

OAI-PMH FOR DUMMIES

Comment mettre en place un dépôt institutionnel avec des ressources limitées ?

Patrice X. CHALON

Knowledge Manager, Centre Fédéral d'Expertise des Soins de Santé (KCE)

Luc HOURLAY

Bibliothécaire/Webmaster, Centre Fédéral d'Expertise des Soins de Santé (KCE)

En 2006, le Centre Fédéral d'Expertise des Soins de Santé (KCE), institution fédérale belge nouvellement créée, a évalué l'opportunité de mettre en place un dépôt institutionnel conforme au protocole OAI-PMH. Compte tenu des ressources limitées, mettre en place un logiciel spécifique n'a pas été l'option retenue ; l'utilisation du processus de description des rapports dans le système intégré de gestion de bibliothèque (SIGB) a été identifiée comme l'option offrant le meilleur rapport coût/bénéfice. En 2010, PMB, le SIGB libre utilisé au KCE, a acquis la fonction de serveur OAI-PMH. Cela a permis au KCE d'exposer ses rapports aux agrégateurs de métadonnées, mais également d'envisager le développement de ses propres agrégateurs pour des domaines spécifiques. Dans cet article, nous décrivons le processus ayant abouti à la création d'un dépôt institutionnel au départ d'un SIGB. Nous discutons également le défi que représente la mise en place d'agrégateurs.

In 2006 evalueerde het Federaal Kenniscentrum voor de Gezondheidszorg (KCE), een recent opgerichte federale instelling, de mogelijkheid om een institutioneel depot te ontwikkelen conform aan de OAI-PMH interoperabiliteitsstandaard. Rekening houdend met de beperkte bronnen werd het opstellen van een specifieke software uiteindelijk niet weerhouden; het gebruik van het beschrijvingsproces van de verslagen binnen het geïntegreerd beheer van het bibliotheekstelsel (in het Frans SIGB genaamd) had de beste kosten-baten verhouding. In 2010 verkreeg het open-source bibliotheekstelsel van het KCE de serverfunctie OAI-PMH; dit stond het KCE toe om haar rapporten open te stellen aan de aggregatoren van metadata, maar eveneens om te overwegen haar eigen aggregatoren voor specifieke domeinen te ontwikkelen. In dit artikel beschrijven we het proces dat tot de ontwikkeling leidde van een institutioneel depot vanuit een SIGB. Eveneens wordt de uitdaging besproken dat de opstelling van aggregatoren met zich meebrengt.

Dépôts institutionnels et OAI-PMH

L'initiative pour les archives ouvertes ("Open Archives Initiative" - OAI) trouve ses racines dans les mouvements "Open Access" et de dépôts institutionnels¹. Un dépôt institutionnel peut contenir une grande variété de documents numériques produits par une institution : articles soumis pour publication ("pre-prints"), articles acceptés pour publication ("post-prints"), actes de col-

loque, rapports, thèses de doctorats et mémoires, ensembles de données ("datasets") résultant de projets de recherche, etc.².

L'OAI conduit plusieurs projets visant à promouvoir l'accès aux ressources numériques ; parmi eux, le développement et le maintien du protocole de collecte de métadonnées de l'initiative pour les archives ouvertes³ ("Open Archives Initiative Protocol for Metadata Harvesting" - OAI-

Tab. 1 : Projets de dépôts institutionnels des institutions fédérales belges.

Aucun projet	Intérêt	Intention	Projet	En production
SPF Emploi, Travail et Concertation sociale	SPF Économie, PME, Classes moyennes et Énergie	SPP Politique scientifique	Ministère de la Défense	KCE
SPF Finances	SPF Personnel et Organisation	Archives générales du Royaume	Institut royal des Sciences naturelles de Belgique	
SPF Justice	Musées royaux d'Art et d'Histoire			
CINEMATEK	Institut National d'Assurance Maladie-Invalidité (INAMI)			
Jardin botanique national de Belgique				

PMH).

OAI-PMH^{4,5} est un protocole qui permet l'interopérabilité entre les dépôts de documents numériques via Internet. Grâce à ce protocole, une première entité – le fournisseur de données ("Data Provider") – expose des métadonnées structurées via OAI-PMH ; et une seconde entité – le fournisseur de services ("Service Provider") – effectue des requêtes de service OAI-PMH pour collecter ("to harvest: moissonner") ces métadonnées et les intégrer à l'index de ses propres bibliothèques numériques, créant ainsi des portails (ou agrégateurs).

En Belgique, de nombreuses institutions académiques et de recherche ont déjà mis en place leur dépôt institutionnel^{6,7}. Au niveau européen et international, des agrégateurs proposent des portails génériques, comme *DRIVER*⁸ ou *OAIster*⁹, ou thématiques, comme *REPEC*¹⁰ dans le domaine de l'économie ou *OpenAIRE*¹¹ qui regroupe les articles en open access résultant de recherches financées par le 7^e programme-cadre européen.

Au niveau des institutions fédérales belges, une

Le KCE étant une institution récente, elle attache une très grande importance à la dissémination des résultats de ces études financées par des fonds publics. Ses analyses se basant sur les résultats d'études réalisées par d'autres institutions, mettre à disposition les rapports du KCE dans un dépôt institutionnel entre donc dans cette stratégie de dissémination.

Mettre en place un dépôt institutionnel

En 2006, le KCE a donc évalué la possibilité de mettre en place un dépôt institutionnel.

Dans ce type de projet, les premières questions portent sur les choix logiciels et de nombreuses solutions compatibles avec le standard OAI-PMH existent actuellement pour mettre en place un dépôt institutionnel, y compris en open source : utilisation d'un logiciel de gestion de ressources numériques¹⁴ (ou digital asset management system - DAMS), d'un logiciel de gestion de bibliothèque (système intégré de gestion de bibliothèque - SIGB) ou, depuis 2007, d'un logiciel de

Tab. 2 : Exemples de solutions logicielles open source permettant la mise en place d'un dépôt institutionnel.

Logiciels de gestion de ressources numériques	Logiciels de gestion de bibliothèque	Logiciels de gestion de contenu Web
<i>DSPACE*</i>	<i>Koha</i>	<i>Drupal</i>
<i>GNU E-Prints*</i>	<i>EverGreen</i>	<i>Plone/Zope</i>
<i>FedoraCommons</i>	<i>NewGenLib</i>	
<i>GreenStone*</i>	<i>Omeka</i>	
	<i>PMB*</i>	

Tableau mis à jour en 2011, les logiciels identifiés en 2006 sont marqué d'une *.

enquête¹² menée en avril 2011 auprès des responsables de bibliothèques nous a montré que, si plusieurs institutions sont intéressées et si deux d'entre elles ont un projet en cours, seul le KCE possède actuellement un dépôt institutionnel en activité.

Le KCE

Le Centre Fédéral d'Expertise des Soins de Santé (KCE) est un organisme fédéral parastatal qui a pour mission de produire des analyses et des rapports pour documenter les pouvoirs publics dans leurs prises de décision en matière de politiques de santé et d'assurance-maladie ; les rapports produits sont disponibles intégralement sur le site Web du KCE¹³.

gestion de contenu Web (Web content management system - WCMS).

Cependant, l'aspect humain s'avère de loin le plus critique : changement de culture, d'organisation, d'habitudes... comptent en effet pour 90% du challenge¹⁵.

Un dépôt institutionnel au KCE

Évaluation des options

Mettre en œuvre un logiciel spécifique au KCE nécessiterait l'ajout d'un processus d'encodage de métadonnées à ceux existant déjà. Dans un contexte de ressources limitées, tant humaines que budgétaires, cette option n'est pas idéale.

Par contre, utiliser un logiciel déjà en place permettrait de réutiliser un encodage existant. Nous avons donc listé les processus de description de nos rapports et évalué les compatibilités au standard OAI-PMH des logiciels utilisés à cet effet.

Aucune option ne pouvait être mise en œuvre directement, mais la plus prometteuse était celle offerte par le logiciel de gestion de bibliothèque. En effet, les rapports y sont déjà encodés par un spécialiste de l'information – gage de métadonnées de qualité – et le format de description bibliographique va même au-delà des exigences des dépôts institutionnels. De plus, il s'agit d'un logiciel géré par le KCE : activer et paramétrer des modules supplémentaires est dès lors aisé.

Par contre, aucune date n'était fixée pour la mise en place du module OAI, et le KCE n'avait pas la possibilité de financer le développement de ce module.

libre de gestion de bibliothèque utilisé au KCE¹⁷ a été pourvu de la fonction serveur OAI-PMH. L'option de mise en place d'un dépôt institutionnel au départ du logiciel de gestion de bibliothèque étant toujours la plus pertinente, le projet fut donc réactivé.

Cette mise en œuvre s'est avérée très simple. Elle se décompose en cinq étapes.

La première étape consiste à mettre en place des "sets" - sélection de documents parmi l'ensemble de ceux décrits dans le catalogue de la bibliothèque - qui seront exposés dans le dépôt institutionnel. Dans notre cas (Fig. 1), trois sets ont été créés sur base d'une requête se basant sur les notions de collection et de sous-collection déjà définies dans le SIGB. Les sets disponibles dans le dépôt institutionnel du KCE correspondent aux trois types de rapports produits par l'institution : GCP pour "Good Clinical Practice" (Guides de bonne pratique clinique), HSR pour "Health Servi-

Tab. 3 : Processus de publications des rapports du KCE.

Outil	Module OAI-PMH
Système de gestion de contenu (CMS)	Non *
Système de gestion intégrée de bibliothèque (SIGB)	Sur la feuille de route dès 2006
Dépôt légal électronique (e-depot)	Disponible mais pas mis en place**
Catalogue collectif fédéral (bib.belgium.be)	Disponible mais pas mis en place**

* Le CMS du KCE a changé en octobre 2011.

** Évalué entre 2006 et 2010, le module existe mais n'est pas mis en place, car il ne correspond pas à la mission de l'outil.

Compte tenu de cette analyse, le projet fut mis en attente. Une veille fut maintenue afin d'identifier les outils susceptibles de répondre au besoin du KCE, et les opportunités de participer à des projets de dépôt institutionnel.

Mise en place

En 2010, la version 3.4 de *PMB*¹⁶, le logiciel

ces Research" (recherche en organisation des soins de santé), et HTA pour "Health Technology Assessment" (évaluation des technologies de santé).

La seconde étape consiste à paramétrer le serveur OAI-PMH. En premier lieu, le dépôt se voit attribuer une description (Fig. 2a), et les sets à exposer sont sélectionnés (Fig. 2b). En second

The screenshot shows a web form for creating a set. The 'Name of the set' field contains 'KCE Reports HTA'. The 'Type' is 'Notice advanced search'. A section titled 'Currently contains the following search:' shows a series of Boolean expressions: 'Series Boolean expression (KCE Reports) And Sub-series Boolean expression (Health Technology Assesment (HTA))'. Below this, there is an 'Edit this search' section with radio buttons for 'Keep the current search' (selected) and 'Copy from a SDI equation' with several options like 'Achat = en traitement', 'Nouveauté_3', 'test_template', 'Achat = sans statut particulier2', and 'HTA Databases'. At the bottom, there is a 'Cache duration' field set to '0 seconds'.

Fig. 1 : Fenêtre de création de set par requête (exemple du set « HTA »).

lieu, il faut autoriser les utilisateurs anonymes (les "Service Providers") à utiliser le connecteur sortant défini comme dépôt institutionnel (Fig. 3).

Ces deux premières étapes sont réalisées en une à deux heures.

La troisième étape consiste à autoriser la réécriture d'URL au niveau du serveur Web hébergeant le SIGB^{18,19}. Notre catalogue étant hébergé chez PMB Services, cette configuration était déjà en place.

La quatrième étape consiste à rendre le dépôt institutionnel visible. En premier lieu, l'information permettant de moissonner le dépôt institutionnel est fournie dans la page d'aide de l'OPAC²⁰. En second lieu, le dépôt est enregistré au niveau des répertoires d'archives ouvertes *OpenDOAR* et *ROAR*. Cet enregistrement permet d'être automatiquement listé dans d'autres répertoires tels que *Repository 66*²¹. En troisième lieu, le dépôt est renseigné auprès des agrégateurs pertinents pour le domaine d'activité de l'institution tels que *DRIVER* et *OAIster*.

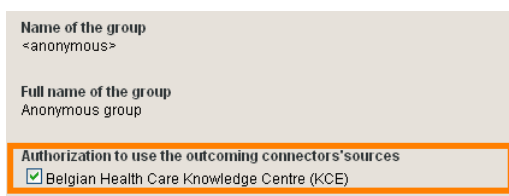


Fig. 3 : Fenêtre d'autorisation du connecteur sortant.

Cette étape nécessite moins d'une heure de travail par cible, mais s'étale sur deux à trois jours compte tenu du processus d'inscription et de validation auprès de chaque répertoire ou agrégateur. Pour ces derniers, la mise à jour de la description des documents de l'institution se fera désormais automatiquement, par synchronisation des métadonnées au rythme défini par l'agrégateur. Le KCE ne doit plus contacter les portails pour voir la description de ses rapports y

apparaître.

La cinquième étape consiste à mettre à jour le dépôt institutionnel. Ceci se fait en 4 clics seulement : 3 pour arriver à la bonne page d'administration (Fig. 4), et un pour mettre à jour le set dans lequel le document préalablement décrit dans le catalogue de bibliothèque doit être ajouté.

La mise en place d'un dépôt institutionnel basé

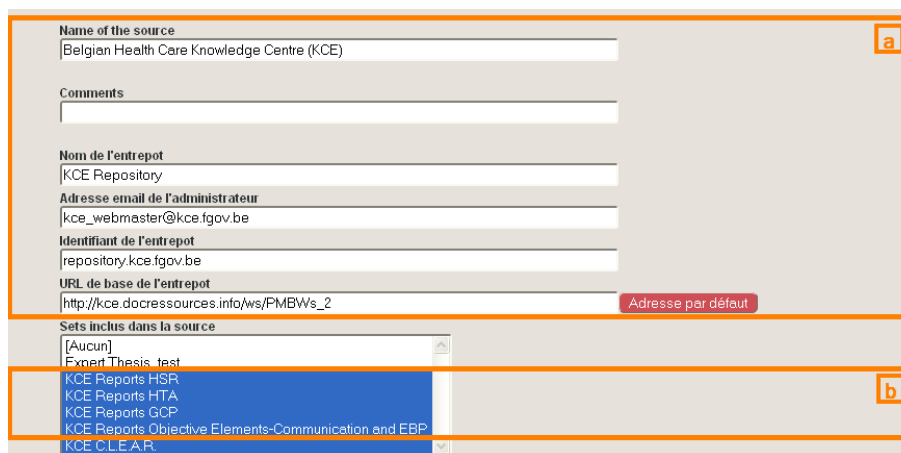


Fig. 2a et b : Fenêtre de paramétrage du serveur OAI-PMH.

sur le logiciel de gestion de bibliothèque s'est avérée une stratégie très pertinente. En effet, avec un investissement très limité, le KCE dispose désormais d'un dépôt institutionnel qui permet de mettre à disposition des agrégateurs la description de ses rapports, y compris le lien vers le texte intégral.

Cette première étape ayant été réalisée avec une grande facilité, le second rôle OAI, "service provider", fut donc envisagé dans le cadre d'un projet européen *EUnetHTA*²². où le KCE est partenaire :

Mettre en place un agrégateur de dépôts institutionnels

En effet, *EunetHTA* regroupe plus de trente agences européennes actives dans l'évaluation de technologies de santé, avec pour objectif de favoriser la collaboration de ces agences pour la production de rapports HTA. L'échange d'information est donc primordial et l'automatisation de ces échanges, réduisant le recours à des ressources

*NOTICE ADVANCED SEARCH				
Name of the set	Type	Information	Latest cache update	Manual update
KCE Reports HSR	Notice advanced search	Currently include 167 notices	06/14/2011 07:16:14	Manual update
KCE Reports HTA	Notice advanced search	Currently include 148 notices	06/15/2011 03:43:16	Manual update
KCE Reports GCP	Notice advanced search	Currently include 98 notices	06/15/2011 03:43:16	Manual update

Fig. 4 : Fenêtre de mise à jour des sets du dépôt institutionnel.

humaines, un facteur de succès du projet.

La première étape de ce type de projet consiste en la sélection de sources de métadonnées (Data Providers). À l'heure actuelle, seules deux institutions partenaires placent leurs rapports dans des logiciels compatibles avec le standard OAI-PMH : le Ludwig Boltzman Institute for Health Technology Assessment (LBI-HTA), qui possède un dépôt sous *E-Print* depuis 2008²³, et le KCE qui possède un dépôt institutionnel basé sur son SIGB *PMB* depuis 2010.

La seconde étape consiste en la mise en place de l'agrégateur. Dans notre cas, nous avons utilisé le SIGB libre *PMB*, que nous connaissons bien, en installant une instance spécifique sur un hébergement mutualisé offrant à faible coût un espace web, des bases de données MySQL et l'interprétation du code PHP. Un connecteur OAI a été paramétré pour chaque cible (ici, LBI-HTA et KCE). À ce stade, l'OPAC de *PMB* peut déjà être utilisé comme interface de découverte.

Cependant, afin de proposer un outil doté d'une meilleure ergonomie, le module *PMB Drupal*²⁴ (actuellement au stade de développement alpha) a été ajouté au site *EUnetHTA News aggregator*, un site agrégeant des flux RSS. Le module *PMB Drupal* permet d'interroger un catalogue *PMB* depuis un site *Drupal* et d'y afficher les résultats, qu'il s'agisse des collections décrites dans le catalogue ou des données moissonnées par *PMB* (Fig. 5). Cette configuration permet d'envisager la mise en place d'un portail lorsque le module sera finalisé.

Mais plus qu'un portail, c'est finalement un démonstrateur qui a été mis en place. En effet, au

sein de *EUnetHTA* seules deux institutions placent actuellement leurs rapports dans des dépôts compatibles OAI-PMH. Ce démonstrateur est désormais utilisé pour montrer aux plus de trente autres partenaires, l'intérêt de placer leurs rapports dans des dépôts institutionnels compatibles avec la norme OAI-PMH et, par là, de "libérer" leurs métadonnées.

Conclusion et perspectives

Mettre en place d'un dépôt institutionnel est une démarche importante pour une institution produisant des rapports scientifiques qui ne donnent pas systématiquement lieu à des publications dans des revues peer-reviewed : elle permet d'ajouter aux résultats de la recherche financée par des fonds publics une voie de dissémination standardisée et accessible. Cependant, il faut constater que peu d'institutions publiques placent actuellement leurs rapports dans des dépôts institutionnels.

Il est donc important de montrer l'intérêt des dépôts institutionnels et les perspectives que cela apporte, mais aussi que la mise en place d'un dépôt institutionnel n'est pas une tâche forcément complexe ou coûteuse, que ce soit en terme de budget ou de ressources humaines.

En effet, si cette mise en place peut paraître complexe de prime abord, s'appuyer sur un processus déjà existant, et plus particulièrement sur le processus de description des ouvrages placés en bibliothèque s'avère une piste extrêmement simple à mettre en œuvre et d'un rapport coût-bénéfice favorable, et ceci d'autant plus que le

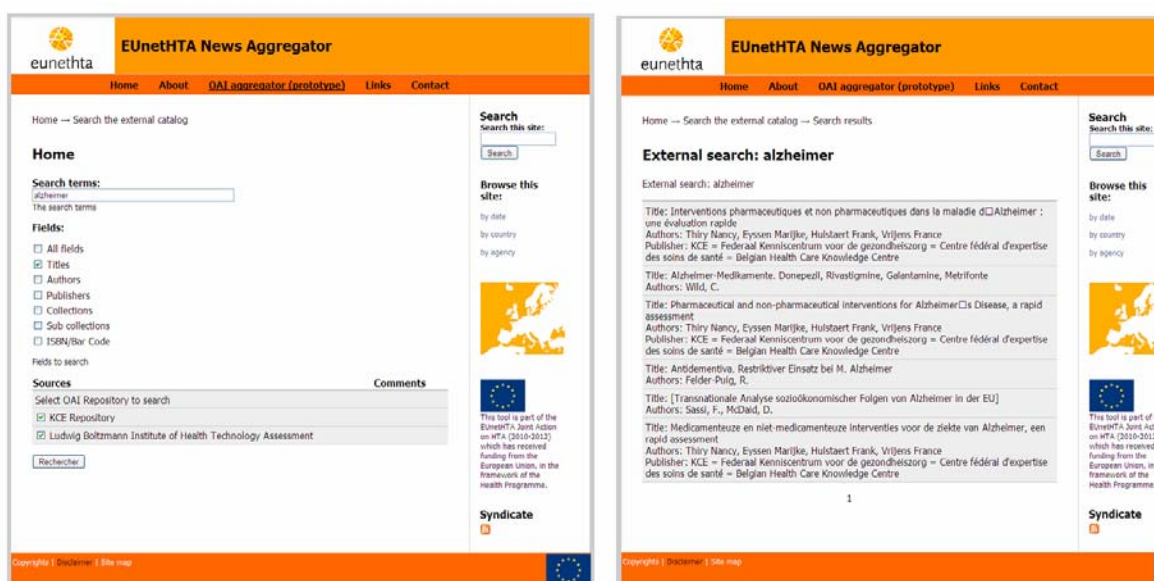


Fig. 5 : Portail de l'agrégateur HTA utilisant Drupal.

SIGB mis en œuvre est un logiciel libre.

Outre le rapport coût/bénéfice, le rapport coût/qualité est également très favorable puisque ce processus s'appuie sur l'encodage de métadonnées réalisé par des spécialistes de l'information rodés à cet exercice.

La mise en place d'un dépôt institutionnel dépend désormais davantage de la prise de conscience par l'institution du potentiel que cela représente. Mettre en place un dépôt institutionnel fournira à l'institution un retour sur investissement immédiat par la signalisation de ses rapports au niveau de portails existants. Mais cette mise en place posera également les bases pour la création de portails spécifiques utiles à ses

propres chercheurs et facilitant l'accès aux résultats de la recherche financée par des fonds publics.

**Patrice X. Chalon
Luc Hourlay**

*Centre Fédéral d'Expertise des Soins
de Santé (KCE)*
Doorbuilding (10th)
Boulevard du Jardin Botanique, 55
1000 Bruxelles
patrice.chalon@kce.fgov.be
luc.hourlay@kce.fgov.be

Juin 2011

La présentation faite par Patrice Chalon lors de l'*Inforum*, le 19 mai 2011, est disponible sur le site de l'ABD-BVD à l'adresse : <<http://www.abd-bvd.be/inforum/2001-Chalon-OAIPMH.pdf>>.

Notes

- 1 The Open Archives Forum (OAF). *2. History and development of OAI-PMH* [en ligne]. <<http://www.oaforum.org/tutorial/english/page2.htm>> (consulté le 13 juin 2011).
- 2 SPARC Europe. *Institutional repositories : A guide to open electronic archives* [en ligne]. <<http://www.sparceurope.org/resources/hot-topics/institutional-repositories>> (consulté le 13 juin 2011).
- 3 Open Archives Initiative. *Open Archives Initiative Protocol for Metadata Harvesting* [en ligne]. <<http://www.openarchives.org/pmh/>> (consulté le 13 juin 2011).
- 4 The Open Archives Forum (OAF). *3. Main Technical Ideas of OAI-PMH* [en ligne]. <<http://www.oaforum.org/tutorial/english/page3.htm>> (consulté le 13 juin 2011).
- 5 Bibliothèque nationale de France. *OAI-PMH* [en ligne]. <http://www.bnf.fr/fr/professionnels/protocoles_echange_donnees/a.proto_oai.html> (consulté le 13 juin 2011).
- 6 University of Nottingham. *Open Access Countries and Organisations - Europe - Belgium* [en ligne]. <<http://www.openoar.org/countrylist.php?cContinent=Europe#Belgium>> (consulté le 13 juin 2011).
- 7 University of Southampton. *Registry of Open Access Repositories: Browse by Country - Belgium* [en ligne]. <<http://roar.eprints.org/view/geoname/geoname=5F2=5FBE.html>> (consulté le 13 juin 2011).
- 8 DRIVER. *DRIVER Search Portal* [en ligne]. <<http://search.driver.research-infrastructures.eu/>> (consulté le 13 juin 2011).
- 9 WorldCat. *OAIster: Find the pearls* [en ligne]. <<http://oaister.worldcat.org/>> (consulté le 13 juin 2011).
- 10 Baum K, Zimmermann C. *RePEc: Research Papers in Economics* [en ligne]. <<http://repec.org/>> (consulté le 13 juin 2011).
- 11 OpenAIRE. *Objectives* [en ligne]. <<http://www.openaire.eu/en/about-openaire/general-information/objectives>> (consulté le 13 juin 2011).
- 12 Chalon P.X. OAI-PMH for dummies : comment mettre en place un dépôt institutionnel avec des ressources limitées ? Une étude de cas, INFORUM 2011 "I&D: for dummies???", Bruxelles (Belgique), 19 mai 2011 (consulté le 13 juin 2011) <<http://www.abd-bvd.be/inforum/2001-Chalon-OAIPMH.pdf>>
- 13 KCE. *À propos du KCE / KCE* [en ligne]. <<http://kce.fgov.be/fr/about-kce>> (consulté le 16 novembre 2011).
- 14 Sarwan N. *Open Source Digital Asset Management* [en ligne]. <<http://www.opensourcedigitalassetmanagement.org/>> (consulté le 13 juin 2011).

- 15 Jakobsson A. Establishing an Institutional Repository: A Step by Step Approach, 10th European Conference of Medical and Health Libraries, Cluj (Romania), 14 septembre 2006.
- 16 PMB Services. *PMB, un SIGB entièrement libre* [en ligne].
<http://www.pmbservices.fr/nouveau_site/pmblelogiciel.html> (consulté le 13 juin 2011).
- 17 Chalon P. X., Alexandre-Joaquim L., Naget- C., Becquart C. Open your mind! Selecting and implementing an integrated library system: the open-source opportunity, 10th European Conference of Medical and Health Libraries, Cluj (Romania), 11-16 septembre 2006 (consulté le 13 juin 2011) <<http://eprints.rclis.org/handle/10760/12135>>
- 18 PMB Services. *Déclarer une réécriture d'URL* [en ligne].
<http://doc.sigb.net/pmb/co/admin_connecteurs_out_3_reecriture_url.html> (consulté le 14 juin 2011).
- 19 The Apache Software Foundation. *URL Rewriting Guide* [en ligne].
<<http://httpd.apache.org/docs/2.0/misc/rewriteguide.html>> (consulté le 14 juin 2011).
- 20 BIBKCE. *Online catalogue - Help* [en ligne].
<<http://kce.docressources.info/opac/index.php?lvl=infopages&pagesid=5>> (consulté le 13 juin 2011).
- 21 Lewis S. *Repository maps* [en ligne]. <<http://maps.repository66.org/>> (consulté le 13 juin 2011).
- 22 EUnetHTA. *EUnetHTA startpage | EUnetHTA* [en ligne]. <<http://www.eunetha.eu/>> (consulté le 13 juin 2011).
- 23 Chalon P. X. Standards to enhance interoperability of information systems and efficiency of information exchange, 7th HTAi Annual Meeting, Dublin (Ireland), 6-9 juin 2010 (consulté le 16 novembre 2011)
<<http://www.slideshare.net/pchalon/htai-7-posterstandardstoehanceinteroperability>>
- 24 Tétart F. ftetart's sandbox: PMB Drupal | drupal.org [en ligne]. <<http://drupal.org/sandbox/pmbdrupal/1343446>> (consulté le 16 novembre 2011).