



La base de connaissances élément central de la gestion des collections électroniques des bibliothèques universitaires

Alexandre Faure

► To cite this version:

Alexandre Faure. La base de connaissances élément central de la gestion des collections électroniques des bibliothèques universitaires. domain_shs.info.docu. 2013. mem_00945638

HAL Id: mem_00945638

https://memsic.ccsd.cnrs.fr/mem_00945638

Submitted on 12 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

Ecole Management et Société

Département Culture Information Technique et Société (CITS)

INTD

MEMOIRE pour obtenir le

Titre professionnel "Chef de projet en ingénierie documentaire" INTD

RNCP niveau I

Présenté et soutenu par

Alexandre Faure

le 01 juillet 2013

La base de connaissances élément central de la gestion des collections électroniques des bibliothèques universitaires

Jury : Ghislaine Chartron, Professeur, Directrice du département Culture, Information, Technique et Société, de l'INTD

Isabelle Boudet, Conservateur général des bibliothèques, Service commun de la documentation, Université Paris Dauphine

A Agnès, sans son soutien, ce mémoire n'aurait jamais pu être écrit.

Remerciements

Je tiens à remercier ici l'ensemble des équipes pédagogiques et administratives de l'INTD ainsi que mes responsables à la bibliothèque de Dauphine pour m'avoir permis de reprendre mes études.

Notice

FAURE Alexandre. La base de connaissances élément central de la gestion des collections électroniques des bibliothèques universitaires. 59 p. Mémoire, sciences de l'information, INTD, 2013.

Avec le développement des résolveurs de liens, des ERMS et dernièrement des outils de découverte, les bases de connaissances ou KB ont pris une place de plus en plus stratégique au sein des systèmes d'information documentaire des bibliothèques universitaires. En effet, ces dispositifs sont essentiels à la mise en place des accès aux documents numériques et à l'enrichissement des catalogues. En s'appuyant sur le cas de la bibliothèque Paris Dauphine, ce mémoire étudie la place des bases de connaissances dans de tels dispositifs. Il s'intéresse ensuite aux projets actuels de création de bases de connaissances nationales et supranationales.

RESSOURCE ELECTRONIQUE ; REVUE ELECTRONIQUE ; ERMS ; OPENURL ; RESOLVEUR DE LIENS ; BIBLIOTHEQUE UNIVERSITAIRE ; BASE DE CONNAISSANCES.

With the development of link resolvers, ERMS and recent discovery tools, knowledge bases or KB took an ever more strategic position in the academic libraries' documentary information systems. As a matter of fact, these systems are essential to the setting of digital documents and catalog improvements access. Relying on the case of the library Paris Dauphine, this study examines the role of knowledge base in the documentary information system. It then focuses on current projects to create databases for national and supranational knowledge.

ELECTRONIC RESSOURCES ; ERMS ; OPENURL ; LINKS RESOLVER . ACADEMIC LIBRARY ; KNOWLEDGE BASE.

Table des matières

Introduction	10
Première partie De nouveaux outils pour répondre aux enjeux de la documentation électronique.....	12
1 Evolution du paysage et des pratiques documentaires	13
1.1 Le poids croissant de la documentation électronique sur les collections des bibliothèques universitaires	13
1.1.1 Qu'est-ce qu'un document électronique ?	13
1.1.2 Le marché de l'édition scientifique électronique	14
1.1.3 Les modèles économiques	15
1.1.4 Impact sur les collections et les budgets des bibliothèques	16
1.2 De nouveaux usages en matière de recherche d'informations	19
1.2.1 Les portails des bibliothèques concurrencés par les moteurs de recherche	19
1.2.2 De nouvelles attentes des lecteurs vis-à-vis des catalogues de bibliothèque	19
2 De nouveaux outils pour de nouveaux besoins	21
2.1 Les ERMS : Gérer la diversité et la complexité.....	21
2.1.1 La gestion de la documentation électronique en bibliothèque	21
2.1.1.1 Un workflow complexe	22
2.1.1.2 Un flux d'informations de gestion important	23
2.1.2 Les ERMS comme outils de gestion du workflow	23
2.1.2.1 Naissance des ERMS	23
2.1.2.2 Standards et protocoles	24
2.2 Offrir un environnement de recherche unifié : de la métarecherche aux outils de découverte	25
2.2.1 Offrir un accès unifié aux collections	25
2.2.2 Qu'est-ce que la recherche fédérée ?	25
2.2.3 Recherche fédérée contre recherche unifiée.....	26
2.2.4 Protocoles d'interrogation et types d'outils existants	26
2.2.4.1 La recherche croisée	26
2.2.4.1.1 Les protocoles Z39.50 et SRU/SRW	26
2.2.4.1.2 L'implantation rapide des métamoteurs de recherche dans les bibliothèques.....	27
2.2.4.1.3 Critique de la recherche fédérée dans son mode synchrone.....	27
2.2.4.2 Protocole asynchrone et outils de découverte	28
2.2.4.2.1 Le protocole asynchrone	28

2.2.4.2.2	Les outils de découverte	28
2.3	Open-URL et résolveurs de liens: gérer la localisation des ressources	29
2.3.1	Naissance de l'Open-Url : interconnecter les publications électroniques	29
2.3.1.1	La stratégie des liens fixes.....	29
2.3.1.2	Les liens dynamiques	29
2.3.1.3	Le DOI	30
2.3.2	La question de la copie appropriée	30
2.3.3	Open-URL et résolveurs de liens	31
	Les bases de connaissances au cœur de ces dispositifs.....	34
1	Qu'est-ce qu'une base de connaissances ?	35
1.1	Définition.....	35
1.2	Acteurs du secteur et processus de constitution.....	36
1.2.1	Circuit des métadonnées	36
1.2.2	Les fournisseurs de contenus.....	37
1.2.3	Les fournisseurs de résolveur de liens	37
1.2.4	Les bibliothèques.....	38
2	La gestion d'une base de connaissances dans un établissement universitaire, le cas Dauphine:	39
2.1	Présentation du contexte	39
2.1.1	Les ressources électroniques à la bibliothèque Dauphine.....	39
2.1.2	Le système d'information documentaire de la bibliothèque	40
2.2	La base de connaissances SFX au cœur du Système d'information documentaire.....	42
2.2.1	Hétérogénéité et qualité des métadonnées	42
2.2.1.1	La question de la qualité des métadonnées liée aux accès aux ressources	42
2.2.1.1.1	Les problèmes d'identifiants incohérents.....	42
2.2.1.1.2	Les couvertures temporelles incorrectes	42
2.2.1.1.3	Les formats de dates incohérents.....	43
2.2.1.1.4	Les problèmes de périodes d'embargo.....	43
2.2.1.1.5	Les données de contenus obsolètes	43
2.2.1.1.6	Le manque de personnalisation.....	43
2.2.1.2	L'articulation avec les métadonnées traditionnelles	44
2.2.2	Maîtriser sa collection.....	44
2.2.2.1	Gérer la volumétrie	44
2.2.2.2	L'impossibilité de couvrir la totalité de la collection dans la base de connaissances	45
2.3	Proposition de solutions au niveau local.....	45
2.3.1	Améliorer l'homogénéisation des métadonnées exposées dans le catalogue	

2.3.2	Gérer les ressources électroniques dans le SIGB ?	46
3	Des projets nationaux et internationaux	47
3.1	Des projets nationaux et internationaux visant à améliorer la qualité des métadonnées.....	47
3.1.1	Le projet KBART.....	47
3.1.2	Qualinca	48
3.2	Des projets de bases de connaissances communautaires	49
3.2.1	Le résolveur allemand EZB	49
3.2.2	Le projet Knowledge Base+.....	49
3.2.3	Le projet GOKb (global open knowledgebase, base de connaissances globale)49	
3.2.4	Les projets français	50
3.2.4.1	Le hub de métadonnées	51
3.2.4.2	Une base de connaissances nationale	52
	Conclusion	53

Liste des tableaux

Tableau 1 : Evolution du coût des abonnements de périodiques de 1980 à 1998	17
Tableau 2 : Les sources utilisées par les doctorants de Bretagne	19
Tableau 3 : Répartition des dépenses documentaires de la Bibliothèque Paris-Dauphine ...	40

Liste des figures

Figure 1 : Coûts des monographies et des périodiques dans les établissements publics membres de l'ARL (1986-2003).....	17
Figure 2 : Évolution du nombre de titres de périodiques électroniques disponibles dans les BU et BIU de 1999 à 2007	18
Figure 3 : Workflow de la gestion des ressources électroniques [4, BARTHET, 2010]	22
Figure 4 : Problème de la copie appropriée	31
Figure 5 : Fonctionnement de l'Open-url	33
Figure 6 : Le workflow de la constitution des bases de connaissances	36
Figure 7 : Schéma du système d'information de la bibliothèque Dauphine : flux des données documentaires	41
Figure 8 : Architecture du futur Hub de métadonnées.....	52

Introduction

Depuis quelques dizaines d'années, face au développement de l'édition numérique, les bibliothèques universitaires ont vu la documentation électronique prendre une place de plus en plus importante au sein de leurs collections. En parallèle, le développement d'Internet a introduit de nouvelles attentes de la part d'une génération d'étudiantss qui a grandi avec la toile et dont les pratiques informationnelles sont largement captées par les grands moteurs de recherche généralistes.

Face à ce double défi, les bibliothèques ont dû mettre en place, au côté du système intégré de gestion de bibliothèques (SIGB), de nouveaux outils afin de gérer et de valoriser la documentation électronique. Ainsi, pour répondre aux problèmes posés par la gestion du cycle de vie des documents électroniques, ont-elles conçu les Electronic Resource Management System (ERMS). Afin de proposer un point d'accès unifié à la totalité des ressources, elles se sont dotées de moteurs de recherche fédérée qui ont évolué ces dernières années en de véritables outils de découverte s'appuyant sur un index de métadonnées centralisé. Enfin, pour permettre à leur public de naviguer entre des sources hétérogènes et faire tomber les barrières liées aux droits d'accès elles ont implémenté des résolveurs de lien.

Dans cet écosystème technique, les bases de connaissances ou KB ont pris une place de plus en plus stratégique. Ce dispositif initialement conçu pour supporter la fonction de localisation des résolveurs de liens est devenu un élément central dans la gestion des ressources électroniques. Il permet d'exposer les notices descriptives des documents électroniques sur le catalogue et d'améliorer la gestion des abonnements au sein de l'ERMS.

La constitution et la gestion de ces entrepôts de métadonnées étant partagées entre fournisseurs de systèmes informatiques documentaires, fournisseurs de contenus et bibliothèques, nous étudierons leur place dans le système informatique documentaire ainsi que l'impact de ces ensembles de métadonnées hétérogènes sur les outils de recherche. Nous nous appuierons sur l'exemple de la bibliothèque de l'université Paris-Dauphine pour étudier les différents problèmes posés par la gestion des métadonnées dans ces bases de connaissances. Ensuite, nous verrons dans quelle mesure des solutions peuvent être proposées dans le cadre de projets nationaux ou internationaux.

Première partie
De nouveaux outils pour répondre
aux enjeux de la documentation
électronique

1 Evolution du paysage et des pratiques documentaires

Le mouvement de l'édition scientifique vers le numérique est antérieur à la création du web. Quelques projets pilotes dans le domaine de la fourniture d'un accès distant au savoir ont été initiés dès les années 1970 (le projet EIES (Electronic Information Exchange System) mené aux États-Unis de 1978 à 1980, le projet anglais BLEND (Birmingham and Loughborough Network Development Project) de 1980 à 1984, le projet ADONIS piloté par les éditeurs scientifiques, puis le projet Tulip d'Elsevier à la fin des années 1990 et enfin, le projet français « Journal revue électronique de Science de l'information » piloté par la MIDIST) [16, CHARTRON, 2007]. Mais c'est l'essor d'Internet qui a contribué au développement de l'offre et de l'usage de la documentation électronique. Il a profondément reconfiguré le paysage documentaire de l'information scientifique et technique et les bibliothèques, depuis les années 1980, ont été amenées à gérer une offre documentaire électronique de plus en plus dense et complexe.

1.1 Le poids croissant de la documentation électronique sur les collections des bibliothèques universitaires

1.1.1 Qu'est-ce qu'un document électronique ?

Le langage courant emploie indifféremment les termes de documents électroniques ou numériques pour désigner les documents fixés sur support électronique nécessitant l'usage d'un matériel pour être consulté [1, ADBS]. Bertrand Calenge dans son ouvrage *Bibliothèque et politique documentaire à l'heure de l'internet* établit une distinction entre les deux termes. Selon sa définition, les documents électroniques sont « les documents digitaux disponibles sur des serveurs extérieurs et mis à disposition sur Internet par d'autres organismes ». À l'inverse, il définit les documents numériques comme « des documents digitaux dont la bibliothèque s'est assurée la pleine propriété, en les numérisant ou en les acquérant » [11, CALENGE, 2008, p. 22-23].

Dans ce travail, nous nous intéresserons aux « documents électroniques » tels qu'ils sont définis par l'auteur. Il nous semble que ce sont ces supports qui posent de véritables enjeux aux bibliothèques universitaires de par le caractère aléatoire de leur disponibilité, les possibilités limitées de manipulation des fichiers et les limites appliquées à l'usage des documents.

Cette dénomination de documents électroniques regroupe une grande variété de types de documents : banques bibliographiques, presse, rapports, revues, collections rétrospectives, banques de données statistiques, banques de données sectorielles et d'information sur les entreprises, contenus divers en accès libre et, plus récemment, livres électroniques.

1.1.2 Le marché de l'édition scientifique électronique

Le nombre de revues scientifiques électroniques disponibles s'est considérablement accru durant la dernière décennie. Ce type de document est devenu l'un des pivots majeurs de l'offre de documentation en bibliothèque de par leur rôle central dans le processus de production et de diffusion de la recherche.

Le nombre de revues académiques à comité de lecture était début 2009 d'environ 25 400 publiant au total environ 1,5 million d'articles par an. En 2008, 96% des titres en STM et 87% en LSHS étaient disponibles au format électronique [13, CARBONE, 2010].

À côté de l'édition commerciale et des sociétés savantes, le mouvement pour le libre accès à la recherche (Open Access-OA), popularisé par l'Initiative de Budapest a contribué au développement des archives institutionnelles et des revues en libre accès. Ces revues en libre accès connaissent une véritable expansion. En 2009 le Directory of Open Access Journals¹ comptait 4 360 titres de revues en libre accès. En 2013, il en compte 9 582. Dans ce mouvement d'ouverture des publications de recherche, les bibliothèques universitaires sont elles-mêmes devenues diffuseurs de contenu en prenant en charge le développement et la gestion des bases d'archives institutionnelles permettant aux chercheurs d'auto-archiver et de diffuser leurs publications.

Quels que soient le mode d'édition et le support, l'offre documentaire en IST, particulièrement des périodiques, devient de plus en plus importante, car poussée par la production documentaire croissante de la recherche et encouragée par le mode d'évaluation des chercheurs et des établissements largement fondé sur le nombre de publications de ces derniers. [30, KLEE, 2006]

Le secteur du livre a abordé plus tardivement et moins rapidement le virage vers l'électronique. Le mouvement a été amorcé à l'initiative des grands éditeurs de revues scientifiques qui ont mis en œuvre des programmes de numérisation. Fin 2009, Springer propose 30 000 titres d'e-books avec un rythme d'augmentation de 5 000 par an ; Elsevier, environ 7 000. Les autres collections importantes sont : IEEE Conference Proceedings (10 600) ; SPIE Conference Proceedings (6 500) ; Safari Textbooks (6 000), Source OECD (3 200). Mais l'offre ne répond pas actuellement à la demande en manuels en ligne pour les étudiants.

Le mouvement vers l'électronique s'est accompagné d'une forte concentration dans l'édition scientifique mondiale. Les 5 grands éditeurs historiques de l'édition scientifique (Reed Elsevier, Thomson Reuters, Wolters Kluwer, Cengage, Wiley et Blackwel) réalisent une part importante de leur chiffre d'affaire sur l'électronique. Le classement 2012 de l'édition mondiale [44, Livres Hebdo, 2012] fournit les informations suivantes :

- Reed Elsevier avec un chiffre d'affaires (CA) de 4 395 millions d'€ est au 2e rang du classement de l'édition mondiale. 63% de l'activité de sa filiale Elsevier (sciences et

¹ <http://www.doaj.org/>

médecine) et 75 % de celle de sa filiale Lexis Nexis (Droit) proviennent du numérique.

- Thomson Reuters : Le numéro 3 mondial réalise 90% de son CA (4 180 millions d'€) sur les produits et services électroniques.
- Wolters Kluwer : Avec un CA de 3 354 millions d'€, le numéro 4 mondial réalise 75% de son activité sur le numérique. Les revenus de cette branche d'activité ont crû de 8% entre 2010 et 2011 alors que l'imprimé a reculé de 5%.
- Cengage : Le numéro 11 mondial réalise 38% de son chiffre d'affaires sur le numérique.
- Wiley : Ce groupe, principal éditeur des sociétés savantes, est classé numéro 12 mondial avec un CA de 1340 millions d'€. Son secteur universitaire a cru de 7% grâce notamment à l'essor du livre électronique.
- Springer : Ce groupe, 19e mondial (CA de 875 M€), mais 2e éditeur mondial de revues scientifiques, techniques et médicales, produit sous 55 marques 2000 revues, 7000 nouveaux ouvrages et des bases de données.

Ces grands éditeurs académiques et juridiques, qui totalisent plus de 30 Md€ de chiffre d'affaires, concentrent une part importante du marché mondial de l'IST, notamment dans le secteur des revues. On dénombre environ 2 000 éditeurs de revues. Les éditeurs anglophones (657) publient en tout 11 550 titres : 477 d'entre eux sont des sociétés à but non lucratif et publient 2 334 titres. En nombre d'articles cités au Journal of Citation Reports, les éditeurs commerciaux pèsent pour 64%, les sociétés savantes 30%, les presses d'université 4% et les autres éditeurs 2%. 95% des éditeurs de revues ne publient qu'un ou deux titres, tandis qu'à l'opposé les 100 premiers concentrent 67% du total des revues, et les 10 premiers 35%. Quatre éditeurs (Elsevier, Springer, Taylor & Francis et Wiley-Blackwell) publient chacun plus de 1000 titres [13, CARBONE, 2010].

Un autre acteur est à prendre en compte dans le secteur, il s'agit des agrégateurs. Ces sociétés qui ont très tôt contribué à structurer l'offre de presse (Factiva, Europresse, Business Source Premier, etc.), se sont peu à peu imposées sur le marché de l'édition scientifique et technique. Leur rôle est de négocier avec les éditeurs les conditions d'exploitation du contenu afin de constituer une offre regroupée sous la forme de bouquet de titres issus de différents éditeurs.

1.1.3 Les modèles économiques

Dans le cadre des revues électroniques, le modèle tarifaire est largement fondé sur celui du papier. Les éditeurs proposent des accès à la version électronique d'un ensemble important de revues, voire même du catalogue complet. Qu'elles soient couplées à un abonnement papier (E-Fee) [15, CAVALIER, 2006] ou tout électronique (E-only), ces offres restent très liées aux abonnements papier. Afin de se prémunir de la concurrence des accès électroniques sur les titres papier, le montant de l'abonnement est calculé sur la base du

chiffre d'affaires réalisé sur les abonnements traditionnels de chaque institution majoré d'un pourcentage pour l'accès à la version électronique des titres non souscrits. En contrepartie, la bibliothèque s'engage à maintenir le chiffre d'affaires réalisé avec l'éditeur. Les désabonnements ne sont ainsi tolérés qu'à la marge des contrats et ne peuvent pas dépasser 1% du chiffre d'affaires réalisé avec l'éditeur. Ces modèles baptisés les « big deals » [25, FRAZIER, 2005] ont permis d'accroître très rapidement l'offre documentaire. Les bibliothèques sont passées rapidement de quelques centaines de titres papier à deux ou trois mille titres disponibles en ligne. Ils sont toutefois fortement décriés et accusés d'enfermer les établissements dans une spirale inflationniste (même si l'augmentation annuelle des tarifs d'abonnement est limitée contractuellement à entre 3 et 6 %) dont il est difficile de sortir puisque l'arrêt d'un abonnement entraîne l'indisponibilité de certains titres [20, COLCANAP, 2009].

Concernant l'abonnement aux bases de données, qu'il s'agisse de bases bibliographiques ou de bases d'agrégateur, le montant de l'abonnement est généralement calculé sur la base de la population desservie et plus rarement sur les usages.

Avec l'émergence du marché des livres électroniques de nouveaux modèles d'acquisition sont apparus. À côté du traditionnel modèle d'abonnement sur titres ou sur bouquets, les éditeurs et agrégateurs proposent aux bibliothèques la possibilité d'acheter des titres à l'unité. Le prix du document dépend alors du nombre d'accès simultanés autorisés. Le fournisseur du document s'engage à maintenir un accès pérenne au fichier.

En parallèle de ce système très proche du modèle traditionnel d'acquisition des livres papier, un nouveau mode d'acquisition est en train d'émerger. Il s'agit du Patron Driven Acquisition. Son principe repose sur l'accès à un catalogue de livres susceptibles d'être acquis en fonction de l'usage constaté. La bibliothèque peut par exemple choisir d'acheter le livre suite à un signalement par un lecteur ou automatiquement après un certain nombre d'utilisations.

1.1.4 Impact sur les collections et les budgets des bibliothèques

Confrontées à la progression rapide des ressources électroniques, les bibliothèques universitaires et de recherche ont dû suivre le mouvement et acquérir de plus en plus de revues au format électronique. Cet essor, conjugué aux effets pervers des modèles économiques en pratique dans le secteur, a entraîné l'accroissement des dépenses documentaires pour les établissements. Le rapport statistique de l'ARL (Association of Research Libraries) pour les années 2002-2003 montre qu'entre 1986 et 2003 le montant des dépenses d'acquisition de périodiques des bibliothèques de recherche américaines avait crû de 258%.

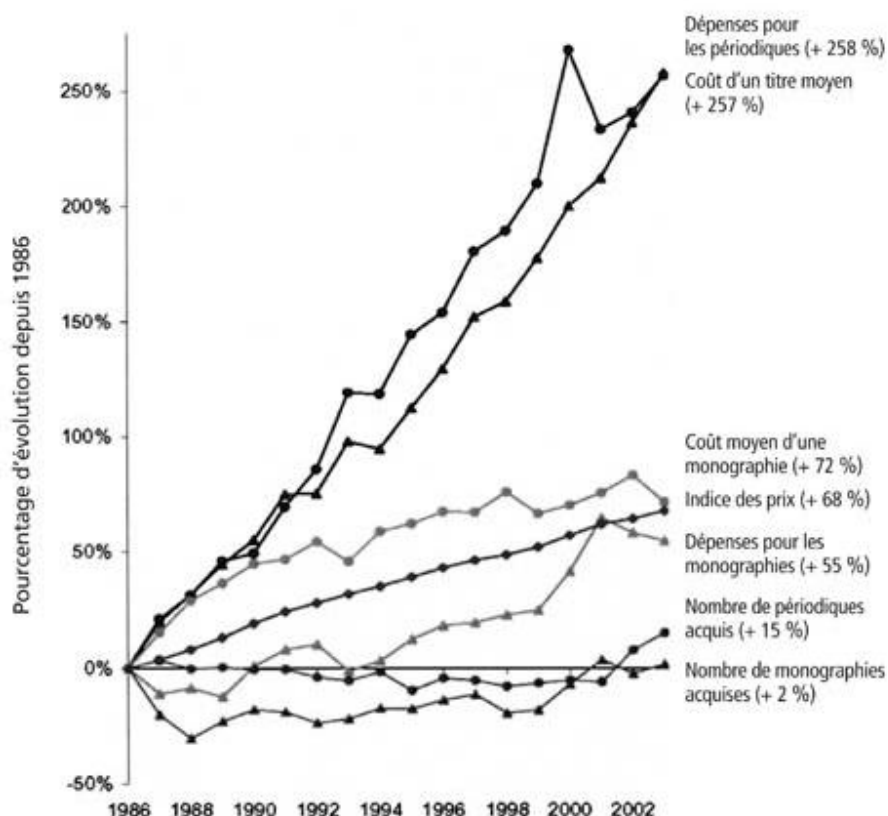


Figure 1 : Coûts des monographies et des périodiques dans les établissements publics membres de l'ARL (1986-2003)

En France, le constat est identique. Une enquête menée par l'Inspection générale des bibliothèques entre 1980 et 1998 auprès de 16 bibliothèques représentatives chacune d'un secteur disciplinaire (CADIST, BU de Paris et de province) a permis de constater les évolutions suivantes [13, CARBONE, 2010] :

Secteurs disciplinaires	Augmentation en %	
	Périodiques français	Périodiques étrangers
Droit-Economie-Gestion	+ 239,62%	+382,87%
Lettres-Sciences humaines	+233,33%	+229,86%
Médecine	+249,57%	+581,48%
Pharmacie	+290,17%	+583,50%
Sciences	222,58%	+439,25%

Tableau 1 : Evolution du coût des abonnements de périodiques de 1980 à 1998

Selon une enquête plus détaillée lancée par la Sous-direction des bibliothèques et de la documentation de la Direction de l'Enseignement Supérieur concernant l'exercice 2002, les universités ont dépensé 11,6M€ en 2002 pour les ressources électroniques.

Cette situation a incité les bibliothèques à se regrouper en consortia afin de mieux négocier les ressources. Cette initiative de structuration a donné naissance au consortium Couperin. Cette association fondée en 1999 et qui regroupe aujourd'hui plus de 200 membres (universités, grandes écoles et organismes de recherche) a pour vocation de négocier les tarifs et les conditions d'achat des ressources académiques électroniques pour ses membres et d'animer un réseau de compétence et d'expertise en matière de documentation électronique.

Grâce à cette entreprise de mutualisation et grâce au « big deals », l'offre documentaire électronique proposée aux usagers dans les bibliothèques de l'enseignement supérieur s'est accrue de façon considérable. Les collections de revues électroniques des bibliothèques sont passées d'une centaine de titres à plusieurs milliers [20, COLCANAP, 2009].

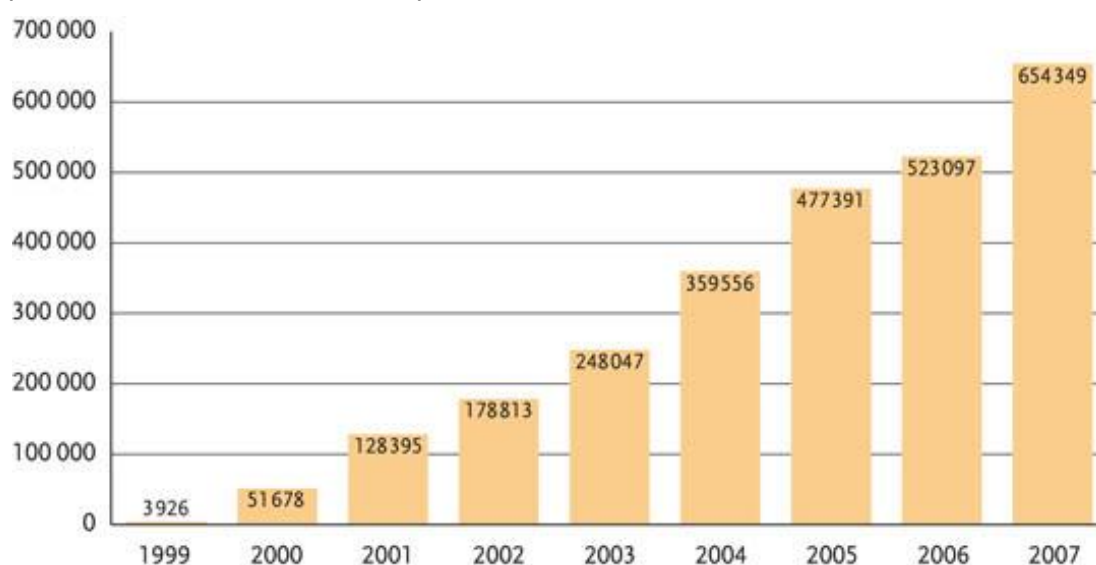


Figure 2 : Évolution du nombre de titres de périodiques électroniques disponibles dans les BU et BIU de 1999 à 2007

En 2012, les différents établissements avaient, en moyenne, accès à 21 bases de données documentaires et à plus de 1 500 titres de périodiques en ligne. Cette moyenne cache une grande hétérogénéité des collections des différents établissements puisque 12 universités offraient l'accès à plus de 3 000 périodiques en ligne et 24 universités en proposaient entre 2 000 et 3 000.

Quoi qu'il en soit, le total des licences d'accès à des périodiques électroniques (150 000) est désormais supérieur au nombre d'abonnements papier (125 000). De ce fait, on rencontre souvent des universités bénéficiant de 1500 périodiques en version papier, mais plus du double en licence d'accès aux périodiques en ligne.

1.2 De nouveaux usages en matière de recherche d'informations

Internet a profondément modifié les usages en matière de recherche d'informations. La possibilité d'accéder à tout moment et de n'importe quel lieu à l'information a modifié les attentes des usagers des bibliothèques. Dans la même dynamique, l'apparition puis l'évolution des moteurs de recherche grand public ont contraint les bibliothèques à réfléchir à leur mode de diffusion de l'information.

1.2.1 Les portails des bibliothèques concurrencés par les moteurs de recherche

Les étudiants actuels appartiennent à la génération Y aussi appelée « Digital natives ». Ces étudiants nés entre 1978 et 1994 ont grandi avec l'informatique et ont connu le développement des nouvelles technologies de l'information et de la communication. Pour eux, Internet est devenu un outil familier et le moteur de recherches le principal outil pour trouver de l'information. L'étude de l'OCLC *Perceptions of Libraries, 2010: Context and Community* publiée en 2010 montre que les étudiants américains utilisent uniquement les moteurs de recherche sur Internet (83% des étudiants interrogés) et Wikipedia (7%) pour débiter leurs recherches d'informations [22, OCLC, 2010].

Au niveau des doctorants, une enquête menée en 2008 auprès des écoles doctorales de Bretagne par l'URFIST de Rennes et le SCD de Bretagne montre que les moteurs de recherche restent largement plus utilisés que les ressources proposées par la bibliothèque [28, HENRIET, 2008].

Les sources les plus utilisées par les doctorants	Les sources rarement ou jamais utilisées
Moteurs de recherche : 96% Catalogue de bibliothèque : 65% Portails spécialisés : 53%	Archives ouvertes : 77% Catalogues étrangers : 74% SUDOC : 71% Bases de données (Pubmed, Francis, Pascal, ...) : 62 %

Tableau 2 : Les sources utilisées par les doctorants de Bretagne

1.2.2 De nouvelles attentes des lecteurs vis-à-vis des catalogues de bibliothèque

Habitué à la simplicité des interfaces des moteurs de recherche ainsi qu'à la puissance des algorithmes de recherche toujours plus performants, les usagers des bibliothèques exigent une expérience de recherche et de navigation simplifiée à partir des catalogues de bibliothèques.

Une étude conduite par l'OCLC en 2010 autour de la perception par les usagers des catalogues des bibliothèques a montré que plus que l'expérience de recherche et de

découverte, c'est l'accès au contenu désiré qui prime. De ce fait, des métadonnées appropriées, précises et fiables sont indispensables à la bonne transition entre les résultats de recherche et l'accès aux documents primaires. En effet, le lecteur exige une transition fluide entre la recherche et l'accès à la ressource, il veut connaître immédiatement la disponibilité de la ressource ainsi que les modalités d'accès.

En situation de recherche, les usagers réclament des informations qui leur permettent d'évaluer les résultats afin de décider de poursuivre ou non leur navigation jusqu'à l'obtention de la ressource voulue. Dans ce but, la présence des tables des matières et des résumés dans les résultats de recherche peut contribuer à l'amélioration de l'expérience de l'utilisateur.

Les usagers ont certaines attentes vis-à-vis des types de résultats que leur recherche produira. Ils s'attendent à ce que le catalogue de la bibliothèque leur renvoie des résultats pertinents. De ce fait, le catalogue doit proposer des fonctions de tri et de classement par pertinence qui rendent les résultats et leur ordre d'affichage évidents pour l'utilisateur.

Si la recherche simple calquée sur les moteurs de recherche reste la norme, les utilisateurs souhaitent tout de même avoir la possibilité de lancer des recherches avancées. Ils réclament des dispositifs permettant de les guider afin de ne pas avoir à parcourir des dizaines de pages pour trouver la ressource qui les intéresse. Ils perçoivent ainsi la restriction des résultats par facette comme une manière efficace de restreindre un nombre de résultats important [12, CALHOUN, 2010].

2 De nouveaux outils pour de nouveaux besoins

Avec le mouvement d'informatisation des bibliothèques entamé depuis le milieu des années 1960, le SIGB est devenu l'outil central de la gestion documentaire dans les bibliothèques universitaires. Mais l'entrée dans les collections de la documentation électronique, la diversité des supports et la masse des documents à traiter ont fait apparaître de nouveaux besoins en fonctionnalités que les systèmes traditionnels construits autour de la gestion de la documentation papier ne sont pas en mesure de prendre pleinement en charge. En parallèle, les nouvelles attentes des lecteurs induites par le développement des moteurs de recherche ont contraint les bibliothèques à se doter de nouveaux outils pour mettre en valeur leurs collections.

2.1 Les ERMS : Gérer la diversité et la complexité

Les ERMS (Electronic Resource Management System) ont été développés pour répondre au besoin de disposer d'un outil capable de gérer et de centraliser l'information liée à la gestion des ressources électroniques.

2.1.1 La gestion de la documentation électronique en bibliothèque

Le circuit traditionnel du traitement des documents dans une bibliothèque est un processus systématique, quasiment linéaire, d'enchaînement de tâches nécessaires à la mise en circulation des documents. Entre la veille et la sélection des documents, l'émission d'un bon de commande, la réception des exemplaires, le catalogage et la mise en rayon des documents, l'ensemble des tâches nécessaires au bon fonctionnement du processus sont entièrement gérées au travers du SIGB dont les différents modules ont été développés pour assurer ces fonctions. [8, BOTHMANN, 2008]. Le système peut ainsi alerter les bibliothécaires quand un exemplaire commandé est arrivé, présenter tous les exemplaires disponibles dans le catalogue ou encore informer le lecteur de la disponibilité d'un document.

En revanche, dans le cadre de la gestion de ressources immatérielles, si la sélection, la commande et la facturation peuvent être aisément traités dans le SIGB, ce dernier peut difficilement informer les personnels et les usagers de la non-disponibilité ou de la disponibilité d'un document commandé. De la même façon, sans objet tangible, il est très difficile d'envoyer le document au catalogage. Ces incapacités viennent du fait que la gestion du workflow dans un SIGB s'appuie sur un élément matériel, l'exemplaire ou le fascicule, qui change d'état en fonction des différentes étapes du processus de traitement.

2.1.1.1 Un workflow complexe

La gestion des ressources électroniques en bibliothèque est un processus complexe qui mobilise différentes compétences et requiert une interaction très forte entre différents services de la bibliothèque. Le schéma ci-dessous illustre cette complexité.

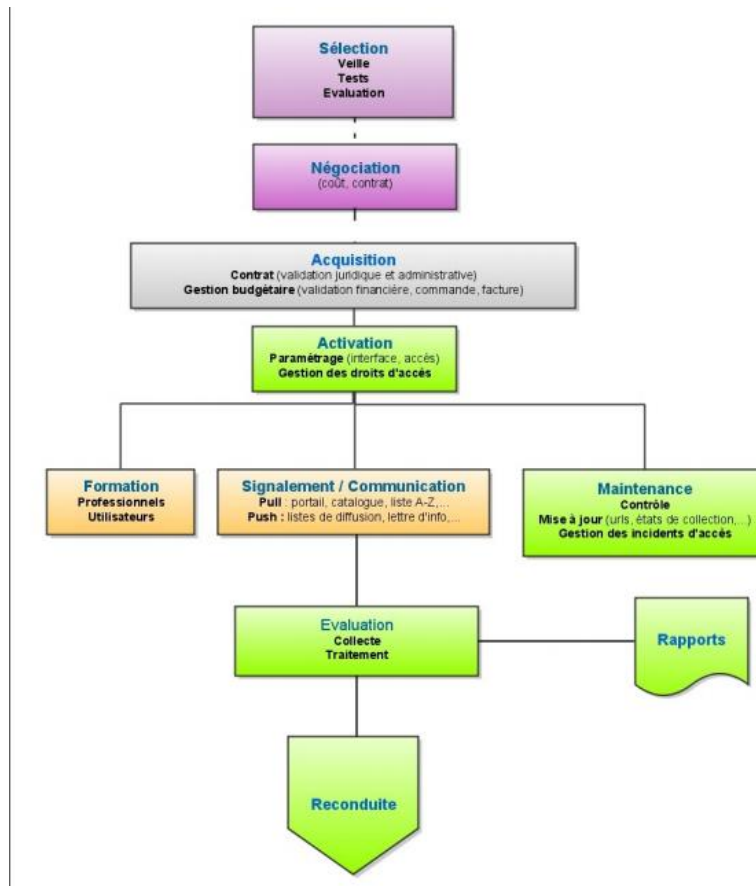


Figure 3 : Workflow de la gestion des ressources électroniques [4, BARTHET, 2010]

Une fois les ressources identifiées, l'évaluation de la ressource permet à l'acquéreur de déterminer le coût et la pertinence de l'acquisition de la ressource au sein des collections. Cette évaluation est généralement conduite à l'issue d'un test proposé à l'usage exclusif des bibliothécaires, d'une partie ou de la totalité des publics. Cette étape pose la question de la mise à disposition de la ressource dans un temps limité, de la limitation des accès à un groupe identifié et de la collecte des impressions des testeurs. Elle constitue une aide majeure au processus d'évaluation du document. Alors que dans le cadre de l'évaluation des ressources traditionnelles, l'examen repose sur des critères financiers et intellectuels, dans le cas des ressources électroniques, d'autres critères sont à prendre en compte : facilité d'accès aux contenus, couverture, fonctionnalités offertes par l'interface, qualité du support technique, méthode de tarification et disposition du contrat de licence, sans parler de l'intégration technique de la ressource dans le système d'information documentaire.

La phase d'acquisition est l'occasion de récolter un certain nombre d'informations qui seront utiles à la gestion du cycle de vie du document : types d'accès pour la mise à disposition, périmètre du bouquet, contact du support et période de reconduction du contrat. Ces informations doivent être accessibles aux divers acteurs de la chaîne de traitement.

Une étape importante dans la mise à disposition des documents électroniques est la phase de paramétrage. Contrairement au document papier, le document électronique, de par son support dématérialisé hébergé chez un fournisseur tiers ne nécessite pas de gestion matérielle. Il convient néanmoins de pouvoir déclarer et configurer le lien d'accès vers ce document. La nature mouvante des collections électroniques nécessite de mettre en place des processus de suivi et de maintenance de ces liens d'accès. En outre, le paramétrage et la configuration de ces accès doivent être en accord avec les termes du contrat de licence passé avec l'éditeur. Par exemple, il convient de mettre en place le dispositif technique nécessaire pour limiter la consultation du document au périmètre fixé par le contrat.

Enfin, afin de juger de l'opportunité de reconduire une ressource, il est nécessaire d'en connaître l'usage qu'il en ait fait. Il s'agit alors pour les bibliothécaires de récolter les données d'usages auprès du fournisseur, de les rassembler et de les consolider.

2.1.1.2 Un flux d'informations de gestion important

Chaque étape de la chaîne de traitement produit des informations utiles à l'ensemble du processus de gestion.

Ainsi, lors du paramétrage des accès, le bibliothécaire système devra pouvoir connaître les conditions d'accès aux différentes bases fixées par les termes des licences d'abonnement. L'acquéreur devra pouvoir aisément vérifier les dates de renouvellement d'abonnement pour éviter les interruptions de connexion. Le gestionnaire des collections devra disposer d'une liste exhaustive des revues et des bouquets dont elles dépendent pour évaluer la nécessité des reconductions. Afin d'évaluer les collections ou de produire des rapports d'activités ou des tableaux de pilotage de l'activité, les bibliothécaires devront pouvoir accéder de façon centralisée aux statistiques d'usages.

On estime ainsi entre 200 et 300 le nombre de métadonnées par ressources à gérer pour assurer la gestion du cycle de vie de ces ressources. [4, BARTHET, 2010]

2.1.2 Les ERMS comme outils de gestion du workflow

2.1.2.1 Naissance des ERMS

Dans les années 2000, les bibliothèques universitaires anglo-saxonnes ont lancé un mouvement de réflexion autour des processus et des modes d'organisations dédiés à la gestion des documents électroniques. Ces études ont révélé que la plupart des établissements avaient développé des outils "maison" ou s'étaient appuyés sur les outils historiques afin de gérer le cycle de vie des ressources électroniques.

En 2002, la Fédération des bibliothèques numériques (DLF - Digital Library Federation) en collaboration avec l'organisme NISO - National Information Standards Organization organise des réunions et ateliers regroupant bibliothécaires et fournisseurs. L'un de ces ateliers dédié à la réflexion autour des standards de la gestion de ressources électroniques débouche sur la création d'un groupe de réflexion baptisé ERMI - Electronic Resource Management Initiative. En 2004 ce groupe publie un rapport qui pose les bases du développement des ERMS. Ce rapport définit les prérequis fonctionnels pour le développement d'un système «capable de prendre en charge la gestion de l'information et des processus nécessaires à la sélection, l'évaluation, l'acquisition, la maintenance et la gestion des accès aux ressources électroniques conformément aux termes des licences d'exploitation » [23, DLF, 2004]. Il établit ainsi les principales fonctions attendues d'un ERMS :

- pouvoir générer et gérer facilement des listes de titres exhaustives pour l'ensemble des abonnements ;
- garder la trace des licences signées et de leurs conditions ;
- gérer l'acquisition de ces bouquets et les renouvellements avec gestion des abonnements
- par consortium ou autres groupements ;
- vérifier la bonne mise à disposition des ressources acquises et signaler les éventuelles interruptions de services aux usagers d'une part et aux fournisseurs de contenus pour réclamation d'autre part ;
- garder trace des relations avec les fournisseurs ;
- assurer la bonne mise à disposition ;
- réaliser des statistiques d'usage

2.1.2.2 Standards et protocoles

Les ERMS ont moins été conçus sur des standards techniques que sur des standards normatifs et les travaux qui ont conduit à leur création ont aussi permis d'établir des normes et protocoles permettant d'améliorer la circulation d'information entre fournisseurs de contenus et bibliothèques.

La norme COUNTER (Counting Online Usage of Networked Electronic Resources) est un code de bonnes pratiques qui permet la récupération structurée des statistiques d'usage auprès des éditeurs de ressources électroniques. Le développement de ce référentiel de pratiques est lié au développement du protocole SUSHI (Standardized Usage Statistics Harvesting Initiative) qui permet le transfert des informations normalisées en COUNTER.

COUNTER et SUSHI permettent de centraliser les données d'usages des ressources au sein des ERMS.

ONIX (Online Information Exchange) est un ensemble de formats gérés par le groupe EdITEUR, qui permet d'exprimer toute la diversité des contrats dans une forme lisible par l'ordinateur. Le format ONIX-PL (Onix for Publication Licenses) est un format xml dédié au

transfert de données liées aux licences. Il est actuellement utilisé par le JISC (Joint Information Systems Committee) pour modéliser une licence type de ressources électroniques. L'idée est de structurer les informations contenues dans les licences de sorte qu'elles soient lisibles par des ordinateurs et que les données ainsi définies fonctionnent comme une base de connaissances compréhensible par d'autres systèmes.

Les ERMS ont connu un grand succès dans les bibliothèques anglo-saxonnes. Malheureusement, ils sont encore assez peu présents sur le territoire français. Pour étendre l'usage de ces systèmes, en 2007, le consortium Couperin a initié un projet d'ERMS mutualisé entre tous ses membres. Après un appel d'offres et la sélection de la société Serials solution en 2010, le projet est abandonné en juin 2011 face à l'incapacité du prestataire à remplir toutes les exigences du cahier des charges.

2.2 Offrir un environnement de recherche unifié : de la métarecherche aux outils de découverte

2.2.1 Offrir un accès unifié aux collections

Avec la multiplication du nombre d'abonnements à des plateformes de ressources électroniques, les bibliothèques se sont retrouvées confrontées au problème de la dispersion des accès aux différentes bases de données. Elles ont, dans un premier temps, valorisé leurs collections électroniques en proposant sur leurs portails des listes de liens vers des plateformes d'interrogation ou des listes de titres de revues électroniques communément appelées A to Z. Ainsi, pour effectuer une recherche au sein de ces sources le lecteur devait-il dans un premier temps identifier la revue ou l'interface de recherche qui devait répondre le mieux à son besoin d'information. Ensuite, il devait se familiariser à l'utilisation des différentes interfaces natives et des modes d'interrogations proposés par les différents fournisseurs. [3, ARSENAULT, 2005]. Ainsi, pouvait-il tout à fait passer à côté d'une ressource pertinente, car non valorisée.

Pour résoudre ce problème de fragmentation de l'offre documentaire proposée, les bibliothèques ont voulu repenser la notion de collection en créant des silos d'information constitués de métadonnées hétérogènes. Les moteurs de recherche fédérés ont permis de réaliser cette unification.

2.2.2 Qu'est-ce que la recherche fédérée ?

Un moteur de recherche fédérée est un système qui permet l'interrogation simultanée de bases de données distantes à partir d'un seul point de recherche. Les bases de données peuvent héberger des contenus très divers (texte intégral, notices de signalements, données brutes, ...) qui peuvent être décrits dans des formats documentaires différents.

2.2.3 Recherche fédérée contre recherche unifiée

On oppose généralement la recherche fédérée à la recherche unifiée. Si la recherche fédérée porte sur des entrepôts externes, la recherche unifiée porte quant à elle sur un entrepôt constitué de métadonnées produites par la bibliothèque. Dans un contexte de diversification des bases produites par les bibliothèques qui ont vu se greffer à leurs catalogues bibliographiques des bases d'archives institutionnelles, des catalogues d'archives ou des plateformes pédagogiques, on a vu apparaître des portails permettant de moissonner des entrepôts internes et d'uniformiser les métadonnées afin de constituer un index de recherche unique. La recherche fédérée est parfois intégrée dans ces portails de nouvelle génération, mais dans la plupart des cas elle fonctionne de manière indépendante.

2.2.4 Protocoles d'interrogation et types d'outils existants

Il existe deux types de protocole d'interrogation pour la recherche fédérée : le premier, appelé recherche « croisée » est le protocole le plus couramment utilisé. Le second protocole, appelé recherche « asynchrone » a été implanté plus récemment.

2.2.4.1 La recherche croisée

La recherche « croisée » repose sur un protocole synchrone qui consiste à interroger les ressources externes à la volée et à renvoyer les résultats en temps réel. À chaque ressource interrogeable est attaché un connecteur. Celui-ci a pour rôle de définir les modalités d'interrogation d'une ressource distante et les modalités de traitement des résultats. Ces connecteurs sont adaptés au protocole d'interrogation autorisé par la ressource distante.

2.2.4.1.1 Les protocoles Z39.50 et SRU/SRW

Plusieurs protocoles existent, mais les plus utilisés sont les protocoles Z39.50, SRU et SRW. Le protocole Z39.50 repose sur le principe client-serveur. Tout client normalisé Z39.50 interroge tout serveur normalisé Z39.50. Le client Z39.50 s'adresse à un ou plusieurs serveurs Z39.50, simultanément ou individuellement. Il intègre des fonctions qui permettent d'élaborer une requête structurée sur des champs spécifiques avec opérateurs booléens et des fonctions de tri de résultats. Les échanges n'utilisent pas l'URL pour transmettre les requêtes et rapatrier les données, mais un langage propre, moins « visible ». Ce standard de recherches inventé en 1982 est surtout utilisé dans le milieu des bibliothèques. Il est petit à petit remplacé par le SRU/SRW plus adapté à l'univers web dans la mesure où celui-ci utilise l'URL pour rapatrier l'information.

Les protocoles SRU (Search/Retrieve via URL) et SRW (Search/Retrieve Web Services) suivent le modèle fonctionnel du Z39.50, mais prennent appui sur l'infrastructure web. Maintenus par la Bibliothèque du Congrès et destinés à faciliter la recherche sur Internet, ils se fondent sur le langage CQL (Common Query Language, standard de syntaxe de requête), qui présente l'avantage d'une (relative) lisibilité. SRU/SRW spécifient tout à la fois la syntaxe des requêtes et celle des réponses données à ces requêtes. SRU/SRW permettent

d'interroger simplement, via le protocole http du web, des bases de données jusqu'alors cachées ou plus difficilement accessibles.

2.2.4.1.2 L'implantation rapide des métamoteurs de recherche dans les bibliothèques

Sur la base de ces protocoles synchrone, dans les années 2000, la métarecherche s'est rapidement déployée dans les bibliothèques universitaires. Le faible coût d'implémentation au regard des bénéfices obtenus en terme de valorisation des collections a contribué à son développement. Dans un comparatif publié dans the *Library Journal Automated Systems*, Marshall Breeding indique que plus de 18 000 bibliothèques avaient à l'époque adopté un moteur de recherche fédéré. Il note qu'en 2009, 600 nouveaux systèmes ont été implantés.

2.2.4.1.3 Critique de la recherche fédérée dans son mode synchrone

Malgré son succès, la métarecherche subit les critiques des professionnels qui pointent les défauts suivants :

- Les temps de réponse trop longs : L'étalon de la recherche d'information pour les usagers est le temps de réponse des moteurs de recherche grand public. Ce dernier est généralement inférieur à la seconde. Or, le délai entre l'envoi de la requête par l'utilisateur et l'affichage des résultats par un moteur de recherche fédérée est dépendant de la source interrogée et du protocole mis en place pour paramétrer cette interrogation. Si une cible interrogée met plus de temps à répondre que l'ensemble des sources interrogées, elle bloquera l'affichage de tous les résultats. Certes, la plupart des systèmes de dernière génération affichent les résultats à la volée pour contrer cet inconvénient, néanmoins, le premier résultat affiché n'est pas forcément le plus pertinent et les temps de réponse dépassent largement ceux d'un moteur de recherche sur Internet.
- Le manque d'exhaustivité : Certaines ressources ne peuvent pas être techniquement fédérables car se pose souvent le problème de l'authentification des accès. Ainsi l'utilisateur pensant effectuer une recherche dans la totalité des ressources proposées par la bibliothèque peut-il passer à côté d'une source pertinente. Jody Condit Fagan dans son article *Federated search is dead* relativise ce défaut en signalant que pour un étudiant écrivant un mémoire à partir d'une quinzaine de sources, les résultats proposés par la recherche fédérée seront suffisamment pertinents pour répondre à son besoin informationnel. Quant à un chercheur, il aura déjà identifié ou identifiera les bases de recherche les mieux adaptées à ses besoins [24, FAGAN, 2011].
- L'impossible dédoublonnage des résultats : Afin d'améliorer les temps de réponse dans le cadre d'une recherche synchrone, les résultats sont envoyés par paquet de 10 ou 20 notices qui sont directement affichées dans le navigateur avant la fin de la requête. Il est donc impossible au système de dédoubler les résultats.

- L'impossible classement par pertinence : Le classement par pertinence est limité voir impossible, car le moteur de recherche ne peut appliquer son algorithme de traitement que sur les données qui sont affichées et non la totalité des données indexées. Par exemple il est impossible de compter le nombre d'occurrences du terme recherché, car la cible interrogée ne renvoie pas toutes les données indexées.
- Difficile maintenance des connecteurs : Comme nous l'avons vu plus haut il existe une variété de connecteurs. Ces derniers sont très difficiles à maintenir et demandent des compétences techniques en matière de programmation. [27, HANE, 2010]

2.2.4.2 Protocole asynchrone et outils de découverte

Afin d'améliorer le fonctionnement de la recherche fédérée, les éditeurs de logiciel pour bibliothèques ont proposé de nouveaux outils baptisés outils de découverte. Ces moteurs de recherche fonctionnent sur la base d'un protocole asynchrone.

2.2.4.2.1 Le protocole asynchrone

À la différence du protocole synchrone qui va interroger les bases en temps réel, le protocole asynchrone se base sur l'interrogation d'un index unique constitué à partir du moissonnage des ressources distantes.

Le moissonnage des ressources distantes repose généralement sur le protocole OAI-PMH. Ce protocole n'est pas un protocole de métarecherche puisqu'il ne supporte pas les fonctionnalités de recherche implantées dans les protocoles SRU-SRW et Z39.50. On parle dans son cas d'un protocole de transfert. Il est la base de la constitution des outils de découverte.

2.2.4.2.2 Les outils de découverte

L'outil de découverte est un index général centralisé composé de données provenant d'éditeurs, d'agrégateurs de contenu et de sources locales (catalogues, dépôts institutionnels, collections numérisées, etc.).

Son but est de permettre aux usagers de découvrir les collections complètes d'une bibliothèque sans distinction de leur localisation, de leur provenance ou de leur format. La possibilité d'élargir la recherche en dehors des collections locales, par exemple en cherchant des répertoires gigantesques disponibles en libre accès, est perçue comme la valeur ajoutée des outils de découverte.

Cet index centralisé permet d'implanter la recherche fédérée au cœur des collections traditionnelles de la bibliothèque. Ainsi, si la recherche fédérée nécessite la mise en place d'un masque de recherche supplémentaire à côté du masque de recherche dédié aux collections de la bibliothèque, les index centraux permettent-ils d'offrir un masque de recherche unique aux utilisateurs. Les temps de réponse sont optimisés, le système peut

effectuer un classement par pertinence puisqu'il a accès à la totalité des données indexées, ce qui améliore l'expérience de recherche [43, YANG, 2010].

2.3 Open-URL et résolveurs de liens: gérer la localisation des ressources

2.3.1 Naissance de l'Open-Url : interconnecter les publications électroniques

Le développement de la publication électronique de la littérature académique a créé très rapidement une demande très importante de la part des chercheurs et des étudiants. Grâce au développement de la documentation électronique, ils peuvent désormais consulter un article à partir de n'importe où et à n'importe quelle heure. Ils ne sont plus contraints de devoir se déplacer à la bibliothèque pour pouvoir accéder au fascicule de la revue dans laquelle l'article qui les intéresse a été publié. Cet usage a créé de nouvelles attentes. Avec le développement du World Wide Web et l'usage de l'hypertexte, les lecteurs ont voulu pouvoir accéder directement à la référence citée par la source en cours de consultation. Ainsi le désir est-il né de pouvoir interconnecter les publications scientifiques publiées au format électronique. [2, APPS, 2006].

2.3.1.1 La stratégie des liens fixes

Ces nouveaux besoins ont conduit dans un premier temps les fournisseurs de contenus à proposer des liens html statiques faisant le lien entre les publications présentes dans un même ensemble, qu'il s'agisse d'une plateforme d'éditeur ou d'un service d'agrégation de contenu. À la suite d'accords entre fournisseurs, des liens de ce type ont été créés entre services de différents éditeurs. Mais cette solution s'est révélée compliquée à maintenir. Une URL est liée à l'emplacement physique du fichier. Elle indique le serveur (nom de domaine), le dossier, voire même le sous-dossier, sur lequel le document est hébergé. Ce lien établi entre deux sources peut aisément être brisé dans le cas où, par exemple, le fichier ciblé est déplacé ou si l'institution qui l'héberge change de nom de domaine sans qu'un système de redirection soit mis en place.

2.3.1.2 Les liens dynamiques

Par la suite, pour alléger la maintenance de ces liens entre documents, les fournisseurs ont créé plusieurs types de syntaxe de construction d'URL. Ces syntaxes permettaient d'identifier une ressource ciblée sur la base d'attributs passés dans la requête HTTP. Cette syntaxe permettait également de rendre l'URL prédictible. Ainsi, un lien vers un journal pouvait-il être créé en utilisant l'ISSN de la publication. Pour atteindre l'article, l'URL pouvait utiliser le volume, le numéro du fascicule et les numéros de page. Cette requête était alors interprétée par un script CGI au niveau du serveur qui redirigeait le lecteur vers la bonne URL du document. Cette solution permettait ainsi l'accès à de multiples articles via la même

passerelle. Cependant, ces liens utilisaient des formats propriétaires qui différaient entre les fournisseurs de ressources et ne permettaient donc pas d'établir des liens entre plateformes.

2.3.1.3 Le DOI

Afin de relier des ressources à l'échelle du web, plusieurs projets de plus ou moins grande ampleur ont tenté de créer des identifiants normalisés pour les ressources électroniques. L'un des plus aboutis aujourd'hui, le Digital Object Identifier (DOI), est un système développé par le CNRI (Corporation for National Research Initiatives) pour le compte de l'Association of American Publishers (AAP) qui voulait définir un identifiant pour l'accès à des portions d'informations numériques afin de permettre la gestion et le recouvrement des droits, la distribution électronique de tout ou partie des œuvres qu'ils publiaient, la navigation hypertexte au travers des citations dans les œuvres, la gestion électronique du copyright, la découverte de documents, l'archivage d'objets électroniques, l'offre de la pérennité sur le Web.

Le DOI prend la forme d'un Uniform Resource Name (URN) c'est-à-dire une syntaxe permettant l'identification des ressources elles-mêmes, indépendamment de leur emplacement². Il s'articule autour d'un identifiant unique d'une portion de publication électronique et d'un mécanisme permettant de localiser le document sur la base de son identifiant.

Le mécanisme de localisation repose sur un répertoire international d'identifiants de tout ou partie de publications électroniques. Son objectif est de permettre la collecte des droits d'utilisation des documents publiés électroniquement sur une base définie par le détenteur des droits. Ainsi, lorsqu'un lien DOI est "cliqué", le DOI est envoyé au résolveur, et l'URL trouvée dans cette base de données est renvoyée au navigateur qui redirigera l'utilisateur vers le document via une requête HTTP.

2.3.2 La question de la copie appropriée

Le système du DOI souffre de quelques limitations. Le mécanisme dans sa forme générale est limité à établir une relation de un à un entre DOI et URL. Le système fonctionnerait bien s'il y avait une seule copie de chaque article de journal électronique. Mais il existe de nombreux cas où plus d'une copie légitime d'un article est disponible :

- Il peut exister des copies locales des documents disponibles sur des serveurs d'institution afin de desservir une communauté universitaire
- Les agrégateurs chargent des copies des articles sur leurs serveurs pour proposer des accès soumis à abonnement.
- Un même article peut exister dans sa version papier

² Contrairement à l'URL qui identifie l'emplacement de la ressource

- Pour des raisons de performance du réseau, les éditeurs utilisent parfois des serveurs miroirs pour héberger les documents.

Dans chacun de ces cas la construction de l'URL d'accès au document dépend du contexte géographique ou institutionnel dans lequel l'utilisateur qui passe la requête de résolution se trouve. C'est le problème de la copie appropriée illustré par le schéma ci-dessous.

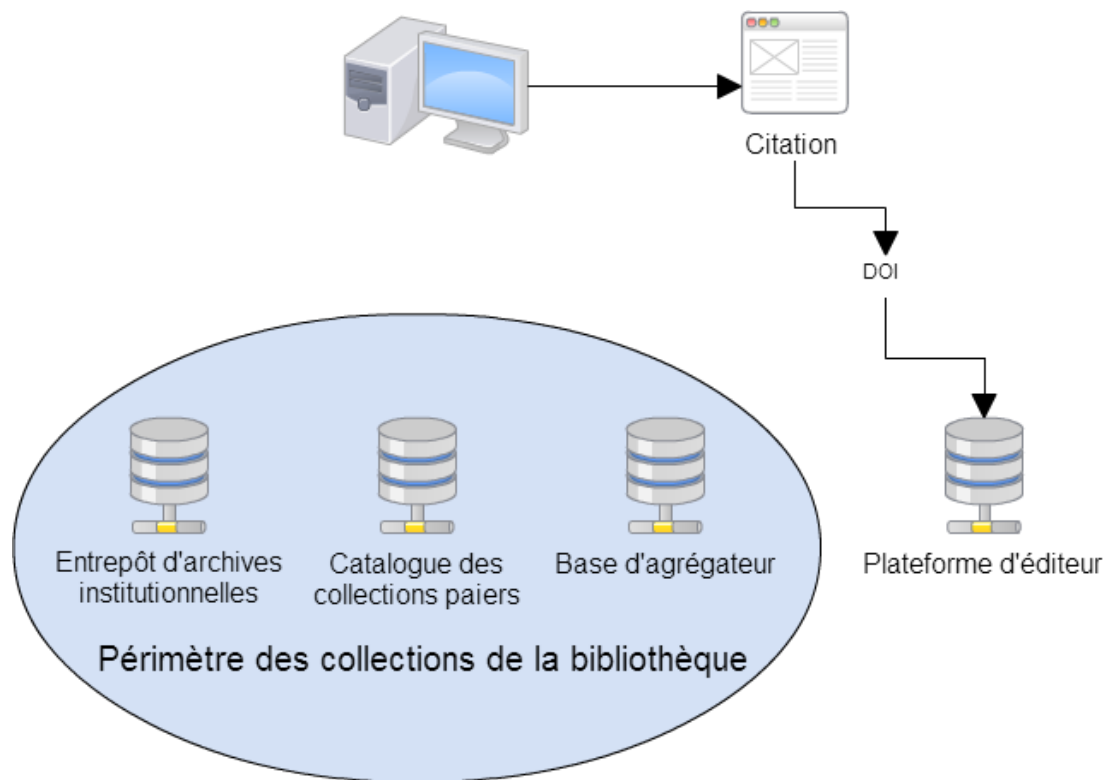


Figure 4 : Problème de la copie appropriée

Les paramètres qui vont influencer sur ces accès sont :

- L'institution à laquelle est rattaché l'utilisateur, liée à son adresse IP ou aux informations de connexion qu'il a fournies pour se connecter à l'environnement numérique de travail,
- Le type d'inscription dont il dispose (selon qu'il est étudiant, chercheur ou autre ...)

Il existe presque autant de combinaisons que d'institutions de rattachement, chaque bibliothèque ayant des abonnements différents, par l'intermédiaire de fournisseurs différents (éditeurs, agrégateurs, bouquets négociés via des consortia). En fonction de tous les intervenants, il est nécessaire de fournir à l'utilisateur le service qui correspondra le mieux à sa recherche. Il faut donc être capable de fournir à l'utilisateur la « copie appropriée » [5, BEIT-ARIE, 2001].

2.3.3 Open-URL et résolveurs de liens

C'est à partir de 1998 que Herbert Van de Sompel a commencé ses travaux qui mèneront à la création de l'OpenURL. Il lance, à l'Université de Gand (Belgique), le projet « SFX, special

effects ». Le logiciel issu de ces recherches est acquis par la société Ex Libris qui le commercialise aujourd'hui encore. Ce modèle rencontre un fort succès, il fait l'objet d'une norme internationale publiée au printemps 2005, la norme Z39.88-2004.

Le protocole sert à créer des liens contextuels. Il permet de lier des métadonnées à la ressource qu'elles décrivent en fonction du contexte de citation de la ressource et des droits dont dispose la personne qui utilise le lien. La norme OpenURL Framework se compose de trois choses :

- le ContextObject : Il s'agit du paquet de métadonnées qui contient les informations sur une ressource et son contexte. Le ContextObject décrit une ressource bibliographique et son contexte. Ce paquet regroupe :
 - le « referent », il s'agit du paquet de métadonnées permettant d'identifier l'objet téléchargé ou affiché
 - le « requester » identifie l'agent qui a initié la requête
 - le « serviceType » identifie le type de service demandé. Par exemple : texte intégral, résumé, demande de PEB...
 - le « referer » désigne l'institution qui a rédigé les métadonnées
 - le « resolver » identifie l'institution qui fournit le service d'accès au document
- l'OpenURL proprement dit qui est le protocole de transport de ces paquets d'informations. Il est basé sur http.
- le Registry qui contient les spécifications des différents éléments constitutifs des ContextObjects et de l'OpenURL.

Concrètement, lorsqu'un éditeur de revues en ligne ou bien une bibliothèque crée un ContextObject, le lecteur voit apparaître, à côté de la référence bibliographique, un bouton qui correspond à ce ContextObject. S'il clique, le ContextObject est aussitôt envoyé sous forme de requête HTTP à un résolveur de lien, qui analyse les métadonnées, les droits de l'utilisateur et le service demandé. Le résolveur effectue un tri entre toutes les références auxquelles le lecteur peut avoir accès et écarte celles auxquelles il n'a pas accès.

Il renvoie au lecteur toutes les références auxquelles il peut avoir accès sous forme de liste de liens.

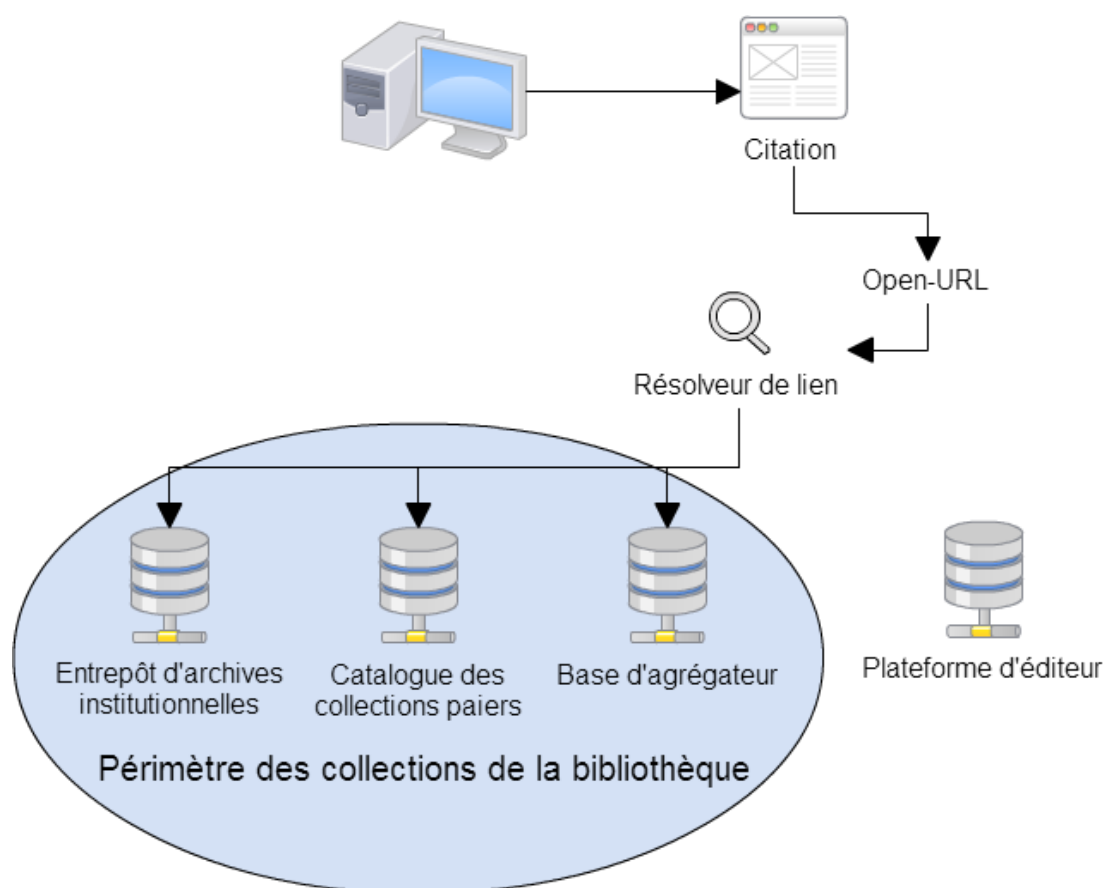


Figure 5 : Fonctionnement de l'Open-url

L'OpenURL permet donc aux résolveurs de liens de lier entre elles les métadonnées, à condition que ces dernières soient encodées dans un format précis

En conclusion, l'OpenURL permet à des résolveurs de liens de lier des métadonnées, pourvu qu'elles soient encodées dans un certain format, à des ressources paramétrées, en tenant compte des droits de l'utilisateur et d'autres paramètres éventuels.

Les bases de connaissances au cœur de ces dispositifs

1 Qu'est-ce qu'une base de connaissances ?

1.1 Définition

La notion de base de connaissances vient des sciences informatiques pour décrire une base de données regroupant des connaissances spécifiques à un domaine spécialisé précis, sous une forme exploitable par un ordinateur. Elle contient des règles, des faits ou d'autres représentations nécessaires à l'exercice d'une activité donnée pour laquelle cette base de connaissances a été développée.

En bibliothèque, la notion de base de connaissances est apparue avec le développement des résolveurs de liens. On parle alors d'une base de données qui rassemble de façon structurée les informations connues sur les ressources électroniques : titres, périmètres des bouquets, états de collection, URL d'accès. Plus récemment, avec l'apparition des ERMS, la base de connaissances s'est enrichie de nouvelles données telles que les termes des licences, les coûts, ou encore les statistiques d'usage.

Une base de connaissances est généralement maintenue par l'organisme fournisseur du système informatique (résolveur de liens ou ERMS) qui l'utilise. Toutefois, une base de connaissances peut tout à fait exister au seul niveau local et être maintenue et administrée par la bibliothèque. Cependant l'intérêt de la base de connaissances est de pouvoir mettre en commun les informations nécessaires à la gestion et à la localisation des documents électroniques et de pouvoir partager ces dernières entre les membres d'un réseau de bibliothèques clientes du même fournisseur ou membres d'un réseau institutionnel de bibliothèques. Ainsi, dans la plupart des cas, les bases de connaissances ont-elles plusieurs instances : une instance globale, pour les informations communautaires et une instance locale, pour les informations propres à la bibliothèque.

Avec le développement des outils de découverte et des index centraux de recherche, une confusion est née dans le monde professionnel. Le terme de base de connaissances se trouve employé pour désigner les index de recherche constitués par les moissonnages des sources d'éditeur. Cette confusion semble naître du fait que les index des outils de découverte sont maintenus par les mêmes acteurs que ceux qui fournissent et maintiennent les bases de connaissances. De plus, les fonctions assurées par ces deux systèmes sont très imbriquées. L'index central va permettre la recherche au sein des collections électroniques et donc gérer les métadonnées d'identification. La base de connaissances, quant à elle, va permettre la localisation des documents en centralisant les métadonnées de gestion. Lorsque l'outil de découverte propose les résultats de recherche, il va interroger le résolveur de liens pour connaître la disponibilité du document et construire un lien OpenURL.

L'implantation d'un résolveur de liens autonome, indépendant d'un index central, a un intérêt limité pour un établissement souhaitant se doter par ailleurs d'un outil de découverte. En effet, si la couverture ouverte dans l'index central est supérieure à celle de la base de

connaissances du résolveur, certains résultats ne donneront pas accès au document primaire. À l'inverse, si la couverture est plus importante dans le résolveur de lien, les ressources ne seront pas toutes indexées dans l'outil de découverte et resteront invisibles.

La base de connaissances sert essentiellement à la construction des liens open-URL. Toutefois, pour permettre un traitement plus aisé des collections électroniques, les bibliothèques chargent les métadonnées qu'elles contiennent au sein de leurs catalogues. La base de connaissances occupe donc une place centrale dans le dispositif de gestion de la documentation, contribue à la gestion des collections électroniques via l'ERMS, la gestion des liens vers les documents via le résolveur de liens et la diffusion des contenus via le portail de la documentation.

1.2 Acteurs du secteur et processus de constitution

L'alimentation et la gestion d'une base de connaissances font intervenir de nombreux acteurs, du producteur de l'information jusqu'au bibliothécaire. Après s'être intéressés aux circuits des métadonnées, nous détaillerons le rôle de chacun d'eux.

1.2.1 Circuit des métadonnées

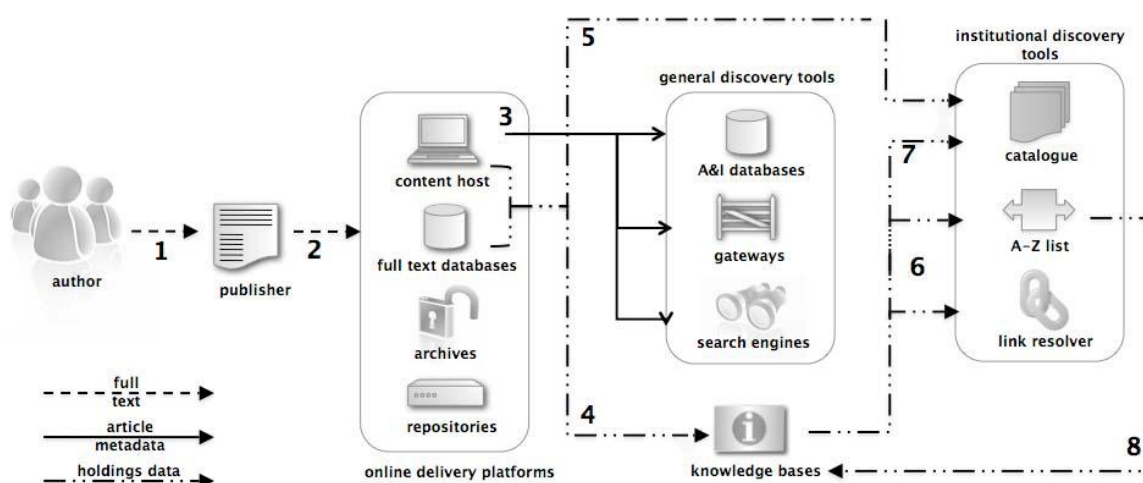


Figure 6 : Le workflow de la constitution des bases de connaissances

« L'auteur fournit des manuscrits à l'éditeur (1). L'éditeur envoie le texte complet de l'article final et des métadonnées associées à un ou plusieurs services de livraison en ligne (les hébergeurs de contenu, les bases de données de texte intégral, les sites d'archives, etc.) (2). L'hébergeur de contenu est souvent responsable de la mise à disposition des métadonnées d'un article auprès d'autres outils de découverte, tels que les bases de données de résumés et d'indexation, les plateformes d'agences d'abonnement, et les

moteurs de recherche (3). Les hébergeurs de contenu et les bases de données en texte intégral fournissent également des informations sur les bouquets, et parfois également des notices MARC, aux développeurs de base de connaissances (4) ou directement à l'institution (5) afin d'informer ces systèmes sur l'étendue du contenu disponible. Des services de la base de connaissances enverront alors les données sur l'état des collections aux listes A-to-Z institutionnelles et aux résolveurs de liens (6) ; l'institution utilise ces informations pour paramétrer sa propre collection. Les développeurs de base de connaissances peuvent également envoyer des notices MARC à l'institution pour qu'elles soient versées dans l'OPAC institutionnel (7).

Dans ce scénario, l'établissement doit fournir à son développeur de base de connaissances des détails sur ses collections pour des contenus tels que les revues électroniques, etc. (8). » [38 , UKSG KBART, 2010]

1.2.2 Les fournisseurs de contenus

Les fournisseurs de contenus peuvent être à la fois source et cible de liens openURL. En tant que source, un fournisseur de contenus doit pouvoir créer des liens OpenURL conformes à leurs références bibliographiques. Les métadonnées transmises utilisées pour construire le Contextobject doivent permettre une identification précise du document cité.

En tant que cible, le fournisseur doit également mettre à la disposition des autres sources OpenURL, notamment les résolveurs de liens, des métadonnées précises décrivant ses collections électroniques. D'autre part, il doit fournir aux bibliothèques et/ou aux fournisseurs de résolveur de liens les données nécessaires à l'identification et à la description des bouquets ainsi que les dates de couverture des différents titres proposés dans le cadre des abonnements. Il n'existe pas de format standard pour de telles données. Ainsi les fournisseurs de contenu doivent-ils s'assurer que les données qui décrivent les contenus sont correctes et régulièrement transmises aux possesseurs de résolveurs de liens, et que les erreurs identifiées sont corrigées le plus rapidement possible.

1.2.3 Les fournisseurs de résolveur de liens

Les fournisseurs du résolveur de liens sont en charge de la technologie du résolveur, de la collecte et de la gestion des données de la base de connaissances globale.

À ce titre, ils reçoivent régulièrement les listes de contenus actualisées de la part des fournisseurs. Avant d'intégrer ces données dans la base de connaissances, ils mettent en place un certain nombre de traitements automatisés ou humains, pour s'assurer de l'uniformité des métadonnées. Ainsi, Serial Solution, éditeur du résolveur de liens Serial 360, assure utiliser 100 000 règles de correction pour garantir la cohérence de la base de connaissances.

Les éditeurs déclarent employer entre 8 et 30 personnes pour administrer leurs bases de connaissances. En tant que responsables de la technologie du résolveur de liens , les éditeurs de logiciel doivent fournir aux bibliothèques les interfaces et les outils nécessaires à l'administration d'une base de connaissances locale.

1.2.4 Les bibliothèques

Le rôle de la bibliothèque est avant tout de proposer l'information au lecteur, l'utilisateur final. Pour ce faire, elle doit s'assurer que les métadonnées proposées sont à la fois correctes et utilisables. Dans le dispositif, la bibliothèque et l'utilisateur final sont les deux acteurs les plus à même de déceler les erreurs dans les données OpenURL ou dans les données concernant les références bibliographiques. Ils sont aussi en première ligne pour signaler ces erreurs aux autres acteurs de la chaîne d'approvisionnement.

La bibliothèque, pour permettre aux fournisseurs de contenu d'être source de liens OpenURL, doit lui déclarer l'adresse de son résolveur de liens. Elle a aussi la charge de personnaliser sa base de connaissances sur laquelle son résolveur de liens repose, de manière à refléter au mieux son propre fonds papier et électronique.

2 La gestion d'une base de connaissances dans un établissement universitaire, le cas Dauphine:

2.1 Présentation du contexte

Bibliothèque de référence reconnue au niveau national à travers ses missions de CADIST (Centre d'acquisition et de diffusion de l'information scientifique et technique) et de pôle associé de la Bibliothèque nationale de France, la bibliothèque de l'université Paris-Dauphine est le premier pôle de documentation pour la Gestion et l'Économie en région Île-de-France.

2.1.1 Les ressources électroniques à la bibliothèque Dauphine

Comme toutes les bibliothèques universitaires francophones, Dauphine n'échappe pas au mouvement national d'accroissement du nombre de ressources électroniques au sein des collections des bibliothèques. Elle met à disposition de ses chercheurs et étudiants 42 bases de données d'une grande variété. La bibliothèque propose des bases de données d'agrégateurs de revues spécialisées ou généralistes (Business Source Complete, Econlit with full text, Cairn,...) des bases de grands éditeurs scientifiques (Springer, Emerald, Blackwell, PsycARTICLES, OECD e-library,...), des bases d'éditeurs institutionnels (OECD e-library, World Bank), des bases de données juridiques (Bibliothèque numérique Dalloz, Jurisclasseur,...), des bases de données d'abstract (Doctrinal +, Delphes, MathSciNet), des bases d'informations financières (Datastream, Thomson One Banker, ...), des bases de données Macro-économiques (Chelem, OECD Library, Worls-Bank Worldbank e-library), des bases d'études sectorielles (Xerfi 700, Passport GMID, ...), un agrégateur de presse (Factiva) et une encyclopédie en ligne (Encyclopedia Universalis). Grâce à cette collection, les membres de l'université peuvent accéder sur place ou à distance à 52 608 revues et 46 988 livres électroniques.

L'acquisition de ce type de ressource pèse énormément sur le budget de la bibliothèque dont les dépenses documentaires représentaient, en 2012, 71% des dépenses totales de l'établissement. Les acquisitions de ressources électroniques représentent 34% des dépenses totales. Ce pourcentage est plus important si l'on considère que l'acquisition de certains abonnements à des revues électroniques est couplée au papier et que dans ce cas la dépense est imputée aux périodiques papiers.

	2009	2010	2011	2012
Périodiques papier	503 578	545 657	550 410	522 085
Monographies	145 938	144 552	165 927	160 438
Documentation électronique	214 582	261 307	258 776	346 971
TOTAL	864 098	951 516	975 113	1 029 494

Tableau 3 : Répartition des dépenses documentaires de la Bibliothèque Paris-Dauphine

2.1.2 Le système d'information documentaire de la bibliothèque

Le système d'information documentaire de la bibliothèque se compose des briques suivantes :

- Le SIGB (ALEPH 500 d'Ex Libris) : Système central de l'activité de la bibliothèque, il permet la gestion des ressources papier de la bibliothèque au travers de ses modules d'acquisition, de catalogage et de circulation. La bibliothèque cataloguant ses documents dans le SUDOC, le module catalogage est peu utilisé.
- Le système de gestion des archives ouvertes (DSPACE) : La base des publications scientifiques de Dauphine (BIRD) recense et donne accès aux travaux de recherche des équipes de l'université. Elle est administrée et valorisée par la bibliothèque à travers le logiciel libre DSPACE qui assure le contrôle et la gestion du workflow de dépôt des documents, d'enrichissement des métadonnées et de diffusion de ces mêmes métadonnées. Les notices de BIRD sont versées au catalogue de la bibliothèque. Elles sont aussi mises à la disposition d'entrepôts externes au moyen d'un entrepôt OAI-PMH.
- Le moteur de recherche fédéré (METALIB d'Ex Libris): Il permet d'offrir aux lecteurs de la bibliothèque un point de recherche unique vers toutes les bases de données qui sont techniquement fédérables. De plus, METALIB permet le catalogage et la diffusion des notices descriptives des bases de données.
- Le résolveur de liens (SFX d'Ex Libris) : La bibliothèque, a fait le choix de ne pas cataloguer ses ressources électroniques dans le SUDOC et dans le SIGB. De ce fait, c'est SFX qui est utilisé comme catalogue des ressources électroniques et les métadonnées d'identification sont moissonnées et envoyées dans PRIMO.
- Le portail documentaire (PRIMO d'Ex Libris) : Ce portail permet d'exposer les métadonnées émanant du SIGB ALEPH, de METALIB, de SFX et de BIRD. Il permet également d'offrir une interface unique de recherche dans toutes les sources gérées par la bibliothèque, ainsi que les sources des fournisseurs extérieurs puisque le système de recherche fédéré est intégré dans l'interface du portail documentaire.

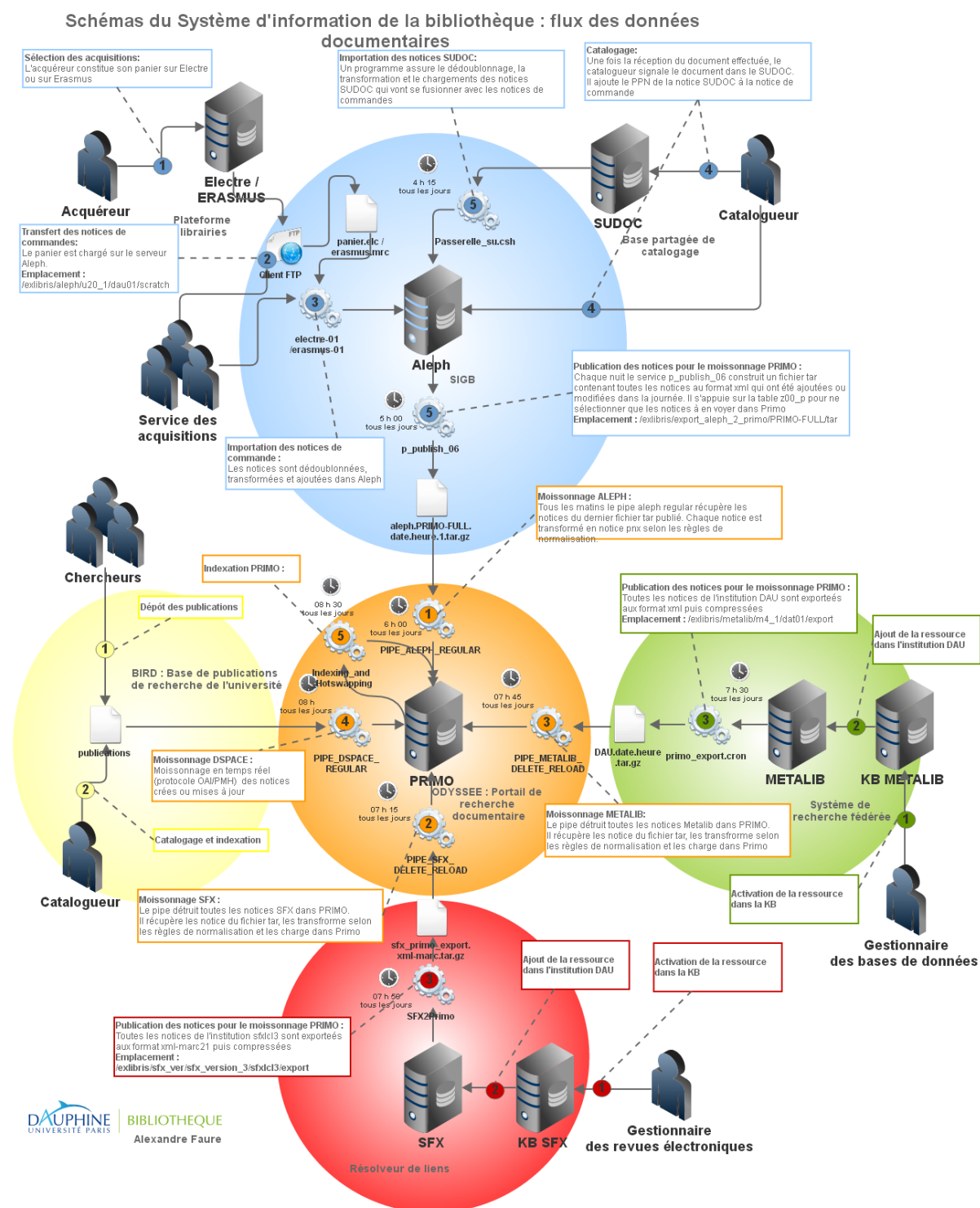


Figure 7 : Schéma du système d'information de la bibliothèque Dauphine : flux des données documentaires

2.2 La base de connaissances SFX au cœur du Système d'information documentaire

Dans cette architecture, SFX joue un rôle important. Il permet le signalement des documents électroniques et la gestion des accès à ces ressources. Pour ce faire, il s'appuie sur une base de connaissances, alimentée et administrée par les équipes d' Ex Libris. Quelques bouquets d'abonnements, non référencés au niveau global de la base de connaissances sont traités au niveau de l'instance locale.

L'architecture de SFX adossée à une base de connaissances partagée offre de grands avantages en termes d'économie de moyens. Elle permet la gestion d'un grand nombre de ressources avec un seul agent. Toutefois, le manque de maîtrise concernant la qualité et la complétude des métadonnées pose de nombreux problèmes.

2.2.1 Hétérogénéité et qualité des métadonnées

2.2.1.1 La question de la qualité des métadonnées liée aux accès aux ressources

Les métadonnées fournies dans le ContextObject des liens OpenURL ou contenues dans la base de connaissances sont essentielles au bon fonctionnement du service de localisation. L'efficacité de l'accès dépend de l'exactitude des informations à tous les niveaux : qualité de la source OpenURL, données générales d'identification des titres, packages ou plateformes, données de localisation et de couverture, paramétrage du résolveur de liens .

En analysant 400 rapports de liens brisés renvoyés par les utilisateurs du résolveur de liens SFX, Xiaotian Chen, de la bibliothèque Cullom-Davis à la Bradley University montre que les principales erreurs sont liées à des problèmes de déclaration de bouquets non conforme à la réalité de l'abonnement et à des problèmes de métadonnées erronées dans la base de connaissances. [18, CHEN, 2012]

KBART dans son guide des bonnes pratiques pour l'alimentation des bases de connaissances des résolveurs de liens identifie neuf types de problèmes récurrents [38, UKSG KBART, 2010].

2.2.1.1.1 Les problèmes d'identifiants incohérents

Lors de la réutilisation d'un ISSN (lors d'un changement de titre de périodique) ou d'un ISBN (quand le même éditeur propose plusieurs titres) le résolveur du lien ne peut identifier convenablement le document. Il se retrouve face à un identifiant théoriquement unique qui ne l'est plus. Concrètement, la syntaxe OpenURL utilise comme composant clé un ISSN unique. Mais si la référence bibliographique (la source) utilise un ISSN exact alors que la base de connaissances utilise un ISSN inexact (l'ISSN d'un titre antérieur ou postérieur par exemple), le résolveur de liens sera dans l'incapacité de rétablir l'erreur et de corriger l'incohérence.

2.2.1.1.2 Les couvertures temporelles incorrectes

Si les fournisseurs de contenu proposent des dates inexactes, ou si les développeurs ne mettent pas à jour les dates dans les bases de connaissances, le résolveur de liens renverra vers des résultats inexacts ou des liens qui ne seront pas valides. L'utilisateur sera alors informé que la ressource qu'il cherche n'est pas disponible même si cette dernière l'est, car, par exemple, le fichier n'aura pas été mis à jour et ne contiendra pas les derniers numéros. À l'inverse, le résolveur de liens peut signaler à l'utilisateur la disponibilité d'une ressource qui ne l'est pas, car elle aura été retirée de la collection sans que la liste ait été mise à jour.

2.2.1.1.3 Les formats de dates incohérents

L'absence de format standard des dates de couverture est un problème majeur qui donne lieu à de nombreuses erreurs d'interprétation. Comment différencier le jour du mois dans une date comme « 20120407 » ? De la même manière, la coexistence de deux dates pour un même document, comme par exemple la date de mise en ligne et la date de publication, dans le cas où ces deux dates sont différentes, peut engendrer un problème de mise en relation entre la base de connaissances et la référence bibliographique.

2.2.1.1.4 Les problèmes de périodes d'embargo

Les conditions d'embargo changeantes peuvent être source d'interrogation pour les développeurs de bases de connaissances. Dans la pratique, les mêmes termes peuvent être entendus différemment par chacun des fournisseurs de contenu. Concrètement, la mention « embargo sur 1 an » peut signifier que l'année civile en cours n'est pas disponible (à compter du 1^{er} janvier de l'année en cours) comme il peut signifier que ce sont les 12 derniers mois de publications qui ne sont pas accessibles. Pour l'utilisateur, il est très difficile de comprendre si la ressource est disponible ou non dans la base de données lorsque la date d'émission coïncide plus ou moins avec la période d'embargo.

2.2.1.1.5 Les données de contenus obsolètes

Sans mise à jour, les données présentes dans la base de connaissances resteront celles qui y ont été versées. De ce fait, les données de contenus obsolètes dirigent vers une présentation incorrecte des données, empêchant l'utilisateur d'accéder à des ressources vis-à-vis desquelles il a le droit d'accès.

2.2.1.1.6 Le manque de personnalisation

Les informations envoyées par les fournisseurs de contenu étant généralement les mêmes pour tous les clients, elles manquent de personnalisation. En règle générale, les clients ont tous accès aux couvertures temporelles les plus larges. De ce fait, les bibliothécaires sont parfois contraints de travailler à la personnalisation de ces listes pour que celles-ci représentent avec plus d'exactitude la réalité de leur abonnement. Ces modifications manuelles sont risquées d'erreur et peuvent alors nuire aux accès.

2.2.1.2 L'articulation avec les métadonnées traditionnelles

Comme beaucoup d'autres établissements, afin de faciliter la gestion des ressources électroniques, nous avons fait le choix de ne pas cataloguer les ressources électroniques. Le problème est que notre résolveur de liens SFX n'offre que peu de métadonnées d'identification (titre, ISSN, éditeur, lien vers les anciens titres et catégories thématiques) et qu'il est impossible d'enrichir en local ces notices. Il y a un appauvrissement de l'information bibliographique encore plus criant lorsque la notice se retrouve à côté d'une notice du même titre dans sa version papier.

D'autre part, l'hétérogénéité des métadonnées moissonnées dans un même entrepôt PRIMO pose le problème de leur uniformisation. En 2012, la bibliothèque a mené un travail d'analyse sur le mode de constitution des facettes proposées par le portail documentaire. Ce dispositif de recherche est essentiel au fonctionnement des portails documentaires modernes. Il permet de proposer à l'utilisateur de filtrer ses résultats en aval de sa recherche.

La bibliothèque voulait revoir les modes de constitution de ses facettes. Elle a donc réalisé des mapping pour éclater les métadonnées des différents entrepôts moissonnés par PRIMO. Ce travail a révélé de grandes disparités pouvant nuire à l'efficacité de ces facettes. Le problème est particulièrement criant au niveau des filtres « sujet ». Par défaut, PRIMO propose à l'utilisateur de filtrer ses résultats en fonction d'une liste de différents sujets émanant des documents affichés. Le problème est que dans chaque entrepôt, le vocabulaire utilisé n'est pas le même. SFX utilise une liste de catégories thématiques « maison » contenant une centaine d'entrées. Dans ALEPH, les documents sont indexés avec l'indexation matière RAMEAU. Dans BIRD, les documents comportent une indexation matière en vocabulaire libre, un indice de la classification du Journal of Economic Literature et les catalogueurs ajoutent parfois des mots-clés RAMEAU pour enrichir les notices. Dans PRIMO, tous ces champs sont utilisés pour constituer la facette « sujet ». Ainsi, certaines facettes sont-elles discriminantes pour l'utilisateur. En effet, si celui-ci veut filtrer sa recherche sur la base d'un code JEL le filtre ne portera que sur des ressources de BIRD sans que l'utilisateur en soit informé. On retrouve ce problème discriminant au niveau de la constitution du nom des auteurs et d'autres métadonnées.

2.2.2 Maîtriser sa collection

2.2.2.1 Gérer la volumétrie

La gestion des ressources dans une base de connaissances peut parfois donner le tournis de par la nature « expansive et mouvante » de la documentation électronique [19, CLARAC, 2013]. Les périmètres des bouquets auxquels les bibliothèques sont abonnées évoluent au gré des accords passés entre éditeurs. En raison des modèles économiques imposés par le secteur et plus particulièrement des fameux *big deals*, les bibliothèques sont amenées à

gérer un volume de collections important sans avoir les ressources nécessaires pour maîtriser complètement cette gestion. Ainsi, l'intérêt des bases de connaissances des résolveurs de liens est-il de pouvoir déléguer la gestion à des tiers. Ces derniers mettent en valeur des interfaces d'administration simples et intuitives qui permettent d'activer en un clic plusieurs centaines de titres d'un bouquet référencé dans leur base de connaissances. Les bouquets sont alimentés automatiquement lorsque le fournisseur, agrégateur ou éditeur, ajoute ou retire un titre dans son offre. Derrière cette apparente simplicité, les professionnels des bibliothèques perdent la maîtrise et la connaissance de l'offre qu'ils proposent aux usagers. Les problèmes d'accès aux ressources ne peuvent pas être anticipés et sont la plupart du temps portés à la connaissance des professionnels par les usagers.

2.2.2.2 L'impossibilité de couvrir la totalité de la collection dans la base de connaissances

Il est actuellement impossible pour notre bibliothèque de gérer la totalité de nos collections électroniques à travers notre base de connaissances. Cette situation est liée à plusieurs facteurs.

Un facteur technique en premier lieu : les éditeurs ou fournisseurs de base de données ne proposent pas une interface technique capable de produire des liens profonds pour leurs ressources. Dans le meilleur des cas, il est toujours possible de signaler ces ressources dans la base de connaissances. Le résolveur de liens renvoie le lecteur vers la page d'accueil de la plateforme. C'est la solution que nous avons adoptée pour signaler les titres de presse offerts dans le cadre de notre abonnement à Factiva. Cette situation n'est pas idéale, elle crée une rupture dans la navigation du lecteur.

Le deuxième facteur est stratégique : certains éditeurs refusent de fournir leurs métadonnées aux éditeurs de bases de connaissances. Le problème est tout particulièrement criant chez les éditeurs juridiques qui commencent à peine à accepter l'intégration des métadonnées dans les bases de connaissances des résolveurs de liens. Cependant, dans le cas où l'éditeur n'est pas présent dans la base de connaissances globale, mais accepte de fournir à la bibliothèque les métadonnées relatives à son abonnement, celle-ci a la possibilité de les intégrer au niveau local de la base de connaissances. Se pose alors le problème de la maintenance de ces métadonnées et du paramétrage des connecteurs qui, selon la structure des liens profonds, peut se révéler être une tâche ardue.

2.3 Proposition de solutions au niveau local

Face à ces problèmes, la bibliothèque mène un certain nombre de réflexions pouvant être rapidement mises en pratique au niveau local visant à lui permettre de reprendre le contrôle sur son offre de documents électroniques et améliorer la cohérence de l'offre existante.

2.3.1 Améliorer l'homogénéisation des métadonnées exposées dans le catalogue

À l'issue du travail d'analyse mené sur les facettes, et afin de mieux uniformiser l'indexation « sujet » dans PRIMO, la bibliothèque a décidé d'entamer un travail de normalisation des indexations matière des différents entrepôts. L'idée est de créer, à côté des indexations matière existantes, une indexation plus généraliste qui permettrait de couvrir la totalité des sources. Un mapping est en train d'être réalisé sur la base des libellés Dewey utilisés dans l'entrepôt BIRD. Au niveau des données de la base de connaissances, nous allons lier chaque entrée de la classification thématique SFX au libellé Dewey correspondant afin de réaliser une table de correspondance qui servira à PRIMO lors du moissonnage des notices pour produire un index thématique supplémentaire. Cet index sera la base d'une facette qui ne sera plus discriminante puisqu'elle permettra d'appliquer un filtre qui portera sur la totalité des ressources.

2.3.2 Gérer les ressources électroniques dans le SIGB ?

Face à la pauvreté des notices de documents présentes dans SFX, la question de la gestion des ressources électroniques dans le SIGB mérite d'être posée. Cette solution permettrait d'enrichir les notices grâce à des référentiels de catalogage maîtrisés et éprouvés et d'utiliser des formats et des vocabulaires documentaires communs. Elle offrirait la possibilité d'avoir une vision globale sur les collections puisque le même système de gestion donnerait accès à la totalité des collections. Collections papier et électroniques seraient mieux articulées.

Concernant le catalogage des documents, de nombreuses notices de revues et de livres électroniques sont produites dans le SUDOC. Il est donc possible de réaliser une économie de moyens en se localisant sur les notices existantes. La notice est importée dans le système de gestion local et suit donc le même circuit que les notices des documents papier.

Mais cette solution pose deux problèmes. La structure actuelle de notre SIGB ne permet pas de gérer les données de localisation. Les collections devraient donc être gérées dans le résolveur de liens SFX en plus de leur catalogage dans le SIGB. Ceci crée deux niveaux de gestion difficiles à articuler. L'autre difficulté vient de la nature mouvante de ces collections. En effet, un titre présent dans un bouquet peut disparaître au gré d'une mésentente entre éditeur et agrégateur et ainsi ne plus appartenir aux collections de la bibliothèque. La question se pose alors de savoir comment la bibliothèque est alertée de cette disparition et nécessite que la bibliothèque se délocalise dans le SUDOC et supprime la notice dans le SIGB. Ainsi, tant que le résolveur de liens n'est pas intégré au SIGB ou en liaison avec ce dernier, permettant ainsi une gestion unifiée des ressources, le catalogage des 100 000 titres électroniques possédés par la bibliothèque n'est pas envisageable.

3 Des projets nationaux et internationaux

Le volume important des métadonnées contenues dans les bases de connaissances et leur rôle capital pour le fonctionnement des outils de localisation et des portails de recherche nécessitent l'adoption de règles de traitement par tous les acteurs de la chaîne de constitution des bases de connaissances ainsi que des outils et des procédures de contrôle de leur qualité. Par ailleurs, la place des fournisseurs de logiciels documentaires dans le processus de traitement des métadonnées d'identification et de localisation des documents électroniques pose le problème de la dépendance des bibliothèques vis-à-vis de ces opérateurs privés. Face à cette situation, des projets de constitution de bases de connaissances nationales ou internationales ont vu le jour ces dernières années.

3.1 Des projets nationaux et internationaux visant à améliorer la qualité des métadonnées

3.1.1 Le projet KBART

Ces dernières années, la chaîne d'approvisionnement des ressources électronique a pris à bras le corps la question de l'intégration des ressources au sein des résolveurs de liens afin d'améliorer la qualité des liens OpenURL. Cependant, la question de la disponibilité de métadonnées pertinentes et à jour pour soutenir la création d' OpenURL via les bases de connaissances des résolveurs de liens reste d'actualité pour les acteurs du secteur.

C'est ainsi que le projet KBART est né, créé par l'UKSG et le NISO (the National Information Standards Organisation). KBART (Knowledge Bases and Related Tools) a été mis en place pour répondre à un rapport commandé par l'UKSG qui recommandait l'adoption de bonnes pratiques reconnues internationalement par les fournisseurs de contenu et les résolveurs de liens afin de rendre la fourniture de métadonnées de ressources électroniques plus précise, complète et efficace.

Le groupe à l'origine de ce projet était composé d'éditeurs, d'agrégateurs de contenu, de fournisseurs de résolveurs de lien, de bibliothèques et de consortia de bibliothèques. Il se voulait représentatif des acteurs de la chaîne d'approvisionnement qui participent activement à l'amélioration de la transmission des ressources électroniques.

Le groupe a publié son rapport en janvier 2010 sous le titre « NISO Recommended Practice » [38, UKSG KBART, 2010].

Les champs concernés sont l'identifiant, le titre, la date de couverture électronique, les embargos.

Ces pratiques recommandent également un format de données et des options de transmission pour les fournisseurs de résolveurs de liens.

De plus en plus, les consortia exigent dans leurs négociations avec les éditeurs que ces derniers leur fournissent des métadonnées conformes aux bonnes pratiques KBART. Par exemple, le JISC (l'agence nationale de licence et négociation e-deal au Royaume-Uni) exige la compatibilité KBART dans ses licences 2012.

Le groupe KBART est maintenant entré dans une deuxième phase de réflexion, se concentrant sur une redéfinition des règles originales afin de pouvoir y inclure certaines informations plus spécifiques concernant les e-books et les actes de conférence, les métadonnées spécifiques de consortium et les revues en open access. [35, PEARSON, 2011]

3.1.2 Qualinca

Qualinca (Qualité et interopérabilité de grands catalogues documentaires) est un projet de recherche fondamentale labélisé par l'ANR dans le cadre du programme Contenus Numériques et Interactions (CONTIN). Le consortium regroupe cinq partenaires complémentaires : deux acteurs nationaux majeurs des systèmes documentaires (L'Agence Bibliographique de l'Enseignement Supérieur (ABES) et l'Institut National de l'Audiovisuel (Ina)) et trois équipes de chercheurs en informatique (GraphIK (*Graphs for Inferences on Knowledge*), l'équipe IASI (*Artificial Intelligence and Inference Systems*) du Laboratoire de Recherche en Informatique de l'université Paris Sud et l'équipe GraphIK (*Graphs for Inferences on Knowledge*)).

Ce projet part du constat que les grands catalogues documentaires sont en train de passer de l'ère de la gestion des métadonnées dans des formats spécifiques aux bibliothèques à l'utilisation de standards du web.

Il vise à développer des mécanismes permettant de qualifier le niveau de qualité d'une base documentaire existante, d'améliorer ce niveau de qualité et de maintenir un niveau de qualité donné en contrôlant les opérations de mises à jour de ces bases.

Les objectifs scientifiques du projet sont de :

- développer un cadre logique pour qualifier la qualité d'une base de connaissances documentaire vis-à-vis de l'identification des entités individuelles et des liens entre entités individuelles ;
- proposer des principes/méthodes/outils pour passer d'un certain niveau de qualité à un meilleur niveau. Ces méthodes s'appuieront sur les approches logiques et numériques utilisées en réconciliation et fusion de références qui, d'une part, seront adaptées aux nouvelles problématiques engendrées par ces bases de connaissances documentaires et, d'autre part, seront étendues pour permettre la prise en compte de connaissances incertaines.

Les résultats de ces recherches pourront fournir aux bibliothèques et fournisseurs de résolveurs les outils et les méthodes nécessaires à l'évaluation de la qualité des métadonnées de leurs catalogues et de leur base de connaissances.

3.2 Des projets de bases de connaissances communautaires

3.2.1 Le résolveur allemand EZB

Le résolveur de liens allemand EZB qui s'appuie sur les données du catalogue collectif de revues électroniques du même nom offre un service de localisation au niveau national. Construit et mis à jour grâce à la collaboration de 430 bibliothèques, le catalogue sert à alimenter la base de connaissances sur laquelle le résolveur s'appuie.

Le service peut être utilisé par des bibliothèques qui ne possèdent pas de résolveur pour servir directement leurs utilisateurs finaux, mais le résolveur allemand peut être utilisé par des bibliothèques ayant leur propre système, soit en le déclarant comme cible de leur résolveur, soit en important les données de la base de connaissances nationale dans leur propre base de connaissances.

3.2.2 Le projet Knowledge Base+

Knowledge Base+ est un projet initié par le JISC Collections en août 2011 pour créer une base de connaissances nationale permettant une gestion plus efficace des ressources électroniques.

Les recommandations KBART et ONIX-PL ont servi de points de départ au travail de formalisation des trois types d'informations accessibles d'ores et déjà sous forme de fichiers publiés en CSV sur le site KB+³ :

- des données bibliographiques : titre, bouquet, plateforme, couverture pour de grands accords nationaux
- des informations d'abonnement : dates de renouvellement, périodes de préavis, etc.
- des informations sur les licences : conditions d'accès distant, walk-in users, accès post-désabonnement, partenariats, etc.

Le projet devrait aboutir courant 2014.

3.2.3 Le projet GOKb (global open knowledgebase, base de connaissances globale)

Le projet GOKb est le fruit d'une collaboration entre l'équipe Kuali OLE (équipe qui regroupe des bibliothèques américaines ayant pour objectif de livrer un ensemble logiciel pour les bibliothèques académiques et de recherche) et le JISC. Son objectif est de créer une base de connaissances utilisant une architecture normalisée et une licence *Creative Commons*.

Cet entrepôt de données librement accessibles contiendra les informations de publication essentielles sur les ressources électroniques comme elles sont représentées dans la chaîne de traitement entre les éditeurs, les fournisseurs de contenus et les bibliothèques. Il s'appuie pour cela sur trois principes :

³ <http://www.kbplus.ac.uk/kbplus/>

- Ouverture des données pour permettre leur usage et réutilisation ;
- Management communautaire pour encourager les bibliothécaires à contribuer, éditer et tenir à jour les données ;
- Collaboration avec les différents acteurs de la chaîne de traitement.

3.2.4 Les projets français

Le projet de plateforme nationale de signalement des ressources électroniques est né dans le cadre du projet Bibliothèque Scientifique Numérique. Ce programme piloté par les grands acteurs de l'Enseignement Supérieur et de la Recherche vise à améliorer l'offre documentaire nationale proposée aux chercheurs et la visibilité de la recherche française. Le projet de plateforme nationale de signalement des ressources électroniques vise à répondre aux recommandations fixées par son troisième segment d'activité (BSN 3) qui propose :

- « d'établir un continuum entre les outils existants de signalement et les plateformes d'accès au texte intégral
- de créer un dispositif fédérant le signalement de toute la documentation accessible sur le territoire et de faire le lien avec la ressource dans le cadre du contrôle des droits » [10, BSN, 2012]

Ces recommandations ont fait l'objet d'un plan d'action dans le cadre d'ISTEX. À côté de la plateforme ISTEX, dédiée au signalement des ressources acquises dans le cadre des licences nationales, le projet de signalement des ressources électroniques propose d'identifier les ressources courantes acquises par les établissements. L'objectif est donc de fournir un outil de découverte basé sur un index centralisé construit à partir du moissonnage des contenus des éditeurs et des agrégateurs ainsi que les ressources des bibliothèques (SUDOC, archives ouvertes).

En mars 2013, l'ABES a publié le résultat d'une étude sur le signalement national des ressources électroniques. Cette étude visait à explorer trois scénarios différents pour la réalisation d'un outil de découverte :

« Un scénario de réalisation interne, élaborant un service de découverte objet d'un développement nouveau : ce scénario s'appuie sur Trove, le moteur de recherche en ligne mis au point par la Bibliothèque nationale d'Australie. Dans ce cas, l'outil de découverte français ferait l'objet d'un développement nouveau utilisant des logiciels libres : les métadonnées et l'index du texte intégral seraient extraits au niveau de chaque éditeur. Les utilisateurs membres d'une bibliothèque abonnée à cette ressource bénéficieraient d'un accès immédiat. Pour les autres, un dispositif de distribution devrait être proposé.

Un scénario utilisant des outils de découverte présents sur le marché : dans ce cas, il s'agirait d'un outil de découverte de ressources national similaire qui reposerait sur une adaptation d'outils existants, comme Summon (Serial Solutions), Primo (Ex Libris), EBSCO Discovery (EBSCO) ou OCLC WorldCat local.

Un scénario qui établirait une collaboration avec Google Scholar : pour ce scénario, Google Scholar est l'élément central, proposant une interface publique ainsi que le moteur de recherche. L'accès au texte intégral s'effectuerait via un résolveur de liens . »

Parmi ces trois scénarios, l'ABES semble avoir adopté une voie médiane : celle d'articuler l'outil de découverte sur le hub de métadonnées et de développer une base de connaissances nationale.

3.2.4.1 Le hub de métadonnées

L'objectif initial du hub de métadonnées est de mettre les données sur le web après les avoir enrichies. Son objectif est de constituer une collection de corpus de métadonnées décrivant des ressources documentaires destinées à la communauté de l'enseignement supérieur et de la recherche. Les données concernées seront uniformisées via une collection de traitements automatiques ou semi-automatiques. Cet entrepôt permettra aux partenaires de l'agence de récupérer ou d'exploiter ces données. Le hub de métadonnées s'intéresse aux ressources électroniques acquises sur le territoire national (dont celles issues des licences nationales), mais aussi le papier, l'open access et le courant.

Le hub de métadonnées doit permettre :

- D'augmenter le signalement des ressources documentaires en multipliant les canaux et modes d'exploitation possibles des données ;
- De valoriser des ressources non-traitées par les bases de connaissances commerciales ;
- D'améliorer les métadonnées des éditeurs ;
- De multiplier le corpus lié à IDREF et donc au web de données.

Le hub constituera une plateforme de traitement qui permettra de diagnostiquer la qualité des métadonnées livrées, de convertir les formats et les vocabulaires, de nettoyer les données, les normaliser, les enrichir et les lier.

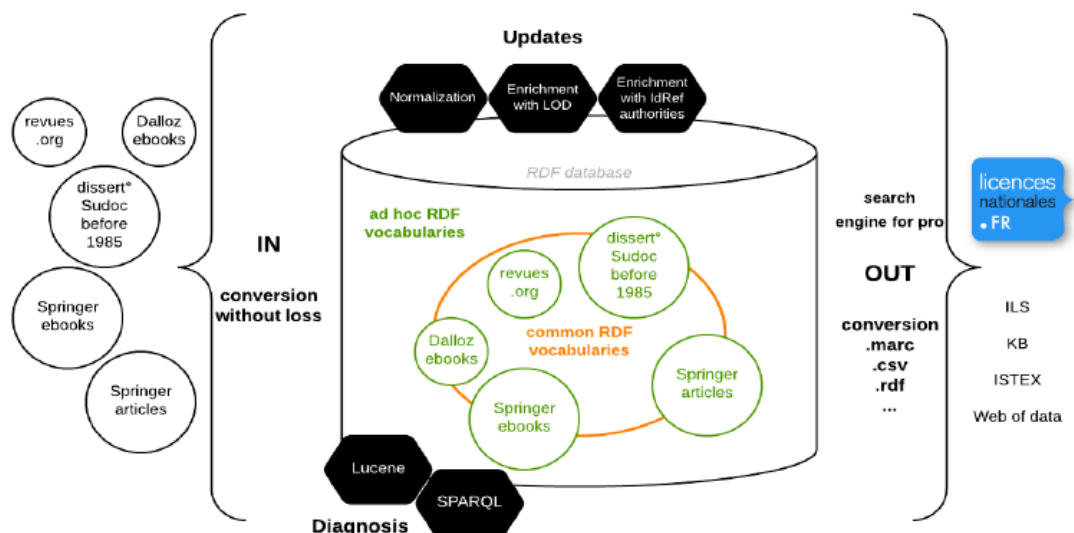


Figure 8 : Architecture du futur Hub de métadonnées

Par conséquent, la future plateforme de découverte s'appuiera sur ce hub de métadonnées pour fournir une interface de recherche unique sur un corpus de données nationales.

3.2.4.2 Une base de connaissances nationale

Faisant le constat qu'une solution de découverte, sans outil de localisation, serait inopérante, l'ABES propose de développer un résolveur de liens s'appuyant sur une base de connaissances nationale en s'inspirant du résolveur EZB pour la constitution de cette base de connaissances. Concernant le résolveur de liens, l'ABES propose d'explorer les options possibles pour se raccorder au projet GOKb.

Cette base de connaissances aura pour premier objectif de signaler l'ensemble des packages de ressources électroniques des établissements d'enseignement supérieur et de recherche. Pour chaque package, elle décrira les ressources et les titres (titre, identifiants, package, plate-forme) et permettra leur localisation (quelles bibliothèques donnent accès à ce titre, couverture pour chaque bibliothèque, licence).

Le projet ne vise pas l'exhaustivité, mais travaille sur des niches intéressant plus particulièrement la communauté académique française (par exemple les données des licences nationales).

Dans le contexte national, où un grand nombre de ressources électroniques sont acquises via le consortium Couperin, et certainement à l'avenir dans le cadre des licences nationales, une base de connaissances permettant de centraliser les informations sur le contenu, les états de collections et les licences relatives à chaque abonnement serait un grand gain pour la gestion des ressources électroniques dans les bibliothèques universitaires françaises.

Conclusion

Les bases de connaissances ont pris une importance croissante dans les dispositifs informatiques des bibliothèques universitaires françaises ces dernières années. Face à l'augmentation croissante des ressources électroniques, le système de gestion partagée entre fournisseurs de contenu, éditeurs de logiciel et bibliothèque semble être la seule alternative pour permettre aux bibliothèques de signaler, de manière optimale, leurs collections électroniques. Cependant, les bibliothèques universitaires, dans un contexte de manque de compétence technique en la matière, perdent la maîtrise des métadonnées produites dans les bases de connaissances.

Le projet de résolveur de liens national, s'il repose sur un système de catalogage et de signalement partagé des ressources comme c'est déjà le cas pour le SUDOC, permettra sans doute aux bibliothèques françaises de retrouver la maîtrise de leurs collections électroniques. L'arrivée sur le marché d'une nouvelle génération de systèmes de gestion de bibliothèque pourrait également produire les mêmes effets. Ces systèmes, conçus pour gérer conjointement les collections papier et électroniques, ne sont plus construits autour d'un catalogue, mais d'une base de connaissances. Ils seront le meilleur moyen de construire une base de gestion unique des ressources imprimées et numériques.

Reste que le volume important et croissant des métadonnées nécessaires à la gestion et au signalement des collections nécessitent un changement culturel au sein des bibliothèques. Les processus de catalogage devront être simplifiés et en grande partie automatisés. Les catalogueurs auront pour fonction moins de produire des métadonnées que de s'assurer de la qualité de celles-ci et deviendront, de fait, des gestionnaires de métadonnées.

Bibliographie

- [1]ADBS, Éd. Document électronique [En ligne]. Vocabulaire de la documentation. Disponible sur : < http://www.adbs.fr/document-electronique-16869.htm?RH=OUTILS_VOC > (consulté le 24 juin 2013)
- [2]APPS A., MACINTYRE R. « Why OpenURL? » D-Lib Magazine [En ligne]. mai 2006. Vol. 12, n°5,. Disponible sur : < <http://dx.doi.org/10.1045/may2006-apps> > (consulté le 3 juin 2013)
- [3]ARSENAULT C., PARÉ F.-X. « Les portails de bibliothèque: Nouvelles fonctionnalités, nouveaux défis ». 2005. Disponible sur : < <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/136> > (consulté le 16 juin 2013)
- [4]BARTHET E., LOUEMBÉ K. Projet ERMS partagé de Couperin : focus sur l'articulation des besoins du consortium et de ses membres [En ligne]. 2010. Disponible sur : < <http://catalogue-intd.cnam.fr/Record.htm?idlist=1&record=19100759124919289319> > (consulté le 18 août 2012)
- [5]BEIT-ARIE O., BLAKE M., CAPLAN P., FLECKER D., INGOLDSBY T., LANNOM L. W., MISCHO W. H., PENTZ E., ROGERS S., VAN DE SOMPEL H. « Linking to the Appropriate Copy ». D-Lib Magazine [En ligne]. septembre 2001. Vol. 7, n°9,. Disponible sur : < <http://dx.doi.org/10.1045/september2001-caplan> > (consulté le 23 mai 2013)
- [6]BERARD R. B. (DIRECTEUR DE. Le signalement des acquisitions numériques à l'échelle nationale : le rôle du hub de métadonnées scénarios et prototype [En ligne]. 16 novembre 2012. Disponible sur : < http://webcast.in2p3.fr/videos-le_role_du_hub_de_metadonnees > (consulté le 25 mai 2013)
- [7]BOTHMANN R. L., HLMBERG M. « Selecting, Acquiring and Renewing Electronic Resources ». In : YU H, BREIVOLD S, Éd. Electronic resource management in libraries research and practice. Hershey, PA : Information Science Reference, 2008. ISBN : 9781599048925 1599048922 1599048914 9781599048918.
- [8]BOTHMANN R. L., HLMBERG M. « Strategic Planning for Electronic Resource Management ». In : YU H, BREIVOLD S, Éd. Electronic resource management in libraries research and practice. Hershey, PA : Information Science Reference, 2008. ISBN : 9781599048925 1599048922 1599048914 9781599048918.
- [9]BREEDING M. Knowledge Base and Link Resolver Study: General Findings. [s.l.] : [s.n.], 2012. (Library Technology Guides).
- [10]BSN. Bibliothèque scientifique numérique : stratégie 2012-2015. [s.l.] : [s.n.], 2012.
- [11]CALENGE B. Bibliothèques et politiques documentaires à l'heure d'Internet. Paris, France : Éditions du Cercle de la librairie, 2008. 264 p.(Bibliothèques (Paris. 1978), ISSN 0184-0886). ISBN : 978-2-7654-0962-5.
- [12]CALHOUN K., CANTRELL J., GALLAGHER M., GAUDER B., LIMES R., CELLENTANI D., OCLC. Online catalogs: what users and librarians want: an OCLC report. Dublin, Ohio : OCLC, 2009. ISBN : 1556534116 9781556534119.

- [13]CARBONE P. Coûts, bénéfices et contraintes de la mutualisation des ressources électroniques: éléments de comparaison internationale et propositions [En ligne]. [s.l.] : [s.n.], 2010. Disponible sur : < <http://www.enssib.fr/bibliotheque-numerique/notice-48727> > (consulté le 1 juin 2013)
- [14]CASWELL J. V., WYNSTRA J. D. « Improving the search experience: federated search and the library gateway ». *Library Hi Tech* [En ligne]. 7 septembre 2010. Vol. 28, n°3, p. 391-401. Disponible sur : < <http://dx.doi.org/10.1108/07378831011076648> > (consulté le 6 juin 2013)
- [15]CAVALIER F. « Les acquisitions dans le contexte électronique ». *bbf*. 1 janvier 2006. Vol. 51, n°6, p. 62-65.
- [16]CHARTRON G. Evolution de l'édition scientifique, 15 ans après [En ligne]. 7 novembre 2007. Disponible sur : < http://archivesic.ccsd.cnrs.fr/sic_00186675 > (consulté le 15 juin 2013)
- [17]CHARTRON G., JOSEPH M., STOTZENBACH C., MINON M., BILBAULT R., GASNAULT J. « Une offre en pleine expansion ». *Documentaliste-Sciences de l'Information* [En ligne]. 2008. Vol. 45, n°2, p. 28. Disponible sur : < <http://dx.doi.org/10.3917/docsi.452.0028> > (consulté le 17 septembre 2012)
- [18]CHEN X. « Broken-Link Reports from SFX Users: How Publishers, Vendors and Libraries Can Do Better ». *Serials Review* [En ligne]. décembre 2012. Vol. 38, n°4, p. 222-227. Disponible sur : < <http://dx.doi.org/10.1016/j.serrev.2012.09.002> > (consulté le 4 juin 2013)
- [19]CLARAC C. « Les bases de connaissances à la croisée des chemins ». *Arabesques*. avril-mai-juin 2013. n°70,.
- [20]COLCANAP G. « Acquérir la documentation électronique pour l'enseignement supérieur et la recherche ». *BBF*. 1 janvier 2009. n°6, p. p. 18-21.
- [21]COUPERIN. Pilote ERMS : arrêt du projet.
- [22]DE ROSA C., CANTRELL J., CARLSON M., STURTZ C., GAUDER B., CELLENTANI D., DALRYMPLE T., OLSZEWSKI L. J., OCLC. *Perceptions of libraries, 2010: context and community: a report to the OCLC membership*. Dublin, Ohio : OCLC, 2011. ISBN : 1556533950 9781556533952.
- [23]DLF. *Electronic Resource Management. The Report of the DLF Initiative*. [En ligne]. [s.l.] : [s.n.], 2004. Disponible sur : < <http://old.diglib.org/pubs/dlf102/> > (consulté le 29 août 2012)
- [24]FAGAN J. C. « Federated Search Is Dead—and Good Riddance! » *Journal of Web Librarianship* [En ligne]. 2011. Vol. 5, n°2, p. 77-79. Disponible sur : < <http://dx.doi.org/10.1080/19322909.2011.573533> > (consulté le 25 mai 2013)
- [25]FRAZIER K. « What's the Big Deal? » *The Serials Librarian* [En ligne]. 23 mai 2005. Vol. 48, n°1-2, p. 49-59. Disponible sur : < http://dx.doi.org/10.1300/J123v48n01_06 > (consulté le 16 juin 2013)

- [26]GIBSON I., GODDARD L., GORDON S. « One box to search them all: Implementing federated search at an academic library ». Library Hi Tech [En ligne]. 6 mars 2009. Vol. 27, n°1, p. 118-133. Disponible sur : < <http://dx.doi.org/10.1108/07378830910942973> > (consulté le 6 juin 2013)
- [27]HANE P. J. « The Truth About Federated Searching ». InformationToday [En ligne]. octobre 2003. Vol. 20, n°10,. Disponible sur : < <http://www.infoday.com/it/oct03/hane1.shtml> > (consulté le 6 juin 2013)
- [28]HENRIET O., MALINGRE M.-L., SERRES A. Enquête sur les besoins de formation des doctorants à la maîtrise de l'information scientifique dans les Ecoles doctorales de Bretagne [En ligne]. [s.l.] : [s.n.], 2008. Disponible sur : < <http://www.enssib.fr/bibliotheque-numerique/notice-1793> > (consulté le 24 juin 2013)
- [29]JOLLY C. « Documentation électronique à l'université ». bbf. 1 janvier 2003. Vol. 48, n°4, p. 5-8.
- [30]KLEE L. « Couperin ». BBF. 1 janvier 2006. Vol. 51, n°1, p. 66-71.
- [31]MACHEFERT S. L'OpenURL dans les institutions françaises. Une chance pour la valorisation des ressources électroniques [En ligne]. Mémoires INTD. [s.l.] : [s.n.], 2007. Disponible sur : < <http://catalogue-intd.cnam.fr/Record.htm?idlist=21&record=19108358124919265309> > (consulté le 18 août 2012)
- [32]MAISONNEUVE M., TOUITOU C. Une nouvelle famille d'Opac [En ligne]. [s.l.] : [s.n.], 2007. Disponible sur : < <http://bbf.enssib.fr/consulter/bbf-2007-06-0012-002> > (consulté le 26 mai 2013)
- [33]MARTINE COPPET (DIRECTRICE-ADJOINTE DU SERVICE COMMUN DE DOCUMENTATION R. DU D. DU S. D'INFORMATION D. profiter des bases de connaissances [En ligne]. [s.l.] : [s.n.], 2012. Disponible sur : < <http://webcast.in2p3.fr/videos-4001> > (consulté le 25 mai 2013)
- [34]NICOLAS MORIN. Du signalement à la KB [En ligne]. Disponible sur : < <http://fr.slideshare.net/nicomor/du-signalment-la-kb-11100283> > (consulté le 1 juin 2013)
- [35]PEARSON S. « The Role of Knowledge Bases in Improving Discoverability Now and in the Future – Why National and International Collaboration Is Key. » [s.l.] : [s.n.], 2011. Disponible sur : < http://concert.stpi.narl.org.tw/Images/upload/CONCERT2011/PPT/Sara_Abstracts.pdf > (consulté le 18 août 2012)
- [36]RÉSEAU DES BIBLIOTHÈQUES DE L'UNIVERSITÉ DE MONCTON. Les outils de découverte : Rapport du groupe de travail du Réseau des bibliothèques de l'Université de Moncton. [s.l.] : Réseau des bibliothèques de l'Université de Moncton, [s.d.].

[37]STOHN C., EWING S., MEARES S., MOSS P. « Building and Maintaining Knowledge Bases for Open URL Link Resolves - Processes, Procedures, and Challenges ». Against the Grain. février 2011. Vol. 23, n°1, p. 26.

[38]UKSG KBART. KBART: Bases de connaissances et outils associés/Rapport préparé par le groupe de travail NISO/UKSG KBART Janvier 2010 [En ligne]. [s.l.] : United Kingdom Serials Group, [s.d.]. Disponible sur : < http://www.couperin.org/images/stories/KBART/kbart_couperin_fr.pdf > (consulté le 23 mai 2013)

[39]VAN DE SOMPEL H., BEIT-ARIE O. « Open Linking in the Scholarly Information Environment Using the OpenURL Framework ». D-Lib Magazine [En ligne]. mars 2001. Vol. 7, n°3,. Disponible sur : < <http://dx.doi.org/10.1045/march2001-vandesompe> > (consulté le 23 mai 2013)

[40]VILLENEUVE A. Réflexion sur la mise en place d'un système de gestion de ressources électroniques (type ERMS): exemple de la Bpi [En ligne]. Mémoires INTD. [s.l.] : [s.n.], 2010. Disponible sur : < <http://catalogue-intd.cnam.fr/Record.htm?idlist=14&record=19111062124919392449> > (consulté le 18 août 2012)

[41]WAKIMOTO J. C., WALKER D. S., DABBOUR K. S. « The Myths and Realities of SFX in Academic Libraries ». The Journal of Academic Librarianship [En ligne]. mars 2006. Vol. 32, n°2, p. 127-136. Disponible sur : < <http://dx.doi.org/10.1016/j.acalib.2005.12.008> > (consulté le 4 juin 2013)

[42]WALKER J. « New resource discovery mechanisms (2) ». In : The E-Resources Management Handbook [En ligne]. [s.l.] : [s.n.], 2008. Disponible sur : < <http://dx.doi.org/10.1629/9552448-0-3.8.2> > (consulté le 30 août 2012)

[43]YANG S. Q., WAGNER K. « Evaluating and comparing discovery tools: how close are we towards next generation catalog? » Library Hi Tech [En ligne]. 23 novembre 2010. Vol. 28, n°4, p. 690-709. Disponible sur : < <http://dx.doi.org/10.1108/07378831011096312> > (consulté le 6 juin 2013)

[44]« Le classement Livres Hebdo 2012 de l'édition mondiale ». Livres Hebdo. 22 juin 2012. n°915,.