. Les projets précurseurs en bibliothèque

Depuis l'Antiquité, de nombreux intellectuels ont noté l'importance d'une bonne organisation des livres dans les bibliothèques. Toutefois, en ce qui concerne les véritables travaux modernes sur ce sujet, on peut citer le philosophe Francis Bacon (1561-1626), le bibliothécaire Gabriel Naudé (1600-1653) et enfin le bibliothécaire Anthony Panizzi (1797-1879)<sup>16</sup>.

---

15 GUTMAN J. *Rapport de recherche bibliographique : Les applications de l'intelligence artificielle à l'assistance ou l'automatisation du catalogage* [En ligne]. Villeurbanne : ENSSIB, 1996. Disponible sur : [www.enssib.fr/bibliotheque-numerique/document-1621](http://www.enssib.fr/bibliotheque-numerique/document-1621) (Consulté le 3/07/2011)

16 FOSKETT D. « Indexation ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/indexation/> (Consulté le 3/07/2011)

Cette période marque donc la naissance de nouveaux outils de gestion de la documentation rassemblés sous l'appellation de systèmes d'organisation des connaissances (voir tableau ci-dessous).

Listes de termes	Classifications et catégories	Listes de relations
Fichiers d'autorité	Vedettes-matière	Thésaurus
Glossaires	Schémes de classification*	Réseaux sémantiques
Dictionnaires	Taxonomies*	Ontologies
Répertoires géographiques	Schémes de catégorisation*	

(\* Ces trois termes/expressions sont souvent utilisés l'un pour l'autre)

Tab. 1 : *Typologie de systèmes d'organisation des connaissances*<sup>17</sup>

## 2.1. Les systèmes d'organisation des connaissances (Soc)

Dans son ouvrage, *Systems of Knowledge Organization for Digital libraries*, Gail Hodge propose une définition détaillée des Soc. Pour elle, ce terme « englobe tous les types de schèmes permettant d'organiser des informations et de promouvoir la gestion des connaissances. [...] [Les Soc] comprennent les schèmes de classification et de catégorisation qui organisent des matériaux à un niveau général, des vedettes-matière qui fournissent un accès plus détaillé, et les fichiers d'autorité qui contrôlent les variantes des informations clés telles que les noms géographiques et les noms de personnes [...], les vocabulaires très structurés, tels que les thésaurus, et les schèmes moins traditionnels, tels que les réseaux sémantiques et les ontologies »<sup>18</sup>.

Ces différents schèmes, ou systèmes, ont donc contribué au développement du web de données grâce à leur complémentarité. En effet, chacun de ces schèmes sera une des étapes nécessaires au bon fonctionnement du web de données.

17 GIBOIN A, ZACKLAD M (dir). « Applications à base de Soc hétérogènes : Thésaurus, ontologies, folksonomies... ». *Document numérique*, mai-août 2010, n°2, vol. 13.

18 HODGE G. *Systems of Knowledge Organization for Digital libraries : Beyond traditional authority files*. Washington, DC : the Council on Library and Information resources, 2000.

### **Les « listes de termes »**

Comme vu précédemment, on peut admettre que les lexiques sont nés avec le début de l'écriture. À partir du XV<sup>e</sup> siècle, la redécouverte des anciens écrits et savoirs dans le sud de l'Europe a contribué à développer de nouvelles façon de voir le monde chez les penseurs. C'est la naissance de l'Humanisme et des dictionnaire<sup>19</sup>.

Ces nouveaux types d'ouvrages deviennent une véritable aide à la compréhension des textes nouveaux et anciens. Contrairement aux lexiques, ils contribuent à formaliser les usages des mots et permettent ainsi aux savants de se baser sur les mêmes définitions de termes pour leurs analyses. La mise en commun de ces analyses, et donc des connaissances, est alors plus aisée.

Par exemple, au niveau des logiciels de bibliothéconomie, la distinction et le rassemblement des connaissances peuvent se faire grâce aux fichiers d'autorité. Ces derniers servent à éviter de recréer manuellement des vedettes, ou termes, et à empêcher ainsi les fautes d'orthographe, de frappe, etc. Une seule vedette permettra alors de regrouper les connaissances (ex. tous les ouvrages de Sartre) sous une seule vedette dite d'autorité (ex. Sartre). Toutes les formes possibles de la vedette seront tout de même répertoriées dans le fichier d'autorité (ex. Sartre ou SARTRE) pour être renvoyées à la vedette d'autorité<sup>20</sup>.

### **Les « classifications et catégories »**

Le deuxième type de schème concerne notamment l'indexation des connaissances. On peut citer l'utilisation d'outils préconçus comme la CDU (Classification Décimale Universelle), créée en 1905, et les vedettes-matière. Selon la définition de l'Afnor, les vedettes matières sont un « Ensemble d'un ou plusieurs descripteurs exprimant et précisant le sujet d'un document. Chaque vedette-matière correspond à un seul sujet, simple ou complexe. » Le but de ces outils est alors de trier et de rassembler les connaissances sous des libellés généraux. La liste de vedettes-matière française la plus connue et la plus utilisée

---

19 FURNO M. « De l'elementarium au thesaurus : l'émancipation des lexiques latins monolingues aux XV<sup>e</sup> et XVI<sup>e</sup> siècles ». *Histoire Épistémologie Langage*, 1997, Tome 19, fascicule 1, pp. 151-175

20 BENZAADI F. « Approche méthodologique pour la création de fichiers d'autorité nationaux ». *Revue d'Information Scientifique & Technique*, 1995, n°2, vol. 5, pp. 29-32

est Rameau (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié)<sup>21</sup> (Cf. Annexe n°1 : Exemple de notice Rameau, p.96).

Les taxonomies, ou taxinomies, sont également très utilisées en information-documentation. Apparues en 1735 avec le naturaliste Carl von Linné et appartenant tout d'abord au domaine de la biologie, elles servaient principalement à classer les organismes vivants. Aujourd'hui, elles ont été adoptées comme méthode de classification par le monde de la documentation.

Toutefois, ces deux premiers schèmes sont de plus en plus délaissés. En effet, utilisés seuls, ils deviennent peu à peu obsolètes par rapport au désir de mise en réseau des données.

### **Les « listes de relations »**

Ce dernier type de langage documentaire a pris son essor grâce aux nouvelles possibilités qu'offre l'informatique en matière de gestion des connaissances. Ainsi, dès les années 1950, les thésaurus se sont développés dans l'univers de la documentation.

Le thésaurus est « un langage structuré, contrôlé et combinatoire »<sup>22</sup>. Il est alors un outil hiérarchique et organisé permettant surtout la mise en relation des connaissances. La recherche d'information est alors plus pertinente car elle prend en compte les relations entre les termes recherchés. Toutefois, cet outil est très peu utilisé dans le milieu des bibliothèques car il est souvent spécialisé sur un domaine, une entreprise, etc. Il est donc plus utilisé par les documentalistes (ex. le thésaurus Motbis spécialisé dans l'éducation – Cf. Annexe n°2 : Le terme « élève » dans le thésaurus Motbis, p.97). La complexité et l'utilisation restreinte au monde de la documentation spécialisée marginalise cet outil peu ergonomique. De plus, les thésaurus sont parfois mal exploités car ils sont traités comme de simples mots-clés. La mise en relation des termes, qui constitue le caractère principal de cet outil, est alors inutilisée.

Enfin, les ontologies, nées de la rencontre entre les domaines des sciences cognitives et de l'intelligence artificielle, sont entièrement liées aux travaux sur les réseaux sémantiques et la représentation des connaissances. « le terme d'ontologie est aujourd'hui utilisé de manière bien plus large et imprécise pour désigner toute

---

21 MENON B. « Les langages documentaires : Un panorama, quelques remarques critiques et un essai de bilan ». Documentaliste Sciences de l'information, 2007, vol. 44, pp. 18-28

22 MENON, 2007, *op. cit.*

classification aisément partageable sur le web grâce à l'usage des standards du W3C [World Wide Web Consortium] »<sup>23</sup>.

Ces schèmes sont toutefois victimes soit de leur ancienneté, soit de leur complexité. Ils sont encore très peu remis en question ou utilisés. De plus, ils restent encore restreints au monde de l'information-documentation et aux sciences cognitives.

Enfin, les bibliothèques restent encore récalcitrante à utiliser des schèmes tels que les thésaurus et les ontologies. Elles prônent encore l'utilisation des classifications telles que la CDU ou la CDD (Classification Décimale de Dewey). Cependant, les listes de relations pourraient permettre de mieux respecter certaines valeurs du *Code de déontologie du bibliothécaire* (Cf. Annexe n°3 : Code de déontologie du bibliothécaire, p.98), telles que « permettre un accès à l'information respectant la plus grande ouverture possible » ou encore « faire connaître et mettre en valeur les collection », et de contribuer ainsi à des projets tels que « La bibliothèque hors les murs ».

## **2.2. La bibliothèque et l'accès à la connaissance**

Le projet de « La bibliothèque hors les murs », tout d'abord appelé « bibliothèque de rue », est apparu dans les années 1960. Ces bibliothèques ont pour but de diffuser le savoir aux populations en difficulté<sup>24</sup>. L'essor d'une politique voulant favoriser l'accès à la culture dans les années 1980, le *Manifeste de l'UNESCO sur la bibliothèque publique 1994* (Cf. Annexe n°4 : Manifeste de l'Unesco, p.100) et la création de la Charte des bibliothécaires en 1992 (Cf. Annexe n°5 : Charte des bibliothécaires, p.103) ont largement participé à l'essor de ce projet.

Ce projet est la preuve de l'importance que les bibliothèques accordent à l'utilisateur et à la diffusion des connaissances. En matière de diffusion des connaissances, les bibliothèques s'efforcent aujourd'hui d'être actives sur de nombreux moyens de communication. Elles tentent de se séparer de leur image

23 ZACKLAD M. « Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI) » [en ligne]. *35e Congrès annuel de l'Association Canadienne des Sciences de l'Information. Partage de l'information dans un monde fragmenté : Franchir les frontières*. Montréal : CAIS/ACSI, 2007. Disponible sur [http://www.cais-acsi.ca/proceedings/2007/zacklad\\_2007.pdf](http://www.cais-acsi.ca/proceedings/2007/zacklad_2007.pdf) (Consulté le 3/07/2011)

24 BONACCORSI J. « Le livre déplacé, une bibliothèque hors-les-murs ». *Communication et langage*, 2001, n°127, vol. 127, pp. 21-34

peu attrayante en allant vers le public, en aménageant au mieux les espaces de consultation (internes et externes à la bibliothèque) et en diversifiant ses activités (formations, interventions d'associations, etc.)<sup>25</sup>.

Le rôle du bibliothécaire est donc également en mutation pour répondre au mieux aux objectifs professionnels du service public et de la communication aux usagers malgré la distance. L'adaptation continue du métier face aux évolutions techniques et aux demandes des lecteurs est capitale au maintien de l'activité de la bibliothèque sur le long terme.

Toutefois, il ne faut pas oublier que ces projets ne sont pas uniquement élaborés en fonction des besoins des usagers et des désirs des bibliothèques. En effet, dans le cas des bibliothèques hors les murs, plusieurs institutions ont pris part au projet. Par exemple, l'État et différents ministères, pour soutenir leur politique de diffusion de la culture à tous, ou encore des associations luttant en particulier contre l'illettrisme (les Centres d'entraînement aux méthodes d'éducation active, la Ligue française de l'enseignement et de l'éducation permanente, les MJC, etc.).

À Lyon, on peut citer un exemple phare de ce désir de diffuser la connaissance au plus grand nombre. En effet, dans le quartier de La Part-Dieu, on peut constater une mixité entre monde professionnel, vie quotidienne et culture avec la présence dans un même lieu de bureaux, d'une gare, d'un centre commercial et, surtout, d'un auditorium et d'une bibliothèque réputée pour être l'une des plus importantes de France.

Enfin, certaines réalisations tels que les SCD (Service Commun de Documentation), le Peb (Prêt entre bibliothèques) ou encore le Sudoc (Système Universitaire de Documentation) sont beaucoup plus spécifiques au monde des bibliothèques universitaires.

En ce qui concerne les SCD, ils permettent d'harmoniser la politique documentaire (acquisitions, budget, etc.) entre les différentes bibliothèques dépendant d'une même université. La visibilité du fonds documentaire et l'organisation des différentes bibliothèques s'en trouvent améliorées.

Avec le Peb, la bibliothèque fait partie d'un réseau de bibliothèques, elle offre à l'utilisateur la possibilité d'obtenir un document qu'elle ne possède en passant par le réseau. Au niveau universitaire, ce réseau s'appelle le Sudoc et il comprend « 1 419 bibliothèques déployées [...] plus de 10 millions de notices

---

25 TABET C. *La bibliothèque « hors les murs »*. Paris : Electre, 2004

bibliographiques, 32 millions de documents localisés et 24 millions d'interrogations publiques en 2010 »<sup>26</sup>.

Toutefois, cette pratique est de moins en moins utilisée par les étudiants face à l'essor des ressources numériques. Ils passent désormais directement par internet et la documentation numérique pour obtenir le document dont ils ont besoin. Cette implication dans le web et l'électronique est une suite logique à l'intérêt que l'information-documentation a toujours eu pour les outils informatiques.

### **2.3. L'informatique en bibliothèque**

À partir des années 1970, on commence à constater le grand retard des bibliothèques universitaires françaises par rapport à leurs voisines européennes grâce au *Livre noir des bibliothèques universitaires* publié par l'ABF (Association des Bibliothécaires de France). Toutefois, il a fallu attendre la publication du rapport Miquel, Ministre de l'éducation nationale, en 1988 pour que la situation des bibliothèques universitaires s'améliore<sup>27</sup>.

En effet, ce rapport commandé par l'État avait pour but de palier les lacunes des bibliothèques en matière de locaux, de services, d'organisation mais également de nouvelles technologies. Ce rapport a donc été bien accueilli par la communauté universitaire qui considère que les bibliothèques doivent permettre un accès au savoir à tous les étudiants, que ce soit au niveau de la documentation ou de l'accès aux nouvelles technologies, et contribuer à leur réussite. Les solutions apportées suite à ce rapport se sont intégrées dans deux plans successifs : le plan U2000 (Université 2000) et le plan U3M (Université du Troisième Millénaire).

En ce qui concerne l'évolution technologique, les résultats du premier plan sont rapides car, en 1996, 65 des 96 bibliothèques universitaires ont été équipées d'un SIGB (Système Intégré de Gestion de Bibliothèque). Aujourd'hui, la mise en place de ce type de progiciel (contraction de « produit » et « logiciel », c'est une suite de logiciels non spécifiques répondant à des besoins professionnels généraux) destiné à la gestion des différentes tâches d'une bibliothèque (gestion des acquisitions, des lecteurs, du prêt/retour, etc.) est devenue une évidence, voire une nécessité. En effet, avec l'apparition des réseaux de bibliothèques universitaires (Cadist, Peb, etc.) et l'augmentation du budget acquisition durant cette même période, le

---

26 ABES. *L'ABES – ABES.fr* [En ligne]. 3 juillet 2011. Disponible sur : <http://www.abes.fr/abes/page.351.labes.html> (Consulté le 3/07/2011)

27 LACHENAUD J-P. *Bibliothèques universitaires : le temps des mutations*. Les Rapports du Sénat, n°59. Paris : Sénat, 1998-1999.



développement de ces logiciels constitue une véritable aide à la gestion des documents.

De plus, ces nouveaux logiciels permettent de créer une véritable bibliothèque virtuelle grâce notamment aux documents numériques. Toutefois, la plupart des SIGB sont aujourd'hui vieillissants. En effet, tous les logiciels propriétaires ont plus de 10 ans, les fonctionnalités proposées évoluent peu et beaucoup de produits sont abandonnés par leurs éditeurs. Cela pose problème pour certaines structures telles que les bibliothèques universitaires qui sont contraintes de suivre les évolutions technologiques au plus près car leurs besoins ainsi que ceux de leurs utilisateurs sont liés aux évolutions des TIC (services et documents en ligne, etc.). Ce besoin a eu pour conséquence de faciliter l'essor des logiciels libres dans le milieu des bibliothèques dès les années 2000.

En effet, les SIGB libres ont comme avantages le respect des standards du web, facilitant ainsi l'échange de données entre bibliothèques, et la présence d'Opac (Online Public Access Catalog) performants, permettant aux usagers de consulter le catalogue de la bibliothèque ainsi que ses documents numériques à partir d'Internet. Parmi les bibliothèques universitaires qui ont récemment fait le choix de s'orienter vers un logiciel libre, on retrouve Lyon 2, Lyon 3 et Saint-Étienne qui font partie d'un consortium pour le développement et le transfert de leurs catalogues sous le logiciel Koha<sup>28</sup>.

Enfin, les bibliothèques universitaires tentent encore aujourd'hui de regrouper leurs données (documents papiers, électroniques, numériques, etc.) afin de créer un catalogue collectif permettant de rechercher, localiser et demander un document. Ce projet est toutefois long à concrétiser car il nécessite également de prendre en compte les droits d'auteur (en particulier les droits de prêts et de reprographie), la pérennité du support et du document, etc.

Avec Internet et l'accessibilité des documents numériques à tous, ces problématiques se complexifient car la garantie d'un usage monoposte de ces documents reste difficile à obtenir. En effet, pour consulter un document numérique, les licences d'utilisation de certains éditeurs nécessitent de conserver les mêmes modalités d'accès que pour un document papier, c'est à dire consultable par une seule personne à la fois. L'arrivée du numérique a donc amené une certaine incertitude juridique et il est important que les bibliothèques mesurent les risques

---

28 SCD Lyon 2, SCD Lyon 3 et al. *Groupe de travail sur les SIGB Libres* [En ligne]. 3 juillet 2011. Disponible sur : <http://www.sigb-libres.info> (Consulté le 3/07/2011)

encourus en diffusant des ressources numériques. Néanmoins, les bibliothèques, qui ont désormais constaté l'utilité de l'informatique et des Tic, continuent leurs recherches dans le domaine d'internet malgré les difficultés techniques, juridiques, etc.

### **3. Bibliothèque virtuelle : du web 1.0 au web 3.0**

Au cours des siècles, les mutations des bibliothèques et les avancées techniques se sont succédées. Cependant, durant ces dix dernières années, ces évolutions ont considérablement accéléré. Cela est notamment dû à l'utilisation de plus en plus fréquente d'Internet dans le monde universitaire<sup>29</sup>. En effet, bien que sa création ne soit pas récente, Arpanet date de 1969, Internet servait principalement à l'échange de mail dans le milieu universitaire jusqu'à fin 1994. Après 1994 et la Conférence internationale du Cern (Conseil Européen pour la Recherche Nucléaire), le web a commencé à servir d'espace vitrine et la bibliothèque est devenue au fur et à mesure une bibliothèque hybride.

#### **3.1. Web 1.0 et la bibliothèque numérique**

Le web 1.0, que j'appellerais le web statique ou vitrine afin d'être plus claire et juste dans mes propos, se situerait entre 1995 et 2003. C'est la période où le web était constitué principalement de pages HTML (Hypertext Markup Language), où un individu publiait seul du contenu que les internautes pouvaient consulter. Ces pages contenaient le plus souvent des informations intemporelles et statiques. Elles étaient repérées sur le web grâce aux premiers Uri (Uniform Resource Identifier), les URL (Uniform Resource Locator)<sup>30</sup>. Ces formats créés par le Cern, dont Tim Berners Lee (actuel directeur du W3C) était membre, sont devenus depuis des standards du web. Ils sont toujours utilisés aujourd'hui et servent même de formats de base au web.

Dans l'imaginaire collectif, la bibliothèque représente un point de référence en matière de savoir. Les individus remettent difficilement en cause la fiabilité des sources présentes en bibliothèque et le bibliothécaire est vu comme un professionnel ayant une connaissance sur tous les domaines existants. Toutefois,

---

29 CALENGE B. *Bibliothèque et politique documentaire à l'heure d'internet*. Paris : Electre, 2008

30 GUILLOU P. « Comprendre le web 1.0 » [En ligne]. *Ideose, numérique pour tous*. 29 novembre 2009. Disponible sur : <http://www.ideose.eu/blog/web-temps-reel/comprendre-le-web-10/> (Consulté le 3/07/2011)

cette image de la bibliothèque n'en fait pas pour autant un lieu privilégié de la recherche d'information.

En effet, avec la création des moteurs de recherche et l'essor d'Internet dans les foyers, le public mène, la plupart du temps, ses recherches sans passer par des outils gérés par des professionnels de l'information. Ainsi, en avril 2011, on constate que la plupart des sites visités par les internautes sont des moteurs de recherches généralistes, des sites d'informations de service (météo, etc.)<sup>31</sup>. La consultation de bases de données spécialisées et de catalogues de bibliothèques en ligne est moindre malgré les efforts déployés par les professionnels de la documentation pour occuper l'espace numérique et aider les usagers à dépasser le bruit informationnel créé par le web.

En effet, durant ces quinze dernières années, les bibliothèques ont développé des accès aux ressources en ligne. En ce qui concerne les bibliothèques universitaires, elles proposent à leurs usagers d'accéder à certains documents électroniques (articles de revues, encyclopédies, ouvrages, etc.) via des bases de données spécialisées auxquelles elles se sont abonnées. Toutefois, les bibliothèques sont encore très peu nombreuses à proposer des documents qu'elles ont numérisés elles-mêmes. Ces quelques documents numérisés sont le plus souvent des thèses ou des documents créés par la bibliothèque elle-même (rapport, etc.). Les bibliothèques proposent parfois aussi des signets, qui sont des sortes de marques pages, afin de faciliter la recherche sur Internet.

Il existe tout de même un projet d'envergure concernant les ressources numériques qui a été lancé en 1997. Il s'agit de la bibliothèque numérique Gallica, créée par la BNF (Bibliothèque de France), qui rassemble aujourd'hui plus de 1 500 000 documents de toutes sortes (livres, images, partitions, etc). Ces ressources ont la particularité d'être en libre accès afin de diffuser les savoir au plus grand nombre<sup>32</sup>. Cependant, la BNF a vu apparaître en 2005 un « concurrent » commercial sur le marché de la numérisation : Google Books.

L'apparition de ces types d'outils simple d'utilisations et permettant souvent l'accès à des contenus intégraux freine l'utilisation de bases de données telles que

31 MÉDIAMÉTRIE. « L'audience de l'internet en France en avril 2011 » [En ligne]. *Médiamétrie, Mesure d'audience Télévision, Radio, Cinéma, Internet, études médias audiovisuels*. 23 mai 2011. Disponible sur : <http://www.mediametrie.fr/presse/communiques/l-audience-de-l-internet-en-france-en-avril-2011.php?id=463> (Consulté le 3/07/2011)

32 BNF. *Gallica - le blog de la bibliothèque numérique de la Bibliothèque nationale de France – BnF* [En ligne]. 29 juillet 2011. Disponible sur : <http://blog.bnf.fr/gallica/> (Consulté le 29/07/2011)

Gallica, Persée, etc. De plus, Google a actuellement le projet de créer un catalogue des bibliothèques du monde entier grâce au partenariat de certaines bibliothèques, dont la bibliothèque municipale de Lyon<sup>33</sup>.

Toutefois, la particularité des bibliothèques face à de grands groupes comme Google est le lien qu'elles entretiennent entre elles via les associations, les institutions mais également dans l'uniformisation de la mise en forme des données qu'elles tentent d'obtenir. Ainsi, avec l'arrivée d'Internet dans les bibliothèques, les premiers formats et normes documentaires sont créés pour faciliter les échanges de notices bibliographiques. Parmi eux, on retrouve le format de notice Marc (MACHINE-Readable Cataloging), le protocole Z39.50 utilisé pour la recherche multi-bases, etc.

Enfin, la naissance des bibliothèques hybrides est avant tout marquée par la simple mise en place d'accès à Internet dans leurs locaux. En effet, les bibliothèques, bien que devant mettre en avant leurs collections, ont aussi pour rôle principal de donner un accès au savoir quel qu'il soit. Cette tâche passe donc aussi par des tentatives de réduction de la fracture numérique. En 2008, seulement 56% de la population française disposait d'une connexion Internet à domicile. Ce chiffre est passé à 64% entre 2008 et 2010<sup>34</sup>.

D'autres projets de grande ampleur et plus spécifiques au monde de la recherche ont aussi vu le jour. Parmi eux, les archives ouvertes pluridisciplinaires Hal (Hyper Articles en Ligne) qui regroupent des articles de chercheurs, publiés ou non dans des revues. Cependant, contrairement à Gallica, le contenu de Hal n'est pas renseigné par des professionnels de la documentation mais par les chercheurs qui y déposent eux-même leurs articles.

### **3.2. Web 2.0 et l'interaction avec l'utilisateur**

Le web 2.0, nommé plus justement le web social ou web collaboratif, se définit par l'action des internautes sur les contenus du web. L'internaute peut donc participer à des wikis (encyclopédies en ligne), créer un blog, déposer du contenu sur des plateformes, commenter d'autres contenus, etc. Il complète le web de

---

33 GOOGLE. *Google Recherche de Livres* [En ligne]. 28 juillet 2011. Disponible sur : <http://books.google.com/googlebooks/about.html> (Consulté le 29/07/2011)

34 GOMBAULT V. « Deux ménages sur trois disposent d'internet chez eux » [En ligne]. *Institut national de la statistique et des études économiques : Accueil*. Mars 2011. Disponible sur : [http://www.insee.fr/fr/themes/document.asp?ref\\_id=ip1340](http://www.insee.fr/fr/themes/document.asp?ref_id=ip1340) (Consulté le 29/07/2011)

consultation en inscrivant l'échange de données et le participatif au centre des pratiques.

Tim O'Reilly, initiateur du terme de web 2.0, décrit les changements inclus dans ce nouveau terme : «

- le web en tant que plateforme ;
- expériences utilisateurs enrichies ;
- exploiter l'intelligence collective ;
- la puissance est dans les données ;
- fin du cycle des versions ;
- dépasser les limites du PC ;
- modèles de programmation allégés. »<sup>35</sup>

Dans cette vision d'Internet, l'internaute devient acteur du web. Le web n'est plus un outil pouvant être géré par quelques individus disposant de connaissances en développement et en programmation informatique. Il est désormais à la portée de tous. Le développement rapide du web ainsi que l'apparition de nouvelles applications et de nouveaux standards sont en partie dus à la forte présence de communautés Open Source sur Internet. En effet, ces communautés ont trouvé en Internet un moyen de communiquer plus facilement entre elles et de mettre facilement à disposition des internautes les technologies qu'elles développent.

Concernant le monde de la documentation, il a profité de certaines de ces évolutions pour tenter d'améliorer ses systèmes d'organisation des connaissances et s'adapter au mieux aux changements des besoins du public.

Durant cette période, certaines bibliothèques ont vu cette nouvelle place de l'utilisateur comme une aide à la gestion de la bibliothèque. Ainsi, de nombreux SIGB ont commencé à intégrer des modules censés aider les bibliothécaires dans certaines de leurs tâches habituelles :

- la folksonomie ou les tags ajoutés directement par les internautes afin d'indexer les documents. Cette technique a tenté de remplacer les taxonomies mais elle a rapidement rencontré ses limites dans le monde de la documentation. Les mots-clés déterminés par les internautes ne sont pas toujours pertinents. Tout d'abord, dans le cas de tags rédigés directement par l'internaute, un même terme peut être écrit sous différentes formes

---

35 PAPY F (dir.). *Technodocumentation : Des machines informationnelles aux bibliothèques numériques*. Paris : Lavoisier, 2009

(pluriel, singulier, masculin, féminin, mots composés, synonymes, etc.). Les mots-clés ne seront alors pas harmonisés avec un risque de répétition d'un même terme écrit différemment. Ensuite, même dans le cas plus courant d'une sélection des tags à partir d'une liste de mots-clés pré-établie par les documentalistes, les mots-clés choisis par l'internaute ne seront pas forcément les plus pertinents ;

- la suggestion d'achats permet aux usagers d'intervenir et d'être acteurs de la gestion du fonds de la bibliothèque ;
- la gestion d'un espace personnel peut permettre à l'utilisateur de modifier directement en ligne des informations le concernant. De plus, il peut ainsi consulter son compte lecteur à distance pour voir les documents qu'il a emprunté, pour éventuellement prolonger ses prêts ou encore pour réserver un document. Dans le milieu des bibliothèques universitaires, ces fonctionnalités sont souvent proposées dans les SIGB ;
- etc.

Parmi les autres outils que se sont appropriés les professionnels de l'information documentation, il existe également les flux RSS. Beaucoup sont utilisés dans le milieu de la veille mais les flux RSS peuvent aussi être présents dans certaines bibliothèques. Ces flux permettent de voir les mises à jour d'un site web automatiquement en passant par un agrégateur. Ainsi, toutes les mises à jour des sites sélectionnés par la bibliothèque seront visibles via l'agrégateur. Parmi les agrégateurs existants, on peut citer notamment Netvibes.

Bien d'autres technologies ont été appropriées par les bibliothèques : actualités montrant les dernières acquisitions, la consultation d'étagère virtuelle, etc. Toutefois, cela ne suffit pas à révolutionner la bibliothèque. En effet, ces outils n'ont pas su attirer suffisamment d'internautes pour affirmer leur pérennité. De plus, l'évolution attendue par le monde de la documentation depuis l'avènement d'internet concernerait plutôt un outil permettant de créer des systèmes parfaits d'organisation des connaissances. Pour beaucoup de professionnels de l'information-documentation, la solution résiderait dans un « web de données »

### **3.3. Web 3.0 et les *linked data***

Le désir d'un format permettant de relier et de faire interagir les données entre-elles date déjà d'une vingtaine d'années dans le monde de la documentation. Ainsi, des thésaurus puis des ontologies ont été créées par les professionnels de

l'information-documentation dans le but d'optimiser l'indexation et la recherche de documents<sup>36</sup>.

En 1999, le concept de web sémantique fut tout d'abord utilisé par Tim Berners Lee, membre et directeur du W3C, dans son ouvrage *Weaving the Web*. Il mentionne des ordinateurs « capables d'analyser toutes les données sur le Web - le contenu, les liens et les transactions entre les personnes et les ordinateurs. [...] les mécanismes d'échange au jour le jour, la bureaucratie et notre vie quotidienne seront traités par des machines qui parlent à d'autres machines ». Par la suite, la diffusion de cette notion passa par des noms plus attractifs mais aussi assez inappropriés tels que le web 3.0 ou encore le web intelligent.

En effet, on ne peut pas parler de web 3.0 (ni de web 1.0 et de web 2.0 d'ailleurs) car ce nom ferait référence à une nouvelle version, ou à une amélioration, du web. Or, en ce qui concerne le web, le versionning est difficilement imaginable puisque l'on utilise encore les mêmes outils et formats des débuts du web tels que le code HTML ou les CSS (Cascading Style Sheets). Les différentes évolutions du web sont donc toutes liées et constituent des compléments du web à ses débuts. Il faut également prendre en compte que, comme le web social, le web de données s'est construit sur les bases du web originel. De plus, le terme de web 3.0 est purement commercial, il s'agit plutôt d'une tentative pour donner un nom au web du futur et attirer le public. Or, on ne peut pas savoir en avance quelle sera la prochaine avancée en matière d'Internet, ni si celle-ci concernera de près ou de loin le web sémantique.

En ce qui concerne les expressions « web sémantique » et « web intelligent », elles ont souvent été critiquées à cause d'une connotation trop proche avec l'intelligence artificielle<sup>37</sup>. Cependant, contrairement à l'intelligence artificielle, les données du web sémantique ne seraient pas organisées automatiquement mais seraient structurées par l'Homme.

L'expression « web de données » a donc été l'appellation la plus largement plébiscitée dans le monde de la documentation. En effet, l'expression « web de données », ou *linked data* en anglais, est très concrète et met principalement en avant la technique et l'interaction des informations plutôt que le résultat. Ce nom a

---

36 MIGNIEN Y. « Bibliothèques Numériques » [En ligne]. *Encyclopaedia Universalis – Education*. 29 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/bibliotheques-numeriques/#5> (Consulté le 29/07/2011)

37 FLORIDI L. « Web 2.0 vs. the Semantic Web: A Philosophical Assessment ». *Episteme*, 2009, n°6, pp. 25-37

l'avantage de rappeler que les informations affichées à l'écran sont loin d'être automatiques mais qu'elles sont gérées par des professionnels de l'informatique et de la documentation via des techniques, des outils et des formats particuliers<sup>38</sup>.

Le principe du web de données, tel qu'il est défini par Tim Berners Lee, est d'enrichir toute information avec des méta-données (des données sur les données), des annotations, etc. L'information ainsi structurée pourrait permettre à toutes les données du web d'être connectées et reliées entre elles. Toutefois, une des conditions à la réussite de ce projet est l'utilisation de formats interopérables et de langages communs (thésaurus, dictionnaires, vocabulaires, etc.). Ainsi, le web de données serait certainement plus proche de la logique et de la représentation des connaissances que de l'informatique en elle-même.

Néanmoins, un projet tel que celui-ci et atteignant le web dans sa totalité reste encore au stade de l'utopie. Les quelques projets déjà réalisés concernent des domaines spécifiques. L'information-documentation est certainement un des domaines les plus impliqués car les problématiques d'organisation des connaissances se trouvent souvent au centre des réflexions. De plus, quelques pratiques déjà existantes en documentation ont été d'une grande aide dans le développement du web de données.

Au niveau des bibliothèques, le web de données pourrait donc être considéré comme une mutation de grande ampleur car il pourrait remettre en cause l'utilisation des catalogues, des SIGB, des OPAC, etc. Les procédures de gestion des documents pourraient elles aussi être modifiées. Les risques et mutations à prendre en compte sont donc nombreux. Cependant, ce projet est important car « il existe un énorme gisement de données enfouies dans tous les ordinateurs de la planète : en les reliant, le Web sémantique permettra d'exploiter cette mine pour améliorer nos connaissances dans de nombreux domaines. »<sup>39</sup>. Pour l'instant, beaucoup de projets en bibliothèques sont encore en cours de réalisation. On peut constater que tous ces projets sont basés sur des formats développés par le W3C.

---

38 BNF. « Web sémantique, Web de données » [En ligne]. *Bibliothèque nationale de France*. 26 octobre 2010. Disponible sur : [http://www.bnf.fr/fr/professionnels/web\\_semantique\\_donnees/s.web\\_semantique\\_intro.html](http://www.bnf.fr/fr/professionnels/web_semantique_donnees/s.web_semantique_intro.html) (Consulté le 29/07/2011)

39 POUPEAU G. « Le Web de données et les bibliothèques » [En ligne]. *Journées d'études AULA sur le Web sémantique*. 27 juin 2008. Disponible sur : <http://www.slideshare.net/lespetitescases/le-web-de-donnees-et-les-bibliothèques-presentation> (Consulté le 29/07/2011)



## **Partie 2 : Application du web de données en bibliothèque**

## 4. Les formats du web de données

Le web de données correspond à une mise en commun de plusieurs technologies et formats. Envisagé comme une évolution capitale pour le web, le W3C s'est chargé de déterminer les formats et les étapes lui paraissant les plus universels possible pour correspondre aux objectifs de ce projet. Ces formats et étapes sont schématisés par la « web semantic stack » (« la pile du web sémantique ») :

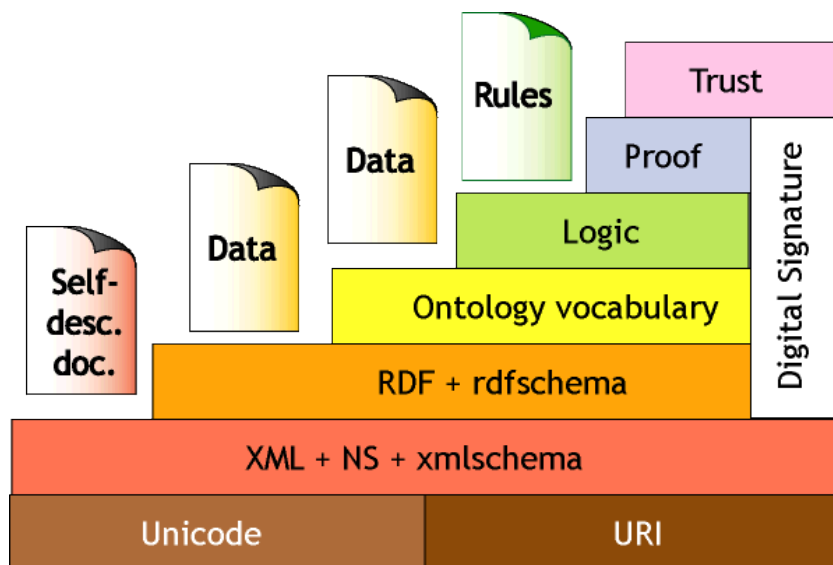


Fig 1. *The web semantic stack*<sup>40</sup>

Ce schéma, créé en 2000 par Tim Berners-Lee, est toujours d'actualité. En effet, il reprend toutes les étapes nécessaires à l'élaboration d'un web de données, hormis peut-être un langage permettant d'interpréter les données de ce web. Il recommande également les standards du web à utiliser pour garantir l'interopérabilité des données. Cette dernière étant une des conditions sine qua non au fonctionnement du web de données.

<sup>40</sup> BERNERS-LEE T. « Semantic Web on XML » [En ligne]. *XML 2000* : Washington DC, 6 décembre 2000. Disponible sur : <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html> (Consulté le 3/08/2011)

## 4.1. La localisation des données : les URI

Les Uri (Uniform Resource Identifier) permettent de localiser les métadonnées. Les URL (Uniform Resource Locator) sont un type d'Uri permettant d'identifier, ou de localiser, l'accès à une ressource électronique.

Afin que le web de données soit opérationnel, il est nécessaire que chaque donnée, appelée aussi terme d'un vocabulaire, dispose d'une Uri spécifique<sup>41</sup>. En effet, le web de données est organisé grâce à des référentiels de valeurs, connus aussi sous le nom *value vocabularies* en anglais, qui sont des sortes de listes ou de vocabulaires contrôlés comportant des propriétés et des classes. Ces listes sont basées sur différents types d'organisation des connaissances : listes de termes, thésaurus, schèmes de classification, taxonomies, listes d'autorités, etc. « value vocabularies often have http URIs assigned to the value, which would appear in a metadata record instead of or in addition to the literal value »<sup>42</sup>. Cette Uri sera donc nécessaire pour identifier tous les éléments d'un vocabulaire. Plusieurs sortes de référentiels de valeurs ont déjà été créés : LCSH (Library of Congress Subject Headings), Vial (Virtual International Authority File), etc.

## 4.2. Les règles d'écriture : le XML

Le format XML (Extensible Markup Language), créé en 1998, est une « évolution du SGML (Standard Generalized Markup Language) pour l'adapter à Internet »<sup>43</sup> qui a été améliorée et normalisée au fur et à mesure par le W3C. Cet outil permet de déterminer la structure d'autres langages informatiques, il est donc considéré comme un métalangage. Il est utilisé pour décrire un document, son contexte, son contenu, etc. Sa principale caractéristique est sa rigidité, jugée comme un frein pour certains tandis que d'autres la voient comme un moyen de structurer et d'homogénéiser les données (Cf. Annexe n°6 : Exemple de structure XML, p.104). Cette impression de rigidité vient principalement de son rôle qui est de déterminer des structures, des vocabulaires et des grammaires. Le XML est vu comme une extension, un complément du HTML qui permet de structurer

---

41 BERMÈS E. « Publier les référentiels sur le Web sémantique » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [http://www.bnf.fr/documents/afnor2011\\_bermes.pdf](http://www.bnf.fr/documents/afnor2011_bermes.pdf) (Consulté le 3/08/2011)

42 BERMÈS, 2011, *op. cit.*

43 RIVIER A. *Aide mémoire d'informatique documentaire*. Paris : Éditions du Cercle de la librairie, 2007

l'information, c'est-à-dire de renseigner des métadonnées (données sur les données)<sup>44</sup>.

Le XML est donc considéré comme une typologie d'écriture du web, utilisée dans divers autres langages et ayant même permis leur création. C'est notamment le cas pour l'OWL (Web Ontology Language) ou pour le RDF (Resource Description Framework) qui sont des piliers du web de données. Le XML permet ainsi de faciliter les échanges et la communication entre les données.

Enfin, le XML permet également de décrire un contenu sans devoir prendre en compte la présentation de la ressource qui est directement gérée par le langage HTML<sup>45</sup>. Ainsi, un document XML peut être lisible quelle que soit l'application utilisée. Ce schéma de données est donc central dans la réalisation d'un web de données ainsi que dans l'interopérabilité qu'il tente d'obtenir.

Les principales caractéristiques du XML ont suscité l'intérêt du W3C qui en a fait une recommandation :

- « La lisibilité : aucune connaissance ne doit théoriquement être nécessaire pour comprendre un contenu d'un document XML
- Autodescriptif et extensible
- Une structure arborescente : permettant de modéliser la majorité des problèmes informatiques
- Universalité et portabilité : les différents jeux de caractères sont pris en compte
- Déployable : il peut être facilement distribué par n'importe quels protocoles à même de transporter du texte, comme HTTP
- Intégrabilité : un document XML est utilisable par toute application pourvue d'un parser (c'est-à-dire un logiciel permettant d'analyser un code XML)

---

44 Dans mes recherches sur le XML, j'ai pu m'aider de l'entretien que j'ai réalisé le 11 février 2011 avec Christian COTE, Maître de conférences à l'Université Lyon 3. Il m'a également fourni le contenu de ses cours de « Structure logique de l'information et XML » et « Représentation des connaissances » du Master 1 Information-Documentation de Lyon 3

45 COMMENT ÇA MARCHE. « Introduction à XML » [En ligne]. *Comment Ça Marche (CCM)* – *Communauté informatique*. 14 octobre 2008. Disponible sur : <http://www.commentcamarche.net/contents/xml/xmlintro.php3> (Consulté le 3/08/2011)

- Extensibilité : un document XML doit pouvoir être utilisable dans tous les domaines d'applications »<sup>46</sup>.

Ce format n'est pas encore très utilisé en bibliothèque. En effet, beaucoup de catalogues et de SIGB répondent encore au format Marc (Machine-Readable Cataloging) qui permet notamment de faciliter l'échange de notices entre les bibliothèques. Toutefois, ce format ne correspond pas aux exigences du web de données.

La Bibliothèque du Congrès a créé plusieurs formats basés sur le langage XML et répondant aux exigences de la bibliothéconomie : indexation, partage de notices, etc. Parmi ces formats, on retrouve notamment les schémas Marc XML et MODS (Metadata Object Description Schema<sup>47</sup>). Du côté de la France, la Mission de la Recherche et de la Technologie du Ministère de la Culture et de la Communication a mis en place le format BiblioML basé lui aussi sur le XML.

Ainsi, cette structure de langage informatique facilite grandement l'échange de données et de documents. De plus, elle a permis de créer le RDF qui est un autre pilier du web de données.

### **4.3. La structure des données : le RDF**

Le RDF (Resource Description Framework) est un type de structure de données. Ce format permet de créer des triplets ou prédicats composés de deux données et d'une relation liant ces deux données. Ainsi, cette structure se construit sous une forme « sujet - verbe - complément » ou « sujet - prédicat - objet ». La structuration des données sous le format RDF permet d'établir des graphes de connaissances, c'est-à-dire une sorte de cartographie mettant en lien les différentes informations d'une base de données. Les objets de ces graphes peuvent être identifiés par une URI. Dans le cas des sujets et des prédicats, cette identification est obligatoire. Les graphes créés ont alors la forme suivante :

---

46 COMMENT ÇA MARCHE, 2008, *op. cit.*

47 CALDERAN L., HIDOINE B., MILLET J (dir.). *Métadonnées : mutations et perspectives : Séminaire INRIA, 29 septembre-3 octobre 2008 - Dijon*. Paris : ADBS, 2008

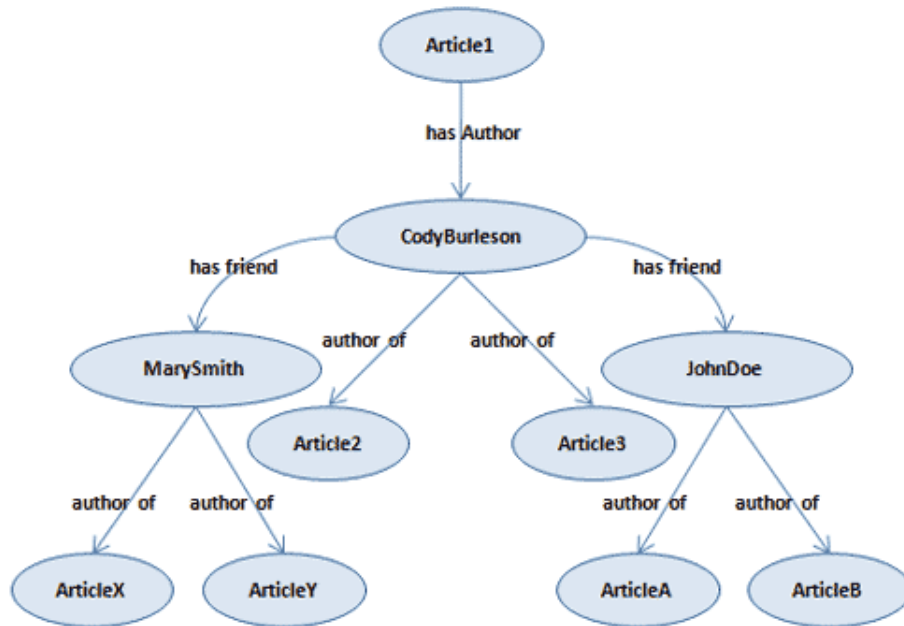


Fig. 2 : Structure de *graphe RDF*<sup>48</sup>

Ce langage est important pour le web de données car il permet de décrire les ressources et les métadonnées sur le web. De plus, il peut s'appliquer sur toutes sortes de documents, structurés ou non.

En ce qui concerne la syntaxe du RDF, elle suit souvent les règles du XML, ce qui a notamment pour conséquence de créer un amalgame entre le RDF, qui désigne le schéma de triplets et les graphes de données, et le RDF/XML, qui désigne la façon dont les données de ces mêmes graphes sont écrites.

Enfin, au niveau du web de données, le RDF permet donc d'aider à la création des vocabulaires et des ontologies.

#### 4.4. Les vocabulaires : le RDFSchema (RDFS) et l'OWL

Le RDFSchema et l'OWL (Web Ontology Language) sont des langages, ou vocabulaires, écrits en XML et basés sur une structure de type RDF. Cependant, il

<sup>48</sup> BURLESON C. « Introduction to the Semantic Web Vision and Technologies : Part 3 - The Resource Description Framework » [En ligne]. *Semantic Focus - The Semantic Web, Semantic Web technology and computational semantics*. 12 octobre 2007. Disponible sur : <http://www.semanticfocus.com/blog/entry/title/introduction-to-the-semantic-web-vision-and-technologies-part-3-the-resource-description-framework/> (Consulté le 3/08/2011)

existe une différence notable entre ces deux formats. En effet, l'OWL est considéré comme plus complet et plus complexe. Il fait partie des ontologies.

Dans sa définition première, l'ontologie est un concept philosophique consistant à étudier « l'être en tant qu'être »<sup>49</sup>. Par la suite, ce terme a été également utilisé en linguistique et en intelligence artificielle. Du point de vue du domaine de l'informatique, le terme d'ontologie désigne la modélisation et la représentation des connaissances. L'ontologie se distingue du vocabulaire par les relations particulières et précises (équivalences, différences, etc.) qu'elle permet de créer entre les termes<sup>50</sup>. Ainsi, « Une ontologie implique ou comprend une certaine vue du monde par rapport à un domaine donné. Cette vue est souvent conçue comme un ensemble de concepts [...], leurs définitions, leurs interrelations et différentes propriétés et contraintes associées. On appelle cela une conceptualisation. Une ontologie peut prendre différentes formes mais elle inclura nécessairement un vocabulaire de termes et une spécification de leur signification »<sup>51</sup>.

Concernant le RDFS, il est ce que l'on appelle un « vocabulaire simple ». En effet, il se contente uniquement d'apporter un système de hiérarchie de classe au graphe RDF<sup>52</sup>. Il permet alors de déterminer si une donnée fait partie d'une sous-classe, d'une sous-catégorie, des propriétés de la ressource, etc. (Cf. Annexe n°7 : Extrait de RDFS, p.105).

L'OWL va plus loin que le RDFS car il permet de créer des liens entre les différents concepts. Il est moins rigide et plus proche de la réalité. Il correspond donc plus à la définition de l'ontologie en informatique. En effet, il peut, par exemple, gérer des liens d'équivalence, de différences, etc.<sup>53</sup> L'OWL permet alors aux ordinateurs de mieux « interpréter » les ressources d'Internet car il offre un vocabulaire plus ample et une représentation sémantique plus juste que le RDFS.

---

49 GARNIER Y (dir.). « ontologie ». *Le Petit Larousse illustré 2009*. Paris : Larousse, 2008

50 DAMERON O. *Modélisation, représentation et partage de connaissances anatomiques sur le cortex cérébral : Thèse*. Rennes : Université de Rennes 1, 2003. Disponible sur : <http://websemantique.org/Ontologie> (Consulté le 3/08/2011)

51 PÉDAUQUE R.T. *La redocumentarisation du monde*. Toulouse : Cépaduès, 2007

52 NEPOTE C. « Web Sémantique : RDFSchema » [En ligne]. *Web Sémantique*. 23 juillet 2011. Disponible sur : <http://websemantique.org/RDFSschema> (Consulté le 3/08/2011)

53 BEN AMEUR H. « Web Sémantique : OWL » [En ligne]. *Web Sémantique*. 23 juillet 2011. Disponible sur : <http://websemantique.org/OWL> (Consulté le 3/08/2011)

## 4.5. Les requêtes : le SparQL

Le SparQL (Simple Protocol And RDF Query Language) est un langage de requête, ou d'interrogation, basé sur la structure en triplet du RDF. Il est une des recommandations du W3C en matière de web de données<sup>54</sup>.

Sa particularité est qu'il crée des « inférences » en se basant sur les graphes RDF et grâce aux termes de la requête, c'est-à-dire qu'il affiche le résultat des requêtes en fonction des liens présents dans les graphes. Ainsi, la requête est traitée dans son ensemble plutôt que terme par terme comme c'est le cas avec le SQL (Structured Query Language). Ce dernier reste actuellement très utilisé pour la gestion des bases de données, et la plupart des moteurs de recherche (Google, Yahoo, etc.).

L'inférence est un raisonnement fait à partir de plusieurs propositions qui permettent d'établir des conclusions. Le SparQL peut utiliser les trois sortes d'inférences existantes :

- l'induction : partir des faits pour obtenir une généralité ;
- la déduction : confirmer ou infirmer une hypothèse par des faits ;
- l'abduction : raisonnement basé sur une première proposition certaine et une seconde proposition incertaine qui est toutefois fortement liée à la première proposition. Cette forme est très proche du syllogisme.

Cette technologie a donc elle aussi une place légitime dans le fonctionnement du web de données et la diffusion des connaissances.

## 4.6. Bilan général

Le nombre important d'applications qui supportent aujourd'hui le XML, le RDF et le SPARQL rend en partie imaginable l'aboutissement d'un web de données. Malgré l'existence de nombreuses bases de données, il serait donc possible d'établir une circulation entre les différentes connaissances présentes sur le web. Toutefois, il ne faut pas oublier que le fonctionnement du web de données est en premier lieu lié à une technique qui a été normalisée et acceptée quasi universellement. Dans cette universalisation, le W3C tient un rôle capital.

De nombreux autres outils existent pour parvenir à appliquer le web de données. Toutefois, ils ne font pas partie des standards et recommandations du W3C. Ils ne sont pas forcément des formats dits ouverts. Ainsi, mon choix s'est

---

<sup>54</sup> COTE, 2011, *op. cit.*



porté sur les exemples les plus connus, génériques et possédant une certaine légitimité dans le web de données.

Enfin, en ce qui concerne le milieu de la documentation et des bibliothèques, ces formats sont étudiés et utilisés dans le but d'améliorer les systèmes d'organisation des connaissances. Que ce soit au niveau de recherches spécifiques au monde de la documentation ou de recherches concernant Internet en général, les professionnels de la documentation jouent un rôle important dans les réalisations du web de données.

## **5. Réalisations pour l'information-documentation**

Beaucoup des nouvelles technologies du web de données sont donc basées sur les formats vus précédemment et recommandés par le W3C. Parmi ces avancées, certaines tentent de s'appliquer au web dans sa globalité sans grand succès. On peut citer, par exemple, les moteurs de recherche sémantique tels que les projets Edelweiss de l'Inria (Institut National de Recherche en Informatique et en Automatique) et Powerset de Microsoft.

D'autres projets plus aboutis existent. Parmi eux, certains deviennent de plus en plus connus, c'est le cas de Dbpedia, créé en 2007, qui tente de structurer toutes les données de Wikipédia afin de les rendre plus facilement interrogeables.

Cependant, les nouveaux outils sur lesquels interviennent les professionnels de la documentation sont bien moins connus du grand public. En effet, pour l'instant, l'utilisation de ces technologies est encore très restreinte aux professionnels de la gestion des connaissances.

### **5.1. Dublin Core et Skos**

Avec le web de données, les bibliothèques ont été confrontées à un problème important de compatibilité des formats avec le reste du web. Le format Marc, qui est couramment utilisé pour la création et l'échange de notices bibliographiques, est incompatible avec le web de données. Cette incompatibilité se situe surtout au niveau du système de triplets et de la séparation entre la ressource et sa description mais également au niveau des vocabulaires.

En effet, le web de données a impliqué la transformation de langages documentaires tels que Rameau et LCSH (Library of Congress Subject

Headings)<sup>55</sup>. Ainsi, la Bibliothèque Nationale de France et la Bibliothèque du Congrès ont adapté ces deux vocabulaires, piliers de l'univers des bibliothèques, en Skos (Simple Knowledge Organization System) afin qu'ils correspondent au fonctionnement en triplet du web de données.

Le Skos est un format, ou vocabulaire, devenu une des recommandations du W3C en 2009. Il sert à représenter divers types de vocabulaires structurés et contrôlés (thésaurus, classifications, etc.) tout en respectant le schéma RDF. Le Skos permet d'encoder les métadonnées, c'est-à-dire qu'il attribue à chaque concept d'un thésaurus une Uri permettant ainsi la communication avec d'autres fichiers RDF. Cependant, bien que ce format soit très orienté vers les métiers de la documentation, il est aussi couramment associé à des formats utilisés par d'autres communautés (informaticiens, logiciens, etc.).

Toutefois, pour être opérationnel, le web de données a également besoin de modèles de métadonnées. Ces modèles sont le plus souvent généralistes pour pouvoir s'appliquer à n'importe quel vocabulaire. L'un des plus connus est le Dublin Core Metadata Initiative (DCMI), appelé plus couramment le Dublin Core (DC)<sup>56</sup>.

Le Dublin Core est un format de description de ressources, créé en 1995, proposant quinze éléments possibles pour la description de ressources : Title, Subject, Description, Source, Language, Relation, Coverage, Creator, Contributor, Publisher, Rights, Date, Type, Format, Identifier<sup>57</sup>(Cf. Annexe n°8 : Les 15 éléments du Dublin Core, p.106). Ces éléments ont été déterminés par un consensus international de professionnels de la documentation (bibliothécaires, conservateurs de musée, etc.), de l'informatique, etc. Le rôle premier du Dublin Core est de décrire des ressources peu complexes. Il est donc bien un modèle censé être complété par ses extensions ou tout autre schéma de métadonnées, en XML de préférence. Les modélisations sont souvent utilisées en ingénierie documentaire pour se rendre compte ou pour créer la structuration d'un système.

---

55 JABES. *Le Web sémantique : un web de métadonnées* [En ligne]. Abes, 2010. Disponible sur : <http://www.slideshare.net/yannn/le-web-smantique-un> (Consulté le 3/08/2011)

56 ABEL M.-H. « Utilisation de normes et standards dans le projet Memorae ». *Distances et savoirs*, avril 2004, vol. 2, pp. 487-511

57 BNF. « Dublin Core » [En ligne]. *Bibliothèque nationale de France*. 6 avril 2011. Disponible sur : [http://www.bnf.fr/fr/professionnels/formats\\_catalogage/a.f\\_dublin\\_core.html](http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html) (Consulté le 3/08/2011)

La frontière est donc très étroite entre les différents formats du web de données. Le but premier de ces formats est d'être interopérables et complémentaires tout en étant capables de décrire les ressources de façon compréhensible pour la machine et pour l'homme. Ainsi, le Dublin Core et le Skos forment un très bon exemple de complémentarité applicable en bibliothèque : le premier permettant de modéliser les concepts et le second servant à les représenter. Ces types de formats ont permis de concrétiser des projets du web de données tel que l'Open Archives Initiative.

## **5.2. L'initiative pour des archives ouvertes : l'OAI**

L'OAI est un projet né dans les années 1990 dont le but est de proposer un « service universel d'archivage des publications scientifiques »<sup>58</sup> mais aussi de valoriser et de permettre l'échange des archives numériques. Ce projet est supporté par des spécialistes des bibliothèques et des sciences de l'information (chercheurs et professionnels). Il a donné naissance en 2001 à un protocole assurant l'interopérabilité entre les différentes bases de données contenant les archives : l'OAI-PMH (Open Archives Initiative's Protocol for Metadata Harvesting).

Les dépôts OAI sont principalement utilisés par les chercheurs pour y déposer les différents articles qu'ils écrivent, qu'ils soient déjà publiés ou non. Pour la recherche, cela constitue une grande avancée car Internet donne ainsi accès à des ressources très récentes. Les chercheurs ne sont plus tributaires des éditeurs et procèdent eux-mêmes à un auto-archivage.

Du côté des bibliothèques, l'OAI permet de faire face et de négocier avec des éditeurs faisant payer leurs abonnements de plus en plus cher. De plus, beaucoup d'éditeurs proposent des revues électroniques qui ne sont pas physiquement à la disposition des bibliothèques mais qui sont le plus souvent conservées dans des bases de données appartenant à l'éditeur. En 2000 et 2001, le CCSD (Centre pour la Communication Scientifique Directe) crée les bases, ou entrepôts OAI, TEL (Thèses En Ligne) et HAL (Hyper Articles en Ligne) permettant aux chercheurs d'y déposer leurs thèses et leurs articles, publiés ou non. Ceci permet au chercheur de diffuser directement son travail de recherche sans devoir attendre qu'il soit publié dans une revue : on appelle ces articles non publiés des pre-prints.

Enfin, au niveau universitaire, Lyon 2 fait partie des précurseurs en matière d'archivage ouvert avec la mise en place en 1998 de l'archivage de ses thèses sous

---

58 SCHMITT J.-P. « L'Open Archives Initiative et la validation des publications scientifiques : Un séminaire sur l'OAI à Genève ». *Documentaliste-Sciences de l'Information*, février 2001, vol. 38, pp. 124-126

Cyberthèses. De plus, les années 2000 et la création du CCSD marquent également la naissance de l'entrepôt ArchiveSic spécialisé dans l'archivage d'articles en sciences de l'information et de la communication. Certaines structures, telles que L'Insa de Lyon et l'Afnor, sont également à l'origine du projet Ori-OAI (Outil de Référencement et d'Indexation - Open Archives Initiatives)<sup>59</sup> repris en 2005 par l'Unit (Université Numérique Ingénierie et Technologie). L'Insa de Lyon participe encore à ce projet en travaillant sur les ressources numériques pédagogiques, scientifiques ou documentaires et leur système de partage. L'Unit cherche à respecter les normes et protocoles internationaux de partage (OAI-PMH), de description (métadonnées) et d'indexation (classifications) dans ses travaux.

C'est donc le foisonnement de tous ces projets qui a permis la création du protocole OAI-PMH. Ce dernier permet d'échanger et de récupérer facilement des données sur le web pour pouvoir ensuite les indexer, les valoriser et les rendre accessibles aux usagers. Cependant, les données collectées devront nécessairement se trouver dans des entrepôts moissonnables OAI en libre accès pour que les échanges soient possibles. Un moissonneur OAI interroge alors les entrepôts moissonnables OAI-PMH à l'aide de requêtes XML puis dépose les données collectées sur une base de données. C'est ensuite les responsables de la base de données qui choisissent de rendre accessible à leur public la totalité ou une partie des données collectées. Ainsi, l'internaute lancera une requête qui interrogera cette base et il pourra retrouver des ressources provenant de différents « entrepôts » et rapatriées sur une même base qui sert de réservoir. Afin de vérifier si la ressource trouvée sur internet est bien un répertoire moissonnable en OAI, le moissonneur a la possibilité de soumettre des requêtes XML aux entrepôts de données. De plus, ces requêtes permettent de définir les caractéristiques de l'exploitation de ces entrepôts par le moissonneur (liste des métadonnées à collecter, fréquence de moissonnage, etc.).

Grâce à ses caractéristiques, l'OAI fait elle aussi partie des projets du web de données. En effet, elle correspond aux recommandations du W3C avec, par exemple, l'utilisation du XML. De plus, elle place les données au centre du web, elle souhaite être interopérable et facilement accessible, etc.

Bien que l'OAI permette une grande ouverture sur les connaissances et intéresse donc fortement le milieu des bibliothèques, elle reste un projet touchant la sphère

---

59 COLMANT Y. « Historique du projet ORI-OAI » [En ligne]. *ORI-OAI.ORG*. 6 mai 2010. Disponible sur : <http://www.ori-oai.org/display/PROJET/Historique> (Consulté le 3/08/2011)

universitaire en priorité. D'autres projets, tel que le FRBR, sont plus spécifiques à la bibliothéconomie.

### **5.3. Un projet spécifique à l'information-documentation : le FRBR**

Le FRBR (Functional Requirements for Bibliographic Records, ou en Français Spécifications Fonctionnelles des Notices Bibliographiques) est un modèle conceptuel de représentation de l'univers bibliographique créé dans les années 1990 et publié en 1998 par l'Ifla (International Federation of Library Associations and Institutions). Il tente de modéliser les informations bibliographiques sous un modèle entités-relation et permet donc de se rendre compte des relations existantes entre les différentes notices bibliographiques.

Les entités du FRBR sont organisées en trois groupes différents :

- Groupe 1 : il décrit le document et ses réalisations
  - œuvre : création intellectuelle ou artistique déterminante, abstraite. « Par exemple : le texte original d'un roman et une traduction de ce roman se rattachent à une même création abstraite »<sup>60</sup>;
  - expression : correspond à la réalisation intellectuelle de ce travail. Par exemple, les versions d'un même document : le pré-print, la version imprimée, la traduction, etc. ;
  - manifestation : identifie les caractéristiques de publication du document : sous quelle forme il se matérialise (PDF, Word, html ou traduit chez X et publié chez Y) ;
  - item : permet d'identifier les exemplaires en cas de copie multiple du document ;

Ces quatre entités nous permettent d'indiquer le fait qu'une œuvre peut avoir plusieurs expressions : chaque œuvre peut avoir plusieurs expressions, chaque expression peut donner lieu à plusieurs manifestations.

---

<sup>60</sup> BNF. « Modèles FRBR, FRAD et FRSAD » [En ligne]. *Bibliothèque nationale de France*. 6 avril 2011. Disponible sur : [http://www.bnf.fr/fr/professionnels/modelisation\\_ontologies/a.modele\\_FRBR.html](http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html) (Consulté le 3/08/2011)

- Groupe 2 : il correspond à la modélisation des autorités. Il met donc en relation le document avec une personne ou une collectivité, physique ou morale, qui a conçu, réalisé, produit ou est en possession du document<sup>61</sup> ;
- Groupe 3 : il décrit le sujet de l'œuvre. Il est décliné en quatre entités : concept, objet, événement et lieu. Une œuvre traite d'un sujet ou une personne traite d'un sujet. Les entités du groupe 3 sont combinables avec celles des groupes 1 et 2<sup>62</sup>.

L'intérêt du FRBR dans le catalogage et le web de données est qu'il permet la modélisation, c'est-à-dire de créer des relations, entre les différentes entités d'une notice. En effet, les relations créées à partir du groupe 3 rappellent fortement les ontologies qui ont une place importante dans le web de données.

Pour certains professionnels, le FRBR est qualifié d'inutile. On peut également penser que si le FRBR éprouve des difficultés à s'appliquer dans les bibliothèques, c'est principalement parce qu'il est arrivé très en retard dans ce milieu<sup>63</sup>. En effet, contrairement aux informaticiens, les bibliothécaires ont tout d'abord créé des formats d'implémentation, de développement (par exemple le Marc) avant de créer des modélisations. Cela pose problème car, lors de la définition de nouveaux standards et de nouveaux besoins, ces formats deviennent peu à peu obsolètes.

Ainsi, grâce au FRBR, les bibliothécaires peuvent se rendre compte des relations existantes entre toutes les informations contenues dans une notice bibliographique. Bien que ce type de relation et de technologie semblent déjà exister dans les catalogues de bibliothèques actuels, avec les systèmes de « rebond » et de recherche à facette, ces systèmes ne sont pas aboutis et comportent de nombreuses lacunes.

Ainsi, malgré des points de vue très divergents sur le FRBR, il est tout de même important de constater que cette technologie représente un grand pas pour la recherche sur le web de données en bibliothèque. De plus, on peut observer une

---

61 LE PAPE P. « RDA : le nouveau paradigme du catalogage. : FRBR, c'est quoi déjà ? » [En ligne]. 5 à 7 de l'*AADS*, 29 mars 2011. 21 juin 2011 Disponible sur : <http://www.adbs.fr/rda-le-nouveau-paradigme-du-catalogage-1-frbr-c-est-quoi-deja--102461.htm> (Consulté le 3/08/2011)

62 Concernant le FRBR, j'ai pu avoir accès au cours de Master 1 Information-Documentation « Logiciels documentaires et Geide » de Mme Fidelia IBEKEWE, Maître de conférences à l'Université Lyon 3

63 BERMÈS E. « Les FRBR, qu'est-ce que c'est ? » [En ligne]. *Figoblog : Un blog sur Internet, la bibliothéconomie et la confiture de figes*. 10 mars 2005. Disponible sur : <http://www.figoblog.org/document594.php> (Consulté le 3/08/2011)

certaine similitude entre le fonctionnement du FRBR et d'autres outils du web de données tels que le Foaf et le Viaf.

#### **5.4. Applications RDF : les vocabulaires Foaf et Viaf**

Le Foaf (Friend Of A Friend) est un vocabulaire RDF permettant de décrire des personnes, leurs connaissances et les relations qu'elles entretiennent entre-elles. Il est donc très proche du fonctionnement du groupe 2 du FRBR (par exemple X est l'auteur de... ou X est la sœur de...). De nombreuses applications Foaf ont déjà été créées (FOAF Bulletin Board, FoaF Explorer, etc.). Toutefois, elles sont plutôt perçues comme faisant partie du web social et non du web de données.

En effet, le Foaf est surtout utilisé par des internautes ne souhaitant pas posséder le simple statut de lecteur sur le web mais voulant être producteurs de données, diffuseurs de contenus et pouvant faire partie de réseaux sociaux, de communautés, de groupes de travail en ligne, etc. Le Foaf comporte de nombreux champs allant des informations de base sur les individus (nom, prénom, etc.) aux groupes auxquels appartient l'individu (entreprise, parti politique, etc.) en passant par ses informations personnelles (connaissances, hobbies, etc.), ses comptes en ligne (adresses email, messageries instantanées, etc.) et ses documents (articles, photos, etc.). On constate alors l'étendue des données pouvant être récoltées via Foaf.

Bien que le Foaf soit très proche du web social, il est tout de même considéré comme l'un des projets précurseurs du web de données. Il fonctionne par rapport au système de triplets recommandé dans le RDF et il est écrit en respectant les balises XML<sup>64</sup>. Toutefois, il n'est pas devenu un standard. En effet, même si cette technologie n'est pas développée à des fins commerciales et que chaque individu est censé gérer son propre fichier Foaf, elle est loin d'être totalement rassurante. Le Foaf, avec toutes les données qu'il peut récolter et mettre en relation sur les individus, relance la polémique sur le respect de la vie privée et l'utilisation des informations personnelles par les entreprises et les spammeurs.

Toutefois, des projets assez similaires ont réussi à s'imposer dans le domaine du web de données et des bibliothèques. On mentionne aujourd'hui plus souvent le Viaf (Virtual International Authority File ou Fichier d'Autorité International Virtuel), développé depuis 2003 par l'OCLC (Online Computer Library Center) et

---

64 FIEVET C. « FOAF, la première brique du Web sémantique ? » [En ligne]. *InternetActu.net*. 2 juin 2004. Disponible sur : <http://www.internetactu.net/2004/06/02/foaf-la-premiere-brique-du-web-smantique/> (Consulté le 3/08/2011)

l'Ifla, spécialisé tout d'abord pour les noms de personnes. Ce projet regroupe 19 bibliothèques nationales du monde entier dont la Bibliothèque Nationale de France<sup>65</sup>. Dans ce projet, « les notices d'autorité produites par des agences bibliographiques nationales du monde entier et représentant la même entité seraient reliées entre elles et mises à la disposition de tous sur Internet »<sup>66</sup>. Cela consiste donc à créer des liens, appelés nœuds, entre les différents fichiers d'autorité dans le but de faciliter les recherches sur un catalogue universel, gratuit et disponible sur Internet quelles que soient la langue, l'écriture (un mot peut avoir différentes formes d'écriture. Par exemple : Johannes P. De Valk et J. P. De Valk) et la graphie de l'utilisateur. Ainsi, cela permet, par exemple, de créer automatiquement et virtuellement un seul fichier d'autorité pour les noms de personnes au niveau international. Cette procédure est appelée appariement et permet de faire une seule recherche dans tous les catalogues faisant partie du Vial.

De plus, le Vial fonctionne de concert avec une autre technologie vue précédemment : l'OAI. En effet, afin de diffuser le Vial sur Internet et de parvenir à l'intégrer parfaitement au web de données, l'OCLC et l'Ifla ont décidé de respecter les protocoles OAI.

Ensuite, il faut préciser que l'efficacité de l'appariement vient du fait qu'il ne prend pas seulement en compte les noms de personnes. En effet, il existe de nombreux homonymes (« environ 10% des paires de noms »<sup>67</sup>). Or, les erreurs dans les fichiers d'autorité Vial s'élèvent à moins de 1%. Cette réussite est possible grâce notamment à la prise en compte des dates de naissances et des ouvrages (ou plutôt des ISBN, International Standard Book Number) liés à la personne. Pour résumer, afin de déterminer un fichier d'autorité, le Vial s'appuie sur les entités du groupe 1 du FRBR vu précédemment.

Aujourd'hui, le Vial tente de se rapprocher encore d'autres technologies telles que l'Isni (International Standard Name Identifier) qui permet d'identifier une

---

65 BOULET V. « VIAF : une brique importante pour le web sémantique » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_vial\\_boulet.pdf](http://www.bnf.fr/documents/afnor2011_vial_boulet.pdf) (Consulté le 3/08/2011)

66 BENNET R., HENGEL-DITTRICH C., al. *VIAF (Virtual International Authority File) : un pont entre les fichiers d'autorité noms de personnes de Die Deutsche Bibliothek et de la Bibliothèque du Congrès* [En ligne]. Seoul : Ifla, 13 juillet 2006. Disponible sur : [http://archive.ifla.org/IV/ifla72/papers/123-Bennett\\_trans-fr.pdf](http://archive.ifla.org/IV/ifla72/papers/123-Bennett_trans-fr.pdf) (Consulté le 3/08/2011)

67 FIEVET, 2004, op. cit.



personne malgré les pseudonymes qu'elle emprunte. Par exemple, Boris VIAN et Vernon SULLIVAN seront identifiés par le même Isni.

Enfin, le Viaf ne cesse de se développer avec de nouvelles applications pour favoriser l'accès aux ressources. La Bibliothèque Nationale de France devrait lancer cet été le site [data.bnf.fr](http://data.bnf.fr). De plus, il continue d'attirer de nouveaux organismes comme, dernièrement, le Sudoc qui devient de plus en plus impliqué dans le web de données et souhaiterait compléter son catalogue collectif grâce au Viaf.

## 5.5. Le Sudoc et l'Idref

Le Sudoc (Système Universitaire de DOcumentation) est un projet lancé en 1992 par le Ministère de l'Enseignement Supérieur et géré par l'Abes (Agence Bibliographique de l'Enseignement Supérieur). Mis en œuvre en 1998, il a été fonctionnel dans toutes les universités participantes seulement à partir de 2002<sup>68</sup>.

Ses objectifs sont de :

- proposer un catalogue collectif des bibliothèques de l'enseignement supérieur accessible à tous sur Internet ;
- permettre aux usagers inscrits d'emprunter des documents grâce aux Peb et, dans certains cas, d'accéder à des documents en ligne ;
- aider les bibliothécaires à la gestion du Peb et diminuer le travail sur les notices grâce à l'échange et au rapatriement de notices bibliographiques ;

Depuis sa création, le Sudoc place les données au centre de son fonctionnement. En effet, dès les années 1990, il s'appuie déjà sur des fichiers d'autorité provenant de la BNF et permettant de déterminer un identifiant unique pour chaque notice bibliographique<sup>69</sup>.

Entre 2006 et 2010, l'intégration du protocole Z39.50, utilisé pour interroger plusieurs catalogues en même temps, dans les SIGB permet aux bibliothèques de récupérer les autorités du Sudoc sur leurs catalogues. De plus, ce protocole permet également au Sudoc de faire le lien et d'intégrer ses autorités aux applications Star, catalogue de thèses, et Calames, catalogue de manuscrits, de l'Abes.

---

68 ABES. « Le réseau Sudoc » [En ligne]. *ABES : Accueil*. Disponible sur : [http://www.abes.fr/abes/page.352\\_le-reseau-sudoc.html](http://www.abes.fr/abes/page.352_le-reseau-sudoc.html) (Consulté le 3/08/2011)

69 NICOLAS Y. « IdRef : Les référentiels comme services. Une politique de l'ABES » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_abes.pdf](http://www.bnf.fr/documents/afnor2011_abes.pdf) (Consulté le 3/08/2011)

Enfin, depuis l'automne 2010, l'Abes a mis en place un nouvel outil lié au Sudoc : IdRef-Autorités (Identifiants et Référentiels Sudoc pour l'enseignement supérieur et la recherche). L'IdRef est une base donnant accès aux autorités du Sudoc. Elle fournit la liste des autorités Sudoc (personnes physiques, collectivités, concepts Rameau et MeSH, etc.) et facilite ainsi l'accès aux ressources en affichant, par exemple, tous les ouvrages d'un auteur. Cette application permet d'interroger les autorités mais aussi de les modifier et de les enrichir. En juillet 2011, l'Abes a prévu d'étendre l'IdRef aux autres applications associées au Sudoc, c'est-à-dire Star et Calames, permettant ainsi l'interrogation des trois bases simultanément.

L'intérêt de l'IdRef est qu'il permet de mettre en relation les données tout en réutilisant un référentiel existant. De plus, il participe activement au web de données en permettant à d'autres applications de récupérer ses référentiels de données et en utilisant le format XML, le RDF et les Uri.

Bien qu'il existe différents types de services et de formats, le Sudoc et, plus largement, les bibliothèques du monde entier tentent d'utiliser les mêmes vocabulaires afin d'assurer l'interopérabilité des données. Toutefois, le seul désir d'interopérabilité et de diffusion des connaissances est insuffisant pour expliquer le succès de ces technologies. Les raisons qui poussent ces bibliothèques à appliquer les mêmes formats sont nombreuses et ne dépendent pas toujours des bibliothèques elles-mêmes. En effet, comme dans beaucoup d'organismes, les bibliothèques doivent aussi répondre aux besoins du public, aux nécessités professionnelles et faire face aux groupes d'influence.

## **6. La prise de décision en bibliothèque<sup>70</sup>**

Les bibliothèques, quel que soit leur rayonnement (national, universitaire, municipal, etc.), dépendent du gouvernement. Comme la plupart des organismes publics, elles sont donc un élément politique attirant certains acteurs. Il est alors évident que le milieu des bibliothèques et celui de l'informatique sont eux aussi, comme beaucoup d'autres domaines, influencés par des lobbyistes. Nous tenterons de mettre en évidence certains de ses acteurs influents.

---

<sup>70</sup> Pour la rédaction de ce chapitre, j'ai pu me baser sur mes différentes expériences professionnelles dans le milieu de la documentation et dans le secteur public

## **6.1. Les groupes d'influences**

Les organismes exerçant une influence sur les bibliothèques peuvent être rassemblés en différentes catégories. En fonction de la catégorie à laquelle ils appartiennent, ces organismes agissent uniquement sur certaines facettes particulières de la bibliothèque avec plus ou moins de force de persuasion. Bien entendu, la liste de ces organismes n'est pas exhaustive mais regroupe essentiellement les acteurs les plus influents dans le domaine des bibliothèques et du web de données.

### **Les financeurs**

Leur appartenance au secteur public contraint les bibliothèques à se plier un minimum aux exigences et à la politique budgétaire publique. Elles sont également souvent soumises à la loi du marché en ce qui concerne les logiciels et les applications qu'elles utilisent.

En effet, bien qu'il existe quelques SIGB libres, très peu de bibliothèques possèdent du personnel compétent pour installer et développer ce type de logiciel. Ainsi, les bibliothèques suivent les évolutions des prestataires qui développent ces logiciels, libres ou propriétaires. Face à un besoin nouveau ou spécifique, les bibliothèques sont donc souvent obligées de payer un surcoût ou bien d'attendre que le développeur intègre les modifications à son logiciel. Dans le deuxième cas, la bibliothèque devra souvent patienter plusieurs années avant que la fonctionnalité désirée soit intégrée dans le logiciel.

Ainsi, il existe encore des logiciels, pour la plupart propriétaires, qui ne peuvent pas être interopérables avec les diverses applications du web de données. Quand aux SIGB libres, ils tentent de respecter les standards du web. Il n'est donc pas rare que les bibliothèques évoluent en fonction du SIGB qu'elles ont choisi<sup>71</sup>.

Ensuite, l'appartenance au secteur public pousse les bibliothèques à se plier aux exigences de l'État en matière de service public et de dépense de l'argent public. Ainsi, la part du budget allouée à l'accessibilité des connaissances par le public est généralement importante. En effet, en tant qu'organismes publics, les bibliothèques doivent faire l'effort de toucher un public le plus large possible, ce qui passe donc par le développement des outils électroniques, l'amélioration des performances, etc. Comme toute structure publique proposant des services, les bibliothèques

---

<sup>71</sup> Cette affirmation est basée sur un précédent mémoire réalisé en 2009 et traitant de l'importance des SIGB dans les structures documentaires.

doivent justifier leurs dépenses et attirer le public grâce à leur réactivité face aux nouvelles technologies et aux nouveaux besoins du public<sup>72</sup>.

Cette appartenance au secteur public pousse également les bibliothèques à respecter les normes et standards décrétés par les différents organismes reconnus.

### **Les standards et normalisations**

Le W3C est un acteur imminent du web de données. En effet, nous avons pu constater que, en partie grâce à lui, le concept de web de données a pris de l'ampleur et qu'un certain nombre de standardisations ont vu le jour.

Cependant, on peut se demander si la présence de Tim Berners-Lee à la direction du W3C ne serait pas également une des raisons principales de l'influence du W3C. En effet, Tim Berners-Lee est considéré comme étant le « créateur du web » puisqu'il est à l'origine de certains langages de base du web. Il peut donc être vu comme le leader charismatique du web. De plus, outre les efforts du W3C pour créer des standards interopérables et les plus pérennes possibles, celui-ci développe aussi des formats permettant la meilleure accessibilité possible pour les personnes ayant un handicap moteur ou mental. Ainsi, le W3C permet de respecter le désir de la diffusion des connaissances à la population la plus large possible recherchée par les bibliothèques.

Ensuite, il existe des organismes de normalisation exerçant une très forte influence sur les bibliothèques : l'Afnor (Association française de normalisation) et l'Iso (International Organization for Standardization) qui tentent de développer des normes communes au niveau international. Ces normes peuvent concerner des points très spécifiques de la documentation. On peut notamment citer la norme sur les thésaurus ISO 25964, prévue pour 2012 et censée permettre de retrouver tout type de ressource grâce à la gestion électronique de thésaurus<sup>73</sup>.

Enfin, l'influence d'autres professionnels est aussi très présente. En effet, la Bibliothèque Nationale de France et la Library of Congress, que j'ai déjà mentionnées dans ce mémoire, sont souvent vues par les autres bibliothèques comme des modèles à suivre. Elles possèdent dans leurs équipes de nombreux

---

72 Cette affirmation est basée sur un précédent mémoire réalisé en 2010. Il concerne les CMS (Content Management Système) et a pour cadre d'étude les collectivités territoriales.

73 RABAULT H., ZYSMAN H. « Les schémas de concepts et le Web sémantique : la norme sur les thésaurus ISO 25964 et le Web sémantique » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_norme\\_thesaurus.pdf](http://www.bnf.fr/documents/afnor2011_norme_thesaurus.pdf) (Consulté le 3/08/2011)

professionnels à la renommée internationale. On observe donc que les groupements de professionnels ont leur importance dans l'orientation des choix dans les bibliothèques. Ces groupements ne sont pas forcément des bibliothèques mais peuvent aussi être des associations de professionnels.

### **Les associations de professionnels**

Il existe en France de nombreuses associations de professionnels : l'ADBS (Association des professionnels de l'information et de la documentation), ADBU (Association des Directeurs des Bibliothèques Universitaires), ABF (Association des Bibliothécaires de France), l'IABD (Interassociation Archives Bibliothèques Documentation, etc. Au niveau international, on retrouve l'Ifla (International Federation of Library Associations and Institutions), l'OCLC (Online Computer Library Center), etc.

« Les associations professionnelles ont été porteuses de l'expression de culture de l'information et continuent de participer à son expansion en tant qu'acteurs de son émergence »<sup>74</sup>. Par culture de l'information, on entend : « la culture de l'information au sens de l'ADBS, [...] l'info-business, la culture des médias, la culture « info-com » et la culture informatique »<sup>75</sup>.

La culture de l'information au sens de l'ADBS est donc celle qui nous intéresse le plus. Proche de l'intelligence économique ou de la gestion des connaissances, la culture de l'information est apparue, en premier lieu, dans le monde de la documentation. En effet, elle consistait tout d'abord à former le public des bibliothèques à la recherche et à la maîtrise de l'information. Ces formations ont évolué au gré de l'apparition des technologies. Cela explique la grande réactivité des bibliothèques, en particulier des bibliothèques universitaires, face aux évolutions.

De plus, pour beaucoup de professionnels, il faudrait également « bâtir une culture informationnelle globale, [...] portant d'une part sur les médias, la documentation et les bibliothèques, d'autre part sur l'informatique et les outils.[...] Il resterait à y ajouter [...] la dimension communicationnelle »<sup>76</sup>.

Cette définition, souvent mise en avant par les associations de professionnels, est donc naturellement vue par les bibliothèques comme un objectif important.

---

74 LE DEUFF O. « La culture de l'information et l'héritage documentaire ». *Documentaliste-Sciences de l'Information*, mars 2010, vol. 47, n°3, pp. 4-11.

75 LE DEUFF, 2010, op. cit.

76 LE DEUFF, 2010, op. cit.

Certaines structures ont d'ailleurs développé de véritables services allant dans ce sens, c'est le cas de l'Enssib avec le service Formist (FORmation à l'Information Scientifique et Technique).

### **La Recherche en France**

Les universités et grandes écoles sont des lieux où on peut trouver beaucoup de chercheurs. Ces chercheurs et, par extension, les établissements et les groupes de travail dont ils font partie sont très influents dans le monde de la documentation.

En effet, les sujets de recherche suscitent souvent l'intérêt des professionnels. La publication de ces recherches permet de faire connaître les nouvelles avancées, de les expliquer et de les analyser. Le statut de chercheur des auteurs de ces publications est considéré comme une preuve de la fiabilité des informations mentionnées. Plus grande sera la notoriété du chercheur, plus ses recherches seront jugées fiables et importantes.

Le monde de la recherche est donc un moyen pour les professionnels de se tenir au courant des technologies les plus récentes. En outre, plus un sujet est traité plus il sera lu et plus il sera source d'intérêt par le public.

Parmi les structures de recherche en information-documentation de l'enseignement supérieur possédant le plus grand rayonnement national, voire international, on trouve l'INTD (Institut National des Techniques de la Documentation), l'Enssib (École Nationale Supérieure des Sciences de l'Information et des Bibliothèques), l'Abes (Agence Bibliographique de l'Enseignement Supérieur) et diverses universités.

Toutefois, prétendre que les bibliothèques sont toujours influencées dans leurs prises de décision serait exagéré. A contrario, il existe aussi un lobbying des bibliothèques ayant lui aussi de l'influence<sup>77</sup>. Aujourd'hui, les bibliothèques créent des réseaux entre-elles afin de faire pression sur les organismes extérieurs. Par exemple, le Couperin (COnsortium Universitaire de PERIodiques Numériques) aide les bibliothèques à faire face aux prix élevés des éditeurs. De plus, en tentant de répondre aux besoins du plus grand nombre, elles disposent souvent du soutien de leurs usagers.

---

77 BATTISTI M. « IFLA 2003 : Congrès mondial des bibliothèques et de l'information » La bibliothèque point d'accès : médias - information – culture ». *Documentaliste-Sciences de l'Information*, avril 2003, vol. 40, n°4-5, pp. 308-315.

## 6.2. La satisfaction de l'utilisateur

Avec la démocratisation d'Internet, le public est passé du silence informationnel, c'est à dire d'une difficulté à trouver les informations, au bruit informationnel qui correspond à une multitude d'informations qui sont rarement très pertinentes. Le public aime ce bruit car il lui permet de trouver beaucoup d'informations et d'affiner au fur et à mesure ses recherches<sup>78</sup>.

Aujourd'hui, on ne cherche plus en sachant clairement ce que l'on veut trouver mais on cherche à l'aveuglette avec des mots-clés généraux pour aller ensuite au particulier. La réflexion se passe après la recherche. La conséquence de cet état d'esprit a été le développement de plus en plus obligatoire de la recherche simple sur tous les champs et de la recherche par facette (les résultats de la recherche sont triés par sujets, auteurs, etc.) dans les SIGB.

De plus, cette facilité d'accès à Internet et à l'information rend l'utilisateur plus distant. Il souhaite se déplacer le moins possible. Il apprécie alors d'avoir les mêmes informations et services que s'il se rendait physiquement dans son centre de documentation. Cela explique le développement des OPAC qui permettent notamment d'avoir accès aux notices des documents et de pouvoir les réserver sur Internet sans avoir de notions en bibliothéconomie.

Aujourd'hui, le public souhaite encore plus de performance de la part des systèmes d'organisation des connaissances. La demande ne concerne pas seulement l'interopérabilité mais également la portabilité et l'exhaustivité des ressources.

Pour ne pas se laisser dépasser par les nouvelles technologies émergentes et perdre leur public, les bibliothèques tentent de développer des outils plus performants que ceux du web en général. Ainsi, on pourrait penser que l'effervescence des bibliothèques autour du web de données serait une réponse à l'émergence récente de projets ayant séduits le grand public tels que le cloud computing. Le cloud computing consiste à héberger sur un serveur distant, privé ou non, des applications habituellement installées sur un ordinateur ou sur des serveurs locaux<sup>79</sup>. Toutefois, cette technologie, appliquée principalement par des

---

78 MAISONNEUVE, Marc. « Dix bonnes raisons de choisir un opac de nouvelle génération ». *Documentaliste - Sciences de l'information*, 2008, vol. 45, n° 3, pp. 16-17

79 BARGE P. « Cloud Computing : l'informatique de demain ». *Le Journal du Net : e-Business, Informatique, Economie et Management*. 21 février 2011. Disponible sur : <http://www.journaldunet.com/developpeur/expert/49652/cloud-computing---l-informatique-de-demain.shtml> (Consulté le 10/08/2011)

groupes tels que Amazon, Google, Apple ou Microsoft, ne correspond pas aux exigences de l'information-documentation. Le système du cloud computing entre dans une pratique purement commerciale qui n'est pas applicable en bibliothèque. Ainsi, le cloud computing s'apparente plutôt à une « autorisation d'accès à distance [au matériel informatique]. On ne vend plus un bien, on alloue temporairement un accès tarifé. On autorise une souscription qui, *in fine*, devient le seul modèle de transaction »<sup>80</sup>.

Il est donc difficile pour une bibliothèque de cerner les besoins de toute la population en matière d'information. En effet, il est difficilement possible de mener des enquêtes sur les besoins de toute la population mais seulement des enquêtes à un niveau plus local. On ne peut donc que supposer les raisons de l'absence d'un public et seulement déduire par l'observation de la société ce qui pourrait l'attirer. Les bibliothécaires prennent alors leurs décisions en se basant sur les expériences passées et des objectifs intemporels.

### **6.3. L'évolution : une nécessité pour les bibliothèques**

Les bibliothèques répondent à des objectifs qui sont, pour certains, les mêmes depuis une trentaine d'années. Outre leur rôle de médiateur de l'information, les bibliothèques ont vu leurs tâches et leurs devoirs s'élargir avec Internet. Aujourd'hui, elles tentent le plus possible de suivre les évolutions d'Internet.

Toutefois, l'apparition des problématiques autour du web de données offre à l'occasion un domaine de la documentation d'être plus performant que le reste du web en matière de traitement et de gestion de l'information.

En effet, le web de données pourrait être une des réponses au désir des professionnels de l'information-documentation de simplifier l'utilisation de leurs outils pour le public. De plus, il permettrait une plus grande exhaustivité et fiabilité des informations ainsi qu'une recherche plus rapide pour l'utilisateur.

Du point de vue de l'organisation des bibliothèques, le web de données pourrait également offrir certains avantages. En effet, face à la multiplication des supports documentaires et des sources d'informations, les bibliothèques se doivent de fournir un ensemble de ressources le plus large possible avec des accès divers. Toutefois, le web se retrouve aujourd'hui avec une masse d'information

---

80 ERTZSCHEID O. « La dialectique de la cage et du nuage ». *affordance.info*. 24 mars 2010. Disponible sur : [http://affordance.typepad.com/mon\\_weblog/2010/03/dialectique-de-la-cage-et-du-nuage.html](http://affordance.typepad.com/mon_weblog/2010/03/dialectique-de-la-cage-et-du-nuage.html) (Consulté le 10/08/2011)



désorganisée qui nuit autant aux usagers des bibliothèques dans leurs recherches qu'aux bibliothécaires eux-mêmes.

En effet, les professionnels de l'information-documentation ont pour rôle de conseiller les usagers, ce qui induit de savoir où trouver l'information. Ils s'efforcent donc aujourd'hui d'« organiser » le web. Ils ont tout d'abord commencé par créer et mettre à disposition de leurs publics des portails, des signets ou encore des systèmes de veille. Les bibliothèques perçoivent donc le web de données comme une suite logique à ces applications. L'organisation des contenus électroniques des bibliothèques et des autres médiateurs de l'information serait alors nettement améliorée et plus compréhensible aussi bien par le public que par les professionnels. De plus, la réactivité des bibliothécaires face aux interrogations des usagers serait considérablement augmentée. Toutefois, l'organisation du web dans sa totalité, bien que souhaitée par les professionnels, reste encore un projet utopique étant donné l'ampleur du travail que cela impliquerait.

Enfin, le web de données comporte également un grand avantage dans le traitement de l'information et notamment le catalogage. La présence de standards, l'interopérabilité et l'uniformisation des notices permettent aux bibliothèques d'échanger encore plus aisément les notices entre elles. Le web de données contribue à accélérer ce travail de catalogage.

Cependant, ces choix ne seront pas sans conséquences pour le monde de l'information-documentation. Les mutations et les remises en question seront aussi nombreuses que lors du passage des bibliothèques au web social.

## **Partie 3 : La bibliothèque hybride, une amorce à la bibliothèque du futur**

## 7. Impacts socio-organisationnels

Avec l'apparition des bibliothèques hybride et l'application du web de données, on se questionne aujourd'hui sur le futur du bibliothécaire : ce métier arrivera-t-il à survivre face au web ainsi qu'à l'accès à distance aux documents, y compris aux livres ? Malgré ces craintes, on peut constater que ce n'est pas la première fois que le métier de bibliothécaire est remis en question. En effet, les bibliothèques ont déjà « survécu », entre autres, à l'informatisation, au web des documents et au web social, le web de données n'est donc qu'une nouvelle évolution.

### 7.1. Les nouvelles compétences des professionnels

Déjà en 2002, d'après une enquête réalisée « auprès de 7 000 enseignants-chercheurs américains en sciences humaines et sociales, la bibliothèque physique occupe la dernière place pour la recherche d'information, loin derrière les moteurs de recherche, portails ou bases de données »<sup>81</sup>. Cette situation s'est par la suite étendue à la France. Les chercheurs tendent aujourd'hui à imaginer un monde sans bibliothèques avec des ressources électroniques accessibles de n'importe quel lieu et pérennes. Cependant, l'ère de l'électronique n'est pas synonyme de la fin des bibliothécaires.

En effet, on le constate en observant les formations continues et initiales proposées aux bibliothécaires et documentalistes. Ces formations proposent de plus en plus de cours sur la documentation numérique, les modèles conceptuels de données et les formats de métadonnées utilisés dans le web de données tels que le Dublin Core.

De plus, les documentalistes sont de plus en plus initiés à la communication essentielle au fonctionnement d'une structure. On y retrouve deux types de communication : interne, qui permet aux membres du personnel de communiquer de manière efficace entre eux (messagerie, intranet, etc.), et externe qui permet de communiquer avec l'extérieur. Cette communication permet aux professionnels de l'information-documentation de mieux comprendre leur public, de le cibler, de promouvoir et/ou de leur proposer des services et des ressources adaptés à leurs besoins. À l'aide de cette communication, les bibliothèques, entre autres, construisent leur image auprès du public mais aussi auprès des autres structures

---

81 SCHOPFEL J. « Bibliothèque sans bibliothécaires ? » [En ligne]. *HAL – INRIA*. 19 juillet 2005. Disponible sur : [http://hal.inria.fr/docs/00/06/26/48/PDF/sic\\_00001524.pdf](http://hal.inria.fr/docs/00/06/26/48/PDF/sic_00001524.pdf) (Consulté le 10/08/2011)

documentaires. Elles utilisent désormais des stratégies de communication en plus des politiques documentaires pour séduire de nouveaux usagers mais aussi pour valoriser la bibliothèque comme lieu privilégié de la transmission des connaissances.

Les professionnels sont également très demandeurs en ce qui concerne l'anglais. En effet, il s'agit de la langue dominante sur le web et donc de la langue la plus utilisée en documentation numérique, que ce soit au niveau de la prise de contact avec les éditeurs ou du traitement du document. Au niveau de la documentation numérique, le droit change et devient de plus en plus important puisque de nouveaux aspects doivent obligatoirement être pris en compte au niveau des bibliothèques : licences d'utilisation de ces documents, respect des accès monopostes aux ressources, droits d'auteur et de copies, etc. Ces formations sont la preuve que les métiers de l'information-documentation évoluent sans cesse et que leurs champs d'action s'élargissent.

On perçoit aussi une mutation des professions grâce aux référentiels métiers proposés par certaines structures telles que le CNRS ou l'ADBS. Ainsi, depuis son recensement en 2002, le CNRS inclut, par exemple, des compétences en numérisation pour les magasiniers ou encore des compétences en conception de systèmes de gestion électronique pour les ingénieurs. Ainsi, on remarque que, pour tous les métiers de l'information-documentation, les compétences exigées en informatique sont de plus en plus importantes. Ces nouvelles compétences requises dans les bibliothèques rendent encore plus mince la distinction entre le bibliothécaire et l'informaticien. Le métier de documentaliste se rapproche des métiers de l'informatique avec la conception et la gestion de base de données, l'administration de sites internet, etc. Ces évolutions expliquent le besoin, quasi-omniprésent, de s'informer et de se former sans cesse aux technologies qui apparaissent en information-documentation. La cartographie des métiers proposée sur le site de l'ADBS schématise très bien cette situation.

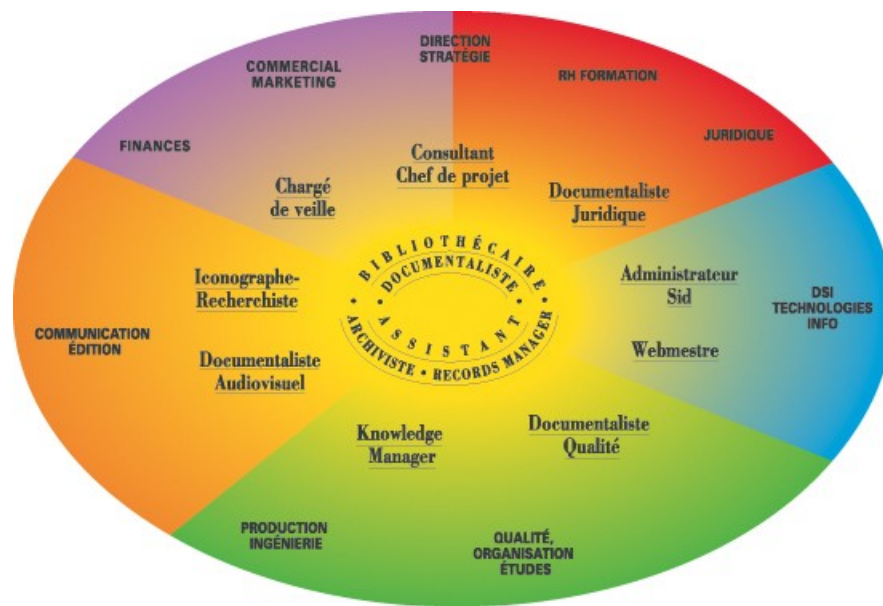


Fig. 3 : Cartographie des métiers de l'information-documentation<sup>82</sup>

Ainsi, en regroupant au centre de son schéma les professions de bibliothécaire, documentaliste, assistant, archiviste et record manager, l'ADBS montre que celles-ci sont complexes et demandent une certaine polyvalence. Pour les bibliothèques, on peut regrouper ces nouvelles compétences exigées en trois types :

- « celles relatives au caractère numérique des objets manipulés [...] il s'agit surtout de problématiques d'intégration des contenus dans les collections » ;
- « celles en « amont » du processus : l'acquisition et la numérisation [...] Il s'agit de compétences nouvelles car elles portent sur les relations commerciales avec de nouveaux prestataires » ;
- « celles portant sur la mise en visibilité des collections dématérialisées, cette communication vise des publics distincts, le public de la bibliothèque »<sup>83</sup>.

<sup>82</sup> ADBS. « Le référentiel des métiers et fonctions » [En ligne]. *Accueil - L'association des professionnels de l'information et de la documentation*. 29 juillet 2010. Disponible sur : [http://www.adbs.fr/le-referentiel-des-metiers-et-fonctions-61928.htm?RH=MET\\_REFMETIER](http://www.adbs.fr/le-referentiel-des-metiers-et-fonctions-61928.htm?RH=MET_REFMETIER) (Consulté le 10/08/2011)

Cependant, la reconnaissance des bibliothécaires et des bibliothèques en tant que source d'information fiable et pertinente est faible. La représentation que le public a de ces professionnels est effectivement très différente de celle que les bibliothécaires ont d'eux-même et qu'ils souhaitent transmettre. Internet rassemble une masse imposante d'informations, ainsi le travail des bibliothécaires et leurs compétences sont souvent sous-évalués malgré leurs efforts pour s'adapter.

## 7.2. Une réorganisation interne

Pour que les bibliothèques puissent s'adapter aux évolutions technologiques, Oliver Ertzscheid souligne quelques mutations nécessaires pour les catalogues<sup>84</sup>. En effet, il parle d'un catalogue hybride qui serait un lien entre les données d'Internet et les compétences en traitement de l'information des bibliothécaires. De plus, face à l'accès aisé aux ressources Internet, il conseille aux bibliothèques de suivre l'exemple du Sudoc et des bibliothèques universitaires, c'est-à-dire de mettre leurs catalogues en libre accès. Ainsi, pour attirer les usagers, les bibliothèques tendent peu à peu vers la transparence.

Les bibliothécaires avaient également pris pour habitude de se focaliser sur la technique et les technologies laissant ainsi de côté les usagers. Ces dernières années, nous avons alors vu apparaître les premiers services de bibliothécaires en ligne qui peuvent être interprétés comme une première démarche vers l'utilisateur. L'utilisateur peut alors trouver sa place dans la bibliothèque mais aussi disposer d'une aide humaine, ce dernier élément étant impossible lors d'une recherche sur un moteur tel que Google. L'humanisme serait alors un deuxième élément essentiel de cette nouvelle bibliothèque.

Aujourd'hui, les bibliothèques, et en particulier les bibliothèques ayant un rayonnement important telles que la BNF, les bibliothèques universitaires ou encore les bibliothèques de grandes municipalités, s'ouvrent aux nouvelles technologies et à l'utilisateur. Les outils sont donc sans cesse en train de s'améliorer en même temps que les compétences et les tâches du personnel des bibliothèques.

---

83 BERTRAND Anne-Marie, EPRON Benoit. « Les ressources numériques : un nouvel enjeu pour les compétences des professionnels des bibliothèques ». *Documentaliste-Sciences de l'information*, 2010, vol. 47, n°2, pp. 65-66.

84 ERTZSCHEID O. « Les « nouveaux » catalogues ou le catalogue en (r)évolutions » [En ligne]. @rchiveSIC :: Accueil. Août 2007. Disponible sur : <http://archivesic.ccsd.cnrs.fr/docs/00/17/80/06/PDF/ArticleABES.pdf> (Consulté le 10/08/2011)

Enfin, le troisième changement significatif nécessitant une réorganisation importante du travail des bibliothécaires est la numérisation des ressources. Bien que la fréquentation des bibliothèques universitaires a plutôt tendance à diminuer, les bibliothèques constatent une augmentation des activités des utilisateurs sur leurs espaces numériques et leurs ressources électroniques<sup>85</sup>.

Toutefois, cette réorganisation est souvent difficile à mettre en place étant donné l'écart existant entre la lenteur administrative due à l'appartenance au secteur public et la rapidité des évolutions du numérique. De plus, ces changements perdent les professionnels eux-mêmes, certains n'arrivent plus à distinguer leurs rôles et places au sein de l'organisme<sup>86</sup>. Cet égarement vient principalement du fait que les bibliothécaires ne s'occupent plus seulement de bibliothéconomie mais également de nouveaux aspects du métier comme l'informatique. Lors de mes passages en bibliothèque, il m'est arrivé de demander au personnel en banque de prêt s'ils éprouvaient des difficultés avec cette informatisation. Pour beaucoup, la réponse a été positive. Les évolutions sont aujourd'hui si rapides que les bibliothécaires ayant une certaine ancienneté éprouvent des difficultés à s'adapter, ils ne sont pas réellement préparés à ces changements.

L'accompagnement au changement et, par extension, la communication interne sont primordiaux pour pouvoir continuer à évoluer face aux réticences et habitudes des professionnels. Ce changement passe donc par une communication réelle auprès du personnel de la bibliothèque qui peut se traduire sous la forme de réunions régulières sur l'avancée du projet, de formations sur l'utilisation d'un nouvel outil, etc. Il est important que chaque membre de la bibliothèque se sente concerné afin de mieux s'appropriier et maîtriser le nouvel outil par la suite. L'apparition de groupes de travail au sein des bibliothèques est l'un des éléments mis en place pour pousser les employés à participer aux changements et ainsi-les motiver davantage. Ces groupes de travail ne sont pas toujours internes aux bibliothèques puisque certaines s'ouvrent vers l'extérieur et créent des groupes où plusieurs bibliothèques, motivées par les mêmes objectifs, interviennent. C'est le

---

85 MERRIEN D. *Colloque international sur les "ressources électroniques académiques : mesures et usages"* [En ligne]. Lille : Couperin, 30 novembre 2009. Disponible sur : [http://www.couperin.org/IMG/pdf/Colloque\\_Mesures\\_usages\\_RE\\_Lille\\_2009.pdf](http://www.couperin.org/IMG/pdf/Colloque_Mesures_usages_RE_Lille_2009.pdf) (Consulté le 10/08/2010)

86 GLEYZE A. « Quelles compétences pour les bibliothécaires ? : Diversité des fonctions et référentiels ». *Documentaliste-Sciences de l'Information*, janvier 2005, vol. 42, n°1, pp. 22-25.

cas du groupe de travail sur les SIGB libres regroupant les bibliothèques universitaires de Lyon 2, Lyon 3 et Saint-Étienne.

### 7.3. Une communication avec l'extérieur

Les accès aux ressources numériques sont essentiellement détenues par quelques grands groupes comme Elsevier, Gallimard, Eyrolles, Google, etc. Les bibliothèques doivent faire face à ces groupes pour tout ce qui concerne l'accès aux ressources numériques, ce qui constitue un nouveau défi pour elles. Les bibliothèques ont alors créé divers regroupements, ou groupes de travail, tels que le Groupe de travail sur les SIGB libres, Couperin, le Sudoc, etc.

Les liens entre les bibliothèques dépendent tout d'abord de la structure dont elles dépendent. Ainsi, les bibliothèques universitaires, qui dépendent d'une université, ont souvent renforcé leurs liens grâce aux différents Pres (Pôle de recherche et d'enseignement supérieur) qui regroupent plusieurs universités et grandes écoles. À Lyon, le Pres a été créé en 2007 sous le nom d'Université de Lyon, il regroupe sous la forme juridique d'un établissement public de coopération scientifique 18 universités, grandes écoles et instituts de Lyon et Saint-Étienne<sup>87</sup>. Ce pôle Pres est l'initiateur de divers projets parmi lesquels on retrouve le projet Brain proposant un catalogue collectif des bibliothèques de l'enseignement supérieur de Lyon et Rhône-Alpes.

Il existe également d'autres groupements universitaires plus axés sur le numérique : les UNT (Universités Numériques Thématiques nationales). Elles « ont pour mission, dans le cadre d'une mutualisation à une échelle nationale, de favoriser la valorisation, la production et la diffusion de ressources pédagogiques numériques validées produites par les établissements d'enseignement supérieur »<sup>88</sup>. Parmi les sept UNT existantes en France, on retrouve l'Unit (Université Numérique Ingénierie Technologie) qui est à l'origine du projet Ori-OAI vu précédemment et qui participe à d'autres projets sur le web de données comme le SupLOMFR<sup>89</sup>.

---

87 UNIVERSITÉ DE LYON. « Présentation de l'Université de Lyon » [En ligne]. *L'Université de Lyon*. 10 août 2011. Disponible sur : <http://www.universite-lyon.fr/1-universite-de-lyon/> (Consulté le 10/08/2011)

88 ÉDUCNET. « Les Universités numériques thématiques » [En ligne]. *Éducnet : Enseignement, TICE, Nouvelles technologies, Ressources pédagogiques numériques*. 3 juillet 2009. Disponible sur : <http://www.educnet.education.fr/superieur/unt> (Consulté le 10/08/2011)

89 « Le SupLOMFR est un format unique de description de métadonnées pour indexer les ressources pédagogiques de l'enseignement supérieur » source : UNT. « Le groupe « Indexation



Enfin, un des derniers projets importants au niveau des catalogues des bibliothèques de l'enseignement supérieur reste le Sudoc déjà mentionné dans ce mémoire.

Le but de tous ces projets va au-delà du simple objectif de fournir un catalogue collectif accessible à tous. En effet, l'utilisation d'outils tels que l'OAI-PMH et par conséquent l'intégration de ces projets dans le web de données ont pour but de faciliter le travail des bibliothécaires en réduisant considérablement les tâches de catalogage grâce à l'échange plus aisé de notices. Par la suite, ces groupes de travail ont également pour souhait de rendre directement interrogeables les métadonnées des catalogues sur des moteurs de recherche comme Google. Ainsi, les bibliothèques pourraient s'adapter aux pratiques du public qui a plutôt tendance à mener ses recherches sur Google.

Le web de données permet alors aux bibliothèques d'améliorer leur façon de communiquer avec le public. Elles tentent de placer le public au cœur de leurs préoccupations avec de nouveaux services destinés à permettre aux usagers d'accéder à toutes les dimensions de l'information. La plupart des bibliothèques proposent à leurs usagers des services leur permettant de travailler soit directement sur place à la bibliothèque, soit à distance de n'importe quel lieu : salles de travail dans la bibliothèque, accès sur place ou à distance aux ressources numériques, etc.

Depuis l'apparition du web social, les bibliothèques ont pris conscience de l'importance de la relation avec l'utilisateur : les moyens d'accéder à l'information évoluent mais les bibliothèques doivent continuer à aider l'utilisateur à trouver l'information désirée. On retrouve ici l'idée du bibliothécaire comme le médiateur entre l'information et le public. Il est important que le métier de bibliothécaire s'adapte sans cesse aux évolutions technologiques afin d'être à même de répondre aux besoins du public et de proposer une information plus pertinente et plus fiable.

Enfin, en ce qui concerne les périodiques, les bibliothèques universitaires se sont regroupées sont en un Consortium Universitaire de Publications Numériques appelé Couperin. Ce consortium a pour objectif de régler les problèmes liés aux prix parfois abusifs de certains éditeurs (EBSCO, Elsevier, Armand Colin, etc.) sur leurs abonnements. En effet, les périodiques numériques sont de plus en plus détenus par de grands groupes d'éditeurs au détriment des petits éditeurs qui ont tendance à disparaître. De plus, cette multiplication des ressources électroniques

---

– SupLOMFR » [En ligne]. *Universités numériques*. 10 août 2011. Disponible sur : <http://www.universites-numeriques.fr/fr/content/le-groupe-%C2%AB-indexation-%E2%80%93-suplomfr-%C2%BB> (Consulté le 10/08/2011)

couplée aux nombreux échanges entre structures et aux mutations professionnelles poussent alors les bibliothèques à revoir leur politique documentaire.

## **8. La politique documentaire**

Aujourd'hui, la politique documentaire en bibliothèque s'apparente de plus en plus à une politique de communication. Le rôle principal des bibliothécaires reste le même : mettre à la disposition et diffuser aux usagers les informations dont ils ont besoin. Toutefois, cette diffusion ne se fait plus seulement grâce aux tâches traditionnelles de bibliothéconomie mais elle passe également par la valorisation, l'organisation des connaissances, le marketing documentaire, etc.

### **8.1. Dématérialisation de la collection**

Les bibliothèques sont de plus en plus des bibliothèques dites hybrides. Ce sont des bibliothèques qui continuent d'assurer les fonctions traditionnelles : prêt, mise à disposition du catalogue et de ressources papiers, etc. Mais qui sont, en plus de ces tâches traditionnelles, des bibliothèques numériques, c'est-à-dire « des organismes qui fournissent les ressources nécessaires, y compris le personnel qualifié, pour sélectionner, permettre l'accès intellectuel, interpréter, distribuer, conserver l'intégrité et assurer la préservation dans le temps de collections de documents numériques afin qu'ils soient disponibles à l'usage d'une ou plusieurs communautés »<sup>90</sup>. Par document numérique, on entend « les documents digitalisés sur lesquels la bibliothèque a la main, qu'elle les ait numérisés ou qu'elle en ait acquis la libre disposition (à l'exception donc des flux électroniques à l'accès distant) »<sup>91</sup>.

On distingue alors trois types de ressources numériques dans une bibliothèque :

- celles produites et mises en ligne par la bibliothèque elle-même incluant sa collection numérisée, les documents électroniques créés par la bibliothèque, etc.
- celles achetées par la bibliothèque : périodiques électroniques, livres numériques payants, bases de données, etc.
- celles extérieures à la bibliothèque, gratuites et mises à la disposition des usagers : archives ouvertes comme Hal (Hyper Articles en Ligne),

---

90 WATERS D. « What are Digital libraries? » [En ligne]. *CLIR Issue*, juillet/août 1998, n°4. Disponible sur : <http://www.clir.org/pubs/issues/issues04.html> (Consulté le 17/08/2011)

91 CALENGE, 2008, op. cit.

blogs, sites web, etc. Elles sont mises à disposition des usagers grâce à des portails, des signets, des agrégateurs de flux RSS comme Netvibes, etc.

Certaines bibliothèques ont donc décidé de rassembler et de créer un véritable fonds de contenus numérisés. En France, le projet le plus représentatif est bien sûr le projet Gallica de la BNF, créé en 1997, qui compte actuellement, en août 2011, plus de 1 500 000 ressources.

L'essor du numérique en bibliothèque permet d'expliquer l'intérêt des applications du web de données. En effet, il offre la possibilité d'organiser ces ressources et de les rendre interopérables. De plus, le respect des standards du web est également un gage de pérennité tout particulièrement pour tout ce qui concerne les formats de métadonnées. Enfin, le numérique reste aussi le moyen de diffuser toute une collection sur l'ensemble du web et de soutenir ainsi l'activité de la bibliothèque.

La numérisation peut être faite soit directement par la bibliothèque, soit par un prestataire externe. Cependant, la numérisation nécessite un budget, des compétences, du personnel et du temps de travail supplémentaires pour les bibliothèques. Les bibliothèques procédant à la numérisation de leur fonds restent donc encore rares bien qu'elles soient tout de même, pour la plupart, en possession d'abonnements électroniques. En effet, la place de la numérisation n'est pas encore évidente pour tous les bibliothécaires, beaucoup d'entre-eux la considère comme une tâche secondaire, et donc non obligatoire, de leur profession. Cette idée est renforcée par deux facteurs : d'une part, par le fait que de nombreux professionnels ne sont pas formés aux techniques et cadres juridiques propres à la numérisation, d'autre part, par le manque d'effectifs de certaines bibliothèques et donc par le surplus de travail exigé par la seule réalisation de tâches traditionnelles.

Concernant leur budget, les bibliothèques dépendent du gouvernement. Elles sont donc intégrées à la politique et doivent donc aussi faire face à la dette publique. Elles souffrent alors, pour la plupart, de la réduction des postes dans la fonction publique et de la réduction ou stagnation budgétaire<sup>92</sup>.

Ces lacunes ont pour conséquences de ralentir considérablement l'évolution des bibliothèques vers le web de données et les nouvelles technologies. Pour palier ces

---

92 SCHÖPFEL J. « L'avenir du métier de bibliothécaire » [En ligne]. *Archimag : Guide pratique 42 : Bibliothèques, les nouveaux usages*, 2011, pp. 62-66

manques, les bibliothèques sont parfois tributaires d'autres structures pour le traitement et la conservation de leurs ressources.

## **8.2. Conservation des documents numériques**

Outre la phase même de numérisation, la conservation de documents numériques sur des espaces de stockage adéquats représente aussi un coût assez élevé. En effet, la bibliothèque a besoin soit de professionnels capables de créer ces espaces, soit de louer ces espaces directement chez des hébergeurs. Par la suite, ce sont les professionnels de la bibliothèque qui devront gérer et utiliser ces espaces de stockage. Quelque soit l'option choisie, le coût et le travail fournis par les professionnels de la bibliothèque ne seront pas négligeables.

En ce qui concerne les périodiques électroniques, le choix du type de stockage n'est pas nécessaire puisqu'ils sont directement archivés sur les plateformes des éditeurs et intégrés dans le prix des abonnements. On retrouve aussi cette situation pour certains livres électroniques payants dont la bibliothèque a acheté les titres. De plus, très peu de bibliothèques disposent d'un espace nécessaire pour héberger leurs propres abonnements électroniques. Toutefois, cet hébergement externalisé chez l'éditeur peut provoquer une perte de données dans de nombreux cas : cessation d'activité, publication arrêtée, désabonnement de la bibliothèque, etc. Dans cette situation, les bibliothèques ne peuvent donc pas garantir la pérennité de ces ressources puisqu'elles dépendent fortement des éditeurs. Par exemple, le SCD Doc'Insa de Lyon a décidé de se désabonner de Numilog<sup>93</sup>, une plateforme de livres numériques, où il a acheté 147 titres de livres dont il reste propriétaire selon le contrat signé avec Numilog. La bibliothèque essaye actuellement de négocier avec Numilog pour récupérer ces 147 titres et ensuite les stocker sur ses propres espaces de stockage tout en respectant les licences d'utilisation de ces livres. Ainsi, nous pouvons voir que les bibliothèques sont de plus en plus poussées à développer leurs propres espaces de stockage pour garantir la pérennité de leurs ressources numériques.

En effet, la pérennité des documents électroniques est très importante. Les documents et le travail réalisé sur les métadonnées doivent être encore lisibles dans les prochaines années malgré l'évolution très rapide du matériel informatique.

---

93 GARY N. « Trop de contraintes : le Doc Insa de Lyon se désabonne de Numilog » [En ligne]. *ActuaLitté - Les univers du livre*. 22 décembre 2010. Disponible sur <http://www.actualitte.com/actualite/23388-lyon-numilog-ebooks-abonnement-contraintes.htm> (Consulté le 17/08/2011).

Les modèles, langages et vocabulaires du web de données constituent alors une grande aide pour assurer la pérennité de la conservation et la meilleure diffusion possible.

Toutefois, l'hébergement local des ressources numériques comporte également des difficultés. Tout d'abord, pour chaque ressource, il est nécessaire de vérifier les licences d'utilisation, les droits pour les stocker puis ensuite les diffuser auprès du public. Ensuite, les bibliothécaires chargés de l'indexation doivent pouvoir utiliser les différents espaces de stockage sans devoir solliciter l'aide d'informaticiens. Enfin, le web de données, surtout le protocole OAI et le modèle OAIS (Open Archival Information System), permettent aux bibliothèques de réduire ces problèmes liés à la conservation des documents. En effet, l'OAI-PMH permet de mutualiser les ressources et leurs données. Ainsi, les bibliothèques disposant d'un moissonneur OAI pourront collecter des informations via les entrepôts OAI d'autres bibliothèques, centres de documentation, etc.

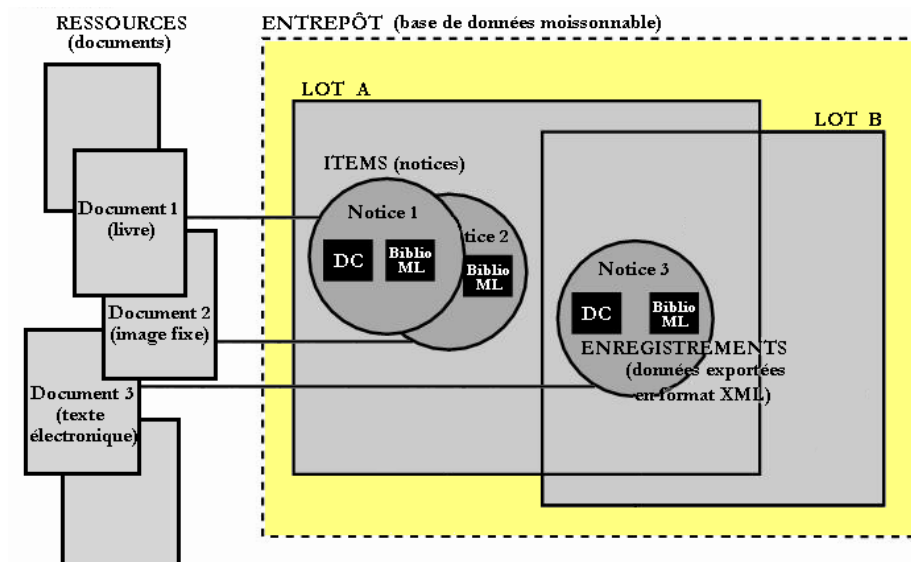


Fig. 4 : Fonctionnement de l'OAI<sup>94</sup>

94 NAWROCKI F. « Le protocole OAI et ses usages en bibliothèque » [En ligne]. *Ministère de la Culture et de la Communication*. 15 février 2005. Disponible sur : <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm> (Consulté le 17/08/2011)

Le travail demandé par l'indexation de ces ressources est donc moins volumineux que prévu pour les bibliothécaires puisqu'il est possible de rapatrier les ressources et leur indexation. De plus, l'hébergement OAI, bien qu'il soit externalisé, dépend de formats ouverts, il est donc pérenne et interopérable. Cette technologie contribue donc à réduire les coûts de la documentation électronique en bibliothèque.

Comme nous avons pu le constater, l'interopérabilité et la pérennité des données sont des problématiques récurrentes en ce qui concerne le web de données et, plus largement, la documentation électronique. Effectivement, en tant que médiateurs de l'information, les bibliothécaires placent au cœur de leur travail ces notions qui assurent la bonne diffusion des ressources.

### **8.3. Accès à l'information**

Le rôle des bibliothèques en tant que médiateur de l'information a évolué. Aujourd'hui, elles donnent accès aux documents eux-mêmes via Internet alors que, peu de temps auparavant, elles se contentaient de diffuser les références des documents sur leurs catalogues. Or, ces mêmes références sont de plus en plus accessibles directement sur le web.<sup>95</sup>

Les bibliothèques doivent donc aujourd'hui se distinguer du web et de l'accès massif à l'information en faisant vivre leurs collections, en anticipant les besoins de leurs publics, en organisant l'accès grâce à des ressources grâce à des portails, signets, etc., en offrant un espace convivial de travail et d'échanges, etc.

La bibliothèque hybride contribue à remplir ces différents objectifs. Elle permet :

- « l'amélioration de la qualité et de la facilité d'utilisation du contenu » ;
- « la multiplication des possibilités d'accès offertes à la population » ;
- « et le renforcement des actions de sensibilisation sur les problèmes de conservation à long terme des données »<sup>96</sup>.

Ces différents objectifs de la bibliothèque hybride sont donc réalisables grâce à l'interopérabilité et à la pérennité des données assurées par le suivi de normes, de modèles, d'éléments juridiques, etc.

---

95 EBSCO. « La Bibliothèque du Futur - Interview du Professeur Gilles Vassal » [En ligne]. EBSCO Information Services. 2011. Disponible sur : <http://www2.ebsco.com/fr-fr/NewsCenter/TC/Pages/Vassal.aspx> (Consulté le 17/08/2011)

96 PÉDAUQUE, 2007, op. cit.

De plus, un document numérique peut être consulté par plusieurs usagers en même temps et des milliers de fois, ce qui permet une très large diffusion. Le numérique résout donc les problèmes posés par le document papier qui ne peut être consulté que par une personne à la fois et qui est limité à un certain nombre d'exemplaires imprimés. Toutefois, dans la pratique, plusieurs problèmes juridiques se posent puisque tous les documents sont soumis à certains droits.

Désormais, les bibliothèques développent et mettent en avant des interfaces créées pour la consultation de ces documents : Opac (Online Public Access Catalog), portails, etc. Elles entrent donc dans une véritable logique de communication pour tenter de vendre leurs services faisant concurrence aux acteurs commerciaux tels que Google et Amazon.

En ce qui concerne les ouvrages, la bibliothèque est devenue aussi un véritable espace virtuel consultable sur place et à distance. Tout d'abord, avec l'apparition des étagères virtuelles permettant de localiser un document dans la bibliothèque ainsi que les documents qui sont classés avant et après lui. Ensuite, avec la bibliothèque en 3D du projet HyperVise qui donne l'illusion à l'utilisateur de se trouver au sein même de la bibliothèque.

Toutefois, la consultation de certains documents placés hors de leur contexte d'origine est parfois difficilement intelligible. Les bibliothécaires font alors souvent appel au web de données pour renseigner les métadonnées ou annoter les documents.

Le rassemblement de ces différentes ressources par les bibliothèques permet aux usagers d'accéder, via une seule plateforme, à des sources hétérogènes déjà triées et dont la fiabilité est certaine. Le but des bibliothèques étant de fournir à terme une interface capable de remplir les exigences des usagers en matière de simplicité et de diversité des documents. Cette interface unique, dépendante du travail des professionnels de l'information-documentation, permet à l'utilisateur d'avoir accès à des ressources souvent beaucoup plus fiables et approfondies que celles obtenues par le biais des moteurs généralistes scrutant uniquement les informations du web visible. Elle évite également aux usagers de devoir naviguer de bases de données en bases de données spécialisées afin de rechercher les informations du web invisible ou celles non indexées par les moteurs de recherche.

Le web de données est donc une des solutions choisies par les bibliothèques pour éviter à l'utilisateur de se perdre dans la masse d'information. Il est aussi très apprécié car il permet d'obtenir des résultats et des ressources plus rapidement

qu'avec les moteurs de recherche généralistes. En effet, ces moteurs nous obligent souvent à scruter les moindres recoins du web pour découvrir des informations pertinentes et originales.

Enfin, les ressources électroniques ont également l'avantage d'être facilement transportables en plus d'être accessibles et copiables aisément. Par exemple, un chercheur ne pouvant se rendre à la bibliothèque et sur Internet pourra enregistrer jusqu'à plusieurs centaines de documents sur un support numérique (CD-Rom, clé USB, etc.). Cette tendance s'est aujourd'hui étendue aux livres et la lecture électronique se banalise relançant une nouvelle fois le débat sur la place des bibliothèques et de la lecture papier.

## **9. La place du livre**

Selon l'enquête menée en 2008 sur les pratiques culturelles des français à l'ère numérique, les français lisent moins et préfèrent utiliser davantage le web et le multimédia<sup>97</sup>. Cette préférence s'explique par l'apparition d'une manipulation et d'une relation différentes aux écrits.

### **9.1. L'essor du livre électronique**

Le livre est certainement l'objet qui, dans l'imaginaire social, représente le mieux la bibliothèque et la diffusion du savoir. En effet, dès le plus jeune âge, les individus sont familiarisés aux livres et apprennent grâce à eux. L'apparition de nouveaux supports, tels que les e-books, change radicalement la vision que certains ont de la lecture.

Le succès récent du livre électronique s'explique par le fait qu'il correspond aux imaginaires, aux représentations que les internautes ont d'Internet : transparence, rejet des limites et des frontières, moins de contraintes physiques, etc. Ainsi, le livre électronique serait pour les usagers à la croisée des courants de pensées libertaires, visant un monde meilleur, et libérales, cherchant à se détacher de toute intervention ou régulation de l'État<sup>98</sup>.

De plus, bien que les premières tablettes de lecture datent du début des années 2000, elles n'ont que récemment suscité l'intérêt du public. Leur insuccès était principalement dû à l'imperfection de ces tablettes qui étaient lourdes, ne

97 DONNAT O. *Les pratiques culturelles des Français à l'ère numérique : enquête 2008*. Paris : La Découverte, 2009

98 REBILLARD F. *Le web 2.0 en perspective : Une analyse socio-économique de l'internet*. Paris : L'Harmattan, 2007



donnaient pas accès à un grand choix de ressources, etc. Ainsi, cette technologie « nouvelle » est encore pour beaucoup d'individus une source de fascination, de mystère et de découverte face à son fonctionnement et à ses caractéristiques. On ne peut pas encore avoir un très grand recul sur ce nouvel outil car la plupart des individus n'ont pas encore dépassé la phase d'appropriation de cette technologie. Le public est alors de plus en plus séduit par la notion d'interactivité, ou relation homme/machine, qui est souvent assimilée aux tablettes de lecture telles que le Reader, le Kindle ou l'Ipad. Aujourd'hui, ces tablettes se multiplient et permettent de transporter sans s'encombrer un, voire plusieurs, livres lisibles grâce à un seul et même support. La portabilité des données désirée par le public concerne donc également les livres.

De plus, les concepteurs de ces tablettes et les éditeurs ont trouvé un intérêt commun, ce qui n'était pas le cas il y a 10 ans, pour développer et rentabiliser ce produit. Désormais, certains éditeurs américains fournissent, dès la publication papier d'un ouvrage, sa publication électronique. Ces nouveaux contrats, signés avec des groupes influents de l'univers du numérique, permettent à ces éditeurs d'améliorer leur chiffre d'affaire en élargissant leur public et en vendant parfois deux exemplaires, un papier et un numérique, à un seul et même client. Les éditeurs fournissant les livres du Kindle touchent environ 70% des recettes.

Enfin, l'attrait du public pour les ressources électroniques vient également du fait que son rôle sur les documents est devenu plus important. En effet, le document peut être copié, des mots peuvent être recherchés dans le texte du document grâce à une recherche en texte intégral, des annotations peuvent être ajoutées facilement, etc.

Toutefois, les avis sur le livre numérique ne sont pas tous aussi positifs. Certains voient les livres numériques comme une menace pour le livre papier. Face à cette croyance, quelques usagers se positionnent tels des défenseurs du livre papier alors que d'autres considèrent la lecture sur écran comme un élément marquant indéniablement le déclin de la lecture papier. Or, ces deux points de vue sont rarement partagés par les bibliothécaires qui voient plutôt ces deux types de lectures comme étant complémentaires.

## **9.2. Une autre lecture : le cas des étudiants de Lyon 2**

Du point de vue des bibliothèques, il faut nuancer l'importance du livre numérique dans la lecture et la diffusion des connaissances. Bien que cette

technologie apporte de nombreux avantages tels que la recherche de termes dans le texte, la portabilité, etc., elle ne remplace pas le livre papier. En effet, elle propose surtout une autre façon de lire.

En 2006, l'Université Lyon 2 a chargé Claire Belisle, psychosociologue et ingénieur au CNRS, de diriger une enquête auprès d'étudiants de première et deuxième année de Licence sur le thème de la lecture numérique. Cette étude a permis de distinguer les principales différences existantes entre la lecture papier et la lecture électronique.

Tout d'abord, la principale divergence se trouve dans le mode de lecture. Pour la lecture papier, il est plutôt linéaire tandis que les documents numériques favorisent une lecture de survolage, à rebond, souvent gênée par des hyperliens, des liens menant à une autre partie du document ou à un autre document. Malgré ce survolage, Dario Salvucci, professeur à la Drexel University, démontre que les mouvements de l'œil lors de la lecture électronique ne sont pas linéaires et sont saccadés, ce qui rend cette lecture plus lente que la lecture papier. En effet, la lecture sur papier est 10% à 30% plus rapide<sup>99</sup>. Ainsi, sur papier, la lecture en diagonale est plus aisée, les mots sautés et les retours en arrière sont moins nombreux, etc. De plus, avec le document électronique, l'utilisateur ne lit pas seulement des informations textuelles, il interprète également un ensemble de signaux non-verbaux présents dans l'interface de lecture, tels que le curseur ou les boutons, qui entraînent pour l'utilisateur une surcharge cognitive durant la lecture.

Ensuite, les documents numériques sont souvent dénués de contexte, le lecteur est alors plus facilement perdu. Le contenu du document papier trouvé dans une bibliothèque pourra, sans être lu, être deviné par l'utilisateur grâce notamment à l'étagère sur laquelle il se trouve et sa quatrième de couverture. Ce contexte permet au lecteur de savoir de quelle manière le document doit être lu. Ainsi, il saura qu'un dictionnaire ne doit pas être lu linéairement et qu'il sera utile pour rechercher la définition d'un terme tandis qu'un roman devra être lu page par page. En ce qui concerne les documents électroniques, ils sont tous lus sur un seul et même support (écran d'ordinateur, tablette de lecture, etc.), ce qui ne permet donc pas de les différencier.

---

99 GHARBI Z. *Analyse des pratiques de lecture sur livres électroniques chez les étudiants universitaires : Thèse présentée en vue de l'obtention du grade de Philosophiae Doctor en sciences de l'information* [En ligne]. Montréal : Faculté des études supérieures, Janvier 2006. Disponible sur : [http://www.theses.umontreal.ca/theses/nouv/gharbi\\_z/these.pdf](http://www.theses.umontreal.ca/theses/nouv/gharbi_z/these.pdf) (Consulté le 17/08/2011)

L'enquête menée par l'Université Lyon 2 montre que près de 80% des étudiants de l'université utilisent tout de même les ressources numériques mises à leur disposition sur l'ENT (Environnement Numérique de Travail) de Lyon 2<sup>100</sup>. Cependant, ce chiffre doit être nuancé car, à cause d'une lecture sur écran qu'ils jugent difficile et fatigante, près de la moitié de ces mêmes étudiants impriment les textes dont ils ont besoin. Le document électronique, rapidement consulté, est alors remplacé par un document papier qui sera lu de manière plus approfondie par l'étudiant. De plus, 37,9% des étudiants se servent du numérique pour rassembler rapidement et facilement une masse d'informations, de documents, qu'ils liront plus tard.

Les étudiants impriment également les textes car ils apprécient de pouvoir travailler directement sur le contenu et de les manipuler. En effet, lors de leurs lectures, les étudiants ont pour habitudes de noter des questions en marge du texte, de mettre des annotations, de surligner des passages, etc. De plus, le document papier offre la possibilité d'étaler plusieurs textes à consulter en même temps et il est également plus facile à feuilleter. Enfin, les étudiants ne disposent pas tous du matériel adéquat pour lire un document électronique quel que soit le lieu où ils se trouvent. Le document imprimé permet alors de travailler entre deux cours, dans les transports en commun, dans un train, etc.

Cependant, ces mêmes étudiants mettent également en avant le caractère manipulable des textes numériques. Le travail peut être facilité grâce aux liens, ou hyperliens, présents sur les textes et permettant d'élargir les recherches mais aussi grâce à la recherche de termes permettant de localiser les données importantes du texte. De plus, la lecture numérique est jugée plus interactive car elle ne met plus seulement en scène du texte mais également des images, du son et toutes sortes d'actions déclenchées par le lecteur. Cette interactivité est alors, pour certains étudiants, une manière de faciliter l'apprentissage tandis que d'autres la voient comme un frein à la lecture.

Bien que très utile, la lecture sur support électronique est une lecture superficielle. Paradoxalement, les étudiants, qui sont près de 40% à avoir une lecture en diagonale des documents électroniques, considèrent presque

---

100 SAEMMER A. « La lecture sur supports numériques : diversification d'une activité complexe » [En ligne]. *3ème Journée Couperin sur le livre électronique : Le livre électronique : quels lecteurs ?*. Rennes : Couperin, octobre 2006. Disponible sur : <http://www.couperin.org/fr/groupe-de-travail-et-projets-deap/ebook/journee-du-livre-electronique/item/454-3e-journee-couperin-sur-le-livre-electronique> (Consulté le 17/08/2011)

unanimentement que la lecture numérique est une vraie lecture. On peut expliquer ce phénomène par le fait que la lecture numérique et l'hypertextualité permettent tout de même de créer des connexions entre les informations menant ainsi à une « représentation mentale de la matière textuelle parcourue »<sup>101</sup>.

Cette même hypertextualité est certainement à l'origine de la fatigue visuelle des étudiants et de leur lecture en diagonale. En effet, le numérique est synonyme pour les étudiants de rapidité et les liens hypertextes sont une tentation de plus pour naviguer de textes en textes. La surabondance de ces liens perd le lecteur qui se demande s'il doit les suivre ou non. Pour ne pas se perdre, il ne faut donc pas sombrer dans l'engrenage des hyperliens, or encore très peu d'utilisateurs y parviennent.

Enfin, même si les étudiants apprécient les documents électroniques pour réaliser leurs travaux, ils continuent d'aimer les textes papiers non seulement pour des raisons pratiques mais également pour des raisons historiques, morales, etc. Ainsi, lors l'étude menée à l'Université Lyon 2, certains étudiants ont déclaré que « un texte sur écran n'a pas de vécu comme un vrai livre », « Le texte sur support numérique n'a pas d'âme, car celle-ci réside dans le papier », etc.

Ainsi, on constate que le document numérique ne remplace pas le document papier mais le complète. En effet, d'une part, la lecture numérique est surtout considérée comme étant une lecture « contemplative », où le texte est survolé, et « immersive », où le lecteur procède à des actions sur le texte, d'autre part, la lecture papier est principalement « réflexive ». En outre, les textes papiers restent encore très présents face aux documents numériques qui rencontrent encore des limites techniques, juridiques, etc.

### **9.3. Les limites du livres numérique**

Le risque de perte de données est très important dans le cas des ressources électroniques présentes sur des supports physiques tels que les CD-Rom. En effet, le CD-Rom parvient difficilement, voire rarement, à atteindre l'âge de dix ans alors que le livre papier peut avoir une longévité de plusieurs dizaines d'années. Cela est aussi valable pour tous les types de supports électroniques car ils sont souvent fragiles : tablettes de lecture, clés USB, ordinateurs, etc. Un livre papier endommagé, qui a des tâches, auquel il manque une page, sera toujours lisible, ce qui n'est pas le cas d'un support numérique endommagé. De plus, les technologies

---

101 SAEMMER, 2006, op. cit.

évoluent rapidement, on ne peut donc pas être certains que le format du livre électronique acheté en 2010 disposera encore de lecteurs compatibles en 2020.

Toutefois, les limites actuelles les plus importantes des livres numériques ne résident pas dans les problèmes techniques mais dans les restrictions juridiques et les peurs économiques des éditeurs. Actuellement, la vente de livres électroniques ne représente que 15% du chiffre d'affaire d'un éditeur<sup>102</sup>. L'investissement est alors encore trop élevé pour les petits éditeurs qui vendront peu. De plus, même si la recette obtenue avec les livres numériques est plus élevée grâce à une diminution du nombre d'intermédiaires par rapport au livre papier (libraires, diffuseurs, etc.), elle subit une TVA à 19,5% contre 5,5% pour le papier, ce qui rend les livres numériques moins attractifs.

La seconde peur des éditeurs vient principalement du piratage qui est déjà récurrent avec les autres types de documents ou applications numériques : musiques, films, logiciels, jeux, etc. Entre les années 1990 et 2000, avec l'essor d'Internet et l'arrivée du haut débit, les échanges de fichiers ont fortement augmenté<sup>103</sup>. En 2002, des groupes tels que Microsoft, Apple et Sony ont alors mis en place les DRM (Digital Rights Management) qui permettent de contrôler l'accès aux fichiers et d'empêcher leur large diffusion sur Internet.

Grâce à ce système, ces groupes parviennent à rendre leurs ressources rares sur le web et ainsi à augmenter le coût de leur accès. Certains, comme MSN Music, ont même rendus ces ressources lisibles uniquement par des outils spécifiques, forçant ainsi l'utilisateur à acheter certains produits.

En outre, les DRM seraient un frein à la vente car, selon Tim O'Reilly, « la piraterie est une sorte d'impôt progressif, qui réduit peut-être de quelques points le pourcentage des ventes d'artistes célèbres (et je dis « peut-être », parce que même ce point n'est pas démontré), mais induit en contrepartie d'immenses bénéfices

---

102 CAZIN X. « Modèles économiques, DRM et piratage : halte aux fantasmes ! » [En ligne]. *4ème Journée Couperin sur le livre électronique : L'e-book en action(s) : de l'acquisition à la diffusion*. Bordeaux : Couperin, mars 2008. Disponible sur : <http://www.couperin.org/fr/groupe-de-travail-et-projets-deap/ebook/journee-du-livre-electronique/item/537-4e-journee-couperin-sur-le-livre-electronique> (Consulté le 17/08/2011)

103 FAUCILHON J. « Le problème des DRM : Frein aux usages et incitation au piratage » [En ligne]. *5ème Journée Couperin sur le livre électronique : quelles pratiques pour quels usages ?*. Lille : Couperin, 2010. Disponible sur : [http://www.couperin.org/components/com\\_flexicontent/medias/faucilhon\\_ceb.pdf](http://www.couperin.org/components/com_flexicontent/medias/faucilhon_ceb.pdf) (Consulté le 17/08/2011)

pour le nombre beaucoup plus élevé de ceux qui tireront avantage de cette publicité ».

Finalement, les DRM, comme beaucoup de techniques censées stopper le piratage, ont rapidement été contournées par les usagers les plus compétents en informatique. Les DRM ont alors contribué à augmenter le piratage de livres numériques. En France, le catalogue le plus piraté, est celui de l'éditeur Eyrolles : 5% de ses ressources sont piratées. Sur l'ensemble du web, seul 1% des livres sont piratés, ceux-ci concernent les domaines de la philosophie, de la cuisine, de la science-fiction et de l'informatique<sup>104</sup>. De plus, les livres piratés, bien que disponibles les trois quarts du temps en format papier, sont rarement disponibles en format électronique légal (seulement 5%).

Il existe tout de même d'autres solutions mises en place pour consulter ou télécharger des livres numériques légalement. D'une part, sur des interfaces sécurisées de consultation en ligne, telles que le Safari O'Reilly, qui regroupe des éditeurs, ou le projet Gallica de la BNF. Cette méthode pourrait s'apparenter à une sorte de bibliothèque en ligne. D'autre part, certains ouvrages sont désormais édités électroniquement sous licence libre, c'est-à-dire que n'importe quel individu a le droit d'utiliser l'ouvrage quelle que soit la situation, de l'étudier, de le copier et de le modifier.

---

104 DAVAL M. « EbookZ L'offre numérique illégale des livres français sur internet » [En ligne]. *5ème Journée Couperin sur le livre électronique : quelles pratiques pour quels usages ?*. Lille : Couperin, 2010. Disponible sur : [http://www.couperin.org/components/com\\_flexicontent/medias/daval.pdf](http://www.couperin.org/components/com_flexicontent/medias/daval.pdf) (Consulté le 17/08/2011)

## Conclusion

Le web de données offre donc une très grande interopérabilité des formats et des applications facilitant ainsi la recherche, l'échange et la portabilité des données. Nous avons pu démontrer que ces caractéristiques sont fortement appréciées par l'utilisateur qui change souvent d'outil pour lire les informations et qui veut retrouver les mêmes informations quelles que soient les interfaces qu'il utilise. De plus, l'organisation des données numériques est également le moyen pour l'utilisateur de repérer plus facilement ce qu'il recherche face à la masse d'information présente sur Internet. En effet, à l'heure actuelle, on pourrait comparer Internet à une immense bibliothèque désorganisée, où le travail de tri des ressources n'est pas fait correctement. Toutefois, on a également pu remarquer que ce même utilisateur était souvent sceptique face aux documents électroniques qu'il considère plutôt comme un document de substitution du document papier. En effet, on ne peut pas toujours accéder ou transporter aisément un document papier.

Du côté des relations entre les professionnels, le web de données ne les crée pas mais il contribue tout de même à les renforcer. En effet, l'échange d'expériences et de notices bibliographiques entre les bibliothèques ne sont pas si récents. Pour le premier, les premières associations de professionnels contribuant à ouvrir les discussions entre les professionnels sont bien plus anciennes que le web de données. Pour le second, l'apparition des formats Marc et du protocole Z39.50, favorisant l'échange de notices entre les bibliothèques, date déjà de la fin des années 1990. Cette situation est alors renforcée grâce aux groupes de travail regroupant les professionnels de l'information-documentation mais également des professionnels et chercheurs des autres secteurs tels que les informaticiens ou les linguistes. En ce qui concerne l'échange de notices et de données, il est facilité par la présence des standards et d'outils applicables à divers types de documents tels que le XML, le Dublin Core ou les entrepôts OAI.

Concernant, les avantages économiques du web de données, celui-ci permet certes de diminuer le travail des bibliothécaires en matière d'indexation et de catalogage ainsi que de rapatrier automatiquement des données hébergées par une autre structure. Néanmoins, les économies faites sont peu importantes face aux dépenses liées à l'installation, et parfois au développement, des nouveaux outils, à la formation du personnel à ces outils, à l'embauche d'un personnel ayant de plus

en plus de compétences, etc. À court terme, il n'y a donc pas de réels gains économiques.

Toutefois, nous avons pu voir que l'orientation des bibliothèques vers le web de données répond à des problématiques anciennes datant de l'époque sumérienne. L'élaboration de systèmes d'organisation des connaissances permettant de rassembler toutes les connaissances au-delà des frontières de la langue est le rêve de tout bibliothécaire depuis des siècles. Aujourd'hui, le web de données pourrait être pour le bibliothécaire une réponse à ce désir. Néanmoins, nous avons également constaté que les bibliothèques ne disposaient pas réellement d'un libre arbitre. En effet, elles sont influencées, d'un côté, par les organismes menant les recherches, développant les applications et faisant ensuite des retours d'expériences positifs sur ces technologies ainsi que par les organismes qui les financent. D'un autre côté, le public a beaucoup d'influence car les bibliothèques tentent de le satisfaire. Ainsi elles suivent les souhaits du public dont les besoins se créent en fonction des offres du marché, ces besoins ne correspondent donc pas à une réelle nécessité et ils sont plutôt nés grâce à un effet de mode.

Je pense que nous pouvons en conclure que l'application du web de données comporte encore de nombreuses failles et points obscurs. En effet, la plupart des outils du web de données ont été appliqués récemment et à petite échelle. Pour l'instant, le web de données en bibliothèque numérique correspond surtout au respect du XML et du Dublin Core ainsi qu'à la mise en place d'entrepôts OAI.

De plus, ces technologies récentes ne permettent pas d'analyser sur le long terme les conséquences financières et les mutations professionnelles. Je pense tout de même que, s'il doit générer un gain financier, celui-ci sera faible. En effet, avec la documentation numérique, les bibliothèques évoluent rapidement. Ainsi, il est fort probable que les technologies du web de données existantes aujourd'hui ne correspondent plus aux besoins de demain. Enfin, je pense également que cet éternel désir des bibliothécaires à vouloir rassembler et organiser l'ensemble des connaissances est une utopie. Cela est encore plus vrai avec Internet et la masse d'information déposée chaque jour par des internautes lambda. Toutefois, cette utopie permet de motiver les bibliothécaires dans leurs recherches et leurs évolutions afin de s'en rapprocher.



## Table des matières

Remerciements.....	3
Introduction.....	5
Partie 1 : De la préhistoire à l'histoire de la sémantique.....	10
1. Les utopies et travaux sur le langage.....	11
1.1. Babel et les sumériens.....	11
1.2. La philosophie du langage et la critique de la langue.....	13
1.3. La linguistique.....	15
1.4. L'intelligence artificielle.....	17
2. Les projets précurseurs en bibliothèque.....	18
2.1. Les systèmes d'organisation des connaissances (Soc).....	19
2.2. La bibliothèque et l'accès à la connaissance.....	22
2.3. L'informatique en bibliothèque.....	24
3. Bibliothèque virtuelle : du web 1.0 au web 3.0.....	26
3.1. Web 1.0 et la bibliothèque numérique.....	26
3.2. Web 2.0 et l'interaction avec l'utilisateur.....	28
3.3. Web 3.0 et les linked data.....	30
Partie 2 : Application du web de données en bibliothèque.....	33
4. Les formats du web de données.....	34
4.1. La localisation des données : les URI.....	35
4.2. Les règles d'écriture : le XML.....	35
4.3. La structure des données : le RDF.....	37
4.4. Les vocabulaires : le RDFS (RDFS) et l'OWL.....	38
4.5. Les requêtes : le SparQL.....	40
4.6. Bilan général.....	40
5. Réalisations pour l'information-documentation.....	41
5.1. Dublin Core et Skos.....	41
5.2. L'initiative pour des archives ouvertes : l'OAI.....	43
5.3. Un projet spécifique à l'information-documentation : le FRBR.....	45

5.4. Applications RDF : les vocabulaires Foaf et Viaf.....	47
5.5. Le Sudoc et l'Idref.....	49
6. La prise de décision en bibliothèque.....	50
6.1. Les groupes d'influences.....	51
6.2. La satisfaction de l'utilisateur.....	55
6.3. L'évolution : une nécessité pour les bibliothèques.....	56
Partie 3 : La bibliothèque hybride, une amorce à la bibliothèque du futur.....	58
7. Impacts socio-organisationnels.....	59
7.1. Les nouvelles compétences des professionnels.....	59
7.2. Une réorganisation interne.....	62
7.3. Une communication avec l'extérieur.....	64
8. La politique documentaire.....	66
8.1. Dématérialisation de la collection.....	66
8.2. Conservation des documents numériques.....	68
8.3. Accès à l'information.....	70
9. La place du livre.....	72
9.1. L'essor du livre électronique.....	72
9.2. Une autre lecture : le cas des étudiants de Lyon 2.....	73
9.3. Les limites du livre numérique.....	76
Conclusion.....	79
Bibliographie.....	84
Ouvrages.....	84
Articles de presse.....	85
Conférences, congrès et colloques.....	87
Rapports, mémoires et thèses.....	89
Documents en ligne.....	90
Pages web.....	90
Annexes.....	95
Annexe n°1 : Exemple de notice Rameau.....	96
Annexe n°2 : Le terme « élève » dans le thésaurus Motbis.....	97

Annexe n°3 : Code de déontologie du bibliothécaire.....	98
Annexe n°4 : Manifeste de l'Unesco.....	100
Annexe n°5 : Charte des bibliothécaires.....	103
Annexe n°6 : Exemple de structure XML.....	104
Annexe n°7 : Extrait de RDFS.....	105
Annexe n°8 : Les 15 éléments du Dublin Core.....	106

# Bibliographie

## Ouvrages

- BOTTÉRO J. *Babylone : À l'aube de notre culture*. Évreux : Gallimard, 1994
- CALDERAN L., HIDOINE B., MILLET J (dir.). *Métadonnées : mutations et perspectives : Séminaire INRIA, 29 septembre-3 octobre 2008 - Dijon*. Paris : ADBS, 2008
- CALENGE B. *Bibliothèque et politique documentaire à l'heure d'internet*. Paris : Electre, 2008
- CHISS J-L., PUECH C. *Fondations de la linguistique : études d'histoire et d'épistémologie*. Louvain-la-Neuve : Duculot, 1997
- DE VAUX R, BENOIT P et al (dir). « Genèse 11.1-9 ». *La Bible de Jérusalem*. Paris : Éditions du Cerf, 1973. p. 29
- DONNAT O. *Les pratiques culturelles des Français à l'ère numérique : enquête 2008*. Paris : La Découverte, 2009
- ECO U. *La recherche de la langue parfaite dans la culture européenne*. Paris : Éditions du Seuil, 1994
- GARNIER Y (dir.). « ontologie ». *Le Petit Larousse illustré 2009*. Paris : Larousse, 2008
- HOBBS T. *Troisièmes Objections aux Méditations sur la philosophie première de Descartes*. 1640
- HODGE G. *Systems of Knowledge Organization for Digital libraries : Beyond traditional authority files*. Washington, DC : the Council on Library and Information resources, 2000
- LOCKE J. *Essai sur l'entendement humain*. Paris : Ellipses, 2003

PAPY F (dir.). *Technodocumentation : Des machines informationnelles aux bibliothèques numériques*. Paris : Lavoisier, 2009

PÉDAUQUE R.T. *La redocumentarisation du monde*. Toulouse : Cépaduès, 2007

REBILLARD F. *Le web 2.0 en perspective : Une analyse socio-économique de l'internet*. Paris : L'Harmattan, 2007

RIVIER A. *Aide mémoire d'informatique documentaire*. Paris : Éditions du Cercle de la librairie, 2007

TABET C. *La bibliothèque « hors les murs »*. Paris : Electre, 2004

### **Articles de presse**

ABEL M.-H. « Utilisation de normes et standards dans le projet Memora ».  
*Distances et savoirs*, avril 2004, vol. 2, pp. 487-511

BATTISTI M. « IFLA 2003 : Congrès mondial des bibliothèques et de l'information » La bibliothèque point d'accès : médias - information – culture ». *Documentaliste-Sciences de l'Information*, avril 2003, vol. 40, n°4-5, pp. 308-315.

BENSAADI F. « Approche méthodologique pour la création de fichiers d'autorité nationaux ». *Revue d'Information Scientifique & Technique*, 1995, n°2, vol. 5, pp. 29-32

BERTRAND Anne-Marie, EPRON Benoit. « Les ressources numériques : un nouvel enjeu pour les compétences des professionnels des bibliothèques ». *Documentaliste-Sciences de l'information*, 2010, vol. 47, n°2, pp. 65-66.

BONACCORSI J. « Le livre déplacé, une bibliothèque hors-les-murs ». *Communication et langage*, 2001, n°127, vol. 127, pp. 21-34

CONDAMINES A. « Linguistique de corpus et terminologie ». *Langages*, 2005, n°157, pp. 36-47

FLORIDI L. « Web 2.0 vs. the Semantic Web: A Philosophical Assessment ». *Episteme*, 2009, n°6, pp. 25-37

- FURNO M. « De l'elementarium au thésaurus : l'émancipation des lexiques latins monolingues aux XV<sup>e</sup> et XVI<sup>e</sup> siècles ». *Histoire Épistémologie Langage*, 1997, Tome 19, fascicule 1, pp. 151-175
- GIBOIN A, ZACKLAD M (dir). « Applications à base de Soc hétérogènes : Thésaurus, ontologies, folksonomies... ». *Document numérique*, mai-août 2010, n°2, vol. 13.
- GLEYZE A. « Quelles compétences pour les bibliothécaires ? : Diversité des fonctions et référentiels ». *Documentaliste-Sciences de l'Information*, janvier 2005, vol. 42, n°1, pp. 22-25.
- GONDRAN M. « Bilan et perspective de dix ans d'Intelligence Artificielle en entreprise ». *Quaderni*, printemps 1995, n°25, pp. 99-123
- LE DEUFF O. « La culture de l'information et l'héritage documentaire ». *Documentaliste-Sciences de l'Information*, mars 2010, vol. 47, n°3, pp. 4-11.
- MAISONNEUVE, Marc. « Dix bonnes raisons de choisir un opac de nouvelle génération ». *Documentaliste - Sciences de l'information*, 2008, vol. 45, n° 3, pp. 16-17
- MENON B. « Les langages documentaires : Un panorama, quelques remarques critiques et un essai de bilan ». *Documentaliste Sciences de l'information*, 2007, vol. 44, pp. 18-28
- SCHMITT J.-P. « L'Open Archives Initiative et la validation des publications scientifiques : Un séminaire sur l'OAI à Genève ». *Documentaliste-Sciences de l'Information*, février 2001, vol. 38, pp. 124-126
- SCHÖPFEL J. « L'avenir du métier de bibliothécaire » [En ligne]. *Archimag : Guide pratique 42 : Bibliothèques, les nouveaux usages*, 2011, pp. 62-66
- WATERS D. « What are Digital libraries? » [En ligne]. *CLIR Issue*, juillet/août 1998, n°4. Disponible sur : <http://www.clir.org/pubs/issues/issues04.html> (Consulté le 17/08/2011)

## Conférences, congrès et colloques

BENNET R., HENGEL-DITTRICH C., al. *VIAF (Virtual International Authority File) : un pont entre les fichiers d'autorité noms de personnes de Die Deutsche Bibliothek et de la Bibliothèque du Congrès* [En ligne]. Seoul : Ifla, 13 juillet 2006. Disponible sur : [http://archive.ifla.org/IV/ifla72/papers/123-Bennett\\_trans-fr.pdf](http://archive.ifla.org/IV/ifla72/papers/123-Bennett_trans-fr.pdf) (Consulté le 3/08/2011)

BERMÈS E. « Publier les référentiels sur le Web sémantique » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [http://www.bnf.fr/documents/afnor2011\\_bermes.pdf](http://www.bnf.fr/documents/afnor2011_bermes.pdf) (Consulté le 3/08/2011)

BOULET V. « VIAF : une brique importante pour le web sémantique » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_viaf\\_boulet.pdf](http://www.bnf.fr/documents/afnor2011_viaf_boulet.pdf) (Consulté le 3/08/2011)

CAZIN X. « Modèles économiques, DRM et piratage : halte aux fantasmes ! » [En ligne]. *4ème Journée Couperin sur le livre électronique : L'e-book en action(s) : de l'acquisition à la diffusion*. Bordeaux : Couperin, mars 2008. Disponible sur : <http://www.couperin.org/fr/groupes-de-travail-et-projets-deap/ebook/journee-du-livre-electronique/item/537-4e-journee-couperin-sur-le-livre-electronique> (Consulté le 17/08/2011)

DAVAL M. « EbookZ L'offre numérique illégale des livres français sur internet » [En ligne]. *5ème Journée Couperin sur le livre électronique : quelles pratiques pour quels usages ?*. Lille : Couperin, 2010. Disponible sur : [http://www.couperin.org/components/com\\_flexicontent/medias/daval.pdf](http://www.couperin.org/components/com_flexicontent/medias/daval.pdf) (Consulté le 17/08/2011)

FAUCILHON J. « Le problème des DRM : Frein aux usages et incitation au piratage » [En ligne]. *5ème Journée Couperin sur le livre électronique : quelles pratiques pour quels usages ?*. Lille : Couperin, 2010. Disponible sur : [http://www.couperin.org/components/com\\_flexicontent/medias/faucilhon\\_ceb.pdf](http://www.couperin.org/components/com_flexicontent/medias/faucilhon_ceb.pdf) (Consulté le 17/08/2011)

LE PAPE P. « RDA : le nouveau paradigme du catalogage. : FRBR, c'est quoi déjà ? » [En ligne]. 5 à 7 de l'ADBS, 29 mars 2011. 21 juin 2011 Disponible sur : <http://www.adbs.fr/rda-le-nouveau-paradigme-du-catalogage-1-frbr-c-est-quoi-deja--102461.htm> (Consulté le 3/08/2011)

MENON B. « Les systèmes d'organisation des connaissances, de Babel aux linked data ». *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique* [En ligne]. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_intro\\_menon.pdf](http://www.bnf.fr/documents/afnor2011_intro_menon.pdf) (Consulté le 3/07/2011)

MERRIEN D. *Colloque international sur les "ressources électroniques académiques : mesures et usages"* [En ligne]. Lille : Couperin, 30 novembre 2009. Disponible sur : [http://www.couperin.org/IMG/pdf/Colloque\\_Mesures\\_usages\\_RE\\_Lille\\_2009.pdf](http://www.couperin.org/IMG/pdf/Colloque_Mesures_usages_RE_Lille_2009.pdf) (Consulté le 10/08/2010)

NICOLAS Y. « IdRef : Les référentiels comme services. Une politique de l'ABES » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_abes.pdf](http://www.bnf.fr/documents/afnor2011_abes.pdf) (Consulté le 3/08/2011)

POUPEAU G. « Le Web de données et les bibliothèques » [En ligne]. *Journées d'études AULA sur le Web sémantique*. 27 juin 2008. Disponible sur : <http://www.slideshare.net/lespetitescases/le-web-de-donnees-et-les-bibliothques-presentation> (Consulté le 29/07/2011)

RABAULT H., ZYSMAN H. « Les schémas de concepts et le Web sémantique : la norme sur les thésaurus ISO 25964 et le Web sémantique » [En ligne]. *Journée d'étude AFNOR CG46 / BnF : Référentiels et données d'autorités à l'heure du Web sémantique*. Paris : BnF, 27 mai 2011. Disponible sur : [www.bnf.fr/documents/afnor2011\\_norme\\_thesaurus.pdf](http://www.bnf.fr/documents/afnor2011_norme_thesaurus.pdf) (Consulté le 3/08/2011)



SAEMMER A. « La lecture sur supports numériques : diversification d'une activité complexe » [En ligne]. *3ème Journée Couperin sur le livre électronique : Le livre électronique : quels lecteurs ?*. Rennes : Couperin, octobre 2006. Disponible sur : <http://www.couperin.org/fr/groupes-de-travail-et-projets-deap/ebook/journee-du-livre-electronique/item/454-3e-journee-couperin-sur-le-livre-electronique> (Consulté le 17/08/2011)

ZACKLAD M. « Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI) » [en ligne]. *35e Congrès annuel de l'Association Canadienne des Sciences de l'Information. Partage de l'information dans un monde fragmenté : Franchir les frontières*. Montréal : CAIS/ACSI, 2007. Disponible sur [http://www.cais-acs.ca/proceedings/2007/zacklad\\_2007.pdf](http://www.cais-acs.ca/proceedings/2007/zacklad_2007.pdf) (Consulté le 3/07/2011)

### **Rapports, mémoires et thèses**

DAMERON O. *Modélisation, représentation et partage de connaissances anatomiques sur le cortex cérébral : Thèse*. Rennes : Université de Rennes 1, 2003. Disponible sur : <http://websemantique.org/Ontologie> (Consulté le 3/08/2011)

GHARBI Z. *Analyse des pratiques de lecture sur livres électroniques chez les étudiants universitaires : Thèse présentée en vue de l'obtention du grade de Philosophiae Doctor en sciences de l'information* [En ligne]. Montréal : Faculté des études supérieures, Janvier 2006. Disponible sur : [http://www.theses.umontreal.ca/theses/nouv/gharbi\\_z/these.pdf](http://www.theses.umontreal.ca/theses/nouv/gharbi_z/these.pdf) (Consulté le 17/08/2011)

GUTMAN J. *Rapport de recherche bibliographique : Les applications de l'intelligence artificielle à l'assistance ou l'automatisation du catalogage* [En ligne]. Villeurbanne : ENSSIB, 1996. Disponible sur : [www.enssib.fr/bibliotheque-numerique/document-1621](http://www.enssib.fr/bibliotheque-numerique/document-1621) (Consulté le 3/07/2011)

LACHENAUD J-P. *Bibliothèques universitaires : le temps des mutations*. Les Rapports du Sénat, n°59. Paris : Sénat, 1998-1999.

## Documents en ligne

ERTZSCHEID O. « Les « nouveaux » catalogues ou le catalogue en (r)évolutions » [En ligne]. @rchiveSIC :: Accueil. Août 2007. Disponible sur : <http://archivesic.ccsd.cnrs.fr/docs/00/17/80/06/PDF/ArticleABES.pdf> (Consulté le 10/08/2011)

JABES. *Le Web sémantique : un web de métadonnées* [En ligne]. Abes, 2010. Disponible sur : <http://www.slideshare.net/yannn/le-web-smantique-un> (Consulté le 3/08/2011)

SCHOPFEL J. « Bibliothèque sans bibliothécaires ? » [En ligne]. HAL – INRIA. 19 juillet 2005. Disponible sur : [http://hal.inria.fr/docs/00/06/26/48/PDF/sic\\_00001524.pdf](http://hal.inria.fr/docs/00/06/26/48/PDF/sic_00001524.pdf) (Consulté le 10/08/2011)

## Pages web

ABES. *L'ABES – ABES.fr* [En ligne]. 3 juillet 2011. Disponible sur : <http://www.abes.fr/abes/page.351,labes.html> (Consulté le 3/07/2011)

ABES. « Le réseau Sudoc » [En ligne]. ABES : Accueil. Disponible sur : <http://www.abes.fr/abes/page.352,le-reseau-sudoc.html> (Consulté le 3/08/2011)

ADBS. « Le référentiel des métiers et fonctions » [En ligne]. Accueil - L'association des professionnels de l'information et de la documentation. 29 juillet 2010. Disponible sur : [http://www.adbs.fr/le-referentiel-des-metiers-et-fonctions-61928.htm?RH=MET\\_REFMETIER](http://www.adbs.fr/le-referentiel-des-metiers-et-fonctions-61928.htm?RH=MET_REFMETIER) (Consulté le 10/08/2011)

BARGE P. « Cloud Computing : l'informatique de demain ». *Le Journal du Net : e-Business, Informatique, Economie et Management*. 21 février 2011. Disponible sur : <http://www.journaldunet.com/developpeur/expert/49652/cloud-computing---l-informatique-de-demain.shtml> (Consulté le 10/08/2011)

BEN AMEUR H. « Web Sémantique : OWL » [En ligne]. *Web Sémantique*. 23 juillet 2011. Disponible sur : <http://websemantique.org/OWL> (Consulté le 3/08/2011)

- BERMÈS E. « Les FRBR, qu'est-ce que c'est ? » [En ligne]. *Figoblog : Un blog sur Internet, la bibliothéconomie et la confiture de figues*. 10 mars 2005. Disponible sur : <http://www.figoblog.org/document594.php> (Consulté le 3/08/2011)
- BERNERS-LEE T. « Semantic Web on XML » [En ligne]. *XML 2000* : Washington DC, 6 décembre 2000. Disponible sur : <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html> (Consulté le 3/08/2011)
- BNF. « Dublin Core » [En ligne]. *Bibliothèque nationale de France*. 6 avril 2011. Disponible sur : [http://www.bnf.fr/fr/professionnels/formats\\_catalogage/a.f\\_dublin\\_core.html](http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html) (Consulté le 3/08/2011)
- BNF. *Gallica - le blog de la bibliothèque numérique de la Bibliothèque nationale de France – BnF* [En ligne]. 29 juillet 2011. Disponible sur : <http://blog.bnf.fr/gallica/> (Consulté le 29/07/2011)
- BNF. « Modèles FRBR, FRAD et FRISAD » [En ligne]. *Bibliothèque nationale de France*. 6 avril 2011. Disponible sur : [http://www.bnf.fr/fr/professionnels/modelisation\\_ontologies/a.modele\\_FRBR.html](http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html) (Consulté le 3/08/2011)
- BNF. « Web sémantique, Web de données » [En ligne]. *Bibliothèque nationale de France*. 26 octobre 2010. Disponible sur : [http://www.bnf.fr/fr/professionnels/web\\_semantique\\_donnees/s.web\\_semantique\\_intro.html](http://www.bnf.fr/fr/professionnels/web_semantique_donnees/s.web_semantique_intro.html) (Consulté le 29/07/2011)
- BURLESON C. « Introduction to the Semantic Web Vision and Technologies : Part 3 - The Resource Description Framework » [En ligne]. *Semantic Focus - The Semantic Web, Semantic Web technology and computational semantics*. 12 octobre 2007. Disponible sur : <http://www.semanticfocus.com/blog/entry/title/introduction-to-the-semantic-web-vision-and-technologies-part-3-the-resource-description-framework/> (Consulté le 3/08/2011)

- COLMANT Y. « Historique du projet ORI-OAI » [En ligne]. *ORI-OAI.ORG*. 6 mai 2010. Disponible sur : <http://www.ori-oai.org/display/PROJET/Historique> (Consulté le 3/08/2011)
- COMMENT ÇA MARCHE. « Introduction à XML » [En ligne]. *Comment Ça Marche (CCM) – Communauté informatique*. 14 octobre 2008. Disponible sur : <http://www.commentcamarche.net/contents/xml/xmlintro.php3> (Consulté le 3/08/2011)
- EBSCO. « La Bibliothèque du Futur - Interview du Professeur Gilles Vassal » [En ligne]. EBSCO Information Services. 2011. Disponible sur : <http://www2.ebsco.com/fr-fr/NewsCenter/TC/Pages/Vassal.aspx> (Consulté le 17/08/2011)
- ÉDUCNET. « Les Universités numériques thématiques » [En ligne]. *Éducnet : Enseignement, TICE, Nouvelles technologies, Ressources pédagogiques numériques*. 3 juillet 2009. Disponible sur : <http://www.educnet.education.fr/superieur/unt> (Consulté le 10/08/2011)
- ERTZSCHEID O. « La dialectique de la cage et du nuage ». *affordance.info*. 24 mars 2010. Disponible sur : [http://affordance.typepad.com/mon\\_weblog/2010/03/dialectique-de-la-cage-et-du-nuage.html](http://affordance.typepad.com/mon_weblog/2010/03/dialectique-de-la-cage-et-du-nuage.html) (Consulté le 10/08/2011)
- FIEVET C. « FOAF, la première brique du Web sémantique ? » [En ligne]. *InternetActu.net*. 2 juin 2004. Disponible sur : <http://www.internetactu.net/2004/06/02/foaf-la-premiere-brique-du-web-smantique/> (Consulté le 3/08/2011)
- GARY N. « Trop de contraintes : le Doc Insa de Lyon se désabonne de Numilog » [En ligne]. *ActuaLitté - Les univers du livre*. 22 décembre 2010. Disponible sur <http://www.actualitte.com/actualite/23388-lyon-numilog-ebooks-abonnement-contraintes.htm> (Consulté le 17/08/2011).
- GOMBAULT V. « Deux ménages sur trois disposent d'internet chez eux » [En ligne]. *Institut national de la statistique et des études économiques : Accueil*. Mars 2011. Disponible sur : [http://www.insee.fr/fr/themes/document.asp?ref\\_id=ip1340](http://www.insee.fr/fr/themes/document.asp?ref_id=ip1340) (Consulté le 29/07/2011)

- GOOGLE. *Google Recherche de Livres* [En ligne]. 28 juillet 2011. Disponible sur : <http://books.google.com/googlebooks/about.html> (Consulté le 29/07/2011)
- GUILLOU P. « Comprendre le web 1.0 » [En ligne]. *Ideose, numérique pour tous*. 29 novembre 2009. Disponible sur : <http://www.ideose.eu/blog/web-temps-reel/comprendre-le-web-10/> (Consulté le 3/07/2011)
- MÉDIAMÉTRIE. « L'audience de l'internet en France en avril 2011 » [En ligne]. *Médiamétrie, Mesure d'audience Télévision, Radio, Cinéma, Internet, études médias audiovisuels*. 23 mai 2011. Disponible sur : <http://www.mediametrie.fr/presse/communiqués/l-audience-de-l-internet-en-france-en-avril-2011.php?id=463> (Consulté le 3/07/2011)
- NAWROCKI F. « Le protocole OAI et ses usages en bibliothèque » [En ligne]. *Ministère de la Culture et de la Communication*. 15 février 2005 Disponible sur : <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm> (Consulté le 17/08/2011)
- NEPOTE C. « Web Sémantique : RDFSchemata » [En ligne]. *Web Sémantique*. 23 juillet 2011. Disponible sur : <http://websemantique.org/RDFSchemata> (Consulté le 3/08/2011)
- SCD Lyon 2, SCD Lyon 3 et al. *Groupe de travail sur les SIGB Livres* [En ligne]. 3 juillet 2011. Disponible sur : <http://www.sigb-libres.info> (Consulté le 3/07/2011)
- UNIVERSITÉ DE LYON. « Présentation de l'Université de Lyon » [En ligne]. *L'Université de Lyon*. 10 août 2011. Disponible sur : <http://www.universite-lyon.fr/l-universite-de-lyon/> (Consulté le 10/08/2011)
- UNIVERSITY COLLEGE LONDON. « SAMPA - computer readable phonetic alphabet ». *UCL - London's Global University* [En ligne]. 3 juillet 2011. Disponible sur : <http://www.phon.ucl.ac.uk/home/sampa/> (Consulté le 3/07/2011)

UNT. « Le groupe « Indexation – SupLOMFR » [En ligne]. *Universités numériques*. 10 août 2011. Disponible sur : <http://www.universites-numeriques.fr/fr/content/le-groupe-%C2%AB-indexation-%E2%80%93-suplomfr-%C2%BB> (Consulté le 10/08/2011)

### ***Universalis***

ABEILLÉ A. « Traitement automatique des langues » [En ligne]. *Encyclopaedia Universalis – Education*. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/traitement-automatique-des-langues/> (Consulté le 3/07/2011)

CARDASCIA G, LAFFORGUE G. « Babylone ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/babylone/> (Consulté le 3/07/2011)

FOSKETT D. « Indexation ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/indexation/> (Consulté le 3/07/2011)

FUCHS C. « Linguistique ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/linguistique/#> (Consulté le 3/07/2011)

MAIGNIEN Y. « Bibliothèques Numériques » [En ligne]. *Encyclopaedia Universalis – Education*. 29 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/bibliotheques-numeriques/#5> (Consulté le 29/07/2011)

RICOEUR P, SEBESTIK J. « Philosophies du langage ». *Encyclopaedia Universalis – Education* [En ligne]. 3 juillet 2011. Disponible sur : <http://universalis.bibliotheque-nomade.univ-lyon2.fr/encyclopedie/philosophies-du-langage/> (Consulté le 3/07/2011)

# **Annexes**

## Annexe n°1 : Exemple de notice Rameau

**Alimentation -- Besoins** [+ subd. géogr.]

*Velette matière nom commun . S'emploie en tête de velette ou en subdivision*

S'emploie également en subdivision [+ subd. géogr.] aux catégories de personnes, êtres vivants (animaux) et groupes ethniques

**<Employé pour :**

Aliments -- Besoins  
Animaux -- Alimentation -- Besoins  
Animaux domestiques -- Alimentation -- Besoins  
Animaux sauvages -- Alimentation -- Besoins  
Besoins alimentaires  
Besoins énergétiques  
Besoins nutritionnels  
Nutrition -- Besoins

**>><<Terme(s) associé(s) :**

[Faim](#)  
[Fibres dans l'alimentation](#)  
[Glucides dans l'alimentation](#)  
[Lipides dans l'alimentation](#)  
[Minéraux dans l'alimentation](#)  
[Protéines dans l'alimentation](#)  
[Soif](#)  
[Vitamines dans l'alimentation](#)

**>>Terme(s) spécifique(s) :**

[Animaux -- Besoins en eau](#)

**Equiv. LCSH :** Nutrition -- Requirements

**Domaine(s) :** 570

**Notice n° :** FRBNF12107826    **Origine :** Laval RVM  
**Création :** 89/03/24    **Mise à jour :** 04/08/09

<http://catalogue.bnf.fr/ark:/12148/cb121078266/PUBLIC>



## Annexe n°2 : Le terme « élève » dans le thésaurus Motbis

élevage intensif [MT 3510]  
élevage ovin [MT 5005]  
élevage porcin [MT 5005]  
élève EM élève (système éducatif)  
aide à l' élève [MT 2105]  
délégué élève [MT 2120]  
interaction maître- élève EM relation enseignant-apprenant  
parent d' élève [MT 2120]  
projet personnel de l' élève [MT 2105]  
relation maître- élève EM relation enseignant-apprenant  
élève (système éducatif) [MT 2120]  
élève décrocheur [MT 2120]  
élève en difficulté [MT 2120]  
élève étranger EM élève primo-arrivant  
élève handicapé ou malade [MT 2120]  
élève intellectuellement précoce EM élève su..  
élève primo-arrivant [MT 2120]  
élève surdoué [MT 2120]  
association de parents d' élèves EM parent d'élève  
conseil d' élèves [MT 2120]  
conseil des délégués des élèves [MT 2120]  
droits des élèves [MT 2120]  
..amme pour le suivi des acquis des élèves EM évaluation internationale (enseign..  
relation entre élèves [MT 2105]  
travail des élèves [MT 2105]  
réseau d'aide spécialisée aux élèves en difficulté EM aide à l'élève  
CAPA pr la scolarisation des élèves en Situation de Handicap EM CAPA-SH (..  
..e scolaire pour l'intégration des élèves han EM assistant pédagogique  
adaptation et scolarisation des élèves handicapés EM intégration scolaire de..  
..e scolaire pour l'intégration des élèves handicapés EM assistant pédagogique

## Annexe n°3 : Code de déontologie du bibliothécaire

### CODE DE DÉONTOLOGIE DU BIBLIOTHÉCAIRE

Le bibliothécaire est chargé par sa collectivité publique ou privée de répondre aux besoins de la communauté en matière de culture, d'information, de formation et de loisirs. Il constitue à cette fin les collections publiques, en assure la mise en valeur et l'usage citoyens. Conscient des responsabilités qui sont les siennes et appliquant les lois et règlements en vigueur, il s'engage à respecter vis-à-vis de l'utilisateur, des collections, de sa collectivité et de sa profession les principes qui suivent. Ce code déontologique du bibliothécaire, distinct de la charte documentaire propre à chaque établissement et de la charte des bibliothèques du Conseil supérieur des bibliothèques, les complète.

#### 1. L'USAGER

Le bibliothécaire est d'abord au service des usagers de la bibliothèque. L'accès à l'information et à la lecture étant un droit fondamental, le bibliothécaire s'engage dans ses fonctions à :

- Respecter tous les usagers
- Offrir à chacun une égalité de traitement
- Garantir la confidentialité des usages
- Répondre à chaque demande, ou, à défaut, la réorienter
- Assurer les conditions de la liberté intellectuelle par la liberté de lecture
- Assurer le libre accès de l'utilisateur à l'information sans laisser ses propres opinions interférer
- Permettre un accès à l'information respectant la plus grande ouverture possible, libre, égal et gratuit, sans préjuger de son utilisation ultérieure
- Garantir l'autonomie de l'utilisateur, lui faire partager le respect du document, favoriser l'autoformation
- Promouvoir auprès de l'utilisateur une conception de la bibliothèque ouverte, tolérante, conviviale

#### 2. LA COLLECTION

Le bibliothécaire favorise la réflexion de chacun par la constitution de collections répondant à des critères d'objectivité, d'impartialité, de pluralité d'opinion. Dans ce sens, il s'engage dans ses fonctions à :

- Ne pratiquer aucune censure, garantir le pluralisme et l'encyclopédisme intellectuel des collections
- Offrir aux usagers l'ensemble des documents nécessaires à sa compréhension autonome des débats publics, de l'actualité, des grandes questions historiques et philosophiques
- Appliquer les dispositions législatives et réglementaires concernant les collections, ainsi que les décisions de la justice, sans se substituer à celle-ci, notamment celles qui interdisent la promotion de toute discrimination et de toute violence
- Assurer la fiabilité des informations, œuvrer à leur mise à jour permanente et à leur conformité à l'état présent des connaissances scientifiques
- Organiser l'accès aux sources d'informations pour les rendre disponibles, y compris à distance, selon les normes professionnelles en vigueur
- Faire connaître et mettre en valeur les collections, les ressources et les services dans le respect de la neutralité du service public
- Faciliter la libre circulation de l'information

### 3. LA TUTELLE (collectivité publique ou privée)

La tutelle définit dans son domaine de compétence une politique générale. Dans ce cadre, la définition de la politique documentaire, déléguée au bibliothécaire, est précisée dans une charte, validée par la tutelle.

Le bibliothécaire en assure la mise en œuvre au quotidien dans le respect de ce code.

- Le bibliothécaire participe à la définition de la politique culturelle de sa tutelle
- Le bibliothécaire applique la politique de sa tutelle tant que celle-ci ne va pas à l'encontre des lois générales, des missions pérennes et spécifiques de l'établissement, ainsi que des valeurs définies dans ce code
- Le bibliothécaire fait valoir auprès de sa tutelle les nécessités de la formation professionnelle, comme stagiaire ou comme formateur, et plus particulièrement celles liées à sa participation aux journées d'étude, aux voyages d'étude et aux instances statutaires des associations professionnelles. Cette participation est considérée comme temps de travail
- Le bibliothécaire rend compte à sa tutelle, en les évaluant, des services et des activités de l'établissement
- Le bibliothécaire veille à ne pas céder aux groupes de pression politiques, religieux, idéologiques, syndicaux, sociaux qui essaieraient d'influer sur les politiques d'acquisitions par imposition forcée, interdiction ou intimidation, directement ou par le biais de sa tutelle

### 4. LA PROFESSION

Les personnels des bibliothèques forment un corps professionnel solidaire. Au sein de ce corps, le bibliothécaire trouve aide et assistance, et apporte ses connaissances et son expérience.

Dans ce cadre, le bibliothécaire :

- Contribue à l'utilité sociale de la profession
- Exerce son métier sans laisser interférer ses intérêts ou ses opinions personnelles
- Développe son savoir professionnel, se forme et forme afin de maintenir un haut degré de compétence
- Visite des bibliothèques, rencontre des collègues, y compris à l'étranger
- S'implique dans la vie professionnelle en étant membre d'associations professionnelles, participe à des congrès (nationaux et internationaux) et en rend compte
- Publie et transmet, fait avancer la réflexion autour du métier en participant à des publications, à des colloques et journées d'étude
- Encourage la coopération, la mutualisation d'outils, l'appartenance à un réseau de coopération et de partage des savoirs
- Recherche l'amélioration des services par l'innovation
- Milite activement pour le recrutement et la promotion de personnel qualifié
- Elargit les publics
- S'implique professionnellement et intègre son établissement dans la vie de la cité

## **Annexe n°4 : Manifeste de l'Unesco**

### **Une porte ouverte à la connaissance**

La liberté, la prospérité et le développement de la société et des individus sont des valeurs humaines fondamentales. Elles ne peuvent s'acquérir que dans la mesure où les citoyens sont en possession des informations qui leur permettent d'exercer leurs droits démocratiques et de jouer un rôle actif dans la société. Une participation créatrice et le développement de la démocratie dépendent aussi bien d'une éducation satisfaisante que d'un accès libre et illimité à la connaissance, la pensée, la culture et l'information.

La bibliothèque publique, porte locale d'accès à la connaissance, remplit les conditions fondamentales nécessaires à l'apprentissage à tous les âges de la vie, à la prise de décision en toute indépendance et au développement culturel des individus et des groupes sociaux.

Ce Manifeste proclame la confiance que place l'UNESCO dans la bibliothèque publique en tant que force vive au service de l'éducation, de la culture et de l'information, et en tant qu'instrument essentiel du développement de la paix et du progrès spirituel par son action sur l'esprit des hommes et des femmes.

En conséquence, l'UNESCO encourage les autorités locales et nationales à s'engager activement à développer les bibliothèques publiques et à leur apporter le soutien nécessaire.

### **La bibliothèque publique**

La bibliothèque publique est le centre local d'information qui met facilement à la disposition de ses usagers les connaissances et les informations de toute sorte.

Les services de bibliothèque publique sont accessibles à tous, sans distinction d'âge, de race, de sexe, de religion, de nationalité, de langue ou de statut social. Des services et des documents spécifiques doivent être mis à la disposition des utilisateurs qui ne peuvent pas, pour quelque raison que ce soit, faire appel aux services ou documents courants, par exemple, les minorités linguistiques, les personnes handicapées, hospitalisées ou emprisonnées.

Toute personne, quel que soit son âge, doit avoir accès à une documentation adaptée à ses besoins. Les collections et les services doivent faire appel à tous les types de supports et à toutes les technologies modernes, de même qu'à la documentation traditionnelle. Il est essentiel qu'ils soient d'excellente qualité,

répondant aux conditions et besoins locaux. Les collections doivent refléter les tendances contemporaines et l'évolution de la société de même que la mémoire de l'humanité et des produits de son imagination.

Les collections et les services doivent être exempts de toute forme de censure, idéologique, politique ou religieuse, ou de pressions commerciales.

### **Missions de la bibliothèque publique**

Il faut tenir compte des missions-clés de la bibliothèque publique relatives à l'information, l'alphabétisation, l'éducation et la culture, qui sont les suivantes :

1. créer et renforcer l'habitude de lire chez les enfants dès leur plus jeune âge;
2. soutenir à la fois l'auto-formation ainsi que l'enseignement conventionnel à tous les niveaux ;
3. fournir à chaque personne les moyens d'évoluer de manière créative ;
4. stimuler l'imagination et la créativité des enfants et des jeunes ;
5. développer le sens du patrimoine culturel, le goût des arts, des réalisations et des innovations scientifiques ;
6. assurer l'accès aux différentes formes d'expression culturelle des arts du spectacle ;
7. développer le dialogue inter-culturel et favoriser la diversité culturelle ;
8. soutenir la tradition orale ;
9. assurer l'accès des citoyens aux informations de toutes catégories issues des collectivités locales ;
10. fournir aux entreprises locales, aux associations et aux groupes d'intérêt les services d'information adéquats ;
11. faciliter le développement des compétences de base pour utiliser l'information et l'informatique ;
12. soutenir les activités et les programmes d'alphabétisation en faveur de toutes les classes d'âge, y participer, et mettre en oeuvre de telles activités, si nécessaire.

### **Financement, législation et réseaux**

En principe, la bibliothèque publique doit être gratuite.

La bibliothèque publique relève de la responsabilité des autorités locales et nationales. Elle doit être soutenue par des textes législatifs spécifiques et financée

par les autorités nationales et locales. Elle doit être un élément essentiel de toute stratégie à long terme en matière de culture, d'information, d'alphabétisation et d'éducation.

Afin d'assurer une coordination et une coopération des bibliothèques à l'échelle nationale, des textes législatifs et des plans stratégiques devraient également définir et promouvoir un réseau national de bibliothèques selon des normes de service appropriées.

Le réseau de bibliothèques publiques doit être élaboré en tenant compte des bibliothèques nationales, régionales, de recherche et spécialisées, ainsi que des bibliothèques scolaires, collégiales ou universitaires.

### **Fonctionnement et gestion**

Une politique clairement formulée doit définir les objectifs, les priorités et les services selon les besoins des communautés locales. La bibliothèque publique doit être organisée efficacement et des normes professionnelles de fonctionnement doivent être maintenues.

La coopération avec les partenaires concernés - par exemple, des groupes d'utilisateurs et d'autres professionnels à l'échelon local, régional, national, de même qu'au niveau international - doit être assurée.

Les services doivent être physiquement accessibles à tous les membres de la communauté. Ceci suppose que les bâtiments de la bibliothèque soient bien situés, que celle-ci offre de bonnes conditions de lecture et d'étude, de même que les technologies adéquates et des heures d'ouverture convenant à tous les usagers. Ceci implique également des services destinés à ceux qui sont dans l'impossibilité de se rendre à la bibliothèque.

Les services de bibliothèque doivent être adaptés aux différents besoins des communautés vivant en zone rurale et urbaine.

Le bibliothécaire est un intermédiaire actif entre les usagers et les ressources offertes. La formation professionnelle et continue du bibliothécaire est indispensable pour assurer des services adéquats.

Des programme d'assistance et de formation des utilisateurs doivent être fournis pour les aider à bénéficier de toutes les ressources.

## Annexe n°5 : Charte des bibliothécaires

Considérant que :

- professionnalisme et volontariat ne s'opposent pas en matière de bibliothèques, mais s'appuient l'un sur l'autre ;
  - les volontaires sont indispensables au fonctionnement d'un service de lecture publique dans les petites communes et que ce volontariat implique l'acceptation de contraintes qui doivent avoir leur contrepartie ;
  - les professionnels sont indispensables dès que la population de la commune ou du groupement de communes responsable de la bibliothèque atteint 2000 habitants et qu'ils assurent l'assistance technique dont ont besoin les volontaires ;
- le Conseil supérieur des bibliothèques a adopté la présente Charte du bibliothécaire volontaire auprès des bibliothèques départementales de prêt

### Article premier

Le bibliothécaire volontaire affirme son engagement personnel auprès de la collectivité, au sein d'un service public de lecture dont il reconnaît les contraintes et assume les responsabilités.

### Article 2

Le bibliothécaire volontaire propose son temps et sa compétence au service de la collectivité, et reconnaît que l'autorité publique s'exerce sur son activité volontaire. L'autorité publique reconnaît le bibliothécaire volontaire comme concourant au service public.

### Article 3

Le bibliothécaire volontaire collabore avec les bibliothécaires professionnels, dans un esprit de complémentarité au service des usagers actuels, potentiels et futurs de la bibliothèque. Il accepte d'être encadré par ces professionnels. Il a droit à recevoir les responsabilités correspondant à ses compétences.

### Article 4

La formation professionnelle est un droit et un devoir du bibliothécaire volontaire. Des formations doivent être proposées sous les formes les plus appropriées au bibliothécaire volontaire, qui a soin de parfaire sa nécessaire formation initiale par une formation continue.

### Article 5

Le bibliothécaire volontaire a le droit à des conditions de travail correctes, tant en matière de moyens que de sécurité.

### Article 6

Le bibliothécaire volontaire offre son engagement sans contrepartie de rémunération.

### Article 7

Toutefois, il a droit à entière indemnisation pour toutes les dépenses engagées dans le cadre de son activité volontaire, et notamment sa formation, ses frais de déplacement et, le cas échéant, ses frais d'assurance.

### Article 8

Le bibliothécaire volontaire est responsable des biens qui lui sont confiés, et du service dont il a la charge. Il a droit à toute la protection publique contre les risques encourus au cours de son activité volontaire.

### Article 9

Le bibliothécaire volontaire accepte de s'engager pour une durée et une régularité déterminées, en accord avec l'autorité publique. Il ne saurait être écarté sans motif grave ou nécessité de service et sans concertation préalable.

## Annexe n°6 : Exemple de structure XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<biblio>
  <livre>
    <!-- Élément enfant titre -->
    <titre>Les Misérables</titre>
    <auteur>Victor Hugo</auteur>
    <nb_tomes>3</nb_tomes>
  </livre>
  <livre>
    <titre>L'Assommoir</titre>
    <auteur>Émile Zola</auteur>
    <couverture couleur="rouge" />
  </livre>
  <livre lang="en">
    <titre>David Copperfield</titre>
    <auteur>Charles Dickens</auteur>
    <nb_tomes>3</nb_tomes>
  </livre>
</biblio>
```



## Annexe n°7 : Extrait de RDFS

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">

  <owl:Ontology rdf:about="http://www.w3.org/2000/01/rdf-schema#" />

  <rdfs:Class rdf:about="http://www.w3.org/2000/01/rdf-schema#Resource">
    <rdfs:isDefinedBy rdf:resource="http://www.w3.org/2000/01/rdf-schema#" />
    <rdfs:label>Resource</rdfs:label>
    <rdfs:comment>The class resource, everything.</rdfs:comment>
  </rdfs:Class>

  <rdf:Property rdf:about="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">
    <rdfs:isDefinedBy rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#" />
    <rdfs:label>type</rdfs:label>
    <rdfs:comment>The subject is an instance of a class.</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class" />
    <rdfs:domain rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Resource" />
  </rdf:Property>
```

## Annexe n°8 : Les 15 éléments du Dublin Core

Élément	Élément (anglais)	Commentaire
1. Titre (métadonnée)	Title	Titre principal du <a href="#">document</a>
2. Créateur (métadonnée)	Creator	Nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document
3. Sujet (métadonnée) ou mots clés	Subject	Mots-clefs, phrases de résumé, ou codes de classement
4. Description (métadonnée)	Description	Résumé, table des matières, ou texte libre. Raffinements : <a href="#">table des matières</a> , résumé
5. Éditeur	Publisher	Nom de la personne, de l'organisation ou du service à l'origine de la publication du document
6. Contributeur	Contributor	Nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document. Chaque contributeur fait l'objet d'un élément Contributor séparé
7. Date (métadonnée)	Date	Date d'un événement dans le cycle de vie du document
8. Type de ressource	Type	Genre du contenu
9. Format	Format	Type <a href="#">MIME</a> , ou format physique du document
10. Identifiant de la ressource	Identifier	Identificateur non ambigu : il est recommandé d'utiliser un système de référencement précis, afin que l'identifiant soit unique au sein du site, par exemple les <a href="#">URI</a> ou les numéros <a href="#">ISBN</a> . Raffinement : <a href="#">Is Available At</a>
11. Source	Source	Ressource dont dérive le document : le document peut découler en totalité ou en partie de la ressource en question. Il est recommandé d'utiliser une dénomination formelle des ressources, par exemple leur <a href="#">URI</a>
12. Langue (métadonnée)	Language	
13. Relation (métadonnée)	Relation	Lien avec d'autres ressources. De nombreux raffinements permettent d'établir des liens précis, par exemple de version, de chapitres, de standard, etc.
14. Couverture (métadonnée)	Coverage	Couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle
15. Droits (métadonnée)	Rights	Droits de <a href="#">propriété intellectuelle</a> , <a href="#">Copyright</a> , droits de propriété divers