

## Archives institutionnelles et archives centralisatrices: pratiques existantes et bonnes pratiques

Lorraine Filippozzi  
[lorraine.filippozzi@hesge.ch](mailto:lorraine.filippozzi@hesge.ch)  
Haute Ecole de Gestion, Genève

### Résumé

*Cette étude a pour but de fournir une base documentaire solide au projet Info-Net Economy, futur portail thématique qui constituera le volet « Économie » de la bibliothèque électronique suisse E-Lib. Après avoir mis en évidence l'intérêt de proposer des portails centralisateurs et surtout l'opportunité que cela représente pour les professionnels de l'information, cette étude analyse les aspects organisationnels, techniques et marketing de projets similaires afin de mettre en évidence les bonnes pratiques à retenir.*

### Mots-clés

*Archives ouvertes, Open acces, Portail, Bibliothèque électronique, Norme OAI-PMH*

Cette étude s'inscrit dans la première phase du projet Info-Net Economy (1), futur portail thématique qui constituera le volet « Économie » de la bibliothèque électronique suisse E-Lib (2). Ce portail sera notamment composé d'un accès aux publications électroniques issues des dépôts institutionnels d'un réseau de partenaires publics et académiques. Cette étude a pour objectif de retenir les bonnes pratiques en matière de dépôt institutionnel et de portails centralisateurs de ressources thématiques. Pour cela, une revue de la littérature concernant les portails centralisateurs de dépôts institutionnels et les archives ouvertes a été menée. Ces éléments de nature théorique ont été complétés par l'étude de projets similaires. Les bonnes pratiques retenues concernent principalement trois aspects : l'organisation, la technologie et la communication liées au projet. Mais avant de les aborder plus en détail, une description plus détaillée du contexte et une définition de la notion d'Open Access s'imposent.

### Contexte de l'étude : Info-Net Economy

Info-Net Economy, qui fait partie du projet E-Lib, vise à offrir un point d'accès unique aux publications électroniques économiques suisses et à valoriser les ressources économiques provenant d'instituts suisses. Il doit favoriser les échanges entre institutions productrices d'informations économiques. Le portail contiendra les éléments suivants : un répertoire des différents acteurs économiques produisant des études/informations économiques (Hautes écoles, instituts spécialisés, offices spécialisés, etc.) ; un répertoire des publications électroniques de certains de ces organismes intéressés à être partenaires (11 en octobre 2009); un accès plus général à l'ensemble des sources d'information économiques utiles (accroissement de l'annuaire de sources ARESO (3)) ; un espace dédié aux échanges entre professionnels de l'information actifs dans le domaine, destiné à partager les produits documentaires réalisés ainsi que les bonnes pratiques.

La centralisation de ces informations sur un portail permettra de remédier au problème qui se pose actuellement en matière de recherche d'informations économiques validées : les ressources sont dispersées sur les sites des différentes institutions, ce qui nécessite de connaître ces différents acteurs. Même si certains sites proposent un regroupement de ressources, ils sont généralement limités à la sphère universitaire, les ressources d'autres types d'institutions restant méconnues. En outre, les avantages escomptés pour les partenaires sont les suivants : la valorisation de leurs publications ; la connaissance de nouvelles sources d'information et de nouveaux outils, du fait des échanges entre partenaires du réseau ; la garantie d'une meilleure visibilité pour leur service d'information documentaire mais aussi pour leur propre institution. Le réseau de partenaires créé pour Info-Net Economy regroupe des offices publics spécialisés, des instituts universitaires, des hautes écoles et des organisations faîtières qui publient dans le domaine de l'économie en Suisse et qui emploient des professionnels de l'information. En effet, le principal enjeu de ce projet est l'opportunité que représentent la centralisation et la diffusion des publications pour les professionnels de l'information, au niveau de leur institution, mais aussi au niveau plus global. Cette opportunité pourrait même constituer l'une des voies d'avenir de la profession.

En vue de la conception de ce portail spécialisé dans le domaine économique, et parallèlement à la phase de recensement des publications et des acteurs prévue durant la première partie du projet, une étude sur les répertoires institutionnels de recherche s'avère indispensable. De quoi faut-il tenir compte lors de la création d'un tel portail ?

La construction du portail d'accès aux publications spécialisées en économie provenant de notre réseau de partenaires se base sur les dépôts mis en place par ces derniers. Un dépôt institutionnel (*institutional repository*) est la collection numérique de la production intellectuelle d'une université ou d'un institut qui centralise, préserve et rend accessibles les connaissances générées par l'institution (4). Ces dépôts ont vu le jour grâce au principe d'archives ouvertes. Tel qu'entendu dans les universités et les hautes écoles suisses, il s'agit du dépôt des publications scientifiques sur un serveur institutionnel ou de la publication dans une revue fonctionnant selon le modèle *Open Access* (OA) (5). Les usagers ont ainsi un accès en ligne immédiat et gratuit aux articles de recherche : ils peuvent lire, télécharger, copier et créer des liens vers le texte intégral sans réserve mais avec obligation de citation. Ce principe offre de nombreux avantages tant pour le chercheur que pour l'utilisateur : rapidité de publication, conservation des droits d'auteurs, accès gratuit et universel, etc. (6)

Ce principe défini, voici les différents projets similaires étudiés qui serviront de modèles pour mettre en valeur les bonnes pratiques retenues. Il s'agit d'un choix de dépôts institutionnels (ou archives institutionnelles), de dépôts centralisateurs (archives centralisatrices) (7) et de portails thématiques qui ont été sélectionnées en fonction de leur proximité géographique, de leur exemplarité et de leur originalité.

<b>Dépôts Institutionnels</b>	<p><b>Infoscience (8)</b></p> <p>Conçue comme une véritable archive institutionnelle, Infoscience centralise et conserve la production scientifique de l'EPFL. Infoscience est une référence en matière de dépôt institutionnel selon le classement <i>Ranking of world repositories</i> (9), où il est classé en septième position.</p>	<p><b>Forschungsplattform Alexandria (10)</b></p> <p>Plateforme servant de vitrine à la recherche menée à l'Université de Saint-Gall, Alexandria a non seulement pour but de rendre publics les résultats de la recherche menée au sein de l'institution, mais aussi de mettre en valeur les instituts de recherche, leurs chercheurs (profils personnels) et les projets en cours. Alexandria n'est pas une archive institutionnelle au sens strict car il n'y a pas de vision d'archivage. Cette plateforme est néanmoins classée en huitième position dans le classement cité ci-contre.</p>
<b>Dépôts centralisateurs</b>	<p><b>RERO DOC (11)</b></p> <p>RERO-DOC est la bibliothèque numérique de RERO (Réseau des bibliothèques de Suisse occidentale). Dans le cadre du mouvement des archives ouvertes, elle offre une plateforme de dépôt de publications</p>	<p><b>Driver (12)</b></p> <p>Portail multidisciplinaire pour la recherche européenne en libre accès, Driver fédère les répertoires d'archives ouvertes européens. Il compte sur treize partenaires en Europe pour</p>

	pour les chercheurs des universités et autres institutions membres. Cette plateforme constitue donc le dépôt centralisateur des publications des nombreuses institutions suisses qui ne possèdent pas leur propre dépôt institutionnel.	constituer le portail paneuropéen d'accès libre aux publications de recherche et regroupe à ce jour plus de 200 dépôts institutionnels ou thématiques, dont la plateforme d'archives ouvertes française HAL (Hyper-article en ligne). Pour le moment, Driver couvre essentiellement des sujets tels que la biologie, l'anthropologie et l'informatique, avec un accès aux ressources par région géographique.
<b>Portails thématiques</b>	<p><b>Econbiz<sup>(13)</sup></b></p> <p>Conçue comme un portail thématique, la bibliothèque virtuelle des sciences économiques est un projet de collaboration entre la Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW) et l'Universitäts und Stadtbibliothek Köln (USB Köln). Elle contient un répertoire de signets électroniques, un métamoteur de recherche dans les catalogues des bibliothèques générales, des liens vers des bases de données en texte intégral, un agenda des manifestations scientifiques internationales et des services d'information.</p>	<p><b>Economists online <sup>(14)</sup></b></p> <p>NEREUS - Networked Economics Resources for European Scholars - représente un consortium de bibliothèques universitaires européennes, disposant de collections majeures en économie et se proposant de fournir collectivement de nouveaux contenus et de développer des services d'information innovateurs pour l'économiste. Le but est de créer un réseau européen de bibliothèques de recherche et de relier ainsi les ressources issues de la recherche universitaire en Europe. À travers le portail thématique Economists Online, l'accès en ligne à la production des principaux économistes universitaires constitue le service-clé de NEREUS. Ce projet consiste à numériser, organiser, archiver et diffuser la production complète des économistes, visant ainsi à rendre disponibles autant de textes intégraux que possible.</p>

Chacune de ces sources a été étudiée selon la même grille d'analyse, qui contenait les points suivants : contexte de création, objectifs, mode de fonctionnement, structure technique générale, contenu (types de publications), fonctionnalités (y compris web 2.0), interface,

ergonomie, outil de recherche, types de recherche proposés, gestion du multilinguisme, format, récupération et qualité des métadonnées, promotion, et finalement les bonnes idées à retenir. Les conclusions de l'analyse de ces différents points ont été regroupées en trois parties : contexte et organisation, technologies et contenus offerts, et finalement promotion.

## Contexte et organisation

Autant dans les projets de dépôts institutionnels que lors de la création de portails centralisateurs de ressources, le mode de fonctionnement est un facteur de réussite ou d'échec déterminant. Il est important de souligner les différents aspects à prendre en considération pour faire les choix les plus adaptés en fonction du contexte : mode de fonctionnement, type de partenaires, niveau de collaboration et rôle des professionnels en information documentaire dans de tels projets.

### Mode de fonctionnement

Le mode de fonctionnement en réseau est très généralement adopté dans les projets de fédération de dépôts institutionnels (Econbiz, Driver, Economists online). Constitué autour d'un noyau dur de départ, le partenariat entre 2 à 10 institutions donne l'impulsion et assure le suivi du projet. Devenant un véritable centre de compétence, le réseau d'experts mis en place devient parfois même fournisseur de service pour les gestionnaires de dépôts. Les archives institutionnelles se sont structurées en réseaux nationaux, voire internationaux, avec pour objectif de partager les compétences, mais aussi de créer des outils et des services communs ([15](#)).

### Type d'archives ouvertes

Il existe différents types d'archives ouvertes : elles peuvent être institutionnelles (liées à un organisme producteur, par exemple une université) ou centralisatrices (liées à plusieurs organismes producteurs, par exemple RERO-DOC qui sert de dépôt à plusieurs universités et hautes écoles), multidisciplinaires (elles couvrent plusieurs domaines de la connaissance) ou thématiques (focalisées sur un domaine de la connaissance, par exemple l'économie avec Repec (Research Paper in Economics) ([16](#)), ou les sciences de l'information et la bibliothéconomie avec E-lis ([17](#))). L'impulsion de ces archives ouvertes institutionnelles, centralisatrices ou thématiques, provient le plus souvent du milieu universitaire. Mais on trouve également des instituts de recherche tels que le Centre national de la recherche scientifique (CNRS) français, qui a créé le dépôt multidisciplinaire HAL (Hyper Articles on Line) ([18](#)), ont aussi intérêt à rendre publics les publications de leurs chercheurs. La multiplication des projets démontre l'intérêt des différentes communautés scientifiques à créer des archives ouvertes ainsi qu'à être présentes sur des portails centralisateurs dans un effort collaboratif de valorisation de la publication scientifique.

### Niveau de collaboration

Toutefois, comment entretenir cette collaboration sur le long terme ? Comment créer un vrai centre de compétences ? En observant les projets similaires, on constate qu'il est nécessaire que les partenaires du réseau soient soutenus par leur instance dirigeante, qui doit être sensibilisée au principe des archives ouvertes. Les professionnels de l'information ont ici un rôle primordial à jouer. Pour convaincre les autorités de tutelle, ils peuvent s'appuyer sur les chiffres concernant l'Open Access et les dépôts institutionnels. Aux États-Unis, plus de 90% des universités ont institué un système de dépôt institutionnel ou sont en cours de réflexion

pour y parvenir (19). Ce soutien au mouvement des archives ouvertes peut même être élevé au niveau national, comme c'est le cas pour HAL (20), qui est soutenu par le CNRS et auquel participent universités et centres de recherche publics (21). Dans la même optique, Repec (22) est né d'un effort collectif de volontaires provenant de plus de 60 pays dans le but d'améliorer la diffusion de la recherche en économie. Si, à l'heure actuelle, très peu de pays se sont dotés d'une politique gouvernementale de soutien à l'Open Access, celui-ci est très courant un niveau des universités (23). En Suisse, l'étude menée en 2009 par la Hochschule für Technik und Wirtschaft (HTW) de Coire a identifié 9 dépôts de hautes écoles universitaires et 8 dépôts en projet qui verront le jour d'ici un à deux ans (24). Le projet E-lib verra par ailleurs la création de la bibliothèque électronique suisse, portail scientifique national (25).

### Rôle des professionnels en information documentaire

La démocratisation du principe d'archives ouvertes au sein des institutions de recherche académique représente une opportunité pour les bibliothèques scientifiques. En effet, en s'impliquant dans les projets mis en œuvre, elles obtiennent un nouveau rôle au niveau de l'acquisition, de l'évaluation et de la diffusion de la recherche (26). Elles peuvent se positionner en intermédiaire entre les chercheurs et les éditeurs, mais aussi entre les chercheurs et le public. Ce nouveau rôle renforce la proximité entre les bibliothèques et les chercheurs et s'inscrit dans l'évolution globale, voire la mutation de leur mission (27). L'information scientifique représente un enjeu de taille pour les bibliothèques académiques et de recherche car elle soulève des défis importants non seulement quant à sa diffusion et à sa conservation, mais aussi quant à sa maîtrise par les usagers. La formation des usagers revient au cœur de leur mission dans des projets de learning centers, ou centres d'apprentissage (28), qui offrent notamment des formations à la maîtrise de l'information et des technologies de l'information et de la communication. Les professionnels de l'information ont donc un rôle stratégique à jouer dans ce climat de changement de comportement des usagers et des modes de publication. Il est même possible, voire souhaitable, qu'ils obtiennent le leadership de projets liés aux archives ouvertes (29).

Cependant, les bibliothèques doivent parfois faire face à des difficultés d'intégration entre le pôle informatique qui réalise les dépôts institutionnels et les instituts de recherche qui les rempliront. Il devient alors nécessaire pour elles de faire valoir leurs compétences en matière d'information numérique et de prendre une place centrale dans sa conservation et sa diffusion. Elles sont en effet, par leur position d'intermédiaire, les plus à même de relayer les besoins et les demandes de leurs usagers (30). Cette collaboration avec les développeurs, qui ont une vision d'ensemble des différents systèmes en place et assurent la bonne intégration du dépôt au système d'information en place, est un des facteurs de succès de tels projets (31).

Si le soutien politique est nécessaire à la vie des centres de compétences mis en place, il est aussi central dans la diffusion du principe des archives ouvertes. Tous les projets étudiés ont pour base une politique interne régissant le dépôt des publications et la création d'archives institutionnelles. Ces politiques se basent sur la Déclaration de Berlin sur le libre accès à la connaissance (2003), où les signataires s'engagent, entre autres, à inciter les chercheurs à éditer leur travail selon les principes du libre accès et à encourager les établissements culturels à soutenir l'accès aux ressources sur Internet. Depuis, un nombre croissant d'institutions académiques ont mis en place leur propre dépôt et instauré l'obligation pour leurs chercheurs



d'y déposer leurs publications. Il est difficile de recenser le nombre total de dépôts dans le monde, mais il existe un répertoire où les entrepôts d'archivage libre peuvent se déclarer ([32](#)).

Pourtant, malgré le soutien hiérarchique des organes de direction de la recherche, il reste un gros travail à accomplir pour sensibiliser et inciter les chercheurs à publier selon ce modèle. On observe en effet dans certains domaines une méfiance et une certaine résistance au changement. Elles sont peut-être dues à la méconnaissance du principe de l'OA et à la crainte d'une perte de qualité et de renommée. Les bibliothèques doivent contribuer à l'information de leurs utilisateurs dans le but de renforcer leur confiance et s'assurer qu'ils adhèrent aux projets de dépôts institutionnels. Elles doivent aussi offrir un support technique aux instituts pour éviter que chacun fasse comme il pense/peut et faciliter l'harmonisation des pratiques. Pour cela, elles peuvent s'appuyer sur les travaux effectués par des leaders des archives institutionnelles tels que SHERPA ([33](#)), consortium qui promeut la création d'archives dans tous les établissements de recherche et d'enseignement supérieur au Royaume-Uni. Un service particulièrement utile est Sherpa-RoMEO (Rights METadata for Open archiving) : il fournit une liste des licences de publications des éditeurs concernant l'archivage de leurs publications par les auteurs qui permet de savoir si l'éditeur permet ou non le dépôt de la publication dans une archive institutionnelle ([34](#)), et, si oui, quelles en sont les conditions (par exemple s'il y a une période d'embargo suivant la publication). Actuellement, selon les statistiques du site SHERPA-RoMEO, 95 % des périodiques donnent le droit aux auteurs d'auto-archiver, et il a été prouvé que l'autoarchivage augmente la visibilité et donc le lectorat ([35](#)). Communiquer et informer sur les enjeux liés aux dépôts institutionnels est donc la première bonne pratique à retenir lors de projets liés à l'Open Access.

## Technique, contenu et fonctionnalités

Une fois le projet mis sur pied, il s'agit d'étudier les différents projets similaires afin de procéder à certains choix techniques concernant notamment la structure, les contenus et les fonctionnalités souhaités, ainsi que les outils existants et les normes à respecter.

### Structure technique

La structure technique des différents portails similaires au futur portail Info-Net Economy repose sur le principe de la confédération de dépôts institutionnels hébergés localement. Ces dépôts, comme Infoscience ([36](#)) ou Alexandria ([37](#)), sont constitués non seulement des objets eux-mêmes, mais aussi de métadonnées descriptives. Dans une structure de fédération de dépôts, comme Driver ([38](#)), les publications des institutions partenaires restent dans leur dépôt local et le portail ne fait que signaler ces sources par les métadonnées qu'il y a récoltées et pointer vers elles par des liens. Il est ainsi possible de fédérer la recherche dans les métadonnées des différentes institutions productrices tout en leur en laissant la responsabilité de la gestion et la complète propriété.

### Contenus et fonctionnalités

Ces portails offrent divers contenus et fonctionnalités. Au niveau du contenu, on trouve différents types de publications académiques : articles scientifiques en majorité, livres et chapitres de livres, travaux d'étudiants (mémoire de master et thèses de doctorat, parfois aussi mémoires de bachelor), mais aussi travaux de professeurs-chercheurs publiés ou non (pré-publications, working papers, papiers et actes de conférences, etc.) ainsi que, comme dans le

cas de l'Archive Ouverte de l'Université de Genève (39), matériel de cours (supports, enregistrements, etc.). Ces contenus de type académique sont souvent complétés par des publications plus vulgarisées telles que des articles de journaux et des communiqués de presse. Dans certains cas il n'y a pas de limitation à la typologie des documents ; il est ainsi de moins en moins rare de voir des données brutes de recherche ou des contenus multimédia. Le contenu des dépôts est généralement détaillé dans leur politique, outil essentiel permettant de définir de manière précise la portée du dépôt. En plus des publications elles-mêmes, presque tous les sites observés proposent une rubrique « Actualités » qui recense aussi bien les nouveaux projets de recherche que la parution d'ouvrages. Ces différents contenus sont généralement mêlés et ce sont les fonctionnalités de recherche qui permettent de cibler l'un ou l'autre des contenus selon ses besoins. Le défi à relever est celui de l'ergonomie. En effet, comment structurer les différents contenus de manière simple pour rendre la recherche la plus intuitive possible ? Pour répondre aux différents besoins, les fonctions de recherche doivent être complètes. De la barre de recherche simple aux champs de recherche combinés de la recherche avancée, en passant par la navigation dans la classification thématique, les types de documents et les acteurs concernés, sans oublier les options de filtrage des résultats, il s'agit d'offrir une large palette de services. Deux exemples sont à signaler pour la simplicité de la recherche et la bonne exploitation des filtres de recherche : il s'agit d'Infoscience (EPFL) (40) et d'Alexandria (Université de Saint-Gall) (41). L'ergonomie de la fonction de dépôt des publications par les chercheurs est généralement bien soignée, et les marches à suivre sont claires et simples.

Les fonctionnalités de type web 2.0 sont encore peu présentes dans les portails de recherche académiques. Dans les projets observés pour cette étude, l'usage de flux RSS est courant pour signaler les nouveautés (Alexandria) ou les nouvelles publications sur une recherche effectuée (Infoscience). Cependant peu de possibilités d'interactions sont offertes aux usagers des portails, chercheurs ou grand public. Aucun des portails observés, de type académique, ne permet aux usagers de « tagger » eux-mêmes les notices ou de poster des commentaires au sujet des ressources. Les possibilités de recherche sémantique sont donc amoindries. Pourtant le projet Inspire (42), portail de nouvelle génération dans le domaine de la physique, prévoit de faire contribuer les usagers au « taggage » et au commentaire des ressources (43). Cet exemple permettra de démontrer la pertinence d'une telle approche. Il est intéressant de relever qu'Infoscience permet à l'utilisateur, une fois enregistré, de s'approprier les données en créant sa collection personnelle en ligne ou en paramétrant le moteur de recherche selon ses besoins. Par ailleurs, si tous les portails observés offrent des FAQ (*Frequently asked questions* ou Foire aux questions) en ligne, seul Econbiz (44) propose un véritable service de renseignement et de référence en ligne.

Le choix de proposer un site multilingue soulève la question de la gestion du multilinguisme. C'est un aspect particulièrement important dans un projet suisse tel qu'Info-Net Economy. Il s'agit en effet de proposer une interface, un outil de recherche et un affichage des résultats en plusieurs langues sans alourdir la présentation générale du site. La solution la plus couramment adoptée est celle de proposer une interface en plusieurs langues et des métadonnées dans la langue de la publication déposée. Dans ces cas, pour améliorer la recherche, les mots-clés peuvent éventuellement être traduits en plusieurs langues.



## Outils et normes

Au niveau technique toujours, il est important de connaître et de bien comprendre les différents outils servant au dépôt de publications d'une part et à la recherche fédérée dans différents dépôts d'autre part.

Concernant les outils de dépôts, selon le Registry of Open Access Repositories ([45](#)), les deux outils logiciels les plus utilisés dans le monde sont Dspace ([46](#)) (plus de 400 utilisateurs) et Eprints ([47](#)) (plus de 300 utilisateurs, dont l'Université de Zürich). CDS-Ware/Invenio ([48](#)), le système développé par le CERN, arrive quant à lui en septième position et compte parmi ses utilisateurs l'EPFL et RERO-DOC. Il s'agit de systèmes d'entrepôts de données.

Cependant, pour que cette fédération des ressources soit possible, il faut appliquer les mêmes normes. En effet, l'aspect déterminant de ces portails est la recherche, qui passe forcément par les métadonnées, et à ce sujet le mot d'ordre est « standardisation ». On parle de dépôts compatibles OAI. Mais qu'est-ce qu'OAI ? Mouvement de promotion des archives ouvertes, l'Open Archive Initiative développe un ensemble de protocole et de standards d'interopérabilité facilitant la diffusion des contenus des documents. Le protocole central est OAI-PMH, Open Archive Initiative Protocol for Metadata Harvesting, ou « Protocole de collecte de métadonnées de l'Initiative Archives Ouvertes ». Il définit un standard pour transférer des collections de métadonnées. Il a été établi dans le cadre de l'Initiative Archives Ouvertes pour implanter des bases interopérables de pré-publications scientifiques soumises par leurs auteurs. L'objectif de l'Initiative Archives Ouvertes est donc très spécifique aux communautés de chercheurs. Pourtant le protocole se révèle aussi intéressant pour des applications plus patrimoniales.

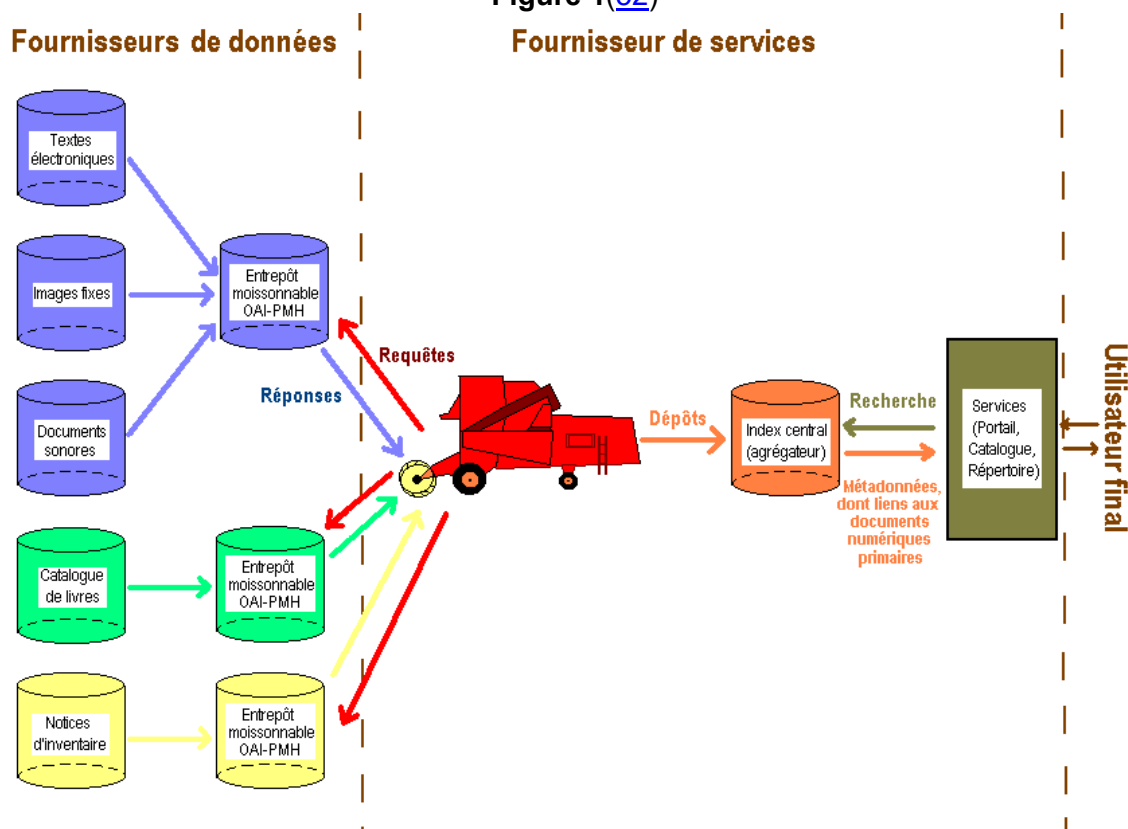
Pour les métadonnées, le protocole OAI-PMH repose sur une norme de description bibliographique, le Dublin Core non qualifié : schéma de métadonnées numériques qui permet de décrire des ressources numériques et d'établir des relations avec d'autres ressources. Il comprend 15 éléments de description formelle (titre, créateur, éditeur), intellectuels (sujet, description, langue) et relatifs à la propriété intellectuelle. On peut convertir les notices primaires en format Dublin Core. De plus, le Dublin Core est facilement encodable en XML.

Un dépôt institutionnel qui applique la norme OAI-PMH possède une base de métadonnées qu'il « expose » aux moissonneurs ([49](#)). Ces métadonnées y sont disponibles dans différents formats pour répondre à différents besoins. Si le format Dublin Core est l'exigence minimale imposée par la Norme OAI-PMH, le format MARCXML permet d'obtenir des métadonnées plus riches à partir de catalogues de bibliothèque par exemple. Ainsi le protocole permet de centraliser les métadonnées référençant diverses ressources tout en laissant ces ressources à leur emplacement initial. Il est alors possible d'accéder à ces ressources en architecture distribuée, en rassemblant les métadonnées et en les exploitant pour les besoins spécifiques d'un service. Dans le cadre de projets de portails tels qu'Info-Net Economy, les ressources sont gérées dans différents établissements. Pour fournir un accès commun à ces ressources, un système permet d'en rassembler les métadonnées dans une base commune. En bref, ce protocole d'échange permet de créer, d'alimenter et de tenir à jour, par des procédures automatisées, des serveurs de métadonnées qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d'origine ([50](#)). Il permet aussi l'agrégation de ressources hétérogènes d'origines diverses. Ce procédé de recherche et d'agrégation des ressources est nommé moissonnage, ou *harvesting* (voir Figure 1).

## Moissonnage OAI-PMH

Le moissonnage OAI-PMH nécessite deux types d'acteurs : d'une part les entrepôts déclarés (fournisseurs de données) ; et d'autre part les moissonneurs (fournisseurs de services). Les fournisseurs de données exposent leurs métadonnées en implémentant le protocole OAI et en donnant accès à leur catalogue au moyen d'un entrepôt OAI, outil chargé de répondre aux requêtes formulées par les fournisseurs de service ; tandis que les moissonneurs parcourent les dépôts et recueillent les métadonnées à l'aide du protocole OAI. Les réponses sont données au format XML et contiennent, selon la requête formulée, des informations sur l'entrepôt, des identifiants, des métadonnées. Par « moissonner » on entend ici le fait de récupérer une copie des métadonnées en local puis de la rendre cherchable comme valeur ajoutée. Un même dépôt peut parfaitement être interrogé par plusieurs moissonneurs. Les moissonneurs permettent d'agréger les ressources provenant de différents dépôts OAI. Ils nourrissent une base XML, qui peut aussi être alimentée par d'autres moyens, y compris en ressources locales et par intervention humaine. Dans ce cas, il faut établir en amont le processus de recherche et de sélection des URL des bases à moissonner (automatique ou manuel, parfois les deux), avec la fréquence de visites et de mises à jour pour chaque base et déterminer, en aval, le stockage, l'indexation, l'exploitation et l'affichage des données (transformation XSLT). En l'absence de dépôt institutionnel moissonnable répondant à la norme OAI-PMH, il est donc possible d'alimenter une base de métadonnées au moyen de flux RSS adaptés pour signaler des publications disponibles sur un site web ou au moyen d'une application web 2.0 basée sur des requêtes URL pré-paramétrées, comme le prévoit le projet RODIN ([51](#)), outil pour l'agrégation et le couplage de sources d'information. Il est par ailleurs aussi possible de récupérer les données de certains dépôts non OAI au moyen de ce même type de requêtes.

Figure 1([52](#))



Le schéma ci-dessus (Figure 1) expose le principe du moissonnage, avec :

- à gauche les différents dépôts institutionnels fournisseurs de données, avec chacun leur entrepôt de métadonnées ;
- au milieu le moteur de moissonnage qui comprend un moteur de moissonnage et un agrégateur de métadonnées (fournisseur de services)
- et à droite l'interface de recherche fédérée proposée à l'utilisateur final.

Il existe des moteurs de moissonnage open source, parmi lesquels les plus utilisés sont OAIster ([53](#)), CiteBase Search ([54](#)), OAIHarvester2 ([55](#)), PKP Harvester ([56](#)), Framework SDX ([57](#)) ou encore Google Scholar ([58](#)). On peut aussi développer sa propre fonction de recherche dans les métadonnées de plusieurs archives sélectionnées ([59](#)), toujours au moyen de requêtes URL dédiées.

Ainsi, les données exposées OAI sont référencées par les moteurs de recherche de documentation scientifique au niveau mondial (comme Google Scholar, Bielefeld academic search engine), mais aussi par les moteurs de recherche simple (Google, Yahoo, MSN, etc.). Les dépôts créés peuvent donc être moissonnés à leur tour et ainsi gagner en visibilité. On constate ainsi que les différents types de dépôts sont complémentaires.

Si le jeu de métadonnées et la manière de les exposer aux moissonneurs est dictée par la norme OAI-PMH, l'indexation thématique de chaque document reste libre. Chaque institution décide donc de sa manière d'indexer, de suivre un thésaurus de vocabulaire contrôlé ou de laisser ce champ libre aux auteurs qui déposent leurs publications. Cela pose problème au moment de la centralisation des métadonnées de plusieurs institutions, car il faut établir des listes de concordances pour faciliter la recherche fédérée par sujet. Driver, portail centralisateur de plus de 200 dépôts, a pris l'option de laisser le champ sujet libre et ne propose donc pas de recherche par sujet ou de navigation dans la classification.

La standardisation des métadonnées concerne non seulement leur moissonnage et leur exposition, mais aussi leur conservation. Si le jeu de métadonnées requis par Dublin Core convient bien à la manipulation des données, il est limité en termes de gestion et de conservation. Des formats internes de gestion et de conservation existent, plus complets et recouvrant les champs requis par Dublin Core : actuellement le plus courant est MARC ([60](#)), plus précisément MARCXML, qui est facilement transformable en différents formats (HTML pour l'affichage direct, Dublin Core ou MODS pour supporter les protocoles de harvesting) ([61](#)).

Il est par ailleurs intéressant de relever l'autre projet de l'Open Archive Initiative, conjoint à OAI-PMH : OAI-ORE (*Open Archives Initiative Object Reuse and Exchange*) ([62](#)). Ce protocole définit un standard pour la description et l'échange de ressources web complexes (exemple : une thèse est un tout composé d'un ensemble de chapitres, elle peut avoir différentes versions, être suivie d'articles, etc.). Tandis qu'OAI-PMH est centré sur les métadonnées, OAI-ORE prend en compte les ressources (les objets de la description). Il s'agit d'une nouvelle manière d'enrichir et d'utiliser les métadonnées décrivant les documents conservés dans des entrepôts OAI en vue de leur réutilisation. Les logiciels de dépôt d'archives Dspace et Eprints ([63](#)) supportent ce nouveau protocole.

En définitive, malgré les standards qui permettent l'agrégation des données, celle-ci ne se fait pas sans adaptations ni ajustements. En effet, il reste souvent un travail d'homogénéisation

des données à effectuer, car la norme laisse une certaine marge de manœuvre au niveau des métadonnées descriptives. En particulier, le Dublin Core n'est pas interprété partout de la même manière. En effet, la « non-qualification » du Dublin Core, si elle laisse une certaine souplesse et une marge de manœuvre dans son utilisation, entraîne de légères disparités dans les métadonnées. Il en résulte un travail d'uniformisation qui peut être important en cas de nombreux dépôts à centraliser. Dans ces cas aussi, il faut veiller à la fusion des termes d'indexation et de la classification des différents dépôts en un thésaurus et un classement communs. Pour pallier à ce genre d'obstacles, Driver propose un guide très précis aux dépôts souhaitant être présents sur le portail. Ces instructions doivent obligatoirement être suivies pour que leur participation soit validée et leur données recherchables sur le portail. En contrepartie, Driver leur offre soutien et retours d'expériences sous forme de mentorat ([64](#)).

## Communication

La communication est un aspect crucial lors de la création d'un dépôt d'archives ouvertes ou d'un portail centralisateur de ressources. Dès le départ, la communication entre partenaires de projets est déterminante du bon déroulement du projet. Ensuite, une fois le dépôt créé, il s'agit non seulement de convaincre les chercheurs de déposer leurs publications, mais aussi de sensibiliser les instances dirigeantes à l'importance de l'auto-archivage et à l'intérêt de figurer dans un portail centralisateur. La visibilité est un enjeu à mettre en avant et à développer.

### Avec les partenaires

Le travail de communication commence donc au sein de l'équipe de projet. Autour d'un noyau dur de professionnels, il s'agit de mettre en place une véritable dynamique de collaboration et de partenariat autour d'objectifs communs. Le partage du savoir-faire est évidemment bénéfique à chacun, et une bonne répartition des tâches en fonction des compétences permet à chacun de s'impliquer activement. Pour les dépôts institutionnels, il est important d'obtenir une masse critique de contenu pour donner l'impulsion aux autres dépositaires potentiels. Pour les portails centralisateurs, une fois l'outil fonctionnel, le noyau dur de départ peut être étendu à d'autres partenaires.

### Avec la hiérarchie

Tout au long du projet, la communication avec les instances de tutelles des institutions est nécessaire, car ce sont elles qui peuvent inciter, voire éventuellement imposer le dépôt obligatoire des publications. Ce soutien « politique » est nécessaire au bon fonctionnement des dépôts. Quant aux portails centralisateurs, il est important de souligner aux instances dirigeantes des partenaires de projets que leur participation constitue un gain en termes de visibilité et de promotion de la recherche.

### Avec les usagers

La communication entre partenaires et avec les instances décisionnelles est certes importante, mais elle l'est aussi vis-à-vis du public. Il s'agit de promouvoir l'outil mis en place auprès des chercheurs et des usagers. Les chercheurs doivent être formés à l'utilisation du dépôt, et une aide à la saisie ainsi qu'une interface simple et conviviale sont nécessaires à leur appropriation de l'outil. Les arguments de promotion sont la visibilité offerte aux chercheurs et la valeur ajoutée par la centralisation des ressources et la recherche fédérée pour le public. Dans

l'environnement actuel, les chercheurs attendent une recherche aisée et efficace, et nous nous devons de leur permettre d'être autonomes en leur offrant des interfaces simples où ils peuvent être autonomes dans leur dépôt d'une part, et obtenir des contenus directement utilisables d'autre part. Il s'agit de mettre à leur disposition non seulement un support technique, organisationnel et institutionnel, mais aussi de leur prodiguer information et formation ainsi que différents services associés à valeur ajoutée : saisie facilitée des références, réutilisation aisée des données, création automatique de listes bibliographiques, interopérabilité avec d'autres systèmes et analyse de l'impact. Infoscience ([65](#)) ou l'Archive ouverte de l'Université de Genève ([66](#)) proposent ainsi à leurs chercheurs une rubrique d'information sur l'Open Access, une aide au dépôt ainsi que des services à valeur ajoutée tels que l'extraction de listes bibliographiques re-exploitable, des flux RSS personnalisés, etc. L'étude menée par Driver ([67](#)) relève l'importance de ces services à valeur ajoutée dans le succès des dépôts observés.

L'effort de « marketing » est aussi important vis-à-vis du large public d'Internet. Il y a donc un vrai travail de référencement et de communication dans différents médias pour toucher le maximum de public. Exposer soi-même ses métadonnées pour se rendre moissonnable à un plus haut niveau est une suite logique au moissonnage des dépôts institutionnels. Il est important que les dépôts moissonnés par de plus grands portails reçoivent un feedback quant à leur interopérabilité et à la qualité des métadonnées qu'ils exposent. Collaborer dans des projets d'ampleur internationale leur permet aussi de gagner en visibilité. Des projets pionniers tels que Driver se veulent multidisciplinaires et collaborent volontiers notamment en partageant leur savoir-faire ([68](#)).

## CONCLUSION

Inscrite dans le cadre du projet de portail Info-Net Economy, dont le cœur sera constitué d'un dépôt de métadonnées OAI-PMH, la présente étude met en évidence les bonnes pratiques à observer lors de la création de portails centralisateurs de ressources. De tels portails sont aujourd'hui nécessaires à la recherche, car les sources sont dispersées et peu valorisées. Le contenu de ces portails étant notamment puisé dans des dépôts institutionnels d'archives ouvertes, nous nous devons d'en explorer les tenants et les aboutissants.

Premièrement, on observe que le mouvement d'Open Access a enclenché une véritable révolution dans le monde de la publication scientifique qui permet un meilleur échange des savoirs dans les milieux académiques. Le succès des dépôts institutionnels, des dépôts centralisateurs de ressources et des portails thématiques repose en partie sur le mode de fonctionnement adopté et l'équipe de professionnels en charge du projet. Il reste cependant encore un gros travail d'information à effectuer auprès des universités pour sensibiliser les directions et les chercheurs. Communiquer et informer sur les enjeux liés aux dépôts institutionnels est donc la première bonne pratique à retenir. Ce rôle peut parfaitement être endossé par les bibliothèques ; cela représente même une véritable opportunité dans l'environnement évolutif de l'information scientifique. Aider les instituts et leurs chercheurs à valoriser leur travail sur le web est une nouvelle mission des bibliothèques académiques. Pour cela elles peuvent s'organiser en partenariats pour échanger les compétences et constituer des réseaux collaboratifs. Au niveau des universités, les bibliothécaires peuvent mettre à profit leurs compétences documentaires dans la création de dépôts institutionnels (comme dans le cas d'Infoscience à l'EPFL) ou proposer aux chercheurs de déposer leurs publications dans un dépôt centralisateur (tel que RERO-DOC en Suisse romande). Une fois le système de dépôt

mis en place, la participation à des portails thématiques tels que Driver est souhaitable car elle augmente la visibilité potentielle des institutions productrices.

Puis, en termes de contenu, la bonne pratique à observer est l'établissement d'une politique claire qui définit ce qui entre dans le cadre du dépôt ou du portail. Techniquement, la centralisation des ressources provenant de différents dépôts est rendue possible par la norme OAI-PMH, elle-même basée sur la norme de description Dublin Core. Le moissonnage des métadonnées permet d'offrir ensuite une recherche fédérée sur tous les dépôts exposés. Ce procédé permet de rendre plus visible des dépôts peu connus, par exemple sur un portail thématique tel qu'Info-Net Economy, qui regroupera les ressources de différents acteurs suisses qui publient de l'information scientifique dans le domaine de l'économie. Appliquer les normes et les outils permettant l'harmonisation et l'échange des métadonnées est la troisième bonne pratique à observer.

Enfin, pour faire l'unanimité, les dépôts et les portails qui les moissonnent se doivent de mettre tous les atouts de leur côté en offrant des interfaces ergonomiques et riches en fonctionnalités à valeur ajoutée. Offrir un cadre de recherche convivial et des services à valeur ajoutée en profitant d'outils novateurs est la quatrième bonne pratique mise en évidence.

En conclusion, on peut souligner l'importance des protocoles et normes d'échanges des données, mais aussi relever que l'organisation et la communication liées aux projets sont tout aussi cruciales pour leur réussite. Réussite que l'on peine encore à mesurer, par manque d'indicateurs éprouvés (69). Si la démocratisation de l'Open Access et le nombre croissant de projets qui voient le jour est un succès en soi, ils doivent être poursuivis et enrichis selon des critères de qualité reconnus dans le milieu de l'information documentaire et de la recherche académique.



## NOTES

- (1) Info-Net Economy [en ligne]. [http://www.e-lib.ch/net\\_economy\\_f.html](http://www.e-lib.ch/net_economy_f.html)
- (2) La bibliothèque électronique suisse, portail pour la communauté scientifique nationale qui proposera un accès centralisé à une vaste offre d'informations scientifiques et de services de bibliothèques dans un point d'accès unique.  
E-lib [en ligne]. [http://www.e-lib.ch/index\\_f.html](http://www.e-lib.ch/index_f.html)
- (3) ARESO : annuaire de ressources économiques de Suisse occidentale créé par la filière Information documentaire de la Haute Ecole de Gestion de Genève en 2008.  
ARESO [en ligne]. <http://campus.hesge.ch/areso/>
- (4) Définition de l'association des bibliothèques canadiennes de recherche, ABCR
- (5) Voir à ce sujet le dossier proposé par l'Université de Genève :  
<http://www.unige.ch/biblio/chercher/openaccess/brochure.pdf>
- (6) Pour plus d'informations et une bibliographie complète sur le sujet des publications en Open Access, voir le site de Charles W. Bailey Jr : <http://www.digital-scholarship.org/>
- (7) Bosc, Hélène. Archives ouvertes : 15 ans d'histoire. Les archives ouvertes : enjeux et pratiques. Paris, ADBS, 2005. pp. 27-54.
- (8) Infoscience [en ligne]. <http://infoscience.epfl.ch/>
- (9) Ce classement a pour but de soutenir les projets d'archives ouvertes en les évaluant selon leur visibilité et leur impact. Ranking web of world repositories [en ligne].  
<http://repositories.webometrics.info/>
- (10) Forschungsplattform Alexandria [en ligne]. <http://www.alexandria.unisg.ch/>
- (11) RERO-DOC[en ligne]. <http://doc.rero.ch/?ln=fr>
- (12) Driver [en ligne]. <http://www.driver-repository.eu/>
- (13) Econbiz [en ligne]. <http://www.econbiz.de/>
- (14) Economists online [en ligne]. <http://www.nereus4economics.info/>
- (15) Muriel Foulonneau, Réseaux d'archives institutionnelles en Europe : logiques de développement et convergences, Archive Ouverte en Sciences de l'Information et de la Communication [en ligne], 2007, [http://archivesic.ccsd.cnrs.fr/sic\\_00205049/en/](http://archivesic.ccsd.cnrs.fr/sic_00205049/en/)
- (16) Repec [en ligne]. <http://repec.org/>
- (17) E-LIS [en ligne]. <http://eprints.rclis.org/>
- (18) HAL [en ligne]. <http://hal.archives-ouvertes.fr/>
- (19) Lynch, Clifford A., Lippincott, Joan K., Institutional Repository Deployment in the United States as of Early 2005. *D-Lib Magazine* [en ligne], vol. 11, no 9, 2005.  
<http://www.dlib.org/dlib/september05/lynch/09lynch.html>
- (20) Op. cit.\_
- (21) Andre, Francis, Charnay, Daniel, *Support of Open Archives at National Level : the HAL experience*, Institutional archives for research : experiences and programs in open access

[en ligne], Rome 30 November – 1 December 2006,

[http://archivesic.ccsd.cnrs.fr/docs/00/18/72/60/PDF/Rome\\_ISS\\_FANDRE.pdf](http://archivesic.ccsd.cnrs.fr/docs/00/18/72/60/PDF/Rome_ISS_FANDRE.pdf)

(22) Ibid.

(23) Chen, Kuang-hua, The unique approach to institutional repository. *The electronic library* [en ligne], vol. 27, no 2, 2009, pp. 204-221,

<http://www.emeraldinsight.com/Insight/viewPDF.jsp?contentType=Article&Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2630270201.pdf>

(24) Pfister, Joachim, Weinhold, Thomas, Zimmermann, Hans-Dieter. *Open Access in der Schweiz : status quo und geplanter Aktivitäten im Bereich von Institutional Repositories bei Hochschul- und Forschungs-einrichtungen in der Schweiz*. In *Information : Droge, Ware oder Commons ? : Wertschöpfungs – und Transformationsprozesse auf den Informationsmärkten* [en ligne]. Boizenburg : Werner Hülsbusch, 2009. Pp. 259-270

(25) Op. cit.

(26) Vezina, Kumiko. Dépôts institutionnels : principaux enjeux [en ligne]. 38ème congrès annuel – perspectives d'avenir : fonctions, réseaux et relations, Corporation des bibliothécaires professionnels du Québec, mai 2007

<http://www.cbpcq.qc.ca/congres/congres2007/Actes/Vezina.pdf>

(27) Blin, Frédéric, Les bibliothèques académiques européennes : Brève synthèse prospective, *BBF* [en ligne], t. 53, no 1, 2008, pp. 12-18, <http://bbf.enssib.fr/consulter/bbf-2008-01-0012-002>

(28) Ibid.

(29) Jones, Catherine. *Institutional repositories: content and culture in an Open Access environment*. Oxford : Chandos Publishing, 2007.

(30) *The research library's role in digital repository services* [en ligne], Association of research libraries, 2009, <http://www.arl.org/bm~doc/repository-services-report.pdf>

(31) Salo, Dorothea. Innkeeper at the roach motel. *Library Trends* [en ligne], vol. 57. No 2, 2008. <http://minds.wisconsin.edu/handle/1793/22088>

(32) University of Nottingham (UK). *OpenDOAR : the Directory of Open Access Repositories* [en ligne]. <http://www.opendoar.org/>

(33) *SHERPA* [en ligne]. <http://www.sherpa.ac.uk/index.html>

(34) *SHERPA-RoMEO* [en ligne]. <http://www.sherpa.ac.uk/romeo/>

(35) Harnad, Steven, Brody, Tim. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine* [en ligne], vol. 10, no 6, 2004. <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

(36) Op. cit.

(37) Op. cit.

(38) Op. cit.

(39) *Archives Ouvertes UNIGE* [en ligne].

<http://www.unige.ch/biblio/chercher/archiveouverte.html>

- (40) Op. cit.
- (41) Op. cit.
- (42) INSPIRE [en ligne]. <http://www.projecthepinpire.net/>
- (43) Brooks, Travis. Giving researchers what they want: SPIRES, high energy physics and subject repositories. Genève, OAI6, juin 2009.  
<http://indico.cern.ch/getFile.py/access?contribId=20&sessionId=6&resId=1&materialId=slides&confId=48321>
- (44) Op. cit.
- (45) *Registry of Open Access Repositories (ROAR)* [en ligne].  
<http://roar.eprints.org/index.php>
- (46) *Dspace* [en ligne]. <http://www.dspace.org/>
- (47) *Eprints* [en ligne]. <http://www.eprints.org/>
- (48) *CDS-Ware/Invenio* [en ligne]. <http://cdsware.cern.ch/invenio/index.html>
- (49) Jones, Catherine. *Institutional repositories: content and culture in an Open Access environment*. Oxford : Chandos Publishing, 2007.
- (50) François Nawrocki, Le protocole OAI et ses usages en bibliothèque [en ligne], Paris, Ministère de la culture et de la communication, 2005.  
<http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>
- (51) *RODIN* [en ligne]. [http://www.e-lib.ch/rodin\\_f.html](http://www.e-lib.ch/rodin_f.html)
- (52) François Nawrocki, *op. cit.*
- (53) *OAster* [en ligne]. <http://www.oaister.org/>
- (54) *Citebase search* [en ligne]. <http://www.citebase.org/>
- (55) *OAI Harvester 2* [en ligne]. <http://www.oclc.org/research/Software/oai/harvester2.htm>
- (56) *Open archive harvester* [en ligne]. <http://pkp.sfu.ca/?q=harvester>
- (57) Sévigny, Martin, Pichot, Malo. *SDX et la moisson OAI* [en ligne]. Paris, Ministère de la culture et de la communication, 2005. <http://www.nongnu.org/sdx/docs/html/doc-sdx2/fr/oai/moissonneur.html>
- (58) *Google scholar* [en ligne]. <http://scholar.google.com/intl/fr/scholar/about.html>
- (59) Boutros, Nader. *Moissonnage des données : exposée général sur les principes*. Carrefour des acteurs de l'édition en archéologie, 1er décembre 2005.
- (60) *MARC Standards* [en ligne]. <http://www.loc.gov/marc/>
- (61) Jones, Catherine. *Institutional repositories: content and culture in an Open Access environment*. Oxford : Chandos Publishing, 2007.
- (62) *OAI-ORE* [en ligne]. <http://www.openarchives.org/ore/>
- (63) Op. cit.

(64) Feijen, Martin, et al. DRIVER : building the network for accessing digital repositories across Europe [en ligne]. *Ariadne*. Issue 53, October 2007.

<http://www.ariadne.ac.uk/issue53/feijen-et-al/>

(65) Op. cit.

(66) Op. cit.

(67) *A DRIVER's Guide to European Repositories : Five studies of important Digital Repository related issues and good Practices* [en ligne], 2007, University of Amsterdam,

<http://dare.uva.nl/aup/nl/record/260224>

(68) *Driver support website* [en ligne]. <http://www.driver-support.eu/index.html>

(69) Bosc, Hélène, Archives Ouvertes : quinze ans d'histoire, In *Les Archives Ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information* [en ligne], C. Aubry, J. Janik (eds.), Paris : ADBS, 2005. pp 27-54 <http://cogprints.org/4408/2/Ouvragearchive.htm>