



HAL
open science

Participation au travail de recherche, de définition et de réalisation d'une offre de gestion des périodiques électroniques dans le cadre de la mise en place d'un portail documentaire par la société Archimed

Paul Belhouchat

► To cite this version:

Paul Belhouchat. Participation au travail de recherche, de définition et de réalisation d'une offre de gestion des périodiques électroniques dans le cadre de la mise en place d'un portail documentaire par la société Archimed. domain_shs.info.edit. 2004. mem_00000445

HAL Id: mem_00000445

https://memic.ccsd.cnrs.fr/mem_00000445v1

Submitted on 22 Mar 2007

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.

DESS Réseaux d'information
et document électronique

**Participation au travail de recherche, de
définition et de réalisation d'une offre de
gestion des périodiques électroniques dans
le cadre de la mise en place d'un portail
documentaire par la société Archimed.**

Paul Belhouchat

Sous la direction de Omar Larouk
Maître de Conférences, Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques (Enssib)

Résumé :

L'évolution des modes de recherche documentaire, associée aux progrès de l'informatique, des réseaux, du web et d'une offre pléthorique de contenus électronique voit la naissance actuellement de besoins informationnels plus précis, orientés vers la recherche fédérée. Les communautés d'étudiants, de chercheurs et globalement tous les utilisateurs d'un système d'information documentaire souhaiteraient avoir la possibilité de réaliser une recherche unique en interrogeant à la fois le catalogue de la bibliothèque où l'on se trouve ainsi que son système de GED, les catalogues d'autres bibliothèques, mais également des ressources distantes et hétérogènes comme des bases de données bibliographiques, factuelles ou en texte intégral, enfin des bases de signets ou des sites web qui pourraient également apporter un complément à une recherche initiale.

Ce rapport fait l'état des nouvelles technologies disponibles et maîtrisées par Archimed afin de permettre la recherche fédérée et présente une analyse de l'offre de portail documentaire imaginé par Archimed. Ce travail donne aussi une vue d'ensemble de ce qui est proposé par les acteurs de l'édition électronique – commerciaux ou libres -, des technologies qu'ils utilisent, des projets voire des partenariats qu'ils ont en direction des portails et de la recherche fédérée.

Descripteurs : Recherche fédérée, Portail documentaire, Protocoles de recherche et de récupération de l'information, Services Web, Z39.50, SRW/SRU, OAI-PMH, OpenURL, édition électronique, mouvement des archives ouvertes, SIM, MASC, Book-Line.

Abstract :

The progress of computer sciences and networks associated to the explosion of the web and an exponential growth of electronic edition has seen an evolution of search behaviour. The different communities of students, researchers and all users of documentary information system would like to use a search tool which offers the opportunity to make federated search (or metasearch). Thank to such a tool, it would be possible to search and retrieve at the same time documents or references in our library catalogue and its internal content management system, in other libraries catalogues, or reach and interrogate remote and heterogeneous resources and databases like bibliographic databases or resources available in full-text. This tool could offer the possibility to complete a initial search by providing web resources like subject gateways.

This report try to draw up a comparative study of the technologies available in order to set up a federated search tool. It analyses Archimed's software offer as regard in documentary portals with integrated metasearch engine. Lastly, this work gives an overview of the position of the protagonists of electronic edition (commercial or open access, as well as their technologies, projects and partnership in connection with federated search and the question of unique access point in order to reach remote and heterogeneous databases and resources.

Keywords : Federated search, metasearch, Portal, information search and retrieval protocols, Web Services, Z39.50, SRW/SRU, OAI-PMH, OpenURL, electronic edition, Open Archive Initiative, SIM, MASC, Book-Line.

Toute reproduction sans accord express de l'auteur à des fins autres que strictement personnelles est prohibée.

Sommaire

INTRODUCTION 7

PARTIE 1 : LES ÉVOLUTIONS DE LA RECHERCHE DOCUMENTAIRE ET LES SOLUTIONS PROPOSÉES PAR ARCHIMED 9

1. Le développement des portails documentaires	9
1.1. Un bref historique : des OPAC aux solutions full web	9
1.2. L'avantage des portails : regrouper un ensemble d'activités au sein d'un seul service	10
2. L'évolution des formes de ressources et de recherche	12
2.1. La recherche catalogue monobase ou multibase	12
2.2. La recherche de ressources hétérogènes locales	12
2.3. La recherche de ressources hétérogènes distantes	13
3. Documentation et formation aux progiciels Archimed	14
3.1. Un important travail de documentation.....	15
3.2. ...et de formation sur les applications.....	15
4. La question des protocoles d'interrogation en vue de la recherche fédérée : évolutions et promesses	16
4.1. Un protocole ancien mais toujours populaire : Z39.50	17
4.1.1. Historique de la norme.....	17
4.1.2. Une norme efficace mais complexe.....	18
4.1.3. Les principaux "services" de Z39.50	19
4.1.4. Les services étendus.....	20
4.1.5. Les profils Z39.50	21
4.1.6. Les formats de réponse.....	22
4.1.7. Les évolutions de Z39.50 : révision du Bib-1 et <i>Bath Profile</i> ..	23
4.1.8. Bilan de Z39.50	24
4.1.9. Schémas explicatifs	25
4.1.10. Principe d'une session Z39.50.....	25
4.1.11. Session Z39.50 et accès multibase	26
4.2. Le standard d'avenir : les Web Services SRW et SRU.....	28
4.2.1. Description du protocole.....	28
4.2.2. Les apports de SRW : HTTP, CQL et XML.	29
4.2.3. Les types de services.....	29
4.3. Un protocole de rebond par URL : OpenURL.....	31
4.3.1. Présentation.....	31
4.3.2. Composition de l'OpenURL	32
4.3.3. Fonctionnement	33
4.3.4. Schéma explicatif de l'OpenURL	35
4.4. Un protocole de métadonnées dédié aux ressources libres : OAI-PMH. 35	
4.4.1. Data Provider et Services Provider	36
4.4.2. Types de requêtes proposées par OAI-PMH	37
4.4.3. Schéma explicatif du protocole OAI	38
4.5. Le souci d'Archimed de faciliter la mise en œuvre et l'utilisation optimale de ces protocoles	39
4.5.1. Le serveur Z39.50 et le connecteur SRW de Book-Line.....	39
4.5.1.1. Schéma du fonctionnement du connecteur SRW.....	40
4.5.2. Présentation du module SARA	42

4.5.2.1. Schéma du fonctionnement de SARA.....	43
4.5.3. Le résolveur OSC-R	44
4.5.4. Le moissonneur Harvey	45
5. Le portail documentaire construit selon Archimed : une offre packagée	46
5.1. La base documentaire multimédia : SIM.....	46
5.2. Le métamoteur de recherche : Book-Line	47
5.2.1. Schéma du fonctionnement de Book-Line	48
5.3. La gestion de portail : MASC.....	49
5.4. Bilan : la proposition d'une offre globale	50
6. Un exemple représentatif de portail de nouvelle génération en France : le portail Sudoc	52
6.1. Historique et prérogatives de l'agence	52
6.2. Genèse du projet	53
6.3. Les outils caractérisant le portail	53
6.4. Les contenus du portail.....	54
6.5. Gestion des formats, des documents et des connexions	54
6.6. Intérêts d'un tel portail.....	55
7. Bilan : Le choix d'Archimed de s'investir dans les nouveaux protocoles en étant implémenteur <i>de facto</i>	56

PARTIE 2 : MÉTHODE ET ORGANISATION DES RECHERCHES 59

1. Problématique du stage : la recherche fédérée et la gestion des périodiques électroniques	60
1.1. Premier constat : l'expression des besoins en matière de recherche fédérée.....	60
1.1.1. Un besoin inconscient chez les étudiants et les novices de la recherche documentaire.....	60
1.1.2. ... mais un besoin réel et exprimé par la communauté des chercheurs	61
1.1.3. Un besoin exprimé par les universités et les bibliothèques en terme de rentabilité de leurs abonnements.....	61
1.2. L'état de l'art sur les solutions mises en œuvres actuellement :	62
1.2.1. Le portail de la bibliothèque universitaire de Paris Dauphine : les limites du revamping HTML	62
1.2.2. Ebauche d'une solution de gestion collective des périodiques électronique : l'APE.....	63
1.2.3. Le Crossref Pilot Search : l'homogénéité de la recherche mais une hétérogénéité dans la présentation des résultats	64
2. Mise en place d'une veille informative, documentaire et concurrentielle	65
2.1. Observation de l'environnement concurrentiel	65
2.2. Veille informative et documentaire	67
2.2.1. Les revues et les sites référentiels en sciences de l'information	67
2.2.2. Weblogs bibliothéconomiques, listes de diffusion et pages professionnelles.....	68

3. Faisabilité de la recherche fédérée : l'attitude des acteurs de l'édition électronique	69
3.1. Du côté des acteurs commerciaux.....	69
3.1.1. La recherche fédérée entre-t-elle dans les intérêts des éditeurs/diffuseurs ?	69
3.1.2. Méthode d'approche pour une mise en contact	70
3.1.2.1. <i>Connaître les technologies utilisées actuellement et éventuellement les projets</i>	71
3.1.2.2. <i>Les sentiments et les projets des acteurs vis-à-vis de la recherche fédérée</i>	72
3.2. Du côté des acteurs de l'Open Access	73
3.2.1. Une vision différente liée au mouvement OAI	73
3.2.2. Consultation de Francis André sur l'Open Access et l'INIST ...	74
3.2.3. Entretien avec Daniel Charnay sur le CCSD.....	75
3.3. Bilan : mise en place d'une base de connaissance.....	77
3.4. Tableau statistique sur l'utilisation des protocoles.....	78
4. Conclusion : perspectives de nouvelles discussions et négociations entre les bibliothèques, les éditeurs de contenus, les éditeurs informatiques et les consortiums	79
CONCLUSION	81
BIBLIOGRAPHIE.....	83
TABLE DES ANNEXES.....	90

Remerciements

Nombreuses sont les personnes auxquelles je tiens à remercier dans le cadre de ce stage d'application :

Eric Ruyffelaere et Olivier Walbecq, responsables de la société Archimed pour m'avoir confié l'étude de cette problématique de la recherche fédérée et des nouveaux protocoles de recherche et de récupération de l'information, problématique associée à la gestion des périodiques électroniques.

Christophe Arnoult, consultant en systèmes d'information chez Archimed pour m'avoir encadré tout au long de ce stage.

Omar Larouk, maître de conférences à l'Enssib, pour son suivi et sa disponibilité.

Laurence Anouilh-André, formatrice chez Archimed et membre du bureau Rhône-Alpes de l'ADBS, pour ses conseils, sa gentillesse et ses prises d'initiatives en organisant deux déplacements d'observation à la Bibliothèque Municipale de Roanne et au SCD de l'Université Lyon 3.

Toute l'équipe de l'agence lyonnaise d'Archimed, ainsi qu'à Paul-Marie Druésne, ingénieur au siège de Lille pour son encadrement tout au long de mes investigations.

Isabelle Suchel-Mercier, conservatrice à la Bibliothèque Municipale de Roanne ainsi qu'à toute son équipe pour m'avoir reçu dans le cadre d'une journée d'étude sur le cycle du document et sur le catalogage.

Jean Bernon, conservateur et directeur du SCD de l'Université Lyon 3 et Jean-Loup Bruschet, ingénieur d'études pour m'avoir accueilli au cours de deux journées consacrées à la gestion des périodiques électroniques en bibliothèque universitaire.

Francis André, responsable du service "prospective et partenariats" et Fabrice Lecocq, ingénieur informatique à l'INIST, ainsi que Daniel Charnay, directeur adjoint du CCSD et chercheur à l'IN2P3 pour avoir pris le temps de m'accorder un entretien sur la présentation de leurs instituts et sur les enjeux des modèles alternatifs à l'édition commerciale au sein de l'édition électronique.

Enfin, je remercie tous mes interlocuteurs français et internationaux des diverses sociétés, instituts, organismes et bibliothèques qui ont bien voulu répondre à mes questions sur les nouveaux enjeux de la recherche documentaire et de l'édition électronique.

Introduction

Ces dernières années ont été marquées par un développement rapide des technologies informatiques et par l'explosion de l'internet et du web qui ont pris une place considérable dans notre quotidien. Cette place est également devenue significative dans le monde des bibliothèques et de toutes les entreprises qui ont à gérer, stocker ou rechercher de l'information. Ces progrès ont vu l'émergence ou plutôt l'évolution de nouveaux désirs des utilisateurs d'un système d'information, que l'utilisateur soit un étudiant, un chercheur ou un ingénieur. Ils souhaitent évidemment retrouver des références pertinentes résultant de leurs requêtes et récupérer des documents en texte intégral ou des liens vers des références complémentaires à leur recherche. Ils souhaitent davantage de confort dans leur travail en menant ces recherches, tout en évitant d'interroger chacune des ressources disponibles dans un portail donné. C'est-à-dire qu'ils veulent travailler sur plusieurs bases en même temps comme s'ils interrogeaient une seule et même base et récupérer, par exemple, les articles en texte intégral des revues auxquelles leur portail est abonné sans forcément passer par l'interface de recherche d'un éditeur.

Créée en 1993, la société Archimed est spécialisée dans l'intégration de solutions progiciels dans le domaine des nouvelles technologies de l'information et de la communication. Outre une offre dans le domaine de la gestion des impressions, de la sécurité et du E-learning, Archimed se pose parmi les leaders sur le marché de la gestion de l'information et de la gestion des portails. Souhaitant répondre aux besoins des usagers des bibliothèques, qu'elles soient municipales ou universitaires, Archimed a mis en place une offre de portail permettant de réaliser des recherches fédérées.

Mon stage s'inscrit donc dans cette question de l'évolution des modes de recherches documentaires et de la recherche fédérée ; question associée à la gestion des périodiques électroniques.

Je vais donc présenter dans une première partie l'évolution qu'a connu la recherche documentaire et les aspects techniques qui permettent de réaliser ces recherches en observant l'évolution et l'apparition de nouveaux protocoles de recherche et de récupération de l'information. Dans un second temps, je vais aborder l'offre de portail documentaire proposée par Archimed et leur choix quant à l'intégration des nouveaux protocoles de recherche permettant d'interroger des bases hétérogènes.

Une seconde partie va porter sur la réflexion et la faisabilité de la recherche fédérée grâce à une enquête que j'ai menée auprès de différents acteurs de l'édition électronique, commerciaux ou issus de l'*Open Access* afin d'avoir leur opinion sur les nouveaux besoins des communautés de chercheurs et d'utilisateurs de toutes sortes et sur les partenariats qu'ils pourraient ou qu'ils ont déjà envisagé en direction du développement de portails de ce type. Ce travail pourra donc servir de base à Archimed à l'intégration de contenus en ligne dans un portail documentaire. Il permet également de comprendre les positions des différents protagonistes de l'édition électronique.

Partie 1 : Les évolutions de la recherche documentaire et les solutions proposées par Archimed

1. Le développement des portails documentaires

1.1. Un bref historique : des OPAC aux solutions full web

L'offre de services proposée par les bibliothèques, médiathèques, bibliothèques universitaires, centres de documentation d'entreprise a considérablement évolué ces dernières années. Ces développements furent évidemment possibles grâce à l'évolution rapide de l'informatique parallèlement aux progrès réalisés dans le domaine des réseaux de communication avec notamment l'arrivée du web au début des années 90.

L'idée d'associer le monde de l'informatique et des réseaux à celui des bibliothèques n'est pas une idée nouvelle. Seulement les technologies ont mis du temps à se mettre véritablement en place ; il faut également reconnaître qu'un bibliothécaire ou un conservateur n'est pas un informaticien et vice-versa ; c'est un choc des cultures professionnelles qui commence petit à petit à porter ses fruits.

Les premiers OPAC (*Online Public Access Catalogue*) remontent à quelques dizaines d'années en arrière, au temps où les plus grandes structures ont informatisé leurs catalogues papier. Il y a encore une dizaine d'années, ce type de service qui apportait l'avantage du gain de temps dans ses recherches connaissait également son lot de désagréments : postes de consultation en nombre insuffisant, lenteur du système, mais surtout une interface spartiate et peu conviviale. A l'origine, l'OPAC est un outil qui permet de consulter un catalogue bibliographique géré par un SIGB (Système Intégré de Gestion de Bibliothèque). L'OPAC peut être défini par ses trois fonctions de base : la recherche d'un

document à partir d'éléments bibliographiques, la localisation des exemplaires, la disponibilité de ces derniers.

Ce n'est vraiment qu'au milieu des années 90 que la recherche documentaire va prendre un nouvel élan grâce au développement d'internet et du web. Caractérisé par son côté attractif et convivial, le web a rapidement incité les bibliothécaires à proposer des sites web de leur bibliothèque, même si ce fut dans un premier temps, comme le rappelle N. Morin, davantage pour suivre la modernité que pour réellement proposer des services particuliers¹.

Il aura donc fallu laisser passer quelques années pour qu'on réalise que le web ainsi que les progrès dans le transfert des données (arrivée du haut débit) permettront aux bibliothèques de proposer des services qui allaient au-delà d'un simple OPAC, même si ce dernier pouvait déjà contenir des services étendus. Notons que cette réflexion n'est pas uniquement celle des bibliothécaires, mais aussi de toutes les entreprises qui doivent gérer de l'information. Cette réflexion, c'est celle des portails documentaires.

1.2. L'avantage des portails : regrouper un ensemble d'activités au sein d'un seul service

Avant d'aborder le sujet des portails documentaires proprement dit, notons de suite l'existence d'une typologie des portails qui sont proposés selon les besoins d'une institution ou d'une entreprise. En général, on distingue cinq types de portails : documentaire, gestion de contenu, collaboratif, décisionnel et applicatif². Dans le cadre du stage et de l'offre proposée par Archimed, la notion de portail documentaire que j'ai eu à approcher ne correspond pas vraiment avec cette typologie dans la mesure où le portail documentaire vu par Archimed touche également à la gestion de contenu (*content management*) ainsi qu'à l'aspect

¹ MORIN N., Contenus et services des sites web des bibliothèques [en ligne] BBF, 2003, t.48, n°4, p.9-13. Disponible sur : <<http://morinn.free.fr/pdf/bbf01.pdf>>

² MEINGAN D., Les portails d'entreprise et la gestion des connaissances. 3/01/2002. http://www.knowledgeboard.com/library/frenchdocs/kbcf_valtech_portails-et-km.doc Concernant la distinction entre le portail documentaire et le portail de gestion de contenu, on peut dire que le premier ajoute à la gestion documentaire un outil de recherche plus ou moins élaboré, un outil de catégorisation et un accès personnalisé. Le second gère l'ensemble du cycle de publication de contenu.

collaboratif. C'est une vision qui vise à répondre aux attentes des bibliothécaires dans le cadre de ce qu'ils souhaitent offrir à leurs usagers.

Brièvement, un portail documentaire doit être caractérisé par des fonctionnalités de base :

- C'est un point d'accès unique aux données et aux services proposés par la bibliothèque ou le centre de documentation.
- Il possède un outil de recherche et/ou d'indexation sur les bases constituant le système d'information de la bibliothèque ou du centre de documentation (SIGB), mais aussi pouvant travailler sur les sources externes telles que des catalogues d'autres bibliothèques, des ressources en lignes diverses.
- Possibilité d'identifier les utilisateurs par la gestion des habilitations et des droits d'accès, pour la personnalisation de la présentation du portail et des services.
- Fonctionnalités et offre de services des OPAC classiques : réservation, panier, consultation du dossier de lecteur.
- Possibilité d'enregistrer les requêtes effectuées, diffusion sélective de l'information.
- Ajoutons également des outils de communication : forums, votes, messages du centre de documentation vers les usagers, etc.
- Espace d'administration, réservé aux bibliothécaires
- Outils statistiques.

De l'OPAC simple au portail documentaire, l'évolution de l'offre de services peut-être vue comme une véritable révolution, bien que toutes les potentialités d'un portail ne soient pas encore exploitées. On pense surtout aux expositions virtuelles, les aspects e-learning qui pourraient faire leur apparition, l'exploitation de dossiers documentaires, l'organisation d'un *chat* sur un thème défini, etc.. N. Morin souligne bien les avantages des portails pour les usagers : la bibliothèque est désormais ouverte pratiquement 24h/24, elle propose des ressources plus vastes,

au-delà de la simple consultation d'un ou plusieurs catalogues et ses ressources sont plus facilement accessibles³.

Avec le développement des portails, on a donc une diversité des ressources qui les composent. Face à cette offre qui peut être souvent pléthorique, nous verrons qu'un besoin va se manifester en direction des outils de recherche fédérée. Le développement de nouvelles formes de recherche est plus ou moins parallèle avec le développement des portails.

2. L'évolution des formes de ressources et de recherche

Forts de ses dix ans d'expérience, Archimed et ses chefs de projets en systèmes d'information documentaire ont pu constater l'évolution des modes de recherche. Selon leur expérience, on peut repérer trois temps ou trois "ères" dans la recherche documentaire ainsi que deux types de recherche documentaire.

2.1. La recherche catalogue monobase ou multibase

Nous sommes dans la vision d'un OPAC classique : aux premiers jours des OPAC, la possibilité était de faire une interrogation au niveau local sur un ou plusieurs catalogues. Puis, avec l'ouverture des réseaux, l'intérêt de l'OPAC a consisté à interroger également de façon distante les catalogues d'autres bibliothèques, via le protocole Z39.50. Ce type de recherche multibase et/ou multicatalogue était déjà une première forme de recherche fédérée.

2.2. La recherche de ressources hétérogènes locales

³ MORIN N., [Contenus et services des sites web des bibliothèques en ligne] BBF, 2003, t.48, n°4, p.9-13. Disponible sur : <<http://morinn.free.fr/pdf/bbf01.pdf>>. Il souligne également que le développement du web et de portails dématérialisent la bibliothèque et donc fait concurrence à ce qu'il appelle la "bibliothèque physique". Enfin, il note une évolution dans la culture professionnelle des bibliothécaires qui doivent s'adapter désormais à une notion de service plus élargie que celle qu'il pouvaient aborder jusqu'alors.

A cette première ère de la recherche documentaire, s'est ajouté par la suite l'apport de nouvelles ressources propres à la bibliothèque comme par exemple un fonds ancien numérisé ou même diverses ressources qu'elles soient sous forme de textes, d'images ou de vidéos, mais aussi des bases de signets. On aura donc d'une part le catalogue de la bibliothèque qui est géré par le SIGB et, d'autre part, des ressources diverses qui seront stockées dans un système de gestion électronique de documents (GED). La problématique de cette seconde étape était donc d'offrir une forme de recherche unique et complètement transparente à l'utilisateur, ce qui lui donnerait l'impression d'interroger une seule source en interrogeant pourtant le catalogue et le système de GED.

Notons toutefois une étape intermédiaire entre le catalogue classique et le système de GED : il s'agit de l'enrichissement multimédia des notices. La forme principale de cet enrichissement est tout simplement le lien 856 du format UNIMARC, c'est-à-dire l'URL de la ressource dans la notice du catalogue. Archimed avait mis en application un module de gestion multimédia intégré à Book-Line qui ne sert pas uniquement à saisir les liens en rapports avec chaque résultat. Ce module est un outil de stockage, de description et de recherche des ressources multimédias elles-mêmes, pour les relier ensuite à n'importe quel résultat de n'importe quelle base interrogée et sans intervention sur les bases elles-mêmes. On va donc au delà du rôle joué par le lien 856 puisqu'on propose d'enrichir sa recherche, une évolution qui sera surtout remarquée avec l'OpenURL.

2.3. La recherche de ressources hétérogènes distantes

Cette troisième "ère" dans la recherche documentaire est celle que nous vivons actuellement. Depuis deux ou trois ans, est apparue la problématique des ressources électroniques issues pour la plupart de bases de données commerciales. Le principe de recherche n'est plus tout à fait le même puisqu'il s'agit d'attaquer une base de documents primaires distante et non plus une base issue d'un système de GED local. On est donc dans une architecture où la recherche peut porter sur le catalogue de la bibliothèque (mais aussi ceux de ses voisins), le système de GED

de la bibliothèque et les ressources hétérogènes distantes. Contrairement aux catalogues, ces bases ne sont pas normalisées de manière unique.

La recherche fédérée devient ici assez problématique, mais pas seulement du fait de l'hétérogénéité des formats, mais également des sources à interroger puisque dans ce type de cas, l'interrogation ne se fera pas forcément qu'en Z39.50.

Cette recherche porte donc sur des documents primaires et sur des références, le but est d'obtenir par exemple le texte intégral de la revue électronique à laquelle la bibliothèque est abonnée et non plus seulement une notice bibliographique extraite d'un catalogue. Cette recherche se déroule dans des environnements différents puisqu'on n'attaque plus uniquement des catalogues, mais des bases de données distantes, caractérisées par des protocoles d'interrogation hétérogènes.

Le projet d'Archimed est donc de pouvoir mettre en place une solution de recherche fédérée et donc complètement transparente à l'utilisateur en étant capable d'interroger de façon homogène un ensemble hétérogène composé d'un catalogue classique, d'un système de GED et de bases distantes.

3. Documentation et formation aux progiciels Archimed

Dans le cadre de son offre de portail documentaire, Archimed a élaboré une architecture technique basée sur la superposition de briques logicielles distinctes. Chaque brique tient un rôle bien précis : la gestion d'un système d'information multimédia pour le SIM, la recherche fédérée via Book-Line et la gestion de portail grâce à MASC. Chaque logiciel est bien sûr indépendant des autres et peut être utilisé seul.

A côté des applications *stricto sensu*, Archimed a également développé des outils permettant d'exploiter au mieux la puissance des protocoles d'interrogation pour les bases distantes et hétérogènes. Outre un connecteur SRW de base dans Book-Line, Archimed a choisi de développer un connecteur nommé SARA qui permet d'interroger des bases de données externes en s'adaptant à leur langage, tout en interprétant cette connexion dans une couche SRW. Un module nommé Harvey permet le moissonnage des métadonnées OAI.

Nous verrons en détail cette architecture en §4 et §5, une architecture certes complexe de part les protocoles et les mécanismes qu'elle met en jeu, mais qui a l'avantage d'être assez claire grâce à cette imbrication de logiciels et de processus.

3.1. Un important travail de documentation...

Les premières semaines de mon stage ont donc été consacrées à une mise à niveau sur les aspects techniques mis en jeu par Archimed dans le cadre de son offre. A cet égard, la consultation de l'ensemble de la documentation technique des logiciels et des connecteurs concernés par la mise en place d'un portail documentaire m'a été accordée.

Dans mon cas, la consultation a porté plus particulièrement sur SIM 1.3, Book-Line 2.6 et MASC 2.5. La lecture de chaque document m'a permis de mieux appréhender le rôle de chaque logiciel ainsi que leur paramétrage.

Par contre, la mise à niveau effectuée par rapports aux protocoles d'interrogation a été plus délicate car la littérature disponible sur le sujet correspond plus à un public d'ingénieurs informaticiens qu'à celle d'un spécialiste de la gestion de l'information. Ici aussi, Archimed m'a donné l'accès à la documentation du connecteur SRW de Book-Line, du connecteur SARA ainsi que de leur résolveur de lien OSC-R. Mon travail a été complété par la consultation de sites spécialisés sur ces protocoles, notamment celui du ZING⁴ ou du NISO⁵, mais aussi via des présentations PowerPoint de divers spécialistes de ces protocoles, comme Hubert Van de Sompel pour l'OpenURL.

3.2. ...et de formation sur les applications

Au cours de cette phase de mise à niveau et de documentation, j'ai pu également bénéficier d'une formation de quatre jours sur le métamoteur de recherche Book-Line 2.6. Cela m'a permis de bien cerner ce qu'est la norme Z39.50 en paramétrant la connexion à une source de ce type, ce qui fut également

⁴ Z39.50 INTERNATIONAL NEXT GENERATION (ZING). Disponible sur : <<http://loc.gov/z3950/agency/zing/>>

⁵ NISO, The OpenURL Framework for Context-Sensitive services [en ligne]. 2003. Disponible sur: <http://www.niso.org/committees/committee_ax.html>

le cas pour une source OpenURL. J'ai pu également me familiariser avec le catalogage Unimarc par le biais du paramétrage du format d'affichage des notices.

Enfin, il y a eu toute une phase de paramétrage sur les index de recherche, les clés de tri et de dédoublonnage, le traitement des zones codées et enfin la création des différents formulaires de recherche.

Dans le cadre de cette formation sur Book-Line, mais aussi en corrélation avec la problématique des périodiques électroniques, j'ai été accueilli dans un premier temps à la médiathèque de Roanne afin de passer une journée d'observation sur la question du catalogage. Dans un second temps, ce fut au cours de deux journées passées au SCD de l'Université Lyon 3, sous la responsabilité de Jean-Loup Bruschet, que j'ai pu avoir un regard concret sur la gestion des périodiques électroniques en bibliothèque universitaire.

Une fois cette étape franchie, nous pouvons dès lors nous pencher sur les différents protocoles mis en jeu dans l'offre de portail proposée par Archimed puis comprendre plus en profondeur le fonctionnement des trois briques logicielles formées par SIM, Book-Line et MASC.

4. La question des protocoles d'interrogation en vue de la recherche fédérée : évolutions et promesses

Avant d'aborder l'étude de ces protocoles, une typologie synthétique s'impose. D'abord seuls Z39.50 et les services web SRW et SRU sont de véritables protocoles de recherche et de récupération de l'information. L'accès à des ressources en Z39.50 ou en SRW/SRU fait partie des liaisons par recherche directe. L'OpenURL fait partie des liaisons par rebond et est un protocole de transport des métadonnées. Enfin, l'accès aux ressources via le protocole OAI-PMH fait partie des liaisons par recherche indirecte. OAI-PMH n'est pas un protocole de recherche mais un protocole de téléchargement de données.

4.1. Un protocole ancien mais toujours populaire : Z39.50⁶

La recherche et la récupération d'informations au sein d'un ou plusieurs catalogues de bibliothèque est indissociable du protocole Z39.50. Cette norme ne permet pas de lancer des recherches simultanées, mais la mise au point d'outils de recherches simultanées qui vont rendre possible l'interrogation de différents catalogues de bibliothèques même si ces dernières utilisent un SIGB différent⁷. Z39.50 a été développée par des bibliothécaires, il s'agit donc d'un protocole créé pour les bibliothèques, même s'il est utilisé par ailleurs pour d'autres systèmes de gestion de base de données.

Pour être synthétique, le principe de Z39.50 consiste à un client d'interroger une base de données via un serveur, en déclarant au préalable la procédure d'interrogation et les formats de réponses qui devront être renvoyés. Basé donc sur un modèle client/serveur, Z.39.50 fonctionne avec le protocole TCP/IP.

4.1.1. Historique de la norme

Cette norme fut pensée par des bibliothécaires américains dès 1984. La 1^{ère} version remonte à 1988, elle est uniquement destinée à l'interrogation des bases bibliographique nord-américaines.

Puis l'agence de normalisation américaine (NISO) publie la seconde version en 1992 qui propose une extension aux bases bibliographique non américaines.

Enfin, en 1995, le groupe de travail de ce protocole appelé ZIG (*Z39.50 Implementors Group* qui est composé de bibliothécaires, d'informaticiens et de constructeurs de SIGB) publie la version 3 qui est étendue aux bases non bibliographiques et qui voit l'apparition de services étendus qui ne se limitent plus

⁶ LAHARY D., La norme Z39.50. [en ligne] support crée en octobre 1998, dernière mise à jour le 18 mars 2003. Disponible sur : <<http://membres.lycos.fr/vacher/profess/cours/mediadix/z3950/>> ; GUILLERAY P.E., Projet de catalogue virtuel accessible par le client Z39.50 du SCD de l'Université Lyon 3 [en ligne]. . Diplôme de Conservateur de Bibliothèque, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (Enssib), Villeurbanne, 2003. [en ligne]. Disponible sur : <<http://www.enssib.fr/bibliotheque/documents/dcb/guilleray.pdf>> : ce dernier fournit une bonne bibliographie sur la norme.

⁷ Il est important de faire la distinction entre un client Z39.50 et un client multibase : le client Z39.50 lance une seule transaction, le client multibase gère une session utilisateur et lance autant de transaction qu'il y a de bases à interroger. A cet égard, Book-Line remplit bien ce rôle en lançant une requête http et les traduit en autant de requêtes Z39.50 qu'il y a de bases à interroger.

à une simple interrogation (DSI, réservation ou commande d'une ressource, etc.). La particularité de la version 3 de Z39.50 est de fonctionner avec internet.

La norme devient internationale en 1998, validée par l'ISO (organisation internationale de normalisation, basée à Genève) sous le nom de ISO 23950. L'administrateur de la norme est la Bibliothèque du Congrès, le ZIG continue de travailler pour faire évoluer la norme comme nous le verrons avec les Web Services SRW et SRU.

4.1.2. Une norme efficace mais complexe

Le grand apport de cette norme réside dans la possibilité de faire des recherches réparties (grâce aux clients Z39.50 développés dans ce but) qui elles-mêmes permettent de créer des catalogues communs virtuels sans que l'utilisateur ait l'impression de chercher dans différents catalogues. On peut également interroger séparément différentes bases bibliographiques dans le cadre d'un réseau local. Z39.50 permet aussi le catalogage partagé.

Les spécialistes sont unanimes pour dire que la norme est malgré tout assez complexe, même si elle propose tout ce que les bibliothécaires souhaitent voir apparaître dans une recherche documentaire. Le texte officiel de la norme dont l'exhaustivité est à l'image de cette dernière (180 pages), décrit l'éventail des possibilités offertes ; des caractéristiques qui ne correspondent pas vraiment à ce qui est réellement utilisé. Ce côté complexe fait passer Z39.50 pour archaïque en comparaison au fonctionnement d'internet, qui recourt à un nombre limité de procédures afin de pouvoir toucher le plus grand nombre d'utilisateurs. Malgré tout, Z39.50 reste une norme incontournable de part sa précision et son efficacité. Cette norme permet une recherche précise sur un grand volume de données. Lourde et difficile à mettre en œuvre, elle donne néanmoins d'excellents résultats par rapports aux meilleurs moteurs de recherche.

Il y aurait également des problèmes d'interopérabilité liés à la différence entre les SIGB, ce qui aboutirait à des incompatibilités lors de requêtes. On pense que ces différences viendraient du fait que de nombreux fabricants de SIGB ne

participent pas aux ateliers du ZIG et feraient fonctionner la norme de façon différente⁸. Nous verrons cette problématique est à l'origine de la naissance du *Bath Profile*.

4.1.3. Les principaux "services" de Z39.50

Le fonctionnement de Z39.50 consiste à proposer une série de services (11 au total) établissant la communication entre le client et le serveur. De ces "*facilities*", 4 représentent les services de base ou "*core services*" (*initialisation, search, retrieval, termination*), les 7 autres sont des services élargis (*result-set-delete, browse, sort, access control, accounting/ressource control, explain, extended services*)

Initialisation permet d'établir le dialogue préalable à l'interrogation entre le client et le serveur afin de savoir ce qu'il sera possible de faire. Client et serveur se déclarent, c'est-à-dire qu'ils présentent la version de Z39.50 qu'ils utilisent. Le serveur annonce aussi son profil, mais aussi les services qu'il propose ou non.

Search : Il s'agit de la formulation de la requête. Le client envoie une requête au serveur conformément au profil présenté par ce dernier, c'est-à-dire qu'on se limite aux champs que le serveur déclare être capable de traiter. Si le serveur utilise le profil bib-1, il est donc possible de faire une interrogation sur l'auteur, le titre, mais aussi sur l'ISBN ou RAMEAU tout en utilisant les opérateurs booléens et un opérateur de proximité (assez rare toutefois). Le service *Search* constitue une première forme de réponse du serveur, ce service ne fonctionne pas sans le service *Present* qui donne ainsi un résultat de la recherche et renvoie ces résultats selon les paramètres demandés par le client lors de la requête (par exemple en UNIMARC).

Retrieval : Il est divisé en deux sous-services : *Present* et *Segment*. *Present* est bien sûr indissociable de *Search* et permet au client de récupérer des résultats qui n'ont pas été envoyés par le serveur dans le cadre de *Search* et ce, par tranche de

⁸ C'est de cette problématique qu'est né le *Bath Profile*.

résultats en spécifiant bien sûr l’affichage souhaité (UNIMARC, LCMARC, SUTRS, etc.). Le client peut donner un nom à cet ensemble en vue d’une utilisation ultérieure. *Segment* permet de segmenter les résultats en ensembles plus réduits, cela permet d’améliorer les transmissions sur le réseau. C’est au niveau de *Search* que l’on définit les segments (par exemple un petit ensemble de résultats sera compris entre 1 et 50 résultats, un grand ensemble sera supérieur à 50 résultats). Les résultats sont donnés dans un format différent pour ainsi gagner de la trame au niveau du réseau.

Termination : C’est la clôture de la session.

4.1.4. Les services étendus

Result-set-delete : le client peut supprimer un ensemble de résultats.

Access control et *accounting-ressource control* : ces deux services sont communs : le serveur peut ainsi contrôler l’accès à ses ressources et donc le cas échéant, interdire le téléchargement de notices. Il peut aussi de gérer une comptabilité dans la consultation des données : enregistrement du nombre de requêtes réalisées, de résultats obtenus, en vue d’une facturation ou d’un débit de crédits à la consultation.

Sort : il permet de trier les résultats de la requête.

Browse (Scan) : c’est la possibilité d’avoir l’affichage d’une liste à partir d’un point de départ donné : on donne par exemple le point d’entrée dans une liste, le nombre d’entrées à afficher et éventuellement le nombre d’entrées à afficher avant le point de départ.

Extended Services : le client peut créer et gérer sur le serveur des packages de tâches (*task packages*) permettant ainsi de sauvegarder des résultats à l’issue d’une session et de les réutiliser par la suite. On peut sauvegarder une requête sur le

serveur (c'est un profil de recherche) : cela permet de définir une recherche périodique qui peut être utilisée dans le cadre d'un service *push*. On peut également faire une commande, dans le cadre d'un PEB. La mise à jour de la base de données est également possible par la modification des notices dans la base. Enfin, on peut définir des spécifications pour l'exportation des résultats, le client pouvant ainsi créer des modèles pour l'exportation des résultats et les utiliser à volonté⁹.

Explain : C'est un service qui existe seulement dans la version 3 de Z39.50. Le serveur crée une base de données interrogeable par le client en indiquant et en décrivant les services qu'il propose, de préciser le contenu des bases de données, de déclarer les services dont l'accès est contrôlé ou payant, etc. Le client trouve plus facilement les informations que ce qu'il pouvait obtenir dans le service *initialisation*. C'est un service assez rare et plutôt complexe à mettre en oeuvre¹⁰.

4.1.5. Les profils Z39.50

Le profil définit les options et les paramètres d'utilisation du protocole quand ces choix sont laissés libres par la norme. Un profil peut spécifier l'utilisation de Z39.50 par un SIGB ou une autre application ou même plusieurs logiciels (un SIGB et un logiciel de prêt entre bibliothèque). Le profil peut être enfin conçu dans le cadre de l'utilisation de Z39.50 avec un autre protocole, comme TCP/IP.

Dans son profil, le serveur annonce son adresse IP et son port (en général, le port 210 ou 2100), le ou les bases de données qu'il propose, sa version de Z39.50, le ou les jeux de caractères qu'il reconnaît. Mais le plus important, c'est l'annonce des ensembles d'attributs (*attribute sets*) et ses formats de réponse (*records syntaxes*). L'agence de maintenance de Z39.50 (*Z39.50 Maintenance Agency*) recense quelques dizaines de profils, le plus courant étant le premier qui est le plus

⁹ Notons qu'il n'y pas de normalisation des informations d'exemplaires dans Z39.50, ce qui a amené le développement format OPAC.

¹⁰ Pour SRW, ces informations sont remontées grâce à une page XML. La norme SRW met en place l'*explain* que Z39.50 n'a pas su faire.

riche et celui développé par les bibliothèques : bib-1, précisément le seul inclus dans la norme.

Il y a donc six ensembles d'attributs :

Les attributs d'usage (*use attributes*) qui définissent les points d'accès pour réaliser une recherche. Z39.50 étant une norme de recherche et de récupération d'information, cet ensemble d'attributs est le plus important. Le profil bib-1, développé à partir des formats MARC, peut offrir une multitude d'attributs d'usage, pouvant être combinés avec des opérateurs booléens ou de proximité. Bien sûr tous ces attributs ne sont pas intégrés dans un seul profil.

Les attributs de relation (*relation attributes*) : ils définissent les relations possibles entre les valeurs des variables de la requête et les valeurs des variables contenues dans la base (plus petit que, plus grand que,...).

Les attributs de troncature (*truncation attributes*) : troncature à droite, à gauche, etc.

Les attributs de complétude (*completeness attributes*) indiquent si le terme de la recherche doit être le seul élément d'un champ ou d'un sous-champ (le champ ou le sous-champ est dit complet ou incomplet).

Les attributs de position (*position attributes*) qui précisent si un terme de la recherche doit être recherché au début d'un champ, à la fin ou n'importe où.

Les attributs de structure (*structure attributes*) précisent la forme de ce qui est demandé (mot, phrase, date, etc.).

4.1.6. Les formats de réponse

Un client Z39.50 peut demander au serveur le format d'affichage qu'il souhaite voir appliquer pour ses réponses. Cela peut être du MARC, SUTRS (*Simple, Unstructured Text*, du texte ASCII présenté en ligne de 72 caractères), GRS-1 (*Generic Record Syntax* : un langage permettant la représentation d'entrées de bases de données), OPAC (un format MARC21 avec inclusion des informations d'exemplaires), XML.

Les spécificités d'un profil tiennent surtout au paramétrage de deux éléments : les attributs d'usage et les formats de réponses. Un serveur qui est paramétré avec un profil bib-1 sera assez performant pour l'exploitation de bases de données bibliographiques traditionnelles dans un des formats MARC, mais à l'origine il ne permettait pas d'exploiter des documents en XML. Inversement, un profil généraliste peut exploiter des ressources différentes, mais sera moins performant à l'image du *bath profile* qui ne permet pas d'interroger sur les mots MeSH ou RAMEAU alors que bib-1 le permet.

4.1.7. Les évolutions de Z39.50 : révision du Bib-1 et *Bath Profile*

Face à la complexité de la norme, mais aussi devant la volonté d'exploiter pleinement une recherche documentaire, le ZIG a mené de nombreuses recherches. Tout d'abord, concernant le profil bib-1, le ZIG a décidé de créer une nouvelle architecture d'attributs en regroupant des communautés d'utilisateurs en leur laissant la possibilité de développer des attributs spécifiques selon leurs besoins. Dans cette nouvelle architecture, deux ensembles d'attributs sont importants, les classes d'attributs (*utility attribute set*) et les attributs interdisciplinaires (*cross domain attribute set*). Pour la classe d'attribut, au lieu de choisir parmi la liste des attributs d'usage de bib-1, la nouvelle classe d'attributs propose un point d'accès. Pour les attributs interdisciplinaires, il s'agit de réduire les points d'accès au plus petit dénominateur commun, on a ramené les points d'accès à un ensemble de 11 éléments tirés du Dublin Core.

Cette logique d'ensemble est de promouvoir l'interdisciplinarité et la généralité de la norme.

Le *Bath profile* est un profil Z39.50 créé en août 1999 et géré par la Bibliothèque Nationale du Canada. Modulable, ce profil identifie trois niveaux de besoins fonctionnels (*functionnal requirements*) :

- Le domaine fonctionnel A (*functional area A*) traite des principales fonctions de recherche et de récupération de données bibliographiques en insistant sur les bibliothèques.
- Le domaine fonctionnel B (*functional area B*) traite des fonctions de recherche et de récupération de données bibliographiques.
- Le domaine fonctionnel C (*functional area C*) traite des fonctions de recherche et de récupération de données au-delà du seul domaine bibliographique pour établir des recherches pluridisciplinaires entre bibliothèques, musées, archives.

Pour chaque domaine, on propose plusieurs niveaux de conformité selon la complexité de la recherche que l'on souhaite réaliser (*conformance levels 0, 1, 2*, le domaine A niveau 0 permet une recherche sur auteur, titre et sujet, le niveau 1 rajoute les mots-clés avec possibilité de troncature).

Le but du *Bath Profile* est de garantir une mise en œuvre propre de Z39.50, ce n'est pas un profil généraliste, mais un profil de mise en conformité de l'implémentation de Z39.50 par les SIGB. Comme Z39.50 est une norme extrêmement riche, il est difficile pour un éditeur de SIGB d'intégrer la totalité de la norme. Le *Bath Profile* propose d'identifier les éléments fondamentaux de la norme, il recommande donc les minima que les SIGB doivent intégrer.

4.1.8. Bilan de Z39.50

Z39.50 est avant tout un protocole qui concerne les bases bibliographiques. Avec la version 3, d'autres disciplines appliquent la norme en précisant les règles de mise en œuvre dans un profil déterminé : les musées disposent du profil CIMI (*Computer Interchange of Museum Information*), GEO (pour les systèmes d'information géographiques) ou le profil GILS (*Government Information Locator Service*). Pour récapituler sur Z39.50 :

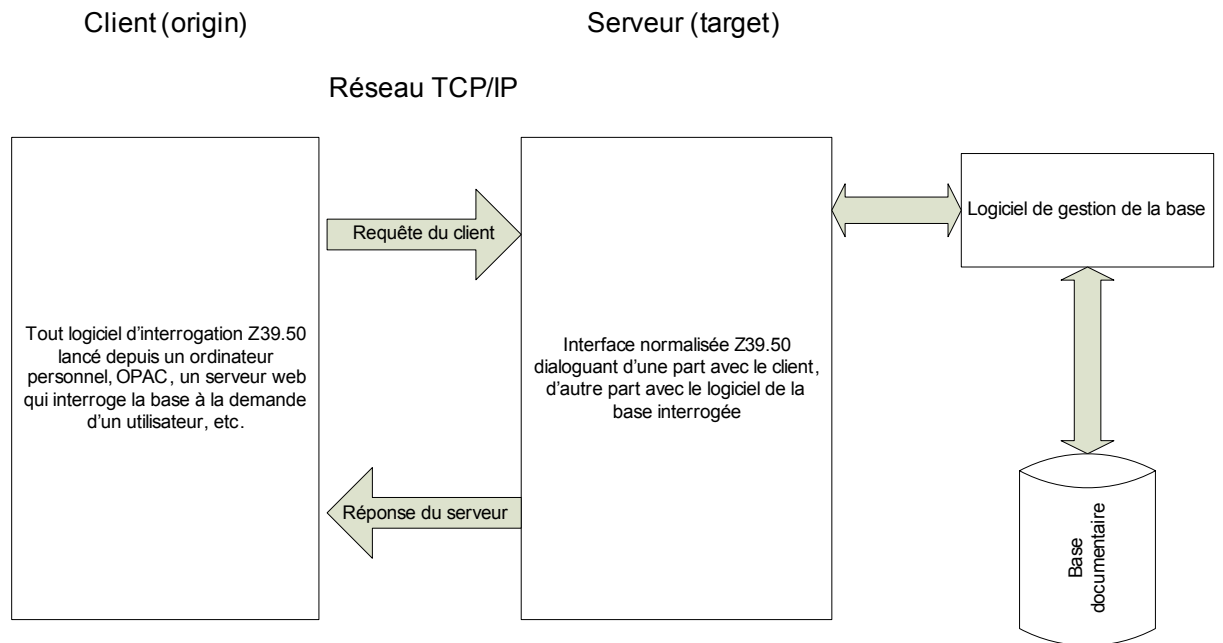
- Il permet une gestion partagée du catalogage bibliographique en donnant la possibilité à un catalogueur de récupérer des notices sans avoir à connaître le mode d'accès spécifique à chaque base.
- Il permet la mise en place d'OPAC et de catalogues collectifs : le lecteur peut interroger une ou plusieurs bases de la même manière et récupère un résultat global et homogène.
- Il offre des services documentaires comme la consultation en ligne des emprunts effectués par un lecteur, la réservation en ligne, l'abonnement à un service de DSI, etc.

Par contre, la complexité de la norme et de sa mise en œuvre est une contrainte non négligeable, de plus que Z39.50 n'est pas un modèle universel pour ceux qui voudraient exploiter la norme en dehors des bases de données bibliographiques. Cette réflexion sur l'ouverture de la norme est celle effectuée par le ZING (*Z39.50 International Next Generation*), groupe de travail rattaché au ZIG. Si l'on veut être ouvert, il faut s'appuyer sur des protocoles faciles à implémenter, par exemple travailler sur la couche http et récupérer les résultats d'une recherche en XML. C'est ce que nous verrons en §4.2 avec SRW/U.

4.1.9. Schémas explicatifs

4.1.10. Principe d'une session Z39.50

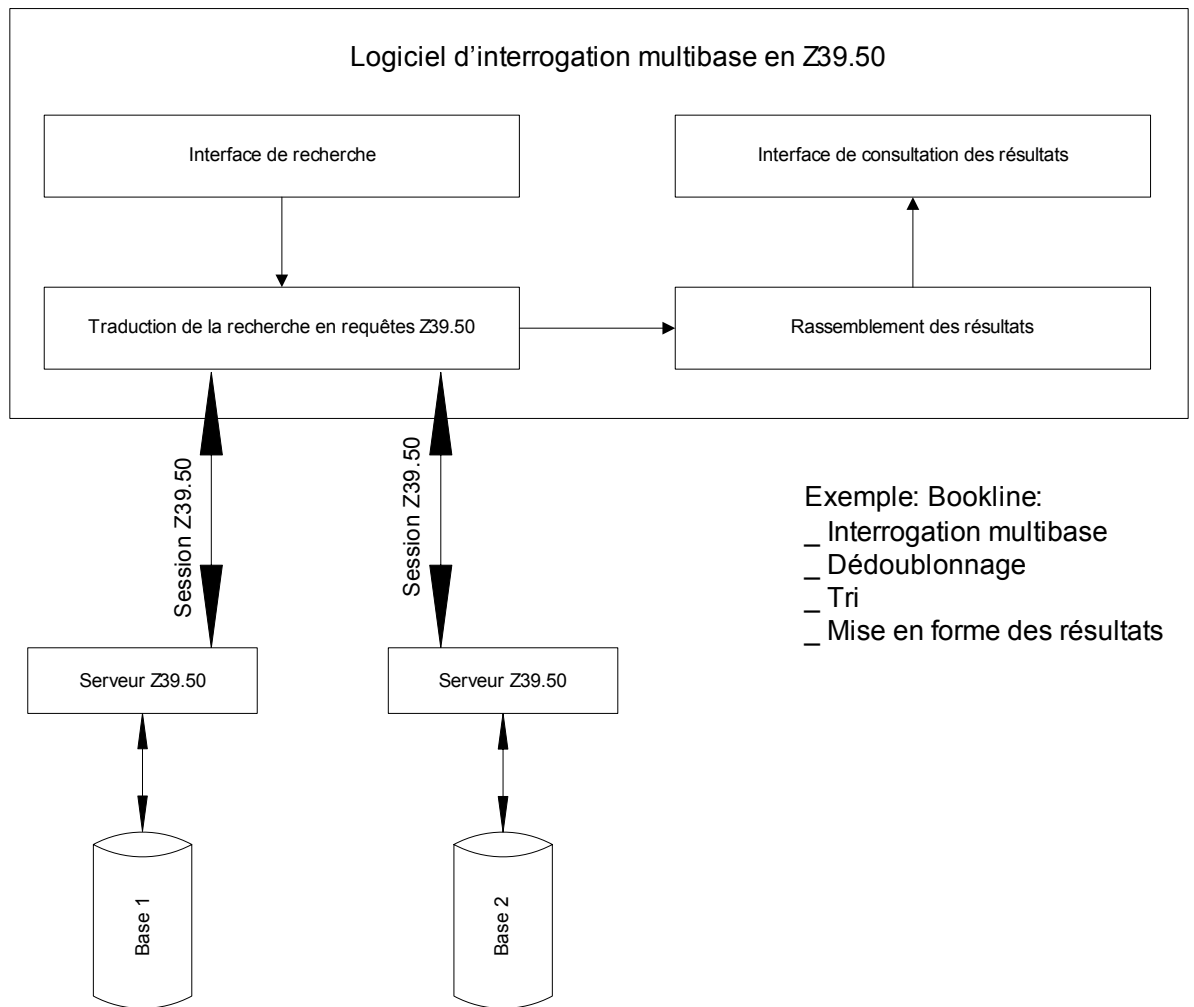
Principe d'une session Z 39.50



Une session ne relie qu'un seul client à un seul serveur

4.1.11. Session Z39.50 et accès multibase

Session Z39.50 et accès multibase



La passerelle multibase gère autant de session Z39.50
que de bases interrogées puis mutualise les résultats

4.2. Le standard d'avenir : les Web Services SRW et SRU¹¹

On l'a vu précédemment pour Z39.50, la fin des années 90 a été caractérisée par la volonté des spécialistes de la recherche et de la récupération d'informations de moderniser ladite norme. Depuis la fin 2001, les questions tournent autour de l'intégration des technologies web (HTTP, SOAP¹²), du XML, de la simplification de Z39.50 de façon à rendre la norme plus lisible (*human readable*) et de l'amélioration des temps de réponses.

Le ZING travaille sur 5 projets : CQL (*Common Query Language*), ZOOM (*Z39.50 Object-Oriented Model*), ezZ39.50 (*Simple Implementation of Z39.50 over SOAP using XER*), ZeeRex (*Z39.50 Explain, Explained and Re-Engineered in XML*) et, en ce qui nous concerne, SRW/SRU¹³.

4.2.1. Description du protocole

SRW (*Search and Retrieve Web Service*) est encore une norme expérimentale dont les premiers tests furent réalisés par l'OCLC et la Bibliothèque du Congrès. La première version publiée de SRW remonte au 13 février 2004. Il s'agit d'un protocole de communication qui permet d'interroger un serveur de données. SRW repose sur le protocole SOAP. Cependant, on peut également se servir des caractéristiques de SRW sans passer nécessairement par la couche SOAP, ce système est alors appelé SRU (*Search and Retrieve URL Web Service*). SRU est donc une variation qui permet pour des clients légers de se passer de SOAP en lisant la requête sur l'URL. Pour simplifier, on peut différencier SRW et SRU par rapport à la méthode de requête utilisée. Pour SRU, la méthode employée est GET (on passe par une URL), pour SRW c'est donc le protocole SOAP qui entre en jeu, la méthode est alors POST puisque l'on travaille directement sur la couche http.

¹¹ Je me suis appuyé surtout sur la documentation technique d'Archimed sur le client/serveur SRW de Book-Line ainsi que le module SARA. Voir également le site officiel du ZING : <<http://www.loc.gov/z3950/agency/zing/srw/>>

¹² *Simple Object Access Protocol*. Il s'agit d'un protocole fondé sur XML afin d'échanger l'information en milieu dit décentralisé. Ce protocole est utilisé pour établir un canal de communication entre services web (invocation à distance via internet de traitements informatiques). Pour plus de détail, voir le site du W3C : <<http://www.w3.org/TR/soap/>>

¹³ Voir le site du ZING: <<http://www.loc.gov/z3950/agency/zing/>>

SRW doit permettre d'interroger des bases de données constituées de métadonnées, mais aussi d'objets textuels ou non, c'est-à-dire des bases contenant tous types de documents.

SRW permet d'exposer une requête et d'obtenir les résultats en XML, de mémoriser ces résultats de façon temporaire, de fournir la possibilité de vérifier les données résultantes, de trier les résultats. On peut également faire un parcours rapide d'un échantillon de résultats, exposer ces possibilités, gérer les versions antérieures de la norme, gérer un token d'authentification, diagnostiquer les problèmes rencontrés et enfin fournir des informations relatives à l'affichage des données.

4.2.2. Les apports de SRW : HTTP, CQL et XML.

Une des différences entre Z39.50 et SRW est le mode de connexion. En Z39.50 la connexion est permanente, on dit qu'on est en mode connecté. En SRW, du fait que l'on passe par HTTP, il n'y a donc par définition pas de connexion : nous sommes en mode déconnecté.

Mais les différences les plus importantes se résument en deux termes : CQL et XML. Z39.50 utilise un langage de requête appelé PQF qui est assez lourd et peu expressif, il n'est pas lisible puisque la requête est codée par des code d'attribut et des chiffres. SRW a recours au CQL (*Common Query Language*), langage plus souple, plus expressif et lisible qui a l'avantage de pouvoir être exprimé en XML, on fait alors du XCQL. SRW utilise donc un langage simplifié pour établir ses requêtes, le renvoi des résultats est également simplifié puisqu'on utilise XML et tout cela se passe dans un environnement réseau standardisé par http ou SOAP..

4.2.3. Les types de services

SRW propose trois principaux services : *Explain*, *Scan* et *Search Retrieve* :

Explain est une demande description du serveur qui renvoie un fichier XML pour se déclarer. Cette description se fait conformément à la norme ZeeRex.

Le *Scan*, à l'instar de Z39.50, permet de parcourir les valeurs d'un index. Le résultat est représenté par quatre valeurs : le libellé du terme, sa forme pour la recherche, un renvoi éventuel vers un autre terme et le nombre de résultats si l'on soumettait ce terme dans une recherche.

La recherche permet donc de récupérer des notices en XML : il est possible de trouver différents types de métadonnées dans ces notices tels que le Dublin Core, Onix, MODS, MarcXml.

Prolongation en XML du standard Z39.50, SRW permet comme Z39.50 de s'adresser au serveur de données et de traiter les résultats obtenus avec les services classiques tels que le dédoublonnage ou le tri. Il n'y a pas de duplication des données, ce qui est appréciable puisque l'information est dès lors pertinente à chaque instant. Par ailleurs, le XML apporte une souplesse et une structuration des recherches et des données récupérées. Le transport sur la couche HTTP via SOAP apporte un côté standard, gage de fiabilité de la norme.

Devant les promesses de SRW, Archimed a orienté les travaux de ses services recherche-développement en direction des Services Web et plus particulièrement de SRW, un protocole qui est susceptible d'apporter des réponses satisfaisantes tant au niveau de la pertinence de la recherche, que de sa rapidité et sa fiabilité. En ce sens, Archimed a donc développé, un connecteur SRW sur son moteur Book-Line et également sur son module SARA.

On le voit bien, SRW est un protocole qui va dans le prolongement de Z39.50 en se voulant être une simplification de ce dernier tout en permettant une égale précision dans les recherches. Comme Z39.50, SRW fait fi des caractéristiques propres à chacune des bases interrogées. SRW s'appuie sur un environnement standardisé, sur http pour le transport des données et sur XML pour la récupération des informations.

4.3. Un protocole de rebond par URL : OpenURL¹⁴

4.3.1. Présentation

Ce protocole, en cours de normalisation NISO, offre la possibilité à un utilisateur de compléter sa recherche initiale par une autre recherche beaucoup plus ouverte en tenant compte du contexte de l'utilisateur. L'OpenURL permet d'étendre une recherche initiale pour retrouver des informations en lien avec une première réponse obtenue sans qu'on ait besoin de spécifier tous les liens qu'on puisse trouver autour de cette première réponse. Il s'agit d'une extension dynamique du résultat, d'où le terme d'Open, mais également d'une extension qui tient compte du contexte de la recherche (*Context Sensitive*).

OpenURL est donc une technologie dont le principe est de créer des liens contextuels qui permettront de créer un lien entre les métadonnées d'une ressource à la ressource elle-même (par exemple les références d'un article à l'article lui-même en texte intégral).

La réflexion qui a amené à la création de ce protocole était de trouver une alternative au problème de la pérennité des URL, en déterminant de façon dynamique et non pas statique un chemin d'accès à une ressource électronique. En quelque sorte, l'OpenURL correspondrait au lien Unimarc 856, mais sans les défauts de ce dernier puisque le lien est dynamique pour l'OpenURL et non statique comme c'est le cas du lien Unimarc 856. Le lien 856 est le même pour tout le monde, il ne répond pas aux critères de recherche dynamique et de prise en compte du contexte. Nous verrons que la solution apportée avec la technologie de l'OpenURL est la mise en place de serveurs de résolutions.

A l'origine, ce protocole fut développé sur la base d'un résolveur de lien de la société Ex-Libris, SFX, il en a résulté la version 0.1 du protocole.

¹⁴ VAN DE SOMPEL H., The OpenURL Framework : Origins, Evolution, Concept (OpenURL Day, New-York, May 13th 2003). [en ligne]. Mai 2003. Disponible sur: <http://palmer.cwpost.liu.edu/csc/van_de_sompel.ppt>

L'OpenURL n'offre donc pas qu'un simple lien, mais un lien contextuel. C'est-à-dire que pour une ressource donnée, nous connaissons le contexte dans lequel elle est citée (la bibliographie qui la cite), décrite (les métadonnées de l'article), utilisée (qui est le lecteur qui recherche, quels droits il dispose et aussi partir de quel système d'information la requête est effectuée). Nous connaissons aussi la façon dont le protocole est utilisé (quel type de résolveur pour quel type de service) afin de lier cette ressource à ce qu'elle décrit, c'est-à-dire l'article lui-même. C'est pourquoi l'on qualifie l'OpenURL de *context sensitive*.

4.3.2. Composition de l'OpenURL

L'OpenURL est en fait composé de trois éléments :

Le *Context Object* représente un paquet de métadonnées qui décrivent une ressource et son contexte. Il regroupe les métadonnées elles-mêmes (*referent*), leur source (*referrer*), la ressource qui contient la référence (*referring entity*, typiquement le résultat d'où est issue la citation) et enfin l'utilisateur qui sollicite le service (*requester*). Par ailleurs, le *Context Object* stipule l'adresse du résolveur de liens utilisé (*resolver*), le type de service demandé par l'utilisateur (*service type entity* ; ce type de service peut consister à obtenir le texte intégral de l'article).

Le *Context Object* décrit donc une ressource bibliographique et son contexte, son format d'origine consistait à le représenter par une paire clef/valeur, c'est le *Key Encoded Value* (KEV). Depuis la publication de la seconde version d'OpenURL, la version 1.0, les *Context Objects* peuvent être décrits en XML avec un schéma qui lui est propre le *XML Context Object Format* (CTX)

Les métadonnées stockées dans le *Context Object* doivent suivre les formats de métadonnées autorisés : par exemple, en KEV il y a des formats pour les monographies, les thèses, les revues, les brevets et chacun de ses formats possède une traduction en XML.

L'OpenURL elle-même, c'est-à-dire un protocole de transport des paquets de métadonnées du *Context Object* basé sur HTTP.

Le *registry* : il contient les spécifications des différents éléments constitutifs des *Context Objects* et de l'OpenURL. C'est un modèle de mise en œuvre de la norme.

4.3.3. Fonctionnement

Dans le cadre d'une recherche documentaire, le *Context Object* est créé par une bibliothèque ou un éditeur d'articles en ligne. A l'issue de sa requête, l'utilisateur voit à côté de sa référence un bouton qui correspond à ce *Context Object* (sur Book-Line, ce bouton est nommé "OpenURL". En activant ce bouton, le *Context Object* est envoyé sous la forme d'une requête HTTP normalisée OpenURL. Les arguments de recherche de la requête sont définis à partir des données contenues dans le résultat de départ : Par exemple Book-Line peut récupérer le titre, l'auteur, le sujet de la référence Z39.50 et les formate dans l'OpenURL.

L'OpenURL est ensuite transmise à un serveur spécialisé appelé résolveur ou serveur de résolution. Ce serveur est paramétré pour savoir interpréter l'OpenURL et la traduire dans le langage spécifique de chaque source web à interroger. Il va traiter les éléments de la requête, le type de service sollicité et les droits de l'utilisateur. Le résolveur va donc trier les références accessibles par l'utilisateur et écarter les autres auxquelles la bibliothèque n'a pas accès, parce qu'elle n'est pas abonnée par exemple à une revue où un article qui correspond à la requête effectuée. Le résultat obtenu peut être par exemple un article en texte intégral dans une revue dont la bibliothèque est abonnée accompagné d'un article issu d'une archive ouverte.

Le résolveur peut être soit un serveur privé ou un serveur mutualisé par plusieurs clients. Archimed met librement à la disposition de ses clients son résolveur, OSC-R.

Depuis SFX d'Ex-Libris, d'autres résolveur ont vu le jour tels que LinkSource (Ebsco), LinkFinder Plus (Endeavor), LinkSolver (Ovid), 1Cate (Openly Informatic). Archimed a également développé son propre résolveur, OSC-R dont nous verrons les caractéristiques en §4.5.3.

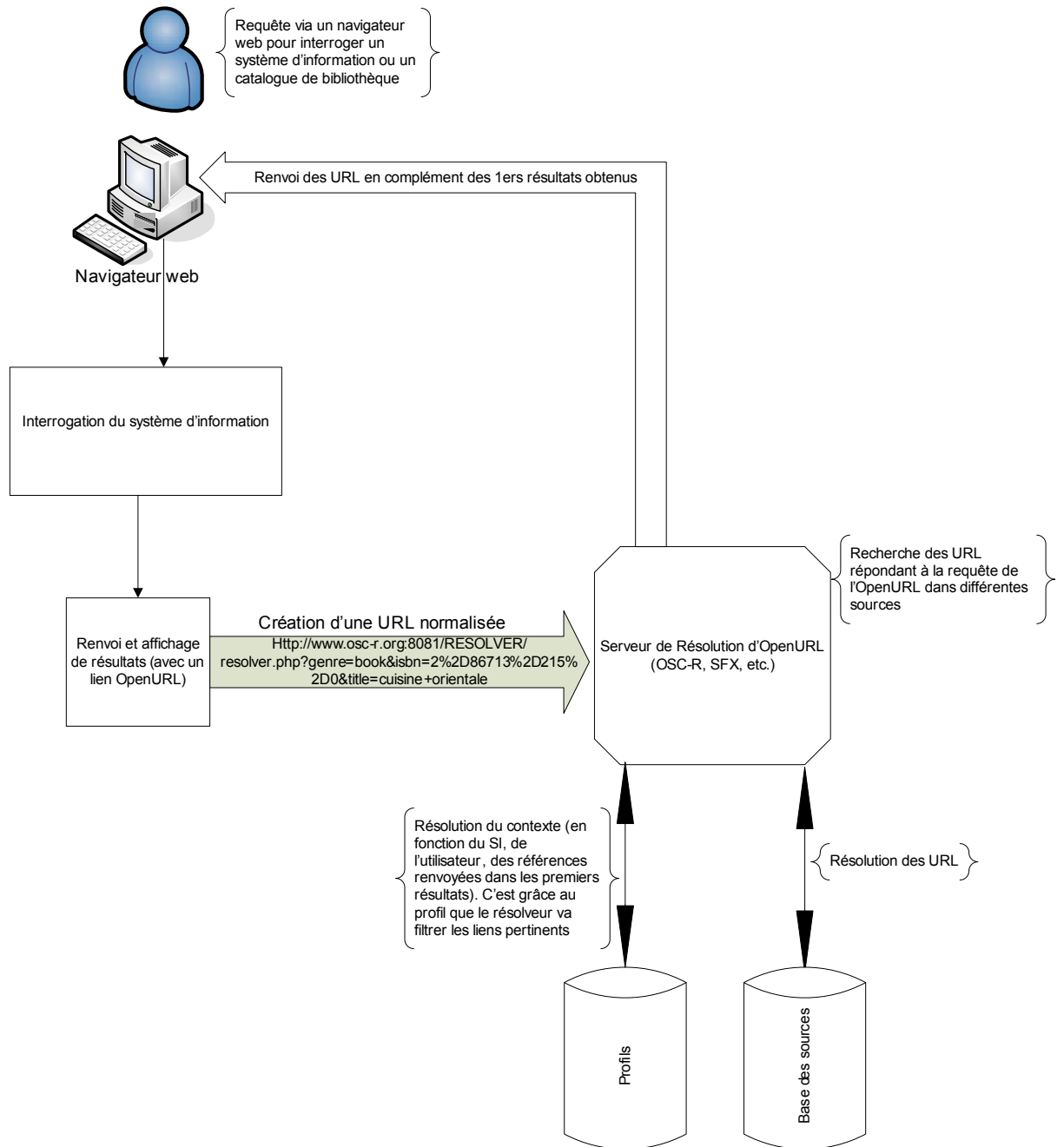
EDITEUR	RESOLVEUR	PORTAIL	TYPE	SITE WEB
Ebsco	LinkSource	Pas de portail	Commercial	http://www.linkresolver.com/
Endeavor	LinkFinder PLUS	EnCompass	Commercial	http://www.endinfosys.com
ExLibris	SFX	Metalib	Commercial	http://www.exlibris-usa.com/products1.html
Fretwell Downing Informatics	OL2	Zportal	Commercial	http://www.fdgroupp.com/fdi/products/about.html
Ovid	LinkSolver	OvidWeb Gateway	Commercial	http://ovid.com/site/products/tools/ovid/linksolver.jsp
Openly Informatics	1Cate	Aucun	Commercial	http://www.openly.com/1cate
Archimed	OSC-R	Portail Archimed	Open Source	http://www.osc-r.org

C'est un protocole caractérisé par une architecture assez légère qui permet de réaliser des recherches autour de cibles disparates. C'est indubitablement un plus pour l'utilisateur.

Par contre, cela nécessite d'avoir une véritable base de métadonnées pour générer la requête pour le résolveur, c'est-à-dire qu'il faudrait être en possession de la notice qui permettrait de générer les OpenURL à partir des métadonnées. On ne peut aussi véritablement considérer OpenURL comme étant un véritable protocole de recherche dans la mesure où l'on ne peut pas agréger des résultats avec d'autres sources pour effectuer des traitements tels que le déboublonnage, le tri, etc.

4.3.4. Schéma explicatif de l'OpenURL

Fonctionnement de l'OpenURL



4.4. Un protocole de métadonnées dédié aux ressources libres : OAI-PMH

Littéralement protocole de moissonnage de métadonnées de l'Initiative des Archives Ouvertes (*Open Archive Initiative Protocol for Metadata Harvesting*), OAI-PMH est né d'un désir de la communauté scientifique de se doter d'outils d'auto archivage afin d'avoir une meilleure accessibilité aux publications scientifiques et techniques, puis vers les autres disciplines en favorisant l'interopérabilité entre les différents dépôts de documents. Ce protocole fut lancé en 1999 lors de la convention de Santa Fé. La première initiative du mouvement des Archives Ouvertes a été l'ouverture du serveur de prépublications de Los Alamos. Répondant à ce souhait de rendre facilement visible les travaux des chercheurs, cette initiative revêt également un enjeu stratégique dans le domaine de l'édition électronique puisqu'elle offre une alternative au modèle des éditeurs commerciaux en gardant la main mise sur la production et la diffusion des travaux.

Avant d'entrer dans la présentation de ce protocole, un mot s'impose sur ce que sont les archives ouvertes. En fait, le terme d'archive ne désigne en rien quelque chose d'ancien, mais plutôt un entrepôt (*repository*) de documents. Le terme *open* fait référence à l'ouverture de l'architecture technique des différents entrepôts grâce à la mise en place de protocoles communs facilitant l'accès au contenu. En aucun cas le terme d'*open* est synonyme de gratuité.

4.4.1. Data Provider et Services Provider

Le protocole OAI s'articule autour de deux étapes et de deux acteurs : le dépôt et la collecte des métadonnées, le fournisseur de service et le fournisseur de données.

C'est le dépositaire qui effectue l'enregistrement du document qu'il souhaite déposer dans l'archive. Pour cela il doit constituer les métadonnées de son document (Dublin Core – qui est le format minimum obligatoire de la norme -, un format MARC ou une DTD spécifique) afin de pouvoir l'identifier et l'indexer. Cette opération se fait grâce à une interface web classique au sein de laquelle on dépose le document dans une archive (*data provider*). Du côté des interfaces, la norme ne préjuge en rien des outils mis en œuvre.

D'autre part, un fournisseur de services ou *services provider* offre à l'utilisateur par le biais d'une interface de consultation l'accès aux différentes bases de documents. Cette consultation peut être réalisée directement en direction des différents entrepôts ou alors après une phase de collecte des métadonnées grâce à un "moissonneur" (*harvester*). Le moissonneur est donc un outil de collecte des métadonnées qui met à jour une base de données locale en allant collecter des métadonnées depuis un ensemble de *repositories* (serveur OAI). Comme le rappelle S.Oulahal, c'est en quelque sorte un cache web¹⁵. Le travail du moissonneur n'est pas sans rappeler celui des robots (*crawlers*) utilisés par les moteurs de recherche afin d'alimenter leurs indexes.

Il est possible d'ajouter un troisième acteur dont le rôle est de rassembler les métadonnées émanant de plusieurs fournisseurs de données et de les rendre accessibles dans un entrepôt OAI. On les appelle assez logiquement les agrégateurs.

L'intégralité du protocole OAI, c'est-à-dire le dépôt et la consultation des documents repose sur le protocole HTTP (mode non connecté comme vu précédemment). En fonction du type de protocole de recherche utilisé pour récupérer des informations dans un réservoir OAI, il serait possible de mener une recherche sous forme de mots-clés avec un retour de données structurées en XML. Ce protocole s'applique à tous types de documents (texte, audio et vidéo).

4.4.2. Types de requêtes proposées par OAI-PMH

OAI-PMH ne spécifie pas de service, par contre il propose 6 types de requêtes pour effectuer l'ensemble des traitements relatifs au dialogue entre le *data provider* et le *services provider*, avec pour chaque réponse une validation du résultat par un schéma en XML. C'est une architecture relativement simple permettant une intégration facile au sein d'un service spécifique.

¹⁵ OULAHAL S., Accès unique à des ressources numériques distribuées [en ligne]. 2003. Disponible sur : <<http://2003.jres.org/actes/paper.125.pdf>>

Getrecord : c'est la demande d'un enregistrement.

Identify : demande d'information sur un entrepôt OAI.

ListIdentifiers : demande la liste des éléments contenus dans un entrepôt.

ListMetadataFormat : demande la liste des formats de métadonnées disponibles dans un entrepôt.

ListRecords : demande des enregistrements de métadonnées contenues dans l'entrepôt/

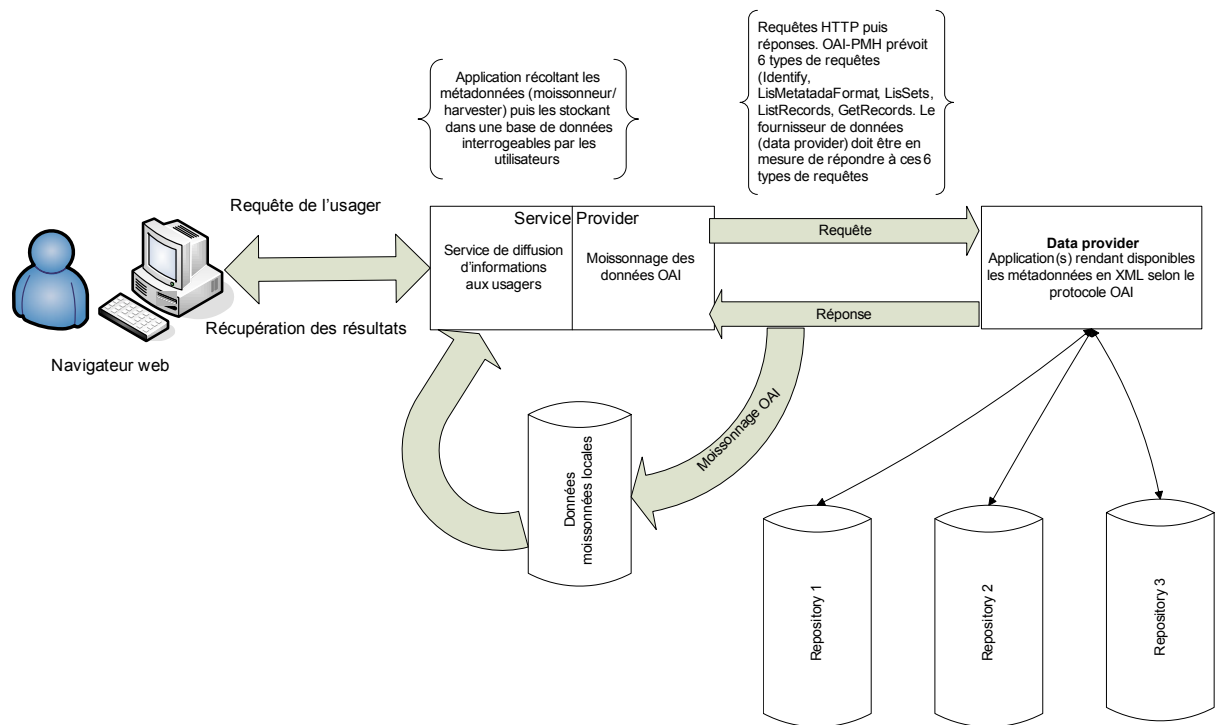
ListSets : demande la liste des ensembles OAI définis dans l'entrepôt.

Surtout, il faut souligner que le protocole OAI n'est en rien un protocole de recherche et de récupération de l'information, il ne gère que les transferts de données et ne compte aucune fonction de recherche. Ce protocole n'est pas inscrit dans une logique de passerelle, mais plutôt dans une logique de moteur puisqu'il y a indexation de métadonnées OAI. OAI est donc un protocole de métadonnées et un protocole de téléchargement.

La principale contrainte d'un tel protocole est qu'il faut conserver les données en cache, c'est une contrainte certes relative car on ne stocke que des métadonnées (donc des notices assez légères). Ce protocole garantit un accès rapide à l'information.

4.4.3. Schéma explicatif du protocole OAI

Fonctionnement du protocole OAI-PMH



4.5. Le souci d'Archimed de faciliter la mise en œuvre et l'utilisation optimale de ces protocoles

La présentation de ces différents protocoles s'avérait nécessaire avant d'aborder ce que propose Archimed dans le cadre de son offre de portail documentaire. Mais avant d'entrer de plein pied dans le triptyque SIM/Book-Line/MASC, il s'agit de se pencher sur la façon dont ces protocoles sont utilisés. A cet égard, Archimed a équipé Book-Line d'un connecteur SRW et a développé trois applications pour l'OpenURL, OAI et SRW.

4.5.1. Le serveur Z39.50 et le connecteur SRW de Book-Line

Book-Line travaille habituellement avec la norme Z39.50 puisqu'il possède un connecteur permettant de mener des recherches grâce à cette norme et c'est

d'ailleurs à cet égard que le progiciel est équipé d'un serveur Z39.50 développé par Archimed, le ZAS (Z39.50 Archimed Serveur). Cette application va donc permettre à d'autres outils de recherche équipés de passerelles d'interrogation Z39.50 d'interroger Book-Line, qui lui-même va mener une recherche auprès d'une ou plusieurs bases de données.

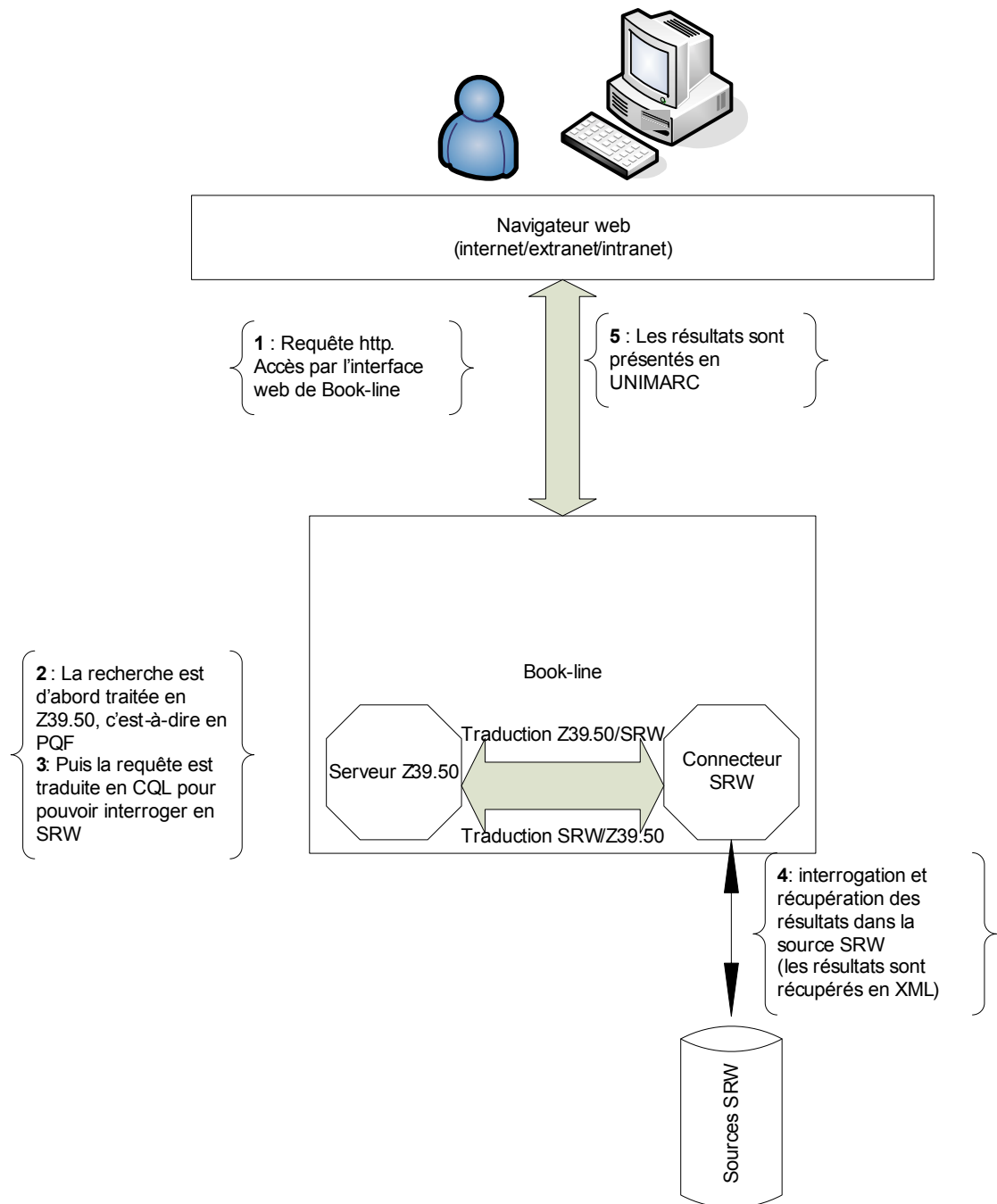
Mais avec l'ouverture de nouveaux protocoles sur le web, Archimed a choisi de se lancer dans l'exploitation du protocole SRW en dotant un connecteur de ce type dans son métamoteur. Le principe est donc de rechercher des ressources dans des bases utilisant le protocole SRW comme protocole d'interrogation.

Un utilisateur va entrer une requête dans Book-Line, cette requête est alors traitée d'abord en Z39.50 et son langage PQF. Une traduction de la requête en CQL est établie par les outils de communication entre le connecteur Z39.50 et le connecteur SRW (toute la difficulté a été de mettre en place ces outils pour faire correspondre Z39.50 et SRW). Une fois la traduction réalisée, le connecteur SRW peut interroger une ou plusieurs bases SRW. Les résultats récupérés sont renvoyés en XML puis sont ensuite traités de façon à être présenté par Book-Line en UNIMARC.

4.5.1.1.

Schéma du fonctionnement du connecteur SRW

Fonctionnement du connecteur SRW de Book-Line



4.5.2. Présentation du module SARA

C'est une première en la matière puisqu'il n'existe pas encore de serveur SRW public sur le web. Archimed veut mettre en place ce qu'elle appelle une "carcasse" ou "squelette" de serveur SRW accompagnée d'une implémentation d'exemple. Cette carcasse devrait permettre à des intégrateurs de réaliser des sources SRW pour Book-Line. Une étape délicate dans cette mise en œuvre sera de faire correspondre SRW et Z39.50 puisqu'ils n'ont pas vraiment les mêmes fonctionnalités, ont un langage, ainsi que les formats de résultats et un paramétrage différents.

Il s'agit en fait d'un serveur SRW, comme le laisse imaginer le développement de son acronyme (SARA=*Search And Retrieve Application*). Connecté à Book-Line, et alimenté par des plug-ins, SARA permet d'interroger n'importe quel type de source en SRW grâce notamment à une architecture technique réalisée entièrement en Microsoft.Net Framework. La stratégie .Net regroupe un ensemble de technologies qui permet l'interconnexion des informations, des individus, des systèmes informatiques et des terminaux, tout en se reposant sur des standards comme XML, HTTP, SOAP, WSDL¹⁶, etc. La technologie .Net Framework propose un environnement d'exécution sécurisé et une plate-forme de développement unifiée. L'environnement .Net Framework est distribué gratuitement sur toutes les versions de Windows et est devenu une alternative à la plateforme Java (Java Development Kit JDK)¹⁷.

Archimed propose pour ses clients un serveur SARA contenant déjà une gamme de plug-ins, mais les ingénieurs d'Archimed peuvent développer d'autres connecteurs selon les exigences de ses clients. Enfin, il est possible d'obtenir un kit de développement si un client désire lui-même créer ses propres plug-ins (SDK *Software Development Kit*).

¹⁶ Web Services Description Language. C'est donc un langage de description de services web qui permet à des applications qui l'utilisent d'auto configurer des échanges entre elles.

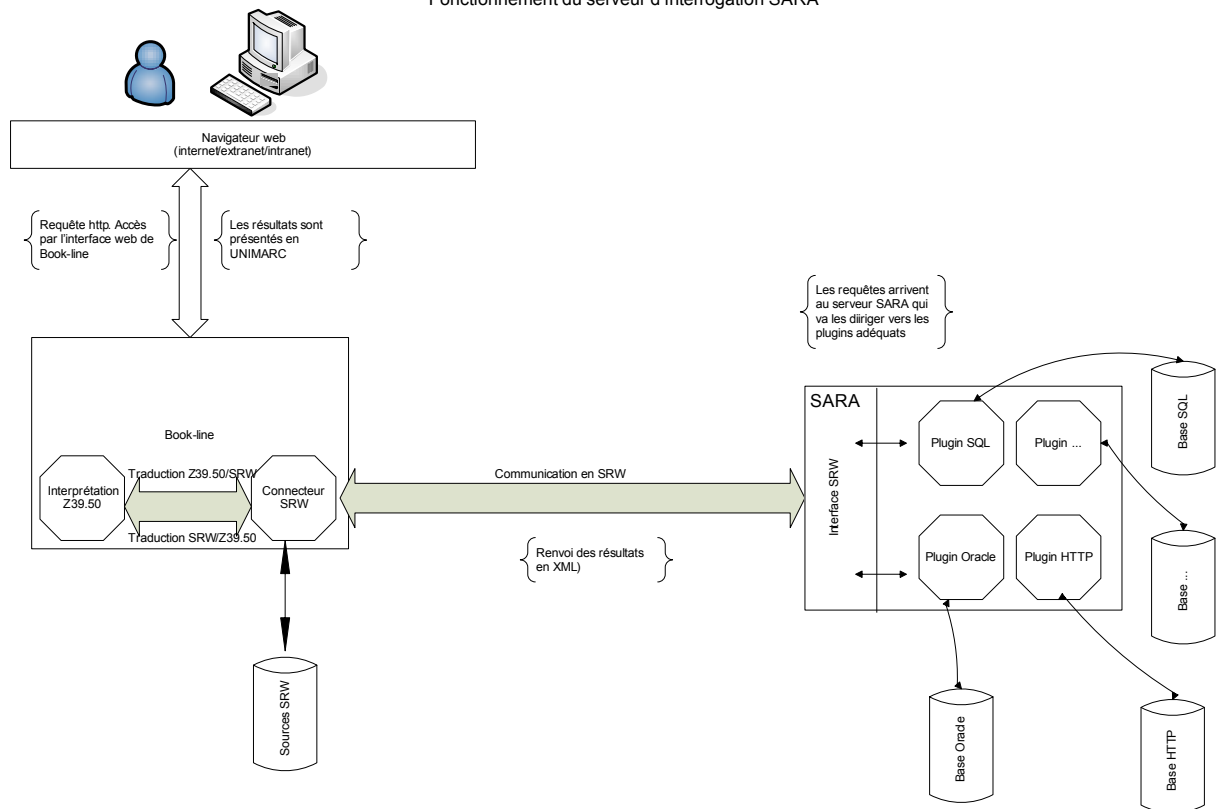
¹⁷ Pour plus de détails, voir la fiche technique de la technologie Microsoft.Net réalisée par la société 3IE : http://www.3ie.org/nouvelles_technologies/fiches.php?techno_id=354.

Ce serveur fonctionne en fait comme un émulateur. On l'a dit, son standard de base est du SRW, mais pour être plus précis il est en SRU natif. Lors d'une requête transmise par le moteur Book-Line (une requête transmise via le protocole Z39.50 ou SRW), SARA va rechercher en fonction des plug-ins qui sont installés en plus de communiquer via SRW/SRU. Si la source à interroger est en HTML, SARA communiquera en HTML grâce à un plug-in adapté, si c'est une base Oracle ou du SQL, SARA s'adaptera à ces modèles pour pouvoir les interroger et faire remonter les résultats vers Book-Line.

Ce qu'il faut retenir ici et ce qui est aussi à l'image du travail d'Archimed, c'est l'effort de rendre la recherche transparente pour l'utilisateur. Derrière toutes ces normes et la complexité technique des systèmes mis en jeu pour pouvoir communiquer avec des ressources différentes via des protocoles différents, c'est en fait de donner l'illusion de faire une interrogation relativement simple en Z39.50. Le schéma représentant cette architecture en annexe est assez évocateur.

4.5.2.1. *Schéma du fonctionnement de SARA*

Fonctionnement du serveur d'interrogation SARA



4.5.3. Le résolveur OSC-R

Dans la continuité des travaux sur les Services Web, mais aussi en souhaitant proposer une solution alternative à l'offre commerciale dans le domaine des résolveurs de liens contextuels, Archimed s'est lancé dans un projet de résolveur en *open source*, sous licence GPL. OSC-R (*Open Source Contextual Resolver*, anciennement appelé OSC@R). Trouvant son origine dans des partenariats avec des universités (Université du Littoral et Université de Technologie de Troyes) et aussi par la prise de conscience des investissements onéreux dans les abonnements aux ressources en ligne et par une faible rentabilité de ces abonnements (on l'expliquera par la suite dans la seconde partie), l'idée est donc de créer une passerelle libre basée sur le standard OpenURL.

Comme nous l'avons vu précédemment, le résolveur est un bon complément de recherche et surtout il dispense un utilisateur de connaître les modalités d'accès à

des ressources hétérogènes. Ici aussi, nous sommes dans une logique de transparence de la recherche.

4.5.4. Le moissonneur Harvey

Enfin, Archimed s'est également impliqué concernant le protocole OAI-PMH. On a vu que cette norme met en scène deux principaux types d'acteurs. D'une part, le fournisseur de données dont le but est de rendre visible les documents entreposés dans son *repository* et rendre possible la récolte de ses métadonnées selon les règles du standard. D'autre part, le fournisseur de services dont les objectifs sont d'effectuer des recherches dans différents réservoirs de documents au sein de serveurs différents et de récolter puis entreposer les métadonnées offertes par les fournisseurs de données. Le fournisseur de services peut également développer une interface permettant la recherche des données récoltées.

Archimed s'est donc positionné en tant que fournisseur de services avec le développement d'un moissonneur OAI nommé Harvey.

Ce chapitre consacré aux différents protocoles utilisés par Archimed montre bien la complexité de la tâche à laquelle font face les chefs de projets en systèmes d'information documentaire. Si l'on veut mener à bien un projet de portail documentaire caractérisé par une recherche fédérée, il va falloir travailler avec différentes normes et certaines affichent plus de promesses que d'autres. Z39.50 est un protocole assez ancien, mais il arrive toujours à s'adapter face aux évolutions technologiques et bibliothéconomiques. L'OpenURL a pu susciter un certain enthousiasme au départ parce que c'est un protocole assez léger à mettre en œuvre, mais il demande un jeu de métadonnées conséquent pour être pleinement efficace mais surtout il ne sert à rien sans référence ou identifiant de référence préalable. Par contre, c'est un protocole qui génère accessoirement beaucoup de bruit du fait que le nombre de ressources interrogeable par OpenURL n'est pas limité. Son efficacité est somme toute relative¹⁸. Le protocole OAI est intéressant à exploiter car il permet d'accéder à des ressources en général libre d'accès. Enfin,

l'activité du ZING et le glissement progressif de Z39.50 vers les Services Web SRU mais surtout SRW laisserait penser que SRW pourrait devenir d'ici quelques années le protocole des SIGB et que les notices bibliographiques seront en XML DC ou en XML Unimarc et non plus en Unimarc ISO 2709 par exemple.

Maintenant, nous pouvons aborder ce qui constitue l'offre logicielle (ou plutôt progicielle) d'Archimed afin de comprendre l'utilisation de ces protocoles dans leur contexte, celui du portail documentaire pensé par Archimed.

5. Le portail documentaire construit selon Archimed : une offre packagée

5.1. La base documentaire multimédia : SIM

Dans le cadre d'une solution de portail comportant l'ensemble du package Archimed, la base de cette architecture commence avec le Système d'Information Multimédia (SIM), solution intégrée de gestion électronique de documents.

Sorte de creuset multimédia, le SIM est capable de gérer des objets de toute nature, qu'il s'agisse de documents textuels, d'images, de documents sonores ou de n'importe quel autre type de document.

Le SIM utilise le standard XML afin d'organiser l'ensemble des informations qu'il gère, mais aussi pour les hiérarchiser et les lier. XML permet également l'ouverture vers différents systèmes d'information. L'organisation du système d'information du SIM repose sur le principe des bases de données relationnelles. Toutes les informations sont modélisées en objets et entretiennent entre elles des relations qui assurent la cohésion du système. Ses autres principales fonctions lui permettent d'éditer les données et de stocker les documents afin de maintenir et enrichir le système d'information. Le SIM est également doté d'un outil de recherche, de fonctions de diffusion pour afficher les informations dans des formats différents et personnalisés ou pour les exporter vers d'autres systèmes.

¹⁸ Efficacité qui sera accrue par le développement des profils d'utilisateurs qui pourront bien mettre en valeur le *context*

Enfin, il peut gérer les droits d'utilisation des fonctions ou des accès aux informations.

On le verra dans la présentation de Book-Line, SIM intègre des protocoles de communication tels que SOAP ou une passerelle Z39.50.

5.2. Le métamoteur de recherche : Book-Line

Le titre de cette partie définit Book-Line comme étant un métamoteur de recherche, ce qui est à la fois vrai et faux. C'est vrai dans la mesure où il réalise des recherches et qui plus est des recherches multibases, mais contrairement aux véritables moteurs de recherche il ne crée pas d'index. Le terme de moteur étant devenu un terme du langage courant pour les personnes qui s'intéressent à la recherche d'information, il semble toutefois que le qualificatif le plus adéquat concernant cet outil soit celui de passerelle. Par commodité et parce qu'il permet une recherche multibase, on peut cependant qualifier ce progiciel de métamoteur.

Book-Line est donc un système fédérateur de recherche et ce, quel que soit le fonctionnement et la structuration des différentes sources sélectionnées, qu'il s'agisse de catalogues de bibliothèques, de systèmes de GED ou de ressources électroniques hétérogènes. Book-Line permet donc d'interroger simultanément plusieurs sources sur une seule et unique interface et de présenter les résultats de façon homogène.

Par ailleurs, le progiciel offre d'autres services dédiés aux bibliothèques tels que l'emprunt ou la réservation des documents.

Comme nous l'avons sous entendu dans le chapitre précédent, il est donc possible d'enrichir les résultats obtenus à l'issue d'une requête grâce à l'OpenURL, mais aussi grâce au module multimédia qui permet d'associer les résultats d'une recherche à une vidéo, un document en texte intégral ou un dossier documentaire. Cette base documentaire contient des objets décrits au format Dublin Core. Par ailleurs, Book-Line utilise XML en qualité de format d'export et d'échange du paramétrage des bases interfacées avec Book-Line, mais aussi comme format de stockage des notices multimédias en Dublin Core (ces notices

senistive de la norme.

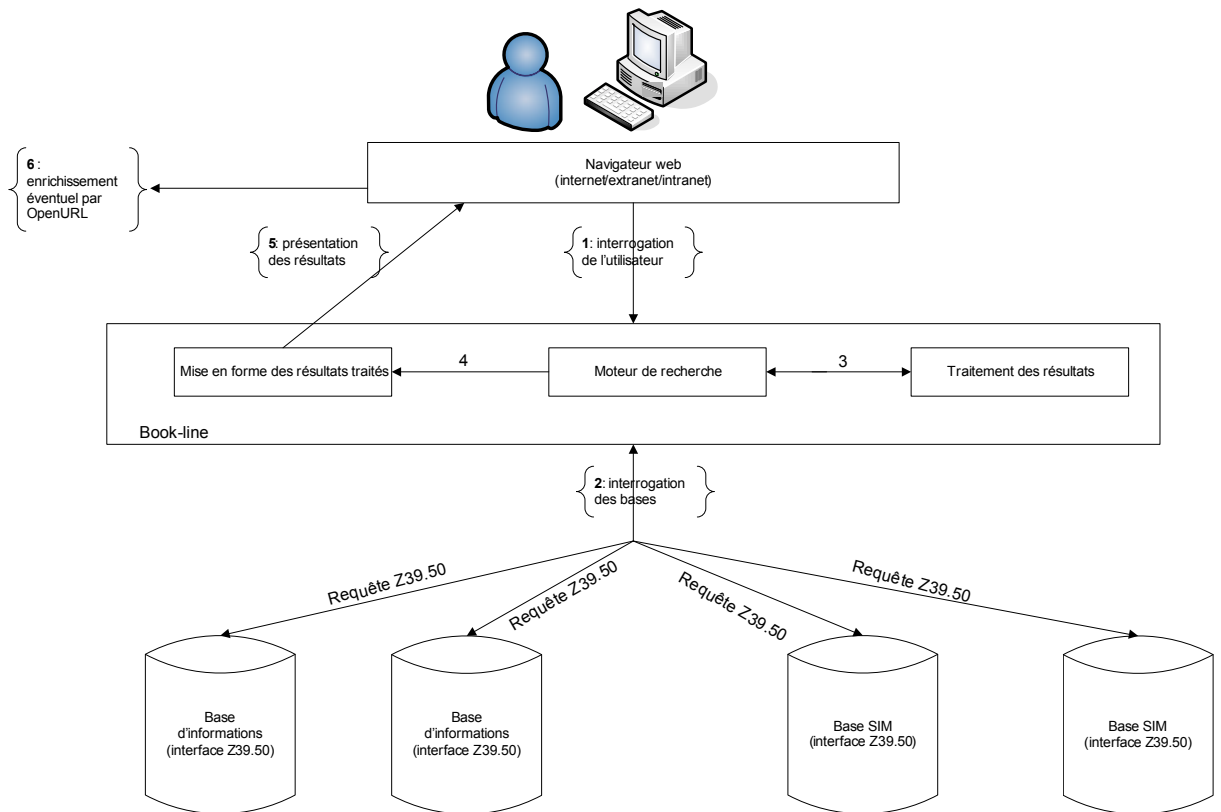
sont donc en XML DC). Mais l'application s'appuie en premier lieu sur le protocole Z39.50.

L'interface web de Book-Line est accessible par n'importe quel navigateur, dès lors l'utilisateur sélectionne les bases qu'il souhaite interroger et saisit sa requête dans le formulaire (en fait il a le choix entre 4 types de recherche : simple, experte, feuilletage, multicritères). Le moteur traduit la requête autant de fois qu'il y a de bases à interroger. Les requêtes se font en Z39.50 au départ puis sont éventuellement traduites si elles ont à traiter des sources SRW/SRU, une base multimédia. A noter que les bases conçues avec le progiciel SIM s'interfacent nativement avec Book-Line via une passerelle Z39.50.

Les résultats sont transmis par chaque base, Book-Line se charge de fusionner l'ensemble des réponses, de les trier et de les dédoubler. Ensuite, la réponse est mise en forme dans l'affichage, d'abord sous la forme d'une liste abrégée. L'utilisateur peut en consulter le détail et éventuellement les stocker dans un panier, les imprimer ou les exporter. Enfin, si l'utilisateur souhaite obtenir davantage de renseignements sur un résultat particulier, il peut déclencher une OpenURL.

5.2.1. Schéma du fonctionnement de Book-Line

Fonctionnement d'une recherche avec Bookline



5.3. La gestion de portail : MASC

La dernière couche progicielle proposée par Archimed dans le cadre d'un portail documentaire concerne précisément un outil de gestion de portail et de travail collaboratif, MASC. Par définition, le portail va permettre un accès unique à l'information, mais il va également donner la possibilité de diffuser l'information de façon personnalisée. Surtout, ce type de progiciel propose une véritable plateforme de travail collaboratif en intégrant en standard un bureau virtuel et des services personnalisés, mettant à la disposition des utilisateurs des outils de communication et de groupware (messagerie, forums, gestionnaire de fichiers, agenda électronique, trombinoscope etc.)

MASC permet de mettre en valeur ses services dans n'importe quel type d'environnement, internet, extranet ou intranet.

A l'image de SIM ou de Book-Line, MASC se base aussi sur des standards puisque le référencement des utilisateurs est conforme à LDAP¹⁹, les données du portail sont stockées en XML, les échanges avec les autres applications s'effectuent sur SOAP : MASC est un outil 100% web et 100% Unicode.

Côté administration, MASC permet de s'occuper facilement de la gestion des applications : l'administrateur déclare l'ensemble des applications du portail et attribue diverses fonctions. Un éditeur de menu permet de présenter et de classer par thèmes toutes les entrées qui donnent un accès aux services. Un gestionnaire d'encarts permet aussi de définir les points d'entrée de la page d'accueil, un encart étant défini comme un résumé d'une fonctionnalité du portail. L'administrateur dispose aussi d'outils pour distribuer une partie des tâches d'administration à certains inscrits, afin d'éditer une partie du menu, d'administrer les forums, de définir les droits d'accès aux fichiers de groupe, etc.

Enfin, MASC possède un module de gestion des inscriptions pour permettre à toute personne qui souhaite accéder aux services du portail d'en faire la demande en ligne. Ce module permet une gestion assez fine du rôle des utilisateurs grâce à la mise en place de profils de type d'utilisateurs pouvant exploiter le portail. Chaque utilisateur a un rôle bien défini sur le portail et des actions autorisées ou non.

5.4. Bilan : la proposition d'une offre globale

Grâce à l'imbrication des trois progiciels SIM, Book-Line et MASC, Archimed est donc en mesure d'offrir une véritable solution globale en matière de portail documentaire. MASC, en qualité de logiciel de gestion de portail va permettre à des utilisateurs, qu'ils soient authentifiés ou non, d'avoir un point d'accès unique vers de nombreux services si leur profil leur donne l'autorisation, même s'il est vrai que pour l'instant il s'avère que cet aspect est encore sous-exploité pour les bibliothèques. L'offre de services en terme de travail collaboratif joue à plein dans un portail de type entreprise ou collectivité locale (cf. §1.2).

¹⁹ *Lightweight Directory Access Protocol* (Protocole simplifié d'accès annuaire). Protocole normalisé de format d'annuaire et d'accès aux données d'annuaire.

La couche inférieure est celle constituée par Book-Line, qui offre le service de recherche multibase. Cette recherche fédérée peut donc être réalisée grâce à la passerelle Z39.50 qui équipe l'application, mais aussi avec les outils additionnels qui vont permettre d'interroger d'autres bases fonctionnant sous différents standards : OSC-R va compléter une recherche grâce à l'OpenURL, le client SRW de Book-Line permettra d'interroger des ressources supportant ce protocole, enfin SARA qui est en quelque sorte un émulateur de standards et de protocoles peut communiquer avec n'importe quelle type de base, du moment que le plug-in adéquat est installé sur le serveur. Enfin Book-Line peut également rechercher des documents de tout type et de tout format dans la troisième couche logicielle qui est le Système d'Information Multimédia qui fera office de base de GED de la bibliothèque ou de l'entreprise. Le SIM stocke donc des ressources GED locales ou des dépôts OAI moissonnés grâce à Harvey.

Plusieurs bibliothèques universitaires ont décidé de s'équiper avec ce type de solution : La Bibliothèque de l'Université du Littoral Côte d'Opale (BULCO) qui s'était associée avec Archimed pour le développement d'OSC-R a mis en place son portail autour de cette solution globale (SIM+Book-Line+MASC)²⁰, idem pour la bibliothèque de Paris Dauphine²¹, l'Université de Technologie de Troyes²² ou dans un cadre un peu différent le Pôle International de la préhistoire²³.

Mais l'exemple le plus représentatif de portail de nouvelle génération est le projet mené conjointement entre Archimed et l'Abes, projet qui met bien en lumière toute la problématique tournant autour de la recherche fédérée sur des ressources hétérogènes distantes.

²⁰ <http://www-bulco.univ-littoral.fr/UnivLittoral/BureauUnivLittoral/Integration/Pages/framesetPortail.asp>

²¹ <http://www.bu.dauphine.fr>

²² <http://www.utt.fr/bibliotheque/BienvenueAuSCD.php?rub=01>

²³ <http://www.pole-prehistoire.com>

6. Un exemple représentatif de portail de nouvelle génération en France : le portail Sudoc²⁴

6.1. Historique et prérogatives de l'agence

L'Agence Bibliographique de l'Enseignement Supérieur (Abes) a été créée par le Ministère de l'Enseignement Supérieur et de la Recherche via un décret ministériel du 24 octobre 1994 (décret n° 94-921). Cet établissement public, rattaché à la Direction de l'Enseignement Supérieur, est chargé de missions axées prioritairement sur la coordination et la coopération. Le but de cette agence est la conception et la mise à disposition de ressources documentaires collectives comme le Système Universitaire de Documentation ou Sudoc. Ses principales prérogatives sont²⁵ :

- Le recensement et la localisation des fonds documentaires des bibliothèques et des centres de documentation de l'enseignement supérieur.
- La coordination du traitement documentaire des collections.
- Le développement des systèmes et des applications nécessaires.
- L'édition sur tout type de support des produits dérivés de la base de données.
- Elle apporte son concours aux établissements d'enseignement supérieur dans le domaine de l'information bibliographique.
- Coopération avec les organismes dont les buts sont similaires, en France ou à l'étranger.

L'Abes n'a pas eu à travailler dans un environnement *ex-nihilo* puisqu'elle s'est appuyée sur des outils documentaires existant préalablement tels que Téléthèses, le Pancatalogue, le CCN-PS, etc. Constatant une hétérogénéité des réseaux et des applications documentaires, le besoin de constituer un catalogue

²⁴ AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Présentation générale du portail documentaire Sudoc [en ligne], janvier 2004. Disponible sur : <www.abes.fr/abes/documents/PortailPrésentation.pdf>

²⁵ BRZUSTOWSKI K., BARRAL S., L'ABES, coordinatrice et animatrice de réseaux [en ligne] BBF, 2003, t.48, n°2, p.32-42. Disponible sur : <http://bbf.enssib.fr/bbf/html/2003_48_2/2003-2-p32-barral.xml.asp>

unique permettant un accès direct et élargi au service de document s'est fait ressentir. C'est la naissance du Sudoc en 1995. Entre janvier 2001 et mars 2002 le Sudoc compte 113 établissements représentant 690 bibliothèques qui forment le "premier cercle" du système.

6.2. Genèse du projet

Depuis plusieurs mois, l'Abes a décidé de compléter les outils qui constituent son offre à ses usagers. Pour répondre aux besoins des utilisateurs, il s'agit d'offrir un accès à différents types de catalogues, mais aussi à des bibliographies et surtout à des documents en textes intégraux. Ce projet n'est possible qu'en enrichissant les outils de recherche en mettant en place un accès multibase via un outil de recherche fédérée, c'est-à-dire un métamoteur. L'Abes souhaite donc accéder à des ressources documentaires distantes et hétérogènes en utilisant une passerelle de recherche basée sur le protocole Z39.50, mais aussi HTTP (Web Services), en constituant une base de métadonnées alimentée par le moissonnage OAI et par l'intégration automatique de sites web, de thèses, etc. dans un cadre strictement contrôlé par une gestion des droits des utilisateurs pour chaque ressource en offrant un accès au texte intégral, aux documents primaires.

En juin 2003, l'Abes désigne Archimed comme maître d'œuvre du portail Sudoc.

6.3. Les outils caractérisant le portail

Le portail Sudoc est avant tout un projet d'outils puisqu'il sera composé :

D'une base XML (SIM) qui référencera les ressources et indexera les documents primaires avec possibilité également de les stocker et de les présenter à l'affichage. Cette base sera alimentée soit de manière directe via un formulaire de saisie ou par moissonnage de métadonnées via le protocole OAI-PMH ou par des outils d'import plus spécifiques.

Une passerelle de recherche fédérée (Book-Line) qui s'appuiera sur l'ensemble des protocoles vus jusqu'alors afin de mener des recherches sur les bases distantes, mais aussi sur la base multimédia XML ainsi que le catalogue Sudoc.

Un outil de contrôle d'accès (MASC) qui permettra de gérer les droits d'accès aux ressources acquises par les établissements. Cela passe par une gestion mutualisée des adresses IP de chaque poste des différents établissements par le portail Sudoc. Cet outil, conformément aux potentialités de MASC, offrira à ces établissements et à leurs utilisateurs un bouquet de services adaptés.

6.4. Les contenus du portail

Ces outils vont travailler autour de différents types de documents regroupant notamment des catalogues de bibliothèques et des bases bibliographiques, des thèses, des travaux de recherches, des fonds numérisés, des bases de signets. Le portail donnera l'accès à des ressources libres, mais également à des ressources dont l'accès est contrôlé ; l'accès sera alors possible soit par authentification de l'utilisateur (login) ou par reconnaissance IP de la machine demandant l'accès à la ressource.

Enfin, le portail permettra d'accéder lorsque cela est possible aux documents primaires et donc de pouvoir réaliser des recherches sur le texte intégral.

6.5. Gestion des formats, des documents et des connexions

Le portail est capable de gérer différents formats bibliographiques et de documents : pour les formats bibliographiques, l'on s'appuie surtout sur MARC, DC, DC AFNOR thèses pour les thèses ; pour les documents primaires, citons les formats propriétaires (PDF, office, etc.) ou XML (TEI, EAD, DocBook, etc.).

Le portail pourra gérer deux types de connexions : les connexions anonymes et celles basées sur une identification. Les accès anonymes permettront d'accéder au catalogue Sudoc et à d'autres catalogues de bibliothèques ou de thèses, des ressources en libre accès, des fonds numérisés. Il sera aussi possible de gérer un

panier pour la session, d'imprimer et de charger ses résultats ou d'en exporter certains.

Dans le cas des connexions par identifications (login ou IP), il est possible d'accéder aux ressources pour lesquelles l'établissement a souscrit un abonnement (qui peut être géré soit par l'établissement ou soit par l'Abes en qualité de coordonnateur de groupement de commandes), à des fonds dont l'accès leur est autorisé, à des services personnalisés si l'utilisateur est identifié par login (compte personnel dans son panier).

Le portail Sudoc est enfin caractérisé par un fonctionnement coopératif, avec des fonctions telles que la suggestion d'une source en accès libre, le dépôt de documents numérisés, etc.

6.6. Intérêts d'un tel portail

L'exemple du portail Sudoc est doublement intéressant à étudier d'un point de vue technique mais aussi bibliothéconomique. En simplifiant, ce portail aurait l'allure de n'importe quel portail documentaire universitaire ; en réalité le déploiement de ce portail n'est pas si aisé qu'il n'y paraît parce qu'il ne concerne pas que le fonds du Sudoc, mais désormais il y a tout ce qui concerne la gestion des droits des autres utilisateurs du portail. Le portail Sudoc s'appuie sur un "principe de subsidiarité" en réalisant au niveau national ce qui est difficile à réaliser localement parce que l'Abes dispose de moyens informatiques conséquents (il s'agit du Cines, Centre Informatique National de l'Enseignement Supérieur), en mutualisant les ressources communes et en valorisant les ressources sans se substituer à l'autonomie des projets universitaires.

Surtout, il faut souligner que l'Abes n'est pas producteur ni acquéreur de ressources documentaires, elle ne peut que diffuser des ressources choisies par les établissements ou le Ministère.

Le portail Sudoc cherchera donc à valoriser des ressources électroniques d'intérêt national, des sources avec abonnements, d'autres ressources électroniques qui peuvent être déjà diffusées par des portails locaux mais dont l'intérêt inciterait

l'Abes à les diffuser plus largement afin de les rendre plus visibles. Le portail Sudoc peut aussi mettre en valeur des fonds numérisés qui ne peuvent être hébergés par des établissements qui ne possèdent pas des systèmes de GED, des sources du web sélectionnées et validées, des réservoirs de travaux universitaires du genre *Open Archive Initiative* qui peuvent promouvoir la recherche nationale (CCSD), etc.

Enfin, certains outils du portail peuvent être intégrés dans les portails locaux. On pense surtout aux encarts de recherche qui pourront être accessibles directement sur le portail d'un établissement, mais aussi d'un encart qui rassemblera toutes les ressources à abonnement accessibles à un établissement.

7. Bilan : Le choix d'Archimed de s'investir dans les nouveaux protocoles en étant implémenteur *de facto*

En guise de conclusion à cette partie, on peut dire que la recherche documentaire a pris un nouvel élan depuis le début des années 2000, un élan qui a aussi fait rentrer davantage les bibliothécaires dans le monde des réseaux, du numérique et du web. Pour l'utilisateur, il revêt un côté relativement simple à aborder et sa convivialité en fait un outil attractif. Pour le bibliothécaire et les sociétés qui œuvrent dans les systèmes d'information, toute l'astuce réside dans cette impression de facilité, d'attractivité et de convivialité qu'il s'agit d'offrir aux usagers.

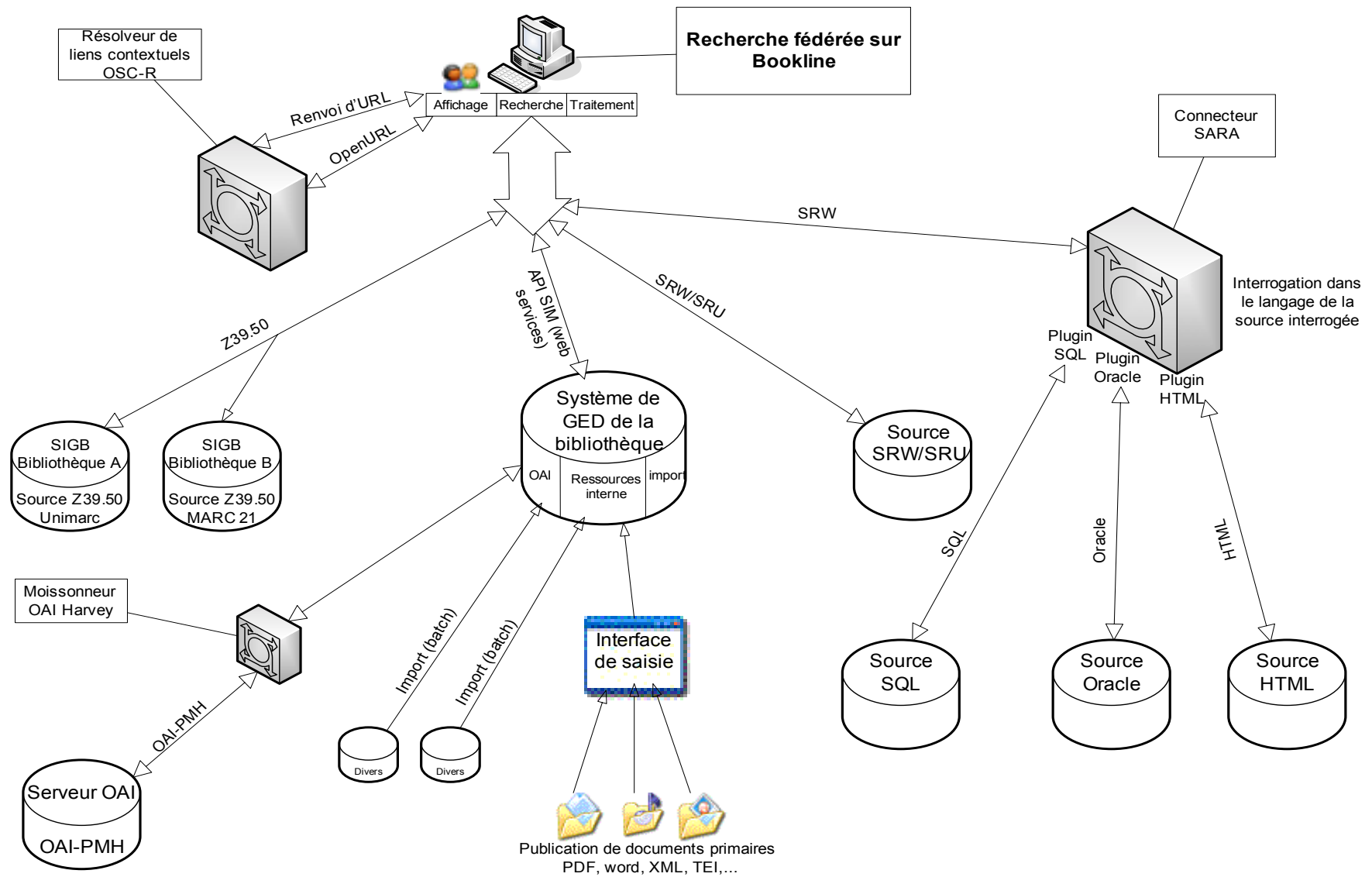
On l'a vu, derrière un portail documentaire et le formulaire de recherche se cache un univers assez complexe, fait de protocoles différents, de documents et de formats variés. Qu'ils s'agissent de Z39.50, de SRW-SRU, d'OpenURL, de ressources OAI, de bases de signets, de dossiers documentaires ou de ressources issues du système de GED de la bibliothèque, le but de la recherche fédérée est de donner l'illusion aux usagers de mener leur recherche de façon unique en leur offrant qui plus est une présentation homogène des résultats obtenus, même si ces derniers viennent de bases différentes.

La complexité réside donc dans la conception des passerelles permettant la communication entre des protocoles comme le fait SARA, de récupérer des notices dans des formats variés (MARC, XML, OAI DC, etc.), pour également les présenter de façon homogène, c'est-à-dire en ce qui nous concerne en Unimarc. Nous le verrons par la suite, les projets de recherche fédérée ne sont pas simples à réaliser, mais ce n'est pas à cause de problèmes techniques. Archimed a donc misé sur le développement des nouveaux protocoles de communication et d'interrogation afin d'offrir à ses clients un produit d'une grande stabilité technique, bâti entièrement sur des standards normalisés.

En plus de proposer une solution progicielle qui peut répondre aux besoins partiels d'un des ses clients (puisque SIM, Book-Line et MASC sont indépendants et peuvent être utilisés seuls) ou de besoins globaux comme nous l'avons vu avec le cas des portails documentaires, Archimed s'est investi dans la recherche sur les nouveaux protocoles avec le développement de projets tels qu'OSC-R. Mais c'est surtout du côté des Services Web qu'Archimed mise car il semblerait que d'ici quelques années, Z39.50 pourrait laisser sa place à SRW. A cet égard, Archimed est implémenteur de fait de cette norme, même si elle n'a pas encore inscrit son nom dans la liste des implémenteurs du ZING.

Maintenant que nous avons une idée sur ce que peut représenter une offre progicielle en matière de portail et les technologies qui sont mises en jeu autour, nous pouvons entrer dans le détail des recherches effectuées au cours de ce stage sur cette question des accès normalisés, notamment vers les périodiques électroniques car c'est là que nous allons trouver matière à discuter, en corrélation avec la recherche fédérée.

Schéma de synthèse : Recherche fédérée avec Book-Line



Partie 2 : Méthode et organisation des recherches

Après avoir vu l'évolution de la recherche documentaire ainsi que l'offre proposée par Archimed parallèlement à l'analyse des différents types de protocoles de recherche et de récupération de l'information et de collecte de métadonnées, nous pouvons enfin nous pencher sur la problématique du stage réalisé au sein de la société lilloise. Comme on a pu le constater dans la partie précédente, la problématique ne se situe pas réellement au niveau technique. Archimed a développé des applications qui ont fait leur preuve notamment au sein des collectivités et des bibliothèques. L'association des trois briques constituées par SIM, Book-Line et MASC représente une solution fiable et dont la place est reconnue sur le marché des systèmes d'information documentaires. Il en est de même avec les nouveaux protocoles de recherche ou Archimed et son équipe d'ingénieurs s'impliquent particulièrement sur les *web services* et sont peut-être les seuls (en qualité de SSII) à miser sur ces nouvelles technologies dont les buts seraient d'effacer progressivement Z39.50, toutefois en continuant de travailler avec.

Le problème se pose en fait non pas en terme de contenant, mais plutôt en terme de contenu : c'est particulièrement au niveau des périodiques électroniques que les questions restent ouvertes, notamment sur la capacité des nouveaux protocoles à s'imposer.

Cette partie va dans un premier temps mettre en lumière cette problématique de la gestion des périodiques électroniques dans le cadre d'un portail fédérateur en étudiant les besoins exprimés par les différentes communautés et en faisant un état de l'art sur certaines solutions mises en œuvre. Ensuite, je vais m'attacher à montrer les sources auxquelles j'ai eu recours dans le cadre de mon stage. Cela est passé par la mise en place d'une veille informative, documentaire et concurrentielle. Enfin, il sera question d'analyser l'attitude des acteurs de l'édition électroniques qu'il aura fallu directement questionner afin de connaître leur

opinion et leurs éventuelles propositions dans le sens de la mise en place d'un portail fédérateur.

1. Problématique du stage : la recherche fédérée et la gestion des périodiques électroniques

1.1. Premier constat : l'expression des besoins en matière de recherche fédérée

1.1.1. Un besoin inconscient chez les étudiants et les novices de la recherche documentaire...

A mon arrivée chez Archimed, je n'étais pas encore capable d'évaluer ce qu'étaient les besoins réels et exprimés en matière de recherche fédérée. Mais il suffit de se remémorer les souvenirs d'étudiant pour se rendre compte qu'un tel outil constituait quelque chose qui serait à la fois confortable et efficace. Confortable parce qu'on ne perdrait pas de temps à rééditer dix fois les mêmes requêtes sur dix moteurs différents (il suffit d'en faire l'expérience avec Copernic par exemple). Efficace parce que les résultats obtenus sont des résultats homogènes, ce qui évite bien sur toute perte d'information si on devait interroger plusieurs moteurs (et bien sûr la pertinence dépend aussi du type de requête que l'on pose...).

Il semble toutefois que ce besoin informationnel n'est pas vraiment conscient au sein de la communauté étudiante (pour le niveau licence première année nous dirons). Un étudiant ou un utilisateur final quelconque n'a peut être pas forcément conscience de l'hétérogénéité des ressources dont il pourrait disposer, mais aussi des systèmes existants. Il cherche simplement à avoir accès à un réservoir de documents ou de références avec un mode de recherche. Maintenant, si l'on fait prendre conscience aux utilisateurs finaux (étudiants ou usagers de bibliothèques) de l'hétérogénéité des ressources, la demande de leur part irait très sûrement dans le sens d'un outil de recherche fédérée.

1.1.2. ... mais un besoin réel et exprimé par la communauté des chercheurs

Par contre, la question ne se pose pas pour les chercheurs et tous les professionnels de l'information. L'outil de recherche fédérée va leur être extrêmement précieux en terme de gain de temps pour la recherche et en terme de qualité avec davantage de documents fiables récupérés (alors qu'avec une recherche classique un certain nombre de références pouvaient échapper au chercheur). D'un point de vue strictement informationnel, la recherche fédérée est d'abord un besoin exprimé par les chercheurs et les spécialistes de l'information.

1.1.3. Un besoin exprimé par les universités et les bibliothèques en terme de rentabilité de leurs abonnements

Enfin, du côté des responsables des bibliothèques, la mise en place d'un outil de recherche fédérée est également pressentie comme un besoin et ce pour une double raison à la fois bibliothéconomique et économique. Il est évident que les bibliothécaires souhaitent offrir dans la mesure de leurs moyens un outil de recherche performant et confortable pour leurs utilisateurs. Si leur bibliothèque offre la possibilité d'interroger différents catalogues, un système de GED et des ressources électroniques auxquelles elle est abonnée, la mise en place d'un outil de recherche fédérée s'impose d'elle-même. D'un point de vue économique, il n'est pas utile de rappeler que les bibliothèques universitaires consacrent une partie considérable de leur budget de fonctionnement aux abonnements aux ressources électroniques. Avec un mode de recherche classique, il n'est pas dit que les étudiants exploitent au maximum les revues électroniques, surtout si l'établissement a souscrit des abonnements auprès de différents éditeurs ou diffuseurs. Les ressources hétérogènes impliquent donc des modes de recherche et de présentation des résultats hétérogènes avec des interfaces qui ne sont pas souvent en français. Il s'avère donc qu'un outil de recherche fédérée pourrait aider à mieux rentabiliser les abonnements souscrits par une université grâce à une meilleure exploitation des ressources auxquelles la bibliothèque est abonnée. Avec une seule interface de recherche, l'information détenue par un établissement

devient alors plus visible et est donc valorisée et valorise par la même le portail de la bibliothèque.

Seulement, voilà : mettre en place une recherche fédérée permettant d'interroger les bases des éditeurs ou diffuseurs en ligne n'est pas une chose aussi aisée même si les technologies permettant de réaliser ce dessein existent. Nous verrons qu'il n'est pas forcément dans l'intérêt des éditeurs commerciaux de laisser leurs bases se faire interroger indépendamment de leur propre moteur de recherche. Nous allons voir qu'il existe cependant des alternatives pour arriver à faire ce type de recherche.

1.2. L'état de l'art sur les solutions mises en œuvres actuellement :

1.2.1. Le portail de la bibliothèque universitaire de Paris Dauphine : les limites du revamping HTML

Si on souhaite utiliser des protocoles normalisés pour mettre en place une recherche fédérée permettant d'interroger des bases de données issues des éditeurs ou diffuseurs de contenus en ligne, il faut donc avoir l'autorisation ou éventuellement avoir passé un accord avec ces derniers. En effet, pour pouvoir interroger leurs bases, il faut connaître leur profil d'interrogation, tout comme cela se fait entre les bibliothèques avec Z39.50. *A priori*, il n'y aurait donc pas d'autres possibilités à l'exception d'une méthode assez empirique : il s'agit du *revamping HTML* que l'on traduit par "rhabillage" ou plus trivialement "ravalement de façade". La technique est relativement simple et consiste en deux étapes :

Il s'agit dans un premier temps d'identifier la méthode de requête utilisée par l'interface de recherche (GET ou POST), cela passe donc par une analyse de la page du formulaire.

La seconde étape est l'analyse et l'interprétation du code source de la page de résultat d'un éditeur qu'on souhaite interroger sans passer par son interface classique. Dans le code HTML, il faut localiser des séquences de balises (BR, TD, SPAN, etc.) qui statistiquement ne se produisent qu'une seule fois pour former un

champ d'interrogation. Une fois ces champs localisés puis "rhabillés", on peut effectivement mener une recherche indépendamment du moteur de recherche initial de l'éditeur. Archimed a expérimenté cette technique avec l'Université de Paris Dauphine, mais on s'est rapidement aperçu que le *revamping HTML* ne saurait être une solution s'inscrivant dans la durée. Ce n'est pas un procédé normalisé et surtout il est à la merci de la moindre modification du formulaire par le webmaster de l'éditeur de contenu. En cas de modification, le *revamping* est à refaire, ce n'est donc pas une technique fiable.

1.2.2. Ebauche d'une solution de gestion collective des périodiques électronique : l'APE

Un autre problème concerne le recensement des périodiques électroniques en France dans la mesure où il n'existe pas encore de bases normalisées de périodiques. L'Abes, en association avec le consortium Couperin et avec l'expertise technique d'Archimed ont décidé de mettre en place une solution pour référencer les milliers de revues auxquelles les universités et divers établissements sont abonnés. Il s'agit d'une base nationale de toutes les revues, ce qui permet de savoir qu'une revue donnée appartient à un éditeur ou un bouquet de service donné et qu'une ou plusieurs universités y sont abonnées. Ce système est là pour faciliter à un utilisateur de retrouver instantanément la liste des abonnements de son université. Ce système est le module d'Accès aux Périodiques Electroniques (APE). L'APE permet de répondre à la question "qui est abonné à quoi ?" et, devant la multiplicité des titres, il répond à la problématique d'un catalogue collectif des abonnements électroniques.

Rappelons toutefois qu'avant l'APE, le système qui prédominait était la Base d'Information Mutualiste sur les Périodiques Electroniques (BIMPE), sorte de réservoir que les établissements pouvaient télécharger et traiter en local (en fait chacun disposait de sa propre base à partir du BIMPE). Aujourd'hui, la BIMPE sert à alimenter l'APE qui gère à peu près 800 établissements (contrôle de l'accès aux abonnements par adresse IP)

Dans le cadre d'une démarche de portail fédérateur, l'APE est une réponse assez satisfaisante, constituant déjà un premier pas en localisant les ressources et en sachant qui y est abonné.

1.2.3. Le Crossref Pilot Search : l'homogénéité de la recherche mais une hétérogénéité dans la présentation des résultats

Dans la mesure où le sujet du stage touche à la problématique de la recherche fédérée, il était impossible de passer sous silence le projet de métamoteur proposé par Crossref en collaboration avec Goggle, le *Crossref Pilot Search*²⁶. Comme son nom l'indique, ce n'est encore qu'un pilote qui fut lancé en avril 2004 avec l'association de neuf éditeurs qui sont bien sur membres de Crossref tels Oxford University Press, Blackwell, Wiley, etc. En août 2004, le pilote donne la possibilité de mener une recherche fédérée auprès de 29 éditeurs.

C'est indubitablement une solution assez séduisante puisque ce produit propose une recherche fédérée alliant la technologie du moteur le plus utilisé de la planète avec un contenu dont la forte valeur ajoutée informationnelle est incontestable.

Si l'on revient dans notre cas de portail fédérateur de bibliothèque universitaire, la solution proposée par Crossref est bien évidemment très intéressante, mais elle n'est qu'à mi-chemin de ce que souhaiterait mettre en place Archimed. Le moteur pilote de Crossref réalise bel et bien une recherche fédérée, les résultats sont présentés sous forme de liste abrégée (comme pour Book-Line ou n'importe quel moteur ou passerelle de recherche), par contre dès que l'on va sur la ressource (à condition d'y avoir accès), nous sommes dans une hétérogénéité totale. On se retrouve sur le site de l'éditeur avec parfois un lien pointant vers les références et le document recherché (en PDF) ou alors avec le texte en HTML sur la page du site. Bien que la solution soit déjà très séduisante, pourquoi ne pas essayer d'aller au bout de la réflexion et de donner accès directement au texte intégral dans un format homogène ? Enfin, on peut aussi souligner les insuffisances liées au fonctionnement des moteurs de recherches qui ne sont pas

encore adaptés à une véritable recherche documentaire puisque la technologie des moteurs ne permet pas encore de faire une interrogation sur les métadonnées comme c'est le cas avec Z39.50. Bien que je n'aie pas vraiment testé le pilote de Crossref dans le détail, il y a fort à parier que les recherches menées sur ce métamoteurs génèrent du bruit, comme c'est effectivement le cas sur Google.

2. Mise en place d'une veille informative, documentaire et concurrentielle

Toute la durée du stage a été caractérisée par un travail de veille sur la problématique de la recherche fédérée ainsi que de l'édition électronique dans ses aspects commerciaux et dans l'optique de l'*Open Access*. Cette veille va consister dans un premier temps à appréhender l'environnement concurrentiel d'Archimed afin de cerner qui est également positionné sur le marché des portails documentaires et qui travaille éventuellement sur des accès normalisés vers des bases de données hétérogènes. Dans un second temps, il s'agira de se tenir au courant de l'actualité bibliothéconomique propre à cette problématique par le biais de sites et de revues spécialisées.

2.1. Observation de l'environnement concurrentiel

Il existe beaucoup de sociétés qui gravitent autour du marché des bibliothèques et plus globalement autour des solutions de portails intégrés. Archimed connaît principalement trois concurrents qu'elle peut croiser au cours d'appels d'offres en France.

Le principal concurrent est la société Ever avec ES-Portal. Ce portail, à l'image de celui d'Archimed permet la recherche de documents à partir de sources hétérogènes avec toutes les fonctionnalités de gestion propre à n'importe quel portail (gestion des droits d'accès, outil de travail collaboratif, etc.). Notons cependant qu'Ever dispose d'un serveur Z39.50 depuis un peu plus de 5 ans.

²⁶ <http://www.google.com/cobrand?restrict=crossref&q=&filter=0&sa=Search%20&cof=AWPID:bbd6d01e9a530922>

Une autre société spécialisée dans le développement des SIGB commence à se positionner sur le marché des portails, il s'agit de l'entreprise canadienne Bibliomondo avec sa solution progicielle Zone-1. On retrouve également la société Cadic qui propose son offre modulaire Ex-Libris (à ne pas confondre avec la société israélienne Ex-libris qui propose notamment Metablib, SFX et Aleph). Enfin, La société Jouve se positionne également sur le marché des portails en France avec une offre également modulaire.

Une fois ce panorama établi, il reste à savoir si ces sociétés s'intéressent au développement des protocoles normalisés afin de pouvoir réellement offrir la possibilité de mettre en place un portail fédérateur. Le parcours des sites web de ces différents acteurs ne m'a pas permis de me faire une idée sur ce qui est réellement fait. Ceci étant, aucun d'entre eux ne fait la promotion du développement des *web Services* ou même de l'OpenURL (pour l'OpenURL, signalons qu'il y a cependant d'autres sociétés qui ont développé des résolveurs de liens contextuels, mais aussi des éditeurs de contenus comme Ebsco ou Ovid). Il est probable que ces sociétés surveillent de près ce qui est fait du côté de ces nouveaux protocoles, mais pour l'instant c'est toujours la norme Z39.50 qui reste la référence en matière de recherche.

Par contre, le site du ZING m'a quand même appris que certaines entreprises s'intéressent de près à la réflexion sur la recherche fédérée en corrélation avec les *Web Services* et notamment SRW. Un atelier datant du 30 octobre 2003 montre que parmi les participants (auditeurs) on retrouve des personnes travaillant pour Dynix, Endeavor ou Ex-Libris en ce qui concerne les maîtres d'œuvre techniques. Du côté des éditeurs ou diffuseurs de contenus, on retrouve Elsevier, Kluwer ou Gale.

Il semble que les spécialistes de ces sociétés soient attentifs aux évolutions de la norme, mais chacun semble attendre que les besoins dans cette voie se fassent clairement ressentir pour commencer à développer des produits permettant de travailler avec SRW²⁷. Notons par ailleurs que la liste des implémenteurs de la norme SRW recensée sur le site du ZING ne compte pas de SSII dans ses rangs. Tous les implémenteurs sont des institutionnels comme la Bibliothèque du Congrès, l'OCLC, la Bibliothèque Royale des Pays-Bas, l'université de Berkeley,

à l'exception de la société danoise spécialisée dans le conseil en systèmes d'information Indexdata. On l'a vu, Archimed est également implémenteur de la norme (avec le connecteur SRW de Book-Line et le connecteur SARA) même si la société nordiste n'a pas encore publié son travail sur le site du ZING. On ne sait pas si d'autres sociétés ont développé le même genre d'application, cela reste dans le secret des services RD des sociétés concurrentes.

2.2. Veille informative et documentaire

Ce travail s'est axé autour des revues référentielles en sciences de l'information d'une part et des sites spécialisés, weblogs et pages professionnelles d'autre part.

2.2.1. Les revues et les sites référentiels en sciences de l'information

Afin de localiser des ressources référentielles pour mon sujet, je me suis orienté en premier lieu vers les sites fédérateurs en sciences de l'information. A défaut de tous les citer, on peut néanmoins insister sur les signets en bibliothéconomie et en sciences de l'information de l'Université de Laval au Québec²⁸, mais aussi le site la page de l'Abes consacrée aux ressources en sciences de l'information²⁹ et bien sûr le site fédérateur de l'Enssib, Sibel.

Parmi les revues accessibles en ligne, mon choix s'est orienté vers la consultation de *D-Lib Magazine*³⁰, *Ariadne Magazine*³¹, la revue de l'*Association of Research Libraries (ARL)*³². Pour celles demandant un abonnement, je me suis penché sur *Archimag*³³ et la *Revue Documentaliste*³⁴. Notons enfin un e-zine récent et particulièrement intéressant, animé par Nicolas Morin du SCD d'Angers et par

²⁷ <http://www.loc.gov/z3950/agency/zing/srw/metasearch-srw.html>

²⁸ <http://www.bibl.ulaval.ca/info/scinfo.html>

²⁹ <http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=7&tabid=17#ressources>

³⁰ <http://dlib.org>

³¹ <http://www.ariadne.ac.uk>

³² <http://www.arl.org>

³³ <http://www.archimag.com>

³⁴ <http://www.adbs.fr/site/publications/documentaliste/index.php>

Marlène Delhayes du SCD d'Aix-Marseille 3 : *BiblioAcid*³⁵ est axé sur les technologies et les bibliothèques, il est aussi constitué de traduction d'articles bibliothéconomiques anglo-saxons, c'est une ressource assez pertinente dans le cadre de ce stage.

2.2.2. Weblogs bibliothéconomiques, listes de diffusion et pages professionnelles

Le phénomène des weblogs, véritables carnets du web permettant une mise à jour quotidienne relativement aisée, s'est étendu depuis peu vers le monde des bibliothèques. Deux nous intéressent particulièrement : le blog de BiblioAcid³⁶ et Figoblog³⁷.

Concernant les listes de diffusion, la veille s'est concentrée sur Biblio-FR et la liste de l'ADBS. Enfin, il existe des sites professionnels mis en place par des spécialistes de l'information, en général des bibliothécaires : on retiendra le site de Nicolas Morin³⁸, celui de Jean-Charles Houpiet³⁹ et de Dominique Lahary⁴⁰ (avec notamment un cours de référence sur Z39.50).

C'est donc autour de toutes ces ressources qu'il aura fallu être tenu informé de ce qui pouvait se passer dans le domaine de la recherche fédérée et éventuellement sur les nouveaux protocoles de recherche et de récupération de l'information. Il faut reconnaître que les informations obtenues arrivèrent au compte-gouttes, notamment sur la question des protocoles qui connaît un nombre encore limité de publications.

C'est donc en contactant directement les acteurs de l'édition électronique que je pouvais espérer obtenir plus d'informations sur l'éventuelle utilisation de ces protocoles et sur la recherche fédérée.

³⁵ <http://www.biblioacid.org/revue/>

³⁶ <http://www.biblioacid.org>

³⁷ <http://figoblog.ouvaton.org/>

³⁸ <http://morinn.free.fr>

³⁹ <http://jhoupier.free.fr/>

⁴⁰ <http://membres.lycos.fr/vacher/>

3. Faisabilité de la recherche fédérée : l'attitude des acteurs de l'édition électronique

Conscients de l'évolution des besoins informationnels des étudiants, des chercheurs et des spécialistes de l'information, l'attitude des acteurs de l'édition économique n'est pas unanime en fonction de leurs objectifs, selon qu'ils s'agissent de prestataires commerciaux ou d'acteurs de *l'Open Access*.

3.1. Du côté des acteurs commerciaux

3.1.1. La recherche fédérée entre-t-elle dans les intérêts des éditeurs/diffuseurs ?

Cette question pourrait être posée autrement, car il ne s'agit pas de savoir si la recherche fédérée proprement dite entre dans les intérêts des acteurs commerciaux de l'édition électronique, mais de savoir si un outil de recherche fédérée indépendant du moteur d'un éditeur est envisageable par ces derniers. On est en face d'un problème d'image car l'utilisateur utilise un moteur différent de celui de l'éditeur pour récupérer une information de qualité (en général) dans la base d'un éditeur donné. Si la recherche est absolument transparente, l'utilisateur obtiendrait uniquement son document en PDF avec éventuellement une mention furtive de l'éditeur détenant les droits sur le document.

La recherche fédérée intéresse les éditeurs, diffuseurs et agrégateurs de contenus, surtout dans le contexte de crise que connaît actuellement l'édition électronique. Les bibliothèques ont du mal à accepter les prix très élevés imposés par les principaux acteurs et certaines se posent la question du renouvellement de certains abonnements dans la mesure où la rentabilité de ces investissements n'est pas forcément avérée⁴¹. On l'a vu, la mise en place d'un système fédérateur dans la recherche de ressources hétérogènes peut être une bonne solution pour

⁴¹ Sur le sujet de la crise des périodiques et de l'alternative au modèle commercial, le premier numéro en ligne de BiblioAcid recense quatre articles : CRAWFORD W., REED C., DAVIS Ph., DELHAYE M., Articles sur la "crise des périodiques" [en ligne]. BiblioAcid, Vol.1, n°1, mars 2004. Disponible sur : http://www.biblioacid.org/revue/pdf/2004_01.pdf

rentabiliser ces abonnements en les rendant facilement accessibles et dont l'information serait davantage plus visible. Certains éditeurs ont franchi le pas vers cette solution, à l'image des 29 partenaires du projet pilote de Crossref et de Google qui devraient être d'ailleurs rejoints par d'autres selon les renseignements obtenus auprès de certains éditeurs comme EDP Sciences⁴². Il semble que d'autres acteurs réfléchissent sur leur position et leur choix vis-à-vis de la recherche fédérée. Afin de mieux saisir leurs points de vue, le mieux a été de les contacter directement.

3.1.2. Méthode d'approche pour une mise en contact

L'étape préliminaire à l'enquête a été bien évidemment de cerner et d'identifier les sociétés qui allaient être contactées. A cet égard, un travail de documentation a été mené pour créer une base de connaissances qui identifie les éditeurs/diffuseurs/agrégateurs avec l'adresse de leur site web et, si cela est possible, l'adresse de leur site en français, leur adresse physique, les coordonnées téléphoniques (standard, coordonnées du responsable informatique) et le mail. Cette base sera également alimentée par l'offre proposée par ces sociétés ou institutions.

L'établissement de cette liste de contacts à prendre a été réalisé selon les priorités d'Archimed et selon les vœux de ses clients. Par la suite, j'ai continué à alimenter cette base grâce notamment aux différents supports de cours de Marie-Pascale Baligand et de Ghislaine Chartron obtenus au cours de l'année de formation à l'Enssib, mais aussi via de nombreuses pages et sites web institutionnels présentant les principaux acteurs de l'édition électronique⁴³. Enfin, je me suis également penché sur les premiers collaborateurs du Crossref Pilot Search et j'ai jugé qu'il était également utile de les interroger vu qu'ils sont entrés de plain-pied dans un projet de recherche fédérée. Je n'ai pas établi de priorités par

⁴² Le dernier recensement des partenaires du Crossref Pilot Search est arrêté au 1^{er} septembre 2004 et est donc susceptible d'évoluer. La liste de ces partenaires est disponible sur le site de Crossref : <<http://www.crossref.org/crossrefsearch.html>>

⁴³ Notamment la page de J.P. Lardy sur les serveurs commerciaux de bases de données bibliographiques : <<http://urfist.univ-lyon1.fr/bd-serv.html>>.

rapport aux nationalités des prestataires, on y trouve des cibles françaises, anglaises, américaines, allemandes, etc.

J'ai donc contacté un échantillon de 97 acteurs de l'édition électronique, composé d'éditeurs, de diffuseurs, d'agrégateurs, qu'ils s'agissent de bases de données bibliographiques, en texte intégral ou de bases factuelles. L'on trouvera aussi les acteurs référentiels de l'Initiative des Archives Ouvertes ainsi que des sites institutionnels de bibliothèques virtuelles ou de catalogues collectifs. A la date du 10 septembre 2004, et suite à de nombreuses relances auprès des différents éditeurs du panel que j'ai constitué, j'ai obtenu 40 réponses de leur part. Pour environ une trentaine de cibles interrogées mais dont la réponse n'est pas encore arrivée, un véritable contact a été établi mais mon enquête doit être soumise aux autorisations de la hiérarchie de ces sociétés ou instituts (notamment les éditeurs juridiques), ce qui a pour résultat une attente parfois très longue qui peut sombrer dans l'oubli si l'on ne relance pas nos interlocuteurs. Malgré tout, les personnes interrogées ont bien voulu m'accorder du temps pour que je leur expose mon travail et je n'ai eu que deux notifications de fin de non-recevoir directes. Je pense que si Archimed continue cette enquête, cette base de connaissances pourrait comporter environ une soixantaine de réponses, ce qui n'est pas négligeable en tenant compte des questions posées sur des technologies utilisées par ces acteurs dont certains considèrent qu'il s'agit d'une information suffisamment sensible pour ne pas être divulguée.

Plusieurs tableaux présentant ces contacts sont joints en annexe 1.

3.1.2.1. *Connaître les technologies utilisées actuellement et éventuellement les projets*

La stratégie employée pour mener mon enquête a été au départ de faire du mailing. Cette technique s'est avérée assez décevante dans la mesure où certains mails étaient envoyés à des adresses du genre [info@...](#) ou [support@....](#), mes mails restaient pour la plupart lettre morte. Par contre j'ai obtenu des résultats assez rapides auprès d'autres cibles, comme le moteur d'Elsevier Scirus, mais j'ai

également rapidement constaté que les questions posées n'étaient pas forcément adéquates et ne répondaient pas aux exigences de mon commanditaire.

De ce fait, j'ai opté d'une part pour un contact téléphonique direct afin d'avoir un interlocuteur clairement identifié comme le responsable du service technique ou informatique de la société ou de l'institution visée. Ici aussi, j'ai eu à faire face à des difficultés, l'interlocuteur espéré n'étant pas toujours joignable et surtout du fait que nous nous trouvions en période de vacances. Ceci étant, j'ai pu obtenir les adresses mails voire les lignes de téléphone direct des responsables informatiques.

La seconde étape est donc de renvoyer un mail, mais cette fois beaucoup plus complet, sans être cependant trop exhaustif mais suffisamment renseigné afin d'éviter de décourager un destinataire qui doit avoir un grand nombre de courriels à traiter quotidiennement.

A partir d'un argumentaire basé sur les nouveaux enjeux et les besoins constitués par la recherche fédérée (illustré notamment par l'exemple du pilote Crossref), la question principale porte sur les technologies utilisées par les éditeurs en terme de protocole de recherche et de récupération de l'information (*information search and retrieval protocols*). Il s'agit de savoir si :

- Les bases visées supportent le protocole Z39.50, l'OpenURL, les *Web services* SRW et SRU ou si elles sont interrogeables via des protocoles propriétaires.
- Une question est posée à propos de l'éventuelle compatibilité des bases avec le protocole OAI-PMH.
- Enfin, il s'agit également de tenter de savoir si les bases de données sont accessibles via une passerelle XML dans le cadre du renvoi d'information grâce à ce format.

3.1.2.2. *Les sentiments et les projets des acteurs vis-à-vis de la recherche fédérée*

Dans un second temps, mon enquête s'est orientée sur un côté plus bibliothéconomique et partenarial. En effet, il s'agit de savoir si les cibles interrogées se sont déjà lancées dans un projet de recherche fédérée et surtout si elles ont déjà autorisé un de leurs clients à pouvoir consulter leurs bases de

données sans passer par leur interface d'interrogation classique (ce qui est le cas de Scirus qui offre ce type de partenariat). Si ce n'est pas le cas, la question est de savoir si ce type de projet peut voir le jour et si la société interrogée pense pouvoir se joindre à un projet de recherche fédérée à l'image de celui de Crossref ou selon ce que Archimed souhaite offrir à ses clients par exemple. Enfin, une dernière question porte sur leur opinion des nouveaux protocoles d'interrogation et sur la place que tiendra SRW dans l'avenir.

3.2. Du côté des acteurs de l'Open Access

3.2.1. Une vision différente liée au mouvement OAI

On l'a sous-entendu précédemment, la logique commerciale et celle de l'Open Access sont très différentes. Si on prend l'exemple du métamoteur de Crossref, on est bien sûr dans une logique commerciale (même si Crossref est un organisme à but non-lucratif) car certaines ressources ne sont pas accessibles gratuitement. L'information est visible sans être pour autant disponible. Pour l'Open Access, c'est la démarche inverse puisque c'est la logique de mise à disposition qui rend l'information visible.

Quoiqu'il en soit, j'ai effectué la même démarche auprès des principaux acteurs de l'Open Access, en sachant bien sûr que tous s'appuient sur le protocole OAI-PMH, mais il était important de savoir également leur opinion sur la recherche fédérée. C'est dans ce but que l'interview de Francis André, le responsable du département prospective et partenariat de l'INIST-CNRS publiée dans le numéro de juin 2004 d'Archimag⁴⁴ a attiré mon attention et l'envie d'en savoir un peu plus. M. André a bien voulu m'accorder un entretien téléphonique. Dans un second temps, c'est Monsieur Daniel Charnay, directeur adjoint du CCSD qui m'a reçu à l'Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) afin de m'expliquer le rôle et la mission du CCSD.

⁴⁴ Entretien réalisé par Miche Remize, in Archimag, juin 2004, n°175.

3.2.2. Consultation de Francis André sur l'Open Access et l'INIST

Cet entretien a permis de mieux comprendre quels étaient les besoins réels, exprimés ou non en matière de recherche fédérée par les différentes personnes qu'elles soient des chercheurs, universitaires ou spécialistes de l'information.

L'objectif de cet entretien était surtout de comprendre qu'elle était la position de l'INIST par rapport à l'Open Access et son rôle.

L'INIST soutient l'Open Access dans ses différentes formes (revues, archives ouvertes), il s'agit d'optimiser ce système pour les besoins de la recherche française, mais c'est également une façon de reprendre en main la production et la diffusion de la publication scientifique. Ce point de vue de l'INIST va donc la faire évoluer dans son fonctionnement puisque l'institut vit en partie sur son offre de services, de bases de données et de documents.

L'INIST travaille avec le CCSD afin de promouvoir leurs activités, de même que l'INRIA a passé un accord avec le CCSD pour mettre en valeur son fonds.

Concernant des collaborations avec des sociétés commerciales, notons les travaux réalisés avec Elsevier sur les aspects de plateforme et d'interopérabilité (avec Science Direct et Science Server le logiciel d'Elsevier). Un travail sur l'interopérabilité a également été réalisé avec la plateforme d'Endeavor, notamment sur le protocole OAI-PMH. Un des produits de l'INIST, `article@inist`, fonctionne avec la technologie du moteur Topic (société Verity). Une API⁴⁵ permet aux partenaires de l'INIST de pouvoir attaquer ses bases de données sans passer par l'interface classique de l'INIST (transfert des résultats en XML, comme pour les partenaires de Scirus qui ont la possibilité d'interroger le moteur sans passer par son formulaire). Ces partenariats sont possibles sur la base d'un accord préalable.

En 1999, l'INIST, en collaboration avec Elsevier, avait établi un lien entre Science Direct et `article@inist`, mais sans réel accès direct. Ce lien permettait à l'utilisateur d'avoir une visibilité au cours de sa recherche vers les deux produits

⁴⁵ *Application Programming Interface*. Interface de programmation d'application qui est en fait un ensemble de commande externes publié par un éditeur donné et qui permet de bénéficier des fonctionnalités d'une application depuis une application différente.

INIST et Science Direct. Francis André qualifie d'ailleurs ce travail de prémisse de résolveur puisqu'on obtient dans un environnement contrôlé une information à partir d'une interrogation, avec une fonction de localisation des documents (le lien vers le texte intégral est possible si l'utilisateur est abonné chez Elsevier). Ce projet a été depuis abandonné par Elsevier, il est clair que les éditeurs préfèrent proposer leurs propres solutions et surtout passer par leur propre interface en développant –s'ils en ont les moyens –leurs propres protocoles.

Enfin, une autre piste développée par l'INIST tourne autour des problèmes d'authentification en relation avec le profilage des utilisateurs. L'INIST miserait sur des certificats, jugés plus fiables que les authentifications par login et mot de passe. Cette voie, qui n'intéresse pas pour l'instant les éditeurs, pourrait gérer l'adéquation des profils d'utilisation avec l'adéquation des ressources, c'est une piste relativement intéressante à explorer, une piste qui d'ailleurs va dans le prolongement de ce qui est souhaité dans le cadre de la mise en place d'une solution globale de recherche fédérée.

3.2.3. Entretien avec Daniel Charnay sur le CCSD

Le Centre pour la Communication Scientifique Directe (CCSD) est relativement récent et date d'un peu plus de quatre ans. Sa mission est la mise en place de bases d'articles scientifiques orientés bien sur vers les sciences fondamentales mais également vers les sciences humaines et cognitives. Le discours de Daniel Charnay sur l'Open Access est en phase avec le point de vue de Francis André dans la mesure où l'idée fondatrice du CCSD est la maîtrise de la chaîne de publication des documents par les chercheurs. Le CCSD se veut être aussi un outil centralisateur pour assurer la pérennité du système d'archives ouvertes et ainsi se donner une assise institutionnelle. L'expérience des chercheurs français en matière d'auto archivage avait montré que les initiatives individuelles que l'on peut qualifier "d'initiatives de laboratoire" (à petite échelle) ne dureraient pas et ne permettraient pas d'offrir une bonne disponibilité des documents archivés.

C'est dans cette optique que le CCSD a passé des accords avec les principaux instituts de recherche nationaux comme l'INRA, l'INRIA, l'INSERM, et entretient des contacts étroits avec l'INIST.

Le reste de l'entretien a porté sur les caractéristiques techniques du CCSD et les problèmes posés par la recherche documentaire dans des documents en sciences dures.

Le CCSD a vu le jour grâce à un démarrage expérimental, notamment avec l'ouverture du réservoir ArchiveSic, mis en place sous l'impulsion de G. Chartron. Par la suite, le CCSD est composé de deux outils, les Article en Ligne (AL) et les Thèses en Ligne (TEL) qui fonctionnent tous deux sous le logiciel e-Print (l'application développée par E. Harnad, le créateur de Arxiv)⁴⁶. Un aspect notable du CCSD réside dans la différenciation des modes de recherche en fonction des types de ressources recherchées. En effet, pour les bases de sciences dures, il n'y a pas de recherche en texte intégral, les scientifiques sont plus axés sur des recherches par formules que par mots-clés. Par contre, les bases en sciences humaines et sociales utilisent le moteur de la société Exalead pour des recherches traditionnelles. Enfin notons le caractère particulier des documents des sciences dures qui comportent des formules et qui sont mis en forme sous un format spécifique, le format Tech. Cela pose des problèmes notamment pour la transformation de ces données en XML dans la mesure où la seule méthode connue à ce jour est de faire glisser les documents Tech vers un format normalisé par le W3C, MathML qui est en fait peu utilisé car trop complexe. De ce fait le consortium W3C réfléchit aux solutions pour permettre une intégration plus aisée des documents scientifiques vers le XML.

L'entrevue avec Daniel Charnay fut fructueuse à plus d'un titre car elle a permis d'appréhender les projets en cours du côté du CCSD et les solutions recherchées pour permettre d'intégrer le XML aux documents scientifiques.

Ces entretiens ont été assez intéressants puisque Francis André et Daniel Charnay m'ont bien expliqué les enjeux de l'Open Access notamment en France, ainsi que ce qui était possible de réaliser dans le cadre de la recherche fédérée. L'IN2P3 a pris conscience que même si l'on optait pour la solution de l'*open access*, il faut travailler vers une voie centralisatrice et institutionnelle pour mettre en valeur de façon pérenne les réservoirs de documents. TEL est d'ailleurs reconnu

par le Ministère de la Recherche et a élargi ses contacts au niveau européen dans le cadre d'un projet de réservoir européen des thèses (European Thesis Online,) .Les projets de l'INIST montrent qu'il n'est pas aisé de travailler de concert avec des éditeurs commerciaux dont les intérêts et la logique ne sont pas les mêmes que ceux d'instituts comme l'INIST ou de bibliothèques. Cela montre que pour réussir ce type de projet, il va falloir trouver un argumentaire, un terrain d'entente et d'après négociations qui peuvent être menées en France par exemple par l'Abes et le consortium Couperin.

3.3. Bilan : mise en place d'une base de connaissance

A l'issue de cette enquête, une base de connaissance sera livrée aux ingénieurs d'Archimed au sein de laquelle ils pourront retrouver les informations concernant les bases qu'il serait possible d'interroger via un des protocoles vu tout au long de ce rapport. Ces contacts permettent d'avoir une vue d'ensemble sur ce qu'il est possible ou non de faire en matière de recherche fédérée. Ils permettent aussi de jauger le degré d'ouverture de certains acteurs par rapport à l'évolution des technologies de recherche documentaire et tout simplement des modes de recherche documentaires eux-mêmes. A l'heure actuelle, les résultats obtenus montrent encore l'utilisation du protocole Z39.50, mais aussi de technologies de recherche simples à mettre en œuvre, des technologies liées au web et à HTTP comme les CGI. L'OpenURL commence à se montrer présent, tout comme OAI-PMH qui est le protocole de base des ressources libres mais qui peut être utilisé par les commerciaux, notamment en ce qui concerne les *abstracts* des articles qu'ils possèdent et qu'ils peuvent mettre en accès libre. Enfin, les *web services* passent encore pour des inconnus auprès de certains éditeurs, d'autres suivent de plus ou moins loin l'évolution de la norme en adaptant une attitude de *wait and see*. Enfin, Elsevier s'est apparemment inspiré des nouveaux standards pour développer un protocole propriétaire dont les fonctions sont similaires à SRW, les

⁴⁶ Ces bases vont bientôt fonctionner sous un logiciels développé par les spécialistes de l'IN2P3.

bases d'Elsevier sont inattaquables puisque qu'elles ne travaillent pas avec des protocoles normalisés. Cela sous-entend des partenariats et donc des négociations.

3.4. Tableau statistique sur l'utilisation des protocoles

SRW/SRU	OPENURL	oai- PHM	Z39.50	XML GATEWAY	PROTOCOLES PROPRIÉTAIRES	ND
2	12	7	5	4	26	4
5.2%	31.5%	18.4%	13.1%	10.5%	68.4%	10.5%

Ce tableau permet d'avoir une idée sur ce qui est utilisé. On constate bien sûr l'utilisation forte de protocoles et de logiciels propriétaires qui permettent l'interrogation et la récupération des données. Les principales sociétés comme Elsevier ont-elles même développé leurs propres protocoles, d'ailleurs Elsevier a mis au point un protocole semblable à SRW, c'est-à-dire qu'il apporte une flexibilité dans l'interrogation des bases grâce à un langage simplifié et qu'il permet le retour des résultats en XML, mais il s'agit d'un protocole entièrement propriétaire. Notons aussi que les sociétés sont épaulées par des prestataires concernant l'interface voire aussi la technologie de recherche :

Le moteur d'Elsevier, Scirus fonctionne avec la technologie du moteur norvégien Alltheweb, lui-même développé par la société Fast.

On retrouve également Jouve avec Brio Directory qui sert d'interface pour L'Année Philologique en ligne ou les Petites Affiches

Les ressources en sciences humaines du CCSD sont interrogeables grâce au moteur Exalead, les éditions EDK ont choisi la société GHS pour leur interface.

Il y aussi d'autres acteurs, notamment les petits éditeurs ou certaines bibliothèques qui n'utilisent qu'une simple interface web, avec parfois un moteur rudimentaire comme HtDig ou simplement un formulaire http qui interrogera par exemple une base SQL.

Les entretiens effectués m'ont aussi permis d'avoir une vue d'ensemble sur les projets autour de ces protocoles. Des efforts sont faits du côté du XML, on trouve aussi quelques projets concernant l'OpenURL. Mais on a peu de choses sur les Web Services et SRW/SRU contrairement à ce que l'on pouvait imaginer avant cette enquête. Certains éditeurs ne connaissent pas ces protocoles, d'autres comme nous l'avons souligné précédemment, surveillent l'évolution du protocole et sont éventuellement près à franchir le cap dans une optique à la fois commerciale et orientée utilisateur final. La plupart des personnes contactées préfèrent être sûre que ces types de protocoles et les nouvelles formes de recherche entrent dans le cadre d'un besoin réellement exprimé. En effet, le changement d'architecture au niveau de la recherche et l'implémentation de nouveaux protocoles engendrerait un coût non négligeable.

4. Conclusion : perspectives de nouvelles discussions et négociations entre les bibliothèques, les éditeurs de contenus, les éditeurs informatiques et les consortiums

Avec l'évolution des modes de recherche et des types de ressources recherchées, nous sommes donc arrivés à un point où il est possible de rechercher des documents ou des références dans un catalogue de bibliothèque, un système de GED ou des ressources électroniques issues d'un ou de plusieurs abonnements. La recherche fédérée et plus précisément l'offre de portail fédérateur proposée par Archimed permet de réaliser ce type de recherche sur tous ces types de ressources, en théorie. En théorie, parce qu'Archimed, bien que disposant d'un outil pleinement opérationnel et démontrant la volonté de proposer un concept de recherche fiable et pérenne en s'appuyant sur des protocoles normalisés, ne peut déployer l'intégralité des possibilités de son portail. En effet, la recherche et surtout la présentation en texte intégral d'articles de périodiques dépend de l'attitude des éditeurs à offrir un accès autre à ses bases de données que par son interface web classique.

L'enquête montre qu'un certain nombre d'éditeurs dont l'objectif est avant tout de se positionner du côté de l'utilisateur final, est prêt à s'investir dans ce type de projet. Pour les éditeurs et diffuseurs de référence, Elsevier, Ebsco, Swets, des partenariats ont été déjà imaginés, notamment entre Swets (Swetswise et Exlibris pour Metalib qui peut donc attaquer les bases de Swetswise indépendamment du moteur classique) : même s'il est difficile d'arriver à un accord, tout est possible. Cela va donc passer par des discussions et des négociations à la fois entre les bibliothèques, les groupements d'établissements (pour la France, il s'agit de l'Abes), les consortiums, les constructeurs de SIGB ou de solutions de portail comme Archimed d'un côté et les éditeurs de l'autre afin de trouver un terrain d'entente pour mener à bien un projet de portail fédérateur. Mais il s'agit aussi de bien discerner les contraintes techniques propres à chacun des partenaires afin de proposer une solution opérationnelle. On pense encore qu'il n'est pas dans l'intérêt des éditeurs de donner un accès autre à leur base de données si ce n'est que grâce à leur propre interface. Mais, on a bien vu que des enjeux à la fois bibliothéconomiques relatifs aux besoins des communautés utilisatrices, des enjeux économiques relatifs à la rentabilité des abonnements souscrits, des enjeux qui s'inscrivent dans un contexte de crise de l'édition électronique, pourraient permettre de changer la donne.

Conclusion

Au jour de finalisation de ce rapport, la base de connaissances montée au cours de ce stage n'est pas encore complétée, mais elle peut néanmoins donner quelques motifs de satisfaction. En effet, ce type d'enquête connaît en moyenne un taux de retour oscillant entre 10 et 20%. Ici, près de la moitié des personnes interrogées ont répondu, certaines ont fait preuve de disponibilité en m'expliquant la position de leur institution ou de leur société et ce qu'ils pourront proposer s'ils ne l'ont pas encore fait à leurs clients en matière de recherche fédérée.

La principale conclusion à tirer de ce stage, c'est qu'Archimed a vraisemblablement une longueur d'avance sur ses concurrents en proposant un produit technologiquement opérationnel. Hormis les compagnies majeures dans les systèmes d'information documentaire comme Ex-Libris, Archimed est peut-être une des seules à avoir implémenter les services web et SRW/SRU, le seul protocole qui puisse réaliser des recherches aussi performante que la norme Z39.50 en apportant en plus la technologie XML dans le retour des résultats.

Le problème, c'est qu'il existe peu de sources interrogeables via ce protocole. On aura observé que Z39.50 est toujours utilisé et qu'il le sera pendant un moment encore. Les sociétés utilisent sinon des technologies plus simples basées sur http, avec notamment l'utilisation de CGI, d'interfaces web simples construites sur des codes propriétaires (ou sur des moteurs de recherche achetés à des prestataires comme Jouve, GHS, Verity, etc.). Ces sociétés peuvent néanmoins offrir un accès différencié aux clients qui le souhaitent en leur envoyant un API qui leur permettra d'interroger leurs bases sans passer par leur interface.

Concernant l'OpenURL, on constatera qu'elle est effectivement présente dans un certain nombre d'acteurs, tout comme le protocole OAI pour les ressources en libre accès (où simplement les résumés d'articles chez les fournisseurs commerciaux).

Les sociétés ou les institutions qui fournissent des contenus en lignes sont néanmoins attentives aux nouveaux protocoles et à la problématique de la

recherche fédérée. Même si leur intérêt est de rendre leurs documents visibles et accessibles via leur interface, elles sont ouvertes aux discussions et aux éventuelles modalités en vue de l'intégration de ressources ou d'un produit dans un portail.

Donc même si Archimed maîtrise les technologies de la recherche documentaire, le problème posé est purement humain. La mise en place d'un tel portail nécessitera des négociations entre bibliothèques, groupements de bibliothèques et consortiums et les éditeurs afin de voir ce qui peut être fait dans cette direction.

L'intérêt envers la recherche fédérée se vérifie toutefois grâce au projet de Crossref qui a donc développé un métamoteur fonctionnant avec la technologie de Google, un produit qui permet bien de rechercher des références dans des bases hétérogènes et qui plus est des références à haute valeur ajoutée, mais qui ne permet pas de faire une véritable recherche documentaire puisqu'on ne peut pas interroger sur des métadonnées comme on peut le faire en Z39.50.

Finalement, même si l'enquête réalisée auprès des éditeurs a permis d'avoir une idée sur leurs aspirations et leurs positions, il est difficile de savoir ce qui va se passer. L'édition électronique, on l'a dit, connaît une crise qui a vu l'émergence d'un modèle alternatif libre, mais un modèle encore insuffisamment poussé dans la mesure où l'OAI n'est qu'un protocole de téléchargement de métadonnées et non pas de recherche et de récupération d'information.

Enfin, les majors de l'édition électronique continuent de développer et de commercialiser des produits assez intéressants, comme le dernier-né d'Elsevier, Scopus qui sera commercialisé à l'automne 2004 (base de données comprenant 12900 titres de 4000 éditeurs proposant le lien vers le texte intégral, mais surtout des fonctions liées aux citations avec un outil d'analyse d'auteurs, et de récupérations de références contenant des références communes avec la référence recherchée initialement).

Bibliographie

1 L'Abes et le Sudoc : portail et partenaires

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Présentation générale du portail documentaire Sudoc [en ligne], janvier 2004. Disponible sur : <www.abes.fr/abes/documents/PortailPrésentation.pdf> (consulté le 15/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Les publications en série en ligne [en ligne]. avril 2004. Disponible sur : <www.abes.fr/abes/DocumentsWebAbes/PSligne.ppt> (consulté le 20/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Le portail documentaire Sudoc [en ligne]. avril 2004. Disponible sur : <www.abes.fr/abes/Documents/portail.ppt> (consulté le 15/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Journées réseau Sudoc, 20 et 21 mars 2003 [en ligne]. 2003. Disponible sur : <www.abes.fr/abes/documents/réseau.ppt> (consulté le 15/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Aspects techniques du Sudoc [en ligne]. 2001. Disponible sur : <www.abes.fr/abes/documents/HStechnique1.ppt> (consulté le 25/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Le Sudoc, état de la base [en ligne], 2003. Disponible sur : <www.abes.fr/abes/documents/base.ppt> (consulté le 22/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Le Sudoc et son réseau [en ligne], 2004. Disponible sur : <www.abes.fr/abes/DocumentsWebAbes/Sudoc.ppt> (consulté le 22/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Le Sudoc : Système universitaire de documentation [en ligne]. 2003. Disponible sur : <<http://web.ccr.jussieu.fr/urfist/SUDOC.ppt>> (consulté le 25/06/2004)

AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPERIEUR (ABES), Aspects techniques du Sudoc et procédures de déploiement [en ligne]. Mars 2004. Disponible sur : <www.abes.fr/abes/DocumentsWebAbes/technique2004.ppt> (consulté le 10/06/2004)

BRZUSTOWSKI K., BARRAL S., L'ABES, coordinatrice et animatrice de réseaux [en ligne] BBF, 2003, t.48, n°2, p.32-42. Disponible sur : <http://bbf.enssib.fr/bbf/html/2003_48_2/2003-2-p32-barral.xml.asp> (consulté le 10/06/2004)

KLEE L., Présentation de Couperin [en ligne]. Janvier 2004. Disponible sur : <http://couperin.cines.fr/Presentation_lk.ppt> (consulté le 13/06/2004)

ROLAND M., Couperin:le département “études & prospectives” [en ligne]. 2003. Disponible sur : <<http://couperin.cines.fr/DeptEP.ppt>> (consulté le 10/07/2004)

2 Recherche documentaire et recherche fédérée

ARCHIMED, Bookline – Passerelle de recherche fédératrice in Journée Couperin « Outils d'accès intégré » 7 juin 2004. Document non disponible.

COALITION FOR NETWORKED INFORMATION (CNI), The CNI Spring 2004 Task Force Meeting

April 15-16, 2004 Alexandria, Virginia [en ligne]. 2004. Disponible sur : <<http://www.cni.org/tfms/2004a.spring/index.html>> (consulté le 3/07/2004)

LAHARY D., Informatique et bibliothèque Vers la banalisation des outils [en ligne] BBF, 2002, t.44, n°1, p.60-66. Disponible sur : <<http://www.enssib.fr/bbf/bbf-2002-1/08-lahary.pdf>> (consulté le 14/06/2004)

LOSSAU N., Search Engine Technology and Digital Libraries. D-Lib Magazine [en ligne]. 2004, vol.10, n°6. Disponible sur : <<http://www.dlib.org/dlib/june04/lossau/06lossau.html>> (consulté le 20/07/2004)

MORIN N., Contenus et services des sites web des bibliothèques [en ligne] BBF, 2003, t.48, n°4, p.9-13. Disponible sur : <<http://morinn.free.fr/pdf/bbf01.pdf>> (consulté le 5/07/2004)

TERRELL J., Cross database searching: information literacy for the real world [en ligne]. 2004. Disponible sur : <<http://lifelonglearning.cqu.edu.au/papers/terrell-133-paper.pdf>> (consulté le 9/08/2004)

URFIST DE STRASBOURG, Comparaison, des logiciels d'interrogation des périodiques électroniques en texte intégral proposés par les éditeurs [en ligne]. 2004. Disponible sur : <<http://www-scd-ulp.u-strasbg.fr/urfist/tableauperintegral.htm>> (consulté le 15/07/2004)

URFIST DE STRASBOURG, Comparaison des logiciels d'interrogation des périodiques électroniques en texte intégral (ACS-Science direct-Ovid-Blackwell) [en ligne]. Mars 2004. Disponible sur : <<http://www-scd-ulp.u-strasbg.fr/urfist/InterfacePeriodElectr.pdf>> (consulté le 15/07/2004)

WALKER J., Metasearch. In The CNI Spring 2004 Task Force Meeting April 15-16, 2004 Alexandria, Virginia [en ligne]. 2004. Disponible sur : <http://www.cni.org/tfms/2004a.spring/presentations/CNI_Walker_Metasearch.ppt> (consulté le 10/07/2004)

3 Les protocoles de recherche et de récupération de l'information

3.1 Z39.50

GUILLERAY P.E., Projet de catalogue virtuel accessible par le client Z39.50 du SCD de l'Université Lyon 3 [en ligne]. . Diplôme de Conservateur de Bibliothèque, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (Enssib), Villeurbanne, 2003. [en ligne]. Disponible sur : <<http://www.enssib.fr/bibliotheque/documents/dcb/guilleray.pdf>> (consulté le 7/06/2004)

INDEXDATA, The Z39.50 Target Directory. [en ligne]. Disponible sur: <<http://www.indexdata.dk/targettest/index.php>> (consulté le 2/06/2004)

LAHARY D., La norme Z39.50. [en ligne] support crée en octobre 1998, dernière mise à jour le 18 mars 2003. Disponible sur : <<http://membres.lycos.fr/vacher/profess/cours/mediadix/z3950/>> (consulté le 10/06/2004)

MORIN N., Z39.50 : une norme en mutation. [en ligne]. 2001. Disponible sur : <<http://morinn.free.fr/articles/z3950.htm>> (consulté le 10/06/2004)

3.2 SRW/SRU

ARCHIMED, SARA, serveur d'interrogation. Guide Technique, juin 2004, 49 p. (Documentation technique Archimed)

ARCHIMED, Bookline Client/Serveur SRW, spécifications techniques, janvier 2004, 33 p. (Documentation technique Archimed)

DENENBERG R., WALKER J., SANDERSON R., SRW and Metasearch. In The CNI Spring 2004 Task Force Meeting April 15-16, 2004 Alexandria, Virginia [en ligne]. 2004. Disponible sur: <<http://www.cni.org/tfms/2004a.spring/abstracts/PB-srw-denenberg.html>> (consulté le 15/07/2004)

DENENBERG R., SRW and Metasearch. Introduction. In The CNI Spring 2004 Task Force Meeting April 15-16, 2004 Alexandria, Virginia [en ligne]. 2004. Disponible sur: <http://www.cni.org/tfms/2004a.spring/presentations/CNI_Denenberg_SRW.ppt> (consulté le 15/07/2004)

LEASE MORGAN E., An introduction to the Search/Retrieve URL Service (SRU). Ariadne [en ligne]. 2004, n°40. Disponible sur : <<http://www.ariadne.ac.uk/issue40/morgan/>> (consulté le 10/08/2004)

SANDERSON R., SRW/U : Re-Introduction. In The CNI Spring 2004 Task Force Meeting April 15-16, 2004 Alexandria, Virginia [en ligne]. 2004. Disponible sur:

<http://www.cni.org/tfms/2004a.spring/presentations/CNI_Sanderson_MetaSRW2.ppt> (consulté le 15/07/2004)

Z39.50 INTERNATIONAL NEXT GENERATION (ZING), SRW-Search/Retrieve Web Service [en ligne]. Février 2004. Disponible sur : <<http://loc.gov/z3950/agency/zing/srw/>> (consulté le 25/06/2004)

Z39.50 INTERNATIONAL NEXT GENERATION (ZING), SRW/Metasearch Meeting, October 30, 2003 [en ligne]. 2003. Disponible sur: <<http://www.loc.gov/z3950/agency/zing/srw/metasearch-srw.html>> (consulté le 25/06/2004)

3.3 OpenURL

ARCHIMED, Oscar in Journée Couperin « Résolveurs de liens » 18 mai 2004 [en ligne]. 2004. Disponible sur : <<http://couperin.cines.fr/040518/040518.html>> (consulté le 20/06/2004)

NISO, The OpenURL Framework for Context-Sensitive services [en ligne]. 2003. Disponible sur: <http://www.niso.org/committees/committee_ax.html> (consulté le 28/06/2004)

NISO, The proposed OpenURL Framework Standard [en ligne]. 2004. Disponible sur : <<http://library.caltech.edu/openurl/Standard.htm>> (consulté le 28/06/2004)

VAN DE SOMPEL H., The OpenURL Framework : Origins, Evolution, Concept (OpenURL Day, New-York, May 13th 2003). [en ligne]. Mai 2003. Disponible sur: <http://palmer.cwpost.liu.edu/csc/van_de_sompel.ppt> (consulté le 4/08/2004)

3.4 OAI-PHM

FOULONNEAU M., Le protocole OAI-PMH : une opportunité pour le patrimoine numérique [en ligne]. Janvier 2003. Disponible sur : <<http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/oai.pdf>> (consulté le 13/07/2004)

FOULONNEAU M., Collaborer pour de nouveaux services culturels en ligne. Le protocole OAI, protocole de collecte de métadonnées de l'Initiative des Archives Ouvertes [en ligne]. Janvier 2004, 61 p. Disponible sur : <http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/guide_oai.pdf> (consulté le 13/07/2004)

OULAHAL S., Accès unique à des ressources numériques distribuées [en ligne]. 2003. Disponible sur : <<http://2003.jres.org/actes/paper.125.pdf>> (consulté le 15/07/2004)

4 Panorama de l'open access

BADOLATO A.M., HAMEAU T., La remise en question des modèles de publications scientifiques [en ligne]. Captain-doc, n°8, juillet-août 2002. Disponible sur : <<http://www.captaindoc.com/lettre-html/lettre08.html#mois1>> (consulté le 3/07/2004)

BOSC H., La communication scientifique revue et corrigée par Internet [en ligne]. 2003 mis à jour le 2/07/2004. Disponible sur <http://www.tours.inra.fr/prc/internet/documentation/communication_scientifique/comsci.htm> (consulté le 9/06/2004)

BOSC H., La Budapest Open Access Initiative (BOAI) pour un libre accès aux résultats de la recherche [en ligne]. 2003. Disponible sur : <http://www.tours.inra.fr/prc/internet/documentation/communication_scientifique/terminal2.htm> (consulté le 9/06/2004)

CHARTRON G., Les archives ouvertes de la communication scientifique [en ligne]. 2003. disponible sur : <<http://web.ccr.jussieu.fr/urfist/archives-ouvertes.htm>> (consulté le 16/06/2004)

GOETTMANN E., L'Initiative de Budapest pour le Libre Accès (Budapest Open Access Initiative) [en ligne]. Captain-doc, n°8, juillet-août 2002. Disponible sur : <<http://www.captaindoc.com/breves/breves08.html>>(consulté le 3/07/2004)

INSTITUT DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE (INIST-CNRS), Le libre accès aux résultats de la recherche [en ligne]. Juin 2004. Disponible sur : <<http://www.inist.fr/openaccess/IMG/pdf/ETUDE.pdf>> (consulté le 20/07/2004)

INSTITUT DES SCIENCES COGNITIVES (ISC CNRS), Archives ouvertes, liens intéressants (mis à jour le 7/10/2003) [en ligne]. 2003. Disponible sur : <http://www.isc.cnrs.fr/liens_internet/rech_biblio/IsidoraOAI.htm> (consulté le 9/07/2004)

URFIST DE PARIS, FOURMI, Archives ouvertes de publications scientifiques [en ligne] dernière mise à jour : juillet 2004. Disponible sur : <<http://web.ccr.jussieu.fr/urfist/fourmi/fourmi-archives.htm>> (consulté le 23/07/2004)

5 Les périodiques électroniques

CENTRE NATIONAL DE RECHERCHE SCIENTIFIQUE (CNRS), Résultats de l'Enquête Périodique, 17 mai 2004. [en ligne]. Disponible sur : <<http://www.cnrs.fr/SHS/actions/periodiques.php>>, deux présentations : <http://www.cnrs.fr/SHS/actions/fichiers/presentation1_170504.ppt> ; <http://www.cnrs.fr/SHS/actions/fichiers/presentation_17052004.ppt> (consulté le 19/07/2004)

CHARTRON G., MARANDIN C., La presse scientifique électronique. Analyse de l'offre des intermédiaires. [en ligne]. BBF, T.43, n°3, 1998, p.28-40. Disponible sur : <http://bbf.enssib.fr/bbf/html/1998_43_3/1998-3-p28-marandin.xml.asp> (consulté le 21/06/2004)

COMITE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, Section 37 du comité National "Economie et société", Classement des revues à comité de lecture en économie et en gestion" [en ligne]. 2003. Disponible sur : <<http://crg.polytechnique.fr/section37/revues.pdf>> (consulté le 25/06/2004)

CRAWFORD W., REED C., DAVIS Ph., DELHAYE M., Articles sur la "crise des périodiques" [en ligne]. BiblioAcid, Vol.1, n°1, mars 2004. Disponible sur : <http://www.biblioacid.org/revue/pdf/2004_01.pdf> (consulté le 23/07/2004)

Enquête sur les ressources électroniques acquises par les s.c.d., s.i.c.d. et bibliothèques d'établissements d'enseignement supérieur [en ligne]. 2004. Disponible sur : <<http://www.sup.adc.education.fr/bib/Acti/Electro/Tab2003.xls>> (consulté le 6/08/2004)

INNEGRAEVE P., La place des périodiques électroniques à l'intérieur des sites internet des SCD et des grands établissements français (tableau récapitulatif) [en ligne]. Mars 2004. Disponible sur : <http://web.ccr.jussieu.fr/urfist/perio_etablissements.xls> (consulté le 14/06/2004)

INNEGRAEVE P., La place des périodiques électroniques à l'intérieur des sites internet des SCD et des grands établissements français (enquête) [en ligne]. Mars 2004. Disponible sur : <http://web.ccr.jussieu.fr/urfist/perio_enquete_scd.doc> (consulté le 14/06/2004)

INNEGRAEVE P., Services électroniques ou études pour les bibliothèques [en ligne] Dernière mise à jour le 24/06/2004. Disponible sur : <<http://web.ccr.jussieu.fr/urfist/supportcours.htm#Gestion>> (consulté le 14/06/2004)

JEANNIN Ph., Enquête sur les revues d'économie [en ligne]. Disponible sur : <<http://www.iut-tarbes.fr/enquete/economie/index.htm>> (consulté le 19/06/2004)

MAHE A., La communication scientifique en (r)évolution (l'intégration des revues électroniques dans les pratiques informationnelles de chercheurs en sciences de la nature comme révélation des mutations du modèle traditionnel de communication scientifique. Thèse de Doctorat, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (Enssib), Villeurbanne, 2002. [en ligne]. Disponible sur : <<http://www.enssib.fr/bibliotheque/documents/theses/mahe/mahe.pdf>> (consulté le 01/07/2004)

MELOT A., Analyse de l'offre des revues en ligne en économie et en gestion, sélection faite par la bibliothèque du groupe HEC [en ligne]. 2004. Disponible

sur : <<http://www.ccr.jussieu.fr/urfist/melot/Urfist971.ppt>> (consulté le 29/06/2004)

MELOT A., Analyse de l'offre des revues en ligne en économie et en gestion, stage Urfist, 17 juin 2004 [**en ligne**]. Disponible sur :

<<http://web.ccr.jussieu.fr/urfist/melot/Urfist971.ppt>> (consulté le 29/06/2004)

SCD UNIVERSITE DE SAINT-ÉTIENNE, Périodiques numériques accessibles librement sur le web [**en ligne**]. Disponible sur : <<http://www.univ-st-etienne.fr/scdoc/periodiques/periodelectr.html>> (consulté le 10/07/2004)

WELLENSTEIN A., Panorama de l'offre éditoriale en matière de revues électroniques en Sciences Humaines et Sociales disponibles via Internet [**en ligne**] mémoire DESS Institut National des Techniques de la documentation (INTD), 2003. Disponible sur :

<<http://pages.globetrotter.net/charro/HERMES10/wellenstein.htm>> (consulté le 20/07/2004)

Table des annexes

**ANNEXE 1 : LES ACTEURS DE L'ÉDITION ÉLECTRONIQUE INTERROGÉS
AU COURS DU STAGE.....I**

**ANNEXE 2 : EXEMPLES DE PROFILS Z39.50 ET PORTAILS
DOCUMENTAIRES ÉQUIPÉS DE PROGICIELS ARCHIMEDANNEXE 2-1 :
EXEMPLES DE PROFILS Z39.50I**

ANNEXE 2-1 : EXEMPLES DE PROFILS Z39.50 II

Annexe 2-2 : Quelques portails documentaires équipés par ArchimedVII

ANNEXE 3 : LE MOUVEMENT OAI ET SES RESSOURCES.....VIII

Annexe 1 : Les acteurs de l'édition électronique interrogés au cours du stage

Série n°1 : les principaux acteurs de l'édition électronique commerciaux, institutionnels et de l'Open Access.

(Les acteurs de l'Open Access et du mouvement OAI sont représentés en jaune ou en clair sur papier).

Nom de la société	Site web international	Site web français	Type d'offre	Adresse physique
ACM (Association for Computing Machinery)	http://www.acm.org/	Pas de site en français	_ Institution spécialisée en informatique. 5 revues, 9 magazines, 16 ACM transactions, 174 conférences, 44 newsletters	USA
ACS American Chemical Society Chemistry.org	http://www.chemistry.org/portal/a/c/s/1/home.html	Pas de site en français	_ American Chemical Society Publications diffuse le texte intégral de 30 revues savantes et magazines dans tous les domaines de la chimie, y compris l'environnement.	USA
AIP (American Institute of Physics)	http://librarians.aip.org	Pas de site en français	_ Publication de 13 revues AIP (sciences physiques)	USA
Annual Reviews	http://www.annualreviews.org/	Pas de site en français	_ 30 revues spécialisées en sciences biomédicales, sciences physiques et sciences humaines	USA
APA (American Psychological Association)	http://www.apa.org/		_ Spécialisé en psychologie avec les bases PsycINFO, PsycINFO Direct, PsycARTICLE, PsycARTICLE Direct, PsycEXTRA, PsycFIRST	USA
APS (American Physical Society)	http://www.aps.org http://prola.aps.org/search (interface de recherche)	Pas de site en français	_ 7 revues (Physical Reviews)	USA
Arxiv	http://arxiv.org/	http://fr.arxiv.org/	_ Articles en sciences physiques, mathématiques, informatique, sciences non linéaires et biologie. Archive publiée par l'Université de Cornell depuis 1991 (TI)	France
Biomed Central	http://www.biomedcentral.com/		_ Propose 135 journaux électroniques en sciences de la vie. Certaines revues ne sont accessibles que par abonnement (TI)	Royaume-Uni
CCSD		http://www.ccsd.cnrs.fr/index.html	_ Articles, cours, thèses et vidéos essentiellement en sciences physiques (TI)	France
CiteSeer	http://citeseer.ist.psu.edu/		_ Recense des publications, communications, techniques, essentiellement en informatique, créée en 1997 (TI)	USA
CogPrints	http://cogprints.ecs.soton.ac.uk/		_ Accès gratuit à plus de 1700 articles en psychologie, linguistique, informatique, philosophie et biologie depuis 1950. En texte intégral	Royaume-Uni

Nom de la société	Site web international	Site web français	Type d'offre	Adresse physique
Crossref	http://www.crossref.org	Pas de site en français	<ul style="list-style-type: none"> _ Service coopératif de liens de références qui permet en cliquant sur la citation d'un journal d'accéder directement au texte intégral de l'article cité. _ Les liens sont réalisés grâce au DOI décrivant les métadonnées spécifiques à chaque article. _ Rassemble plus de 152 éditeurs _ A lancé en avril 2004 un Site pilote avec Google pouvant rechercher simultanément chez 25 éditeurs 	USA
CSA (Cambridge Scientific Abstracts)	http://www.csa.com	Pas de site en français	16 bases de données dans divers domaines (arts, éducation, sociologie, informatique...) dont ERIC, LISA, etc.	France
EBSCO	http://www.ebsco.com	www.ebsco.fr	<ul style="list-style-type: none"> _ Electronic Journal Service (EJS, Ebsco Host) : 9800 titres (concurrent direct de swetswise) _ A to Z : nouveau produit phare d'Ebsco : plateforme unique d'accès aux revues en ligne, statistiques d'utilisation par éditeur. 	France
Editions JB Baillière		http://www.33docpro.com/	_ 8 revues spécialisées en médecine.	France
EDP Science	http://www.edpsciences.org/		_ 32 revues dont ESAIM revues et European Physical Journals (accès gratuits à certaines revues, dont celles de l'INRIA, European Physical Journal, Astronomy et Astrophysics Supplement Series)	France
EINS GEM	http://www.eins.org/	Pas de site en français		France
Electronic Library	http://rzblx1.uni-regensburg.de/ezeit/fl.phtml?bibid=AAAAA&colors=1&lang=en		_ Propose plus de 8200 revues électroniques, en sciences de l'ingénieur, économie et information communication (TI)	Allemagne
Elsevier	http://www.elsevier.com/wps/find/homepage.cws_home	http://www.elsevier.fr/html/index.cfm	<ul style="list-style-type: none"> _ Science Direct : c'est le concurrent direct du Web of Science de l'ISI. 1800 titres en ligne, accès au texte intégral via Crossref. Intégration de 15 BDD bibliographiques. Intégration du moteur Scirus pour filtrer le web scientifique. En français: sur www.elsevier.fr : 40 revues spécialisées en médecine accessibles sur Science Direct. 	USA/Pays-Bas/France

Nom de la société	Site web international	Site web français	Type d'offre	Adresse physique
Emerald	http://www.emeraldinsight.com	Pas de site en français	_ Base de données de périodiques en texte intégral, traitant du management, du marketing, de la gestion des ressources humaines ainsi que des bibliothèques et des services d'information (bibliothèques électroniques...).	Royaume-Uni
Europresse	http://www.europresse.com	site francophone	_ Plus de 30 sources françaises de presse nationale et régionale augmentées de plus de 120 sources internationales _ Un choix de journaux, de publications spécialisées, de fil d'agences de presse et de transcriptions d'émissions radiophoniques _ Plus de 20 années d'archives et plus de 20 millions d'articles disponibles Une mise à jour quotidienne	France
Find Article	http://www.findarticles.com/		_ Propose plus de 700 magazines et journaux depuis 1998 en santé, économie, automobile, informatique, société, loisirs, sport. Mise à jour régulière (TI)	USA
Gale (Thomson)	http://www.galegroup.com	Pas de site en français	_ De nombreuses bases de données dans différentes disciplines, business, économie, droit, médecine, histoire ainsi qu'Ingenta.	USA/Royaume-Uni
Hermès		http://www.e-revues.lavoisier.fr/	_ Revues mises en ligne par Lavoisier	France
Highwire	http://www.highwire.org	Pas de site en français	_ 343 revues (STM en général) en 2003 dont 144 avec des archives ouvertes.	USA
HW Wilson	http://www.hwwilson.com		_ Plus de 50 bases de données en sciences humaines, communication, bibliothéconomie, etc.	USA
IEE (Institution of Electrical Engineers)	http://ioj.iee.org.uk/	Pas de site en français	_ Electronic letters, IEE Proceedings (13 titres), IEE Engineering Journals (6 titres), IEE review, IEE Conference Proceedings, IEE Seminar Digest	Royaume-Uni
IEEE Xplore (Institute of Electrical and Electronics Engineers)	http://ieeexplore.ieee.org	Pas de site en français	_ Service donnant accès au texte intégral de 114 revues, magazines et conférences de l'IEEE depuis 1988, ainsi que les standards de l'IEEE.	USA
Ingenta	http://www.ingenta.com/	Pas de site en français	_ Accès aux titres de 260 éditeurs, 6000 revues. Portail intégré avec commande d'articles à l'unité.	Royaume-Uni

Nom de la société	Site web international	Site web français	Type d'offre	Adresse physique
INIST	http://www.inist.fr	http://www.inist.fr	<ul style="list-style-type: none"> _ Francis : BDD bibliographiques en SHS et éco (2,5M de références) _ Pascal : BDD bibliographique en STM (14 M de référence) _ Article@inist : catalogue du fonds documentaire de l'INIST (7,5M de références) _ Connectsciences : ensemble de ressources et de services de l'INIST: Pascal (3 derniers mois), Francis (1 an), article d'article@inist, Thesa, biblioscience, base de liens, etc.). Pas d'accès au texte intégral 	France
International Union of Crystallography	www.iucr.org	Pas de site en français	_ 7 revues spécialisées en cristallographie	Royaume-Uni
IOP (Institute of Physics)	http://www.iop.org/EJ/	Pas de site en français	_ 40 revues spécialisées en sciences physiques dont 4 électroniques exclusivement	Royaume-Uni
John Libbey Eurotext		http://www.john-libbey-eurotext.fr	_ 20 revues spécialisées en médecine.	France
JSTOR (Journal Storage)	http://www.jstor.org	Pas de site en français	_ JSTOR est un service d'archive pluridisciplinaire de périodiques en texte intégral. Le service propose l'accès à des collections rétrospectives depuis le premier numéro de la revue (y compris des collections commençant au 19 ^{ème} siècle), jusqu'à quelques années précédant l'année courante. Sciences humaines, art, musiques, sciences exactes, etc.	USA
Kluwer	http://www.wkap.nl/	http://www.editionskluwer.be	_ Couverture de 650 titres	Belgique
Lamy		http://www.lamy.fr	<ul style="list-style-type: none"> _ Lamyline Reflex : 300 000 pages de commentaires des experts Lamy, 1,7 million de documents officiels. _ Environ-line : Actualité et veille juridique, textes et jurisprudence, nomenclature des installations classées, ouvrages Lamy, informations pratiques _ LamyDOC.com : Dossiers thématiques, Contrat de travail, licenciement, bail commercial, copropriété, abus de biens sociaux, divorce, risque industriel, pollution de sites, associations... 	France
Lexis Nexis	http://www.lexisnexis.com	http://www.lexisnexis.fr	<ul style="list-style-type: none"> _ Lexis-Nexis : 35 000 sources d'information, actualités économiques et financières françaises et internationales _ Jurisclasseur : fonds documentaires comprenant toutes les sources du droit 	France
Masson		http://www.e2med.com	_ 52 revues spécialisées en médecine.	France
MathSciNet (American Mathematical Society AMS)	http://www.ams.org/mathscinet	possibilité de choix en français	_ 48983 titres ajoutés en 2004; 1799 revues couvertes; liens vers 431138 articles originaux; 10843 rapporteurs actifs; 411487 auteurs indexés en mathématiques	USA

Nom de la société	Site web international	Site web français	Type d'offre	Adresse physique
Nature Publishing Group	http://www.nature.com	Pas de site en français	_ 57 revues dont Nature (sciences médicales, sciences physique, écologie et sciences de la terre)	France
Num Dam	http://www.numdam.org/		_ Numérisation de documents anciens en mathématiques, programme qui met en libre accès les données bibliographiques et articles de 6 revues spécialisées en mathématiques (TI)	France
OCLC	http://www2.oclc.org/oclc/fseco/index.asp	Pas de site en français	_ Electronic Collection Online : 4672 revues, 69 éditeurs tout domaine. Site agrégateur, orienté service d'archivage. Intégration avec les bases FirstSearch.	USA
OCLC firstsearch	http://firstsearch.oclc.org/html/login_fr.html	Pas de site en français	_ 12,7 millions de références issues de 12000 périodiques dans les domaines des affaires, des sciences humaines et sociales, de la technologie, mise à jour quotidienne	USA
Ovid	http://www.ovid.com	Pas de site en français	_ Commercialise des bases de données et propose des accès intégrés à l'information scientifique et technique	France
Oxford University Press	http://www3.oup.co.uk/jnls/	Pas de site en français	_ 190 revues tout domaines confondus. Revues mises en ligne par Highwire Press.	Royaume-Uni
Project MUSE	http://muse.jhu.edu/	Pas de site en français	_ Project MUSE fournit en ligne et en texte intégral plus de 200 journaux internationaux dans le domaine de l'art et de l'humanité, des sciences sociales et des mathématiques. Auteur/Editeur : The Johns Hopkins University Press	USA
ProQuest	http://www.proquest.com/	Pas de site en français	_ ProQuest permet l'interrogation de 4436 publications (périodiques ou journaux), des bases de données ABI/Inform, Periodical Abstracts II et ProQuest Medical Library. En anglais et très souvent en texte intégral	USA
Pubmed Central	http://www.pubmedcentral.nih.gov/		_ Propose 80 titre de journaux électronique en biosciences, sélectionnées par la NLM depuis 1995 (TI)	USA
Reves.org	http://www.revues.org/		_ Accès gratuit à 18 périodiques en SHS. En texte intégral	France
Royal Insitute of British Architect (RIBA)	http://www.architecture.com/go/Architecture/Reference/Library_894.html	Pas de site en français		Royaume-Uni

Nom de la société	Site web international	Site web français	Type d'offre	Adresse physique
Science Magazine	http://www.sciencemag.org http://www.aaas.org	Pas de site en français	_ Revue Science. 3 options: Science Online Premium Package (revue plus autres produits), Science Online complet, Science Current Original Research (2003)	USA
Scirus	http://www.scirus.com	Pas de site en français	_ Outil de recherche spécialisé dans l'information scientifique. Le outil de recherche spécialisé dans l'information scientifique. Le de revues scientifiques, et c'est Fast, la société qui produit le moteur de recherche alltheweb qui a conçu sa technologie	Pays-Bas/USA
Springer	http://www.springer.de http://springerlink.metapress.com http://springerlink.com	http://www.springer-paris.fr/	_ SpringerLink : http://springerlink.com Couverture de 470 titres, à travers 11 libraries: chimie, sciences de l'ingénieur, droit, médecine, informatique, etc.	France/Allemagne
Swets (Swetswise)	http://informationservices.swets.fr http://www.swetsblackwell.com http://www.blackwell-synergy.com	http://informationservices.swets.fr	_ Swetswise : Interface permettant la gestion des abonnements et simplifiant l'accès en ligne aux périodiques électroniques. 7664 titres de 272 éditeurs. 2 collections Blackwell en STM et SHS. Blackwell-Synergy : plus de 380 000 articles issus de 739 revues.	France
Techniques de l'ingénieur		http://www.techniques-ingenieur.fr	_ BDD spécialisée en électronique.	France
Thomson ISI	http://www.isinet.com/ http://www.thomsonderwent.com	Pas de site en français	_ Web of Science (WoS) : 8500 périodiques, possibilité d'accès au texte intégral. Plateforme intégrée pour la recherche (citations, sommaires, bibliographies et outils de gestion, évaluation, veille). C C search (Current Content) : Base bibliographique orientée veille disposant d'une interface Webspirs permettant des recherches retrospectives et multibases. Couverture sur 8000 titres et 2000 ouvrages. C C Connect: 8000 titres et 2000 ouvrages. Journal of Citation Reports (JCR) : outil d'évaluation critique des plus grandes revues mondiales, il couvre plus de 7000 revues. Son indicateur le plus utilisé est le facteur d'impact	Royaume-Uni
UNESCO	http://www.unesco.org/shs/shsdc/journals/shsjournals.html		_ Le centre de documentation de l'UNESCO propose l'accès gratuit à 337 titres de périodiques en sciences sociales (TI)	France
Urbaline	http://www.einsgem.com/databases/shortinfoACOM.htm			Italie
Urbamet		http://urbamet.documentation.equipement.gouv.fr/cgi/index.pl?PROFILE=USER-MET	_ Les 3 dernières années en ligne de données bibliographiques française sur l'aménagement, les villes, l'habitat et le logement, l'architecture, les équipements collectifs, les transports, les collectivités locales etc., propose aujourd'hui plus de 200 000 références de documents.	France
Wiley	http://www3.interscience.wiley.com	Pas de site en français	_ Interscience : couverture de 330 titres et 150 en sciences de la vie et médecine	Royaume-Uni

Série n°2 : catalogues collectifs de bibliothèque et produits spécialisés en droit, sport, économie et marketing.

Produit (revue, base de données ou catalogue collectif)	Editeur	Producteur (si différent de l'éditeur)	Site web international	Site web français	Offre
Ariadna (Biblioteca Nacional)			http://www.bne.es/esp/cat-fra.htm		La bibliothèque nationale espagnole et son catalogue
Bibliotheca Universalis			http://www.kb.nl/gabriel/bibliotheca-universalis/fr/bibliotheca_universalis_accueil.htm		Bibliotheca Universalis a pour ambition de mettre à la disposition du grand public, au moyen des technologies de l'information et de la communication, les œuvres majeures du patrimoine mondial scientifique et culturel, textes, images et son
Catalogue collectif des bibliothèques autrichiennes.			http://www.bibvb.ac.at/verbund-opac.htm		3,9 millions de notices appartenant à 60 bibliothèques
Generalis	Indexpresse			http://www.generalis-indexpresse.com/	Base de données bibliographiques donnant accès à des références d'articles de périodiques généralistes de langue française. Couvre tous les domaines. Résumés de 150 000 articles de presse
German libraries on line			http://www.hbz-nrw.de/produkte_dienstl/qermist/index.html		Une sélection de bibliothèques allemandes diffusant leurs catalogues sur le Web
Heracles	Sportdoc			http://www.sportdoc.unicaen.fr/heracles/	Contient actuellement 102 758 références bibliographiques concernant le domaine du Sport et de l'Education Physique, quel que soit leur support.
HERO (Higher Education & Research Opportunities in the United Kingdom)			http://www.hero.ac.uk/index.cfm		Le premier portail de la recherche universitaire au Royaume Uni. Accès aux catalogues des bibliothèques universitaires anglaises
Juripro	Coface ORT			http://www.juripro.com/WAJurpro/IndexBU.jsp	Sources juridiques officielles françaises et européennes et jurisprudence
Kompass Europe	Kompass International Neuenchwander SA		http://www.kompass.com	http://www.kompass.fr	Banque de données sur les entreprises européennes de l'industrie et des services, avec leur données administratives et juridiques, leurs produits et services

Produit (revue, base de données ou catalogue collectif)	Editeur	Producteur (si différent de l'éditeur)	Site web international	Site web français	Offre
La Revue Banque	La Revue Banque			http://www.revue-banque.fr/	Information des professionnels de la banque et de la finance
L'Année philologique	Société Internationale de Bibliographie classique			http://www.annee-philologique.com	Bibliographie de l'Antiquité gréco-latine depuis 1969 : dépouillement annuel de 1500 périodiques et 500 recueils et actes de colloque
Le Monde	Le Monde			http://www.lemonde.fr	Accès au quotidien
Le Moniteur des travaux publics et du bâtiment	Groupe Moniteur			http://www.lemoniteur-expert.com/	Portail construction du Groupe Moniteur
Les Petites Affiches	Petites Affiches Presses			http://www.petites-affiches.com/	Le quotidien du droit. Actualités rédigées par des journalistes professionnels, articles de doctrine, chroniques, notes de jurisprudence. Base de données de périodiques.
Lexbase	Lexbase			http://www.lexbase.fr	4 bases juridiques thématiques en droit privé, doctrine, sources et jurisprudence, revues
LSA	L'Usine Nouvelle			http://www.lsa.fr/	
Marketing science	Informa PubOnLine		http://pubsonline.informs.org/Library		Accès au TI de ce périodique
Périodiques de la RAS (Russian Academy of Sciences)	Russian Academy of Sciences		http://ojs.aip.org/		Accès à 13 périodiques. Périodiques traduits du russe et diffusés par l'AIP

Produit (revue, base de données ou catalogue collectif)	Editeur	Producteur (si différent de l'éditeur)	Site web international	Site web français	Offre
RedIRIS Catalogue collectif des bibliothèques espagnoles			http://www.rediris.es/recursos/bibliotecas/		Catalogues des bibliothèques espagnoles
RERO (Réseau des Bibliothèques de Suisse Occidentale)			http://www.rero.ch/		Recherche rapide, par mots-clés, par index, par vedettes en mode expert. Différents sous ensembles par types de documents y compris des articles accompagnés d'analyse. Filtres possibles : langue, pays, date. Historique. Interface en français.
Revue EP.S	Editions Revue EP S			http://www.revue-eps.com/fr/dynamique/recherche.htm	Recherche sur le titre, l'auteur, l'année d'un article, d'un ouvrage, d'un produit multimédia... en éducation physique et sportive. Permet notamment de rechercher les articles des titres suivants : La Revue EP.S ; La Revue EPS1 ; La Revue EP.S + EPS1
RLG Research Libraries Group			http://www.rlg.org		Créé en 1974, ce réseau comprend les universités de Columbia, Harvard, and Yale Universities and The New York Public Library. Les membres sont des bibliothèques nationales, universitaires ou possédant des fonds rares. Il contient près de 120 millions de références de livres, près de 7 millions de périodiques, etc. Il offre de nombreuses bases de données et de nombreux services. Par exemple, SCIPIO catalogue collectifs de documents rares en art depuis le 16e siècle
RLG's Eureka (Handbook of Latin American Studies)			http://eureka.rlg.org/Eureka/zgate2.pod		Nouvelle interface Web ouverte à mi-2003 pour les non-professionnels. Elle permet une recherche dans RLG Union Catalog et dans RLG Serials File
Sistema Bibliotecario. Bibliothèques italiennes			http://www.biblio.polito.it/it/documentazione/biblioit.html		Catalogues de bibliothèques italiennes
Site des étudiants de l'UFR de STAPS à Liévin. Le site sportif étudiant				http://membres.lycos.fr/stapslievin/	Regroupe infos et liens utiles dans le domaine des Activités Physiques et Sportives. Propose aussi des cours en ligne
SWITCH (The Swiss Education & Research network)			http://www.switch.ch/libraries/		Catalogues de bibliothèques suisses d'éducation et de recherche
The Economist			http://www.economist.com/		Accès à ce périodique

Série n°3 : Revues scientifiques, informatiques et biomédicales.

Produit (revue, base de données ou catalogue collectif)	Editeur	Producteur (si différent de l'éditeur)	Site web international	Site web français	Offre
American Mathematical Monthly	The Mathematical Association of America (MAA)			http://www.maa.org/pubs/monthly.html	Mensuel de la MAA (mathématiques)
Atmosphère-Océan	SDMM (service des données sur le Milieu Marin) de la Direction du MPO (Ministère fédéral canadien des Pêches et des Océans)		http://www.cm-os.ca/Ao/chroninstr.htm		Résultats de recherche originale, des articles sommaires et des observations sur les documents publiés dans les domaines des sciences de l'atmosphère et de l'océan ainsi qu'en hydrologie. Les articles peuvent être en anglais ou en français.
Bulletin of the Institute of Mathematical Statistics	Institute of Mathematical Statistics (IMS)		http://www.imstat.org/		Mathématiques appliquées aux statistiques
Byte	CMP Media LLC		http://www.byte.com/		Une revue de référence sur l'informatique
Cell	American Society for Cell Biology	Cell Press	http://www.cell.com/		Tous les contenus électroniques de la revue disponibles sur la plateforme Cell Press (biologie, biochimie, neuroscience, biologie moléculaire, génétique, génomique, bioinformatique, etc.) Lien depuis depuis la référence d'un article cité vers son texte intégral s'il appartient à une revue de la plateforme
Communications in Partial Differential Equations	Marcel Dekker		http://www.dekker.com/servlet/product/ProductID/PDE/		Revue spécialisée en mathématiques
Fibonacci Quarterly	The Fibonacci Association		http://www.engineering.sdsate.edu/~fib/		Revue spécialisée en mathématiques
La Recherche	Société d'éditions scientifiques			http://www.larecherche.fr/	Accès en ligne à la revue

Produit (revue, base de données ou catalogue collectif)	Editeur	Producteur (si différent de l'éditeur)	Site web international	Site web français	Offre
Le Doctrinal	Transactive			http://www.doctrinal.fr	Le Doctrinal, mis à jour tous les mois, est un premier point d'entrée dans le droit français. Base de données bibliographiques, le Doctrinal dépouille depuis le 1er janvier 1993 l'intégralité de la doctrine publiée dans plus de 200 publications françaises
Medecine/Sciences	EDK			http://www.edk.fr	Revue française en sciences biomédicales
New Scientist	Reed Business Information Ltd			http://www.newscientist.com/	Accès à la version en ligne de la revue
OECD Source	OCDE		http://www.sourceocde.org/		La bibliothèque en ligne de l'Organisation de Coopération et de Développement Economiques, est accessible à tout internaute désireux d'obtenir des informations sur les activités de l'OCDE. Toutefois, l'accès au texte intégral des livres, périodiques et bases de données statistiques n'est réservé qu'aux abonnés et aux utilisateurs des institutions abonnées. SourceOCDE comporte des livres classés par thèmes (il y en a 20), comprenant l'ensemble de nos monographies et rapports ; 24 périodiques ; des titres de référence ainsi que les bases de données statistiques de l'OCDE.
Revue de l'IEEE : _ IEEE Transactions on computer IEEE- _ Transactions on Knowledge and Data Engineering	IEEE Computer Society		http://computer.org/tc/ http://computer.org/tkde/		Les revues de l'IEE sont spécialisées en informatique
Revue UCP : _ Current Anthropology _ Journal of Labor Economics _ Journal of Political Economy	University of Chicago Press		http://www.journals.uchicago.edu/CA/journal/index.html http://www.journals.uchicago.edu/JOLE/journal/ http://www.journals.uchicago.edu/JP/E/journal/		Revue des presses Universitaires de Chicago.
Traitement du Signal	GRETSI			http://www.lis.inpg.fr/revue/	Les articles publiés sont des résultats originaux de recherche. Ils recouvrent à la fois les domaines fondamentaux, les aspects d'applications et tous les supports pour le signal, l'image et la parole.

***Annexe 2 : exemples de profils Z39.50 et
portails documentaires équipés de progiciels
Archimed***

Annexe 2-1 : exemples de profils Z39.50

Profil z39.50 du catalogue des bibliothèques du SCD de l'Université de Limoges

<<http://www.unilim.fr/scd/infos/z3950.htm>>

Hôte : scd-ab6.unilim.fr

Port : 210 jusqu'au 25 avril 2004 (2100 à partir du 28 avril 2004)

Nom de la base : ABSYSF

Disponibilité : 24/24 H 7/7 J. Fermeture dernière semaine de juillet, 2 premières semaines de d'août et 2 semaines à Noël

Attributs d'usage :

ATTR_USE:1:[AUTP] # PERSONAL NAME
 ATTR_USE:2:[AUTC] # CORPORATE NAME
 ATTR_USE:3:[AUTC] # CONFERENCE NAME
 ATTR_USE:4:[TIT] # TITLE
 ATTR_USE:5:[SERI] # TITLE SERIES
 ATTR_USE:6:[T503] # TITLE UNIFORM
 ATTR_USE:7:[T010] # ISBN
 ATTR_USE:8:[T011] # ISSN
 ATTR_USE:11:[T020] # BGF NUMBER
 ATTR_USE:12:[TITN] # LOCAL NUMBER
 ATTR_USE:13:[T676] # DEWEY CLASIFICATION
 ATTR_USE:14:[T675] # UDC CLASIFICATION
 ATTR_USE:21:[MATE] # SUBJECT HEADING
 ATTR_USE:22:[MATE] # SUBJECT RAMEAU
 ATTR_USE:30:@FEIN # DATE
 ATTR_USE:31:@FEPU # DATE OF PUBLICATION
 ATTR_USE:32:@FEIN # DATE OF ACQUISITION
 ATTR_USE:35:[T510] # TITLE PARALLEL
 ATTR_USE:36:[T512] # TITLE COVER
 ATTR_USE:37:[T513] # TITLE ADDED TITLE PAGE
 ATTR_USE:39:[T515] # TITLE RUNNING
 ATTR_USE:41:[T517] # TITLE OTHER VARIANT
 ATTR_USE:45:[T670] # SUBJECT PRECIS
 ATTR_USE:48:[T020] # N NATL BIBLIO
 ATTR_USE:49:[T021] # N LEGAL DEPOSIT
 ATTR_USE:50:[T022] # N GOVT PUB
 ATTR_USE:51:[T071] # N MUSIC PUBLISHER
 ATTR_USE:52:[TITN] # NUMBER DB
 ATTR_USE:54:[LENG] # CODE LANGUAGE
 ATTR_USE:55:[T660] # CODE GEOGRAPHIC AREA
 ATTR_USE:59:[T210] # PLACE PUBLICATION

ATTR_USE:62:[T330] # ABSTRACT
 ATTR_USE:63:[T300] # NOTE
 ATTR_USE:1003:[AUT] # AUTHOR
 ATTR_USE:1004:[AOTP] # AUTHOR NAME PERSONAL
 ATTR_USE:1005:[AUTC] # AUTHOR NAME CORPORATE
 ATTR_USE:1006:[AUTC] # AUTHOR NAME CONFERENCE
 ATTR_USE:1009:[T600] # SUBJECT NAME PERSONAL
 ATTR_USE:1011:[T005] # DATETIME ADDED TO DB
 ATTR_USE:1012:[T005] # DATETIME LAST MODIFIED
 ATTR_USE:1016: # ANY
 ATTR_USE:1018:[T210] # PUBLISHER
 ATTR_USE:1031:[L030] # MATERIAL TYPE
 ATTR_USE:1035: # ANYWHERE
 ATTR_USE:1036:[T600] # AUTHOR TITLE SUBJECT

NB : ATTR_USE:N:[XXXX] N = attribut de recherche Z39.50
 [XXXX] = champ de la banque BRS

Attributs de relation, position, structure, troncature et complétude :

La société GFI responsable de la distribution du logiciel AB6 en France est dans l'incapacité de nous fournir des informations précises concernant le paramétrage de ces attributs...

On peut vraisemblablement supposer que les valeurs sont au moins :

relation : 3 (=)

position, structure, troncature et complétude : 1 (valeur par défaut)

Ci-dessous les informations parcellaires dont nous disposons via GFI :

structure : phrase

relation : <, <=, >=, >, =, <>

troncature : gauche, droite, les 2

opérateur booléens : AND, OR, NOT

Présentation des données bibliographiques :

Format Unimarc / ISO 2709.

Jeu de caractères ISO 8459-1

Champs de données locales :

Les notices bibliographiques présentées contiennent des champs marc de données locales décrivant les exemplaires du titre en possession de la bibliothèque : localisation précise, cote, règle de prêt, disponibilité.

Ces informations se trouvent dans le champ MARC 998 (extraction à la volée de la base Oracle lors de la requête) pour les ouvrages et dans le champ MARC 966 (dans la notice originelle) pour les collections de périodiques.

OUVRAGES :

998 : champ contenant les exemplaires

«a»[cobarc] code-barre (numérique)

«b»[sudesc] succursale (alphanumérique)

«c»[cldesc] localisation (alphanumérique)

«d»[cpdesc] type d'exemplaire (alphanumérique)

«e»[cosign] cote (alphanumérique)

«f»[cosigs] n° inventaire (alphanumérique)

«g»[costat] situation interne de l'exemplaire (alphabétique ; C = Circulation, P = Perdu, X = Pilonné)

«h»[prfp] date du prêt (Date/Heure)
 «i»[prfdev] date de retour prévue (Date/Heure)

PERIODIQUES :

966 : Etiquette contenant les exemplaires

«c» localisation

«e» état de collection

«a» cote

Profil z39.50 du du SCD de l'Université Lyon 3

< <http://www.indexdata.dk/targettest/xml/193.52.199.5-21210-ADVANCE.xml>>

Ce profil est présenté en XML :

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<explain>

  <serverInfo protocol="Z39.50">
    <host>193.52.199.5</host>
    <port>21210</port>
    <database>ADVANCE</database>
  </serverInfo>

<databaseInfo>
  <title primary="true" lang="en">SCD de l'Université Lyon 3</title>
  <author>SCD de l'Université Lyon 3</author>
</databaseInfo>

<metaInfo>
  <dateModified>2002-04-22T15:07:39ZCEST</dateModified>
</metaInfo>

<indexInfo>
<index search="true">
  <title primary="true" lang="en">Title</title>
```

```

    <map>
      <name>Title</name>
    </map>
    <map primary="true">
      <attr type="1" attributeSet="bib-1">4</attr>
    </map>
  </index>
  <index search="true">
  <index search="true">
  <index search="true">
  <index search="true">
  <index search="true">
  <index search="true">
    <title primary="true" lang="en">Subject heading</title>
    <map>
      <name>Subject heading</name>
    </map>
    <map primary="true">
      <attr type="1" attributeSet="bib-1">21</attr>
    </map>
  </index>
  <index search="true">
  <index search="true">
  <index search="true">
  <index search="true">
  <index search="true">
  <index search="true">
    <title primary="true" lang="en">Author</title>
    <map>
      <name>Author</name>
    </map>
    <map primary="true">

```

```
        <attr type="1" attributeSet="bib-1">1003</attr>
    </map>
</index>
<index search="true">
<index search="true">
<index search="true">
<index search="true">
<index search="true">
<index search="true">
</index>
</indexInfo>

<recordInfo>
    <recordSyntax name="Unimarc">
        <elementSet name="F" />
    </recordSyntax>
    <recordSyntax name="OPAC">
        <elementSet name="F" />
    </recordSyntax>
</recordInfo>

</explain>
```

Je n'ai pas développé tous les index du profil, ici, nous voyons le paramétrage pour les index titre, sujet et auteur.

Annexe 2-2 : Quelques portails documentaires équipés par Archimed

- **Portail Sudoc** : <<http://www.portail-sudoc.abes.fr/>> : Book-Line + SIM + MASC.
- **BULCO** (Bibliothèque de l'Université du Littoral Côte d'Opale) : < <http://www-bulco.univ-littoral.fr>> : Book-Line + SIM + MASC.
- **Bibliothèque de l'Université Paris Dauphine** : <<http://www.bu.dauphine.fr/dauphine/default.asp>> : Book-Line + SIM + MASC.
- **BNUS (Bibliothèque Nationale Universitaire de Strasbourg)** : < <http://www-bnus.u-strasbg.fr/>> : Book-Line.
- **SCD de l'Université de Technologie de Troyes** : < <http://www-scd.utt.fr/MASC25/Default.asp?INSTANCE=UTT>> : Book-Line + SIM + MASC.
- **Réseau des Bibliothèques de Saint-Étienne** : <<http://www.bm-st-etienne>> : Book-Line + MASC.
- **Médiathèque de Dôle** : <<http://www.dole.org/sitedole/MEDIATHEQUE/default.htm>> : Book-Line.
- **Institut National des Jeunes Aveugles** : <<http://www.inja.fr>> : Book-Line + MASC.
- **Pôle International de la Préhistoire** : <<http://www.pole-prehistoire.com>> : Book-Line + SIM + MASC.

Annexe 3 : le mouvement OAI et ses ressources

1 Les principaux réservoirs (repositories)

- **OAI** : <<http://oaisrv.nsd.cornell.edu/Register/BrowseSites.pl>> Liste pluridisciplinaire des archives recensées dans le répertoire OAI (154 repositories)
- **OAF** : <http://www.oaforum.org/oaf_db/list_db/list_repositories.php>: liste des archives recensées dans le répertoire OAF comprenant près de 200 “entrepôts”
- **GNU (eprints.org)** : <<http://software.eprints.org>>: Répertoire GNU, pluridisciplinaire, 129 archives recensées.

2 En France

2.1 Les bases du CCSD (Centre pour la Communication Scientifique Directe) :
<<http://www.ccsd.cnrs.fr>>

- Thèses en ligne : <<http://thesesenligne.ccsd.cnrs.fr/>>: Thèses de doctorats et habilitations notamment en biologie, chimie, géologie, mathématiques, physique, sciences de l'information.
- Articles en ligne : <<http://hal.ccsd.cnrs.fr>> (Hyper Article en Ligne) articles scientifiques en libre accès (serveur aussi pour déposer)
- Archive de l'Institut Jean Nicod : <http://jeannicod.ccsd.cnrs.fr/index_en.html> (laboratoire interdisciplinaire contenant plus de 280 documents en linguistique, philosophie, sciences cognitives, sciences sociales.

2.2 Autres projets

- **Grisemine** (Université Lille 1) : <<http://crdoc.univ-lille1.fr/default.asp?bustl/grisemine>> : C'est une bibliothèque numérique de littérature grise francophone où sont diffusés communications, articles de périodiques, rapports de recherche, travaux d'étudiants, extraits de thèses ou thèses, supports de cours ou de travaux dirigés. 963 documents.
- **Pastel** : <<http://pastel.paristech.org/>>: base regroupant les thèses de dix grandes écoles d'ingénieurs parisiennes. Pastel est né de l'association de ces écoles pour offrir un service d'archivage en ligne des thèses. Pastel représente un potentiel de 500 thèses par an dans les domaines des sciences de l'ingénieur, sciences de la terre, environnement et chimie.

3 Archives internationales

- **ArXiv** : <<http://xxx.lanl.gov>>
 Site miroir français de ArXiv : <<http://fr.arXiv.org>>
 (Plus de 150 000 articles en physique, informatique, mathématiques) 15 miroirs dont celui de Jussieu et du Centre pour la Communication Scientifique Directe (IN2P3)

- **Cogprints** : <<http://cogprints.soton.ac.uk>>
 Plus 1600 articles dans la base biologie, informatique, édition électronique, linguistique, neuroscience, philosophie, psychologie.

- **ResearchIndex** : <<http://citeseer.nj.nec.com>>
 Plus de 7 millions de "pages" dans la base, 5 millions de citations, en "Computer Sciences". Outils développés pour analyser les citations "Autonomous Citation Indexing"
 629 254 auteurs référencés.

- **WoPEc** (Working papers in Economics) : <<http://netec.mcc.ac.uk/WoPEc.html>>
 Archive ouverte de "working papers" en économie, créée en 1996 avec des financements du programme Electronic Libraries(eLib) /UK.
 Environ 80 000 documents (articles de revue + Working papers). Tous les articles sont téléchargeables mais pas nécessairement en libre accès.

- **Index national des prépublications et thèses en mathématiques en France/**
 The Mathematics Preprint Search System (MPRESS)
 <<http://www-mathdoc.ujf-grenoble.fr/>>
 <<http://www-mathdoc.ujf-grenoble.fr/MPRESS/>>
 Archive ouverte distribuée (sur les web des laboratoires) coordonnée par la Cellule MathDoc (Unité mixte de service du CNRS) de Grenoble.
 10 archives sont reliées au niveau international par MPRESS
 Plus de 5000 documents dans la base française, plus de 45 000 documents dans MPRESS.

- **eScholarship** Repository de l'Université de Californie : archive institutionnelle
 <<http://repositories.cdlib.org/escholarship/about.html>>
 Base d'archives de working papers et de prépublications des départements de recherche de l'université de Californie. Plus 1000 documents.

- Le serveur du **CERN** (Centre Européen de la Recherche Nucléaire) :
 <<http://cds.cern.ch/>>: Plus de 650 000 références bibliographiques dont 320 000 en texte intégral relatives à la physique dans sa globalité ainsi qu'à ses disciplines corrélatives.

- **Mathematical Physics Preprint Archive** :
<http://rene.ma.utexas.edu/mp_arc/index.html>
Archives électroniques en Mathématique et Physique et sujets connexes.
Sites miroirs à <[Austin](#)>, <[Genève](#)>, <[Barcelone](#)>, <[Rome](#)>.
- **Fermilab Preprint** :
<http://fnalpubs.fnal.gov/techpubs/fermilab_spires.html>
Recherche possible sur les archives depuis 1970. Domaines couverts: Cosmologie, accélérateurs, physique générale, physique des hautes énergies, particules et champs, mécanique quantique, physique théorique.

4 Les réservoirs de périodiques

- **Directory of Open Access Journal (DOAJ)** : <<http://www.doaj.org/>>: base pluridisciplinaire avec plus de 1100 titres en texte intégral.
- **FreeMedicalJournals** : <<http://www.freemedicaljournals.com>>: rassemble les adresses de 1300 revues offrant des ressources en libre accès
- **Biomed Central** : <<http://www.biomedcentral.com>>: médecine, biologie, indexées dans Medline (ici, c'est l'auteur qui paie pour publier)
- **SciELO (SCientific Electronic Library On line)** : <<http://www.scielo.org>>: sélection de périodiques spécialisés en STM en texte intégral (hispanophones et lusophones).
- **Project Muse** : <<http://muse.jhu.edu/>> : plus de 200 revues en texte intégral dans le domaine des sciences sociales, arts et mathématiques.
- **Erudit** : <<http://www.erudit.org/revue/index.html>> : Site de diffusion de 40 revues universitaires du Québec, rassemblant un éventail de disciplines des sciences humaines et sociales, ainsi que des sciences naturelles. Pour le moment, les revues dans Érudit sont disponibles gratuitement. Érudit compte déjà 6 000 articles, 6 livres numériques et quelques centaines de prépublications (données décembre 2002). L'outil de recherche permet d'interroger l'ensemble de la collection ou certains de ses segments.
- **Project Euclid** : <<http://projecteuclid.org/Dienst/UI/1.0/TitleSort/>>: Le projet Euclid de la Cornell University Library, en partenariat avec SPARC (Scholarly Publishing and Academic Resources Coalition) vise à offrir une plateforme de publications dans le domaine des mathématiques appliquées et des statistiques.
- **MDPI (Molecular Diversity Preservation International)**:
<<http://www.mdpi.net/journals.htm>>: C'est une organisation suisse à but non lucratif qui recensait des dépôts de structures moléculaires ou biomoléculaires. Depuis 1996,

MDPI diffuse en libre accès 8 revues spécialisées dans le domaine de la chimie, mais également des conférences et e-conférences.

5 Les bases de données en lignes gratuites recensées par l'URFIST Lyon 1.

- 716 bases de données sur plus d'une trentaine de sujets. <<http://urfist.univ-lyon1.fr/gratuits/index.html>>