



Difficile convergence des archives ouvertes en SIC

Emma Bester, Ghislaine Chartron

► **To cite this version:**

Emma Bester, Ghislaine Chartron. Difficile convergence des archives ouvertes en SIC. 2010.
sic_00627729

HAL Id: sic_00627729

https://archivesic.ccsd.cnrs.fr/sic_00627729

Submitted on 29 Sep 2011

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.

Difficile convergence des archives ouvertes en SIC

Emma Bester* — Ghislaine Chartron**

* Doctorante, *Laboratoire Dispositifs d'information et de communication à l'ère numérique (DICEN), CNAM*

3 rue Conté, 75003 Paris

emma.bester@cnam.fr

** Professeur CNAM, *Laboratoire Dispositifs d'information et de communication à l'ère numérique (DICEN), CNAM*

3 rue Conté, 75003 Paris

RÉSUMÉ. L'article rend compte d'un test de moissonnage pour quatre archives ouvertes sélectionnées afin de déterminer la faisabilité et les limites de leur interconnexion pour un domaine scientifique particulier: le champ des SIC. Les problèmes rencontrés d'interopérabilité organisationnelle, technique et sémantique sont détaillés et contextualisés à un état de l'art sur la question. La question de la difficile représentation sémantique commune pour ce champ scientifique, manifeste dès l'étape de repérage et d'identification des archives ouvertes SIC sur le répertoire international OpenDOAR, est ensuite approfondie à partir de l'étude détaillée du champ « subject ». L'article conclut à la nécessité de repenser à la source les modes de production de ces dispositifs selon des procédures partagées au sein de la communauté scientifique et entre gestionnaires de ces archives.

ABSTRACT. The article reports a test of harvesting four selected open archives to determine the feasibility and limitations of their interconnection in a particular scientific field: the field of Information Science. It presents organizational, technical and semantic interoperability issues with a state of the art in that question. The difficult issue of common semantic representation for this scientific field, which is obvious from the step of locating and identifying Information Science open archives in the international Directory OpenDOAR, is then deepened with the detailed study of the field "subject". The article concludes with the need to rethink the modes of production of these devices in accordance with procedures shared within the scientific community and among managers of these archives.

MOTS-CLÉS : Archives ouvertes, Moissonnage, Interopérabilité, Interopérabilité sémantique, SIC, Science de l'information, Classification.

KEYWORDS: Open Archive, Harvesting, Interoperability, Semantic interoperability, Information science, Information and Communication Sciences, Classification.

1. Introduction, problématique

Nous considérons dans cette proposition un système particulier d'organisation des connaissances, à savoir les archives ouvertes d'articles scientifiques et nous nous intéressons au croisement possible de ces dispositifs dans le cadre de la diffusion et l'accès à ces savoirs. Les archives ouvertes sont des dispositifs de circulation des textes en croissance dans le cadre du mouvement du libre accès et des politiques de mandat obligatoire, qui se développent par exemple en Europe avec le projet OpenAire¹. Ils supportent la construction de dispositifs d'accès à l'information (recherche fédérée) par les fournisseurs de services OAI², et sont particulièrement attendus pour des champs disciplinaires donnés (Xia, 2007 ; Bherwani *et al.*, 2009 ; Puplett, 2010).

Dans le cadre de ces systèmes d'information, se posent des problèmes de référencement spécifiques, liés à la production de métadonnées. C'est la question de l'interopérabilité des métadonnées qui sera approfondie dans ce travail empirique. Nous nous appuyons sur la pratique du moissonnage d'archives ouvertes afin de déterminer la faisabilité et les limites de leur interconnexion pour un domaine scientifique particulier: le champ des SIC. Notre objectif est d'identifier les éventuelles spécificités du terrain des archives en SIC ou au contraire de montrer que ce sont les mêmes problèmes que l'on rencontre.

On peut faire l'hypothèse de problèmes spécifiques compte-tenu de la définition hétérogène du champ, manifeste dès l'étape de repérage et d'identification des archives ouvertes SIC : nous traitons d'abord du problème de l'interopérabilité mais la difficile représentation sémantique commune est redécouverte à cette occasion. Elle invite à s'interroger sur la pertinence de ces dispositifs comme sources documentaires pour des systèmes d'organisation des connaissances, au regard d'autres dispositifs comme les grandes bases de données Francis ou LISA par exemple.

Les travaux menés par A. Coleman et J. Roback (2005) lors de la mise en place du moissonneur thématique pour les sciences de l'information DL-Harvest³ pointent certaines difficultés, mais nous proposons de développer davantage la problématique liée à l'interopérabilité sémantique.

¹ <http://www.openaire.eu/>

² <http://www.openarchives.org/service/listproviders.html>

³ Le moissonneur thématique pour les sciences de l'information DL-Harvest, ouvert en mai 2005 par l'université d'Arizona, n'est plus en activité.

2. Identification et sélection d'archives ouvertes en science de l'information et de la communication (SIC)

2.1 SIC vs LIS

Pour identifier les dépôts OAI à moissonner dans le cadre de ce travail, nous nous appuyons sur le répertoire international OpenDoar où sont répertoriées 1500 archives ouvertes en janvier 2010, date de l'étude. Contrairement à son corollaire ROAR, il offre seul un accès thématique. Cependant, dès cette étape préalable à la constitution de l'échantillon pour l'étude envisagée, la question se pose de l'identification des SIC en tant que discipline scientifique telle qu'elle est instituée en France par le CNU (71^e section)⁴. OpenDoar ne propose pas de catégorie Sciences de l'information et de la communication (SIC) proprement dite, non plus que de catégories distinctes Science de l'information ou Communication : seule l'entrée Library and Information Science (LIS), relative plus spécifiquement à la bibliothéconomie, est proposée.

Dans un premier temps, il a été décidé de travailler à partir des 55 dépôts indexés LIS⁵, mais ce sont, à l'aune de cette approche ambivalente ou parcellaire, les questions de définition et de représentation des sciences de l'information (Fondin, 2001 ; Zins, 2005) qui surgissent d'entrée.

2.2 Les LIS, aux frontières des Computer and Information Science / Information Technology

L'étude des 55 dépôts indexés LIS aboutit à une grille d'analyse (tab. 1), support de l'échantillonnage (tab. 2). Outre l'instabilité de certains dépôts, on note des limites strictement liées aux pratiques documentaires d'indexation. La grille rend compte également de la grande diversité d'exploitation du système des archives ouvertes⁶ : typologies des dépôts et documentaires variées. Elle met par ailleurs en exergue la proximité des LIS avec les disciplines *Computer and Information Science* et/ou *Information Technology* pour quatre dépôts.

⁴ Section 71 du CNU

<<http://www.cpcnu.fr/documentsSection.htm?option=menuSection&numeroSection=71>>

⁵ <<http://www.opendoar.org>> OpenDOAR: 55 dépôts indexés LIS à la date de l'étude (janvier 2010). Ils sont 63 en janvier 2011.

⁶ Voir par exemple : Armbuster C., Romary L., « Comparing Repository Types - Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication », *working paper*, arXiv:1003.4187v1.

Difficile convergence des archives ouvertes en SIC

<i>Critères d'exclusion de l'échantillon</i>	
Non accessibilité (à la date de l'étude)	7 archives
Collections spécifiques (événement limité dans le temps, nature des objets/ressources ou des porteurs)	18
Pas de documents LIS identifiés sur les dépôts malgré l'indexation sur OpenDOAR	5
Masse critique insuffisante en LIS (moins de 250 dépôts)	9
<i>Critères de sélection pour l'échantillon</i>	
Archives ouvertes multidisciplinaires, dont LIS (clairement identifiés, nb de dépôts supérieur à 250)	3
Dépôts thématiques LIS	9
Dépôts LIS, <i>Computer and Information Science / Information Technology</i>	4

Tableau 1. Grille d'analyse des archives ouvertes indexées LIS dans le répertoire DOAR (55 dépôts en janvier 2010)

2.3 Disparité organisationnelle des archives ouvertes en science de l'information

Parmi les seize dépôts répondant aux critères de sélection pour l'échantillon (tab. 1), quatre ont été retenus pour le test de moissonnage (tab.2).

Archive ouverte	Type de dépôt	Porteur institutionnel	Nombre de documents (janvier 2010)
ArchiveSIC http://archivesic.ccsd.cnrs.fr	National ; thématique : SIC	CCSD (France)	1300
E-LIS http://eprints.rclis.org/	International ; thématique : LIS	AePIC (Italie) CILEA (Espagne)	11048
Bib-Info http://www.bib-info.de/index.html	Institutionnel ; thématique : LIS	Berufsverband Information Bibliothek (Allemagne)	670
QUT http://www.digitalrepositor.y.qut.edu.au/	Institutionnel ; pluridisciplinaire	QUT (Australie)	303 en LIS

Tableau. 2 Echantillon de 4 archives ouvertes en SIC à fin de test de moissonnage

Les critères de masse critique (plus de 250 documents), de fiabilité (porteurs), de diversité linguistique et géographique, et de représentativité ont servi la sélection présentée. Trois archives thématiques LIS, dont une institutionnelle (Bib-Info), une nationale (ArchiveSIC) et une internationale (E-LIS) ont été retenues. Le dernier dépôt, QUT, a été sélectionné car il présente à la fois les caractéristiques d'une archive pluridisciplinaire et celles d'un ancrage dans les *Computers and IT*.

Ces dépôts répondent de modes d'organisation divers :

- ArchiveSIC (Chartron *et al.* 2003) est une archive thématique à vocation majoritairement nationale où seuls les travaux scientifiques sont acceptés. Conçue en 2001 par des chercheurs de la discipline, qui continuent d'assurer le travail de suivi des dépôts (modération a minima), elle procède néanmoins d'un financement institutionnel (CCSD). ArchiveSIC est interrogeable depuis 2006 dans le portail national HAL, mais propose un plan de classement thématique, soit une liste de domaines définie par les concepteurs à l'origine du dispositif ;
- E-LIS (De Robio *et al.* 2005) a, *a contrario*, une vocation internationale et accepte une grande typologie documentaire (documents scientifiques, techniques, publiés ou non, bases de données primaires, etc). Développée en 2003 dans le cadre du projet RCLIS⁷, l'archive, aujourd'hui supportée financièrement ou techniquement par le consortium AePIC-CILEA et DuraSpace, s'appuie sur des bibliothécaires volontaires dans plus de trente pays pour assurer le suivi des dépôts (modération a minima). Le plan de classement hiérarchisé proposé, le JITA, procède d'une fusion entre les classifications du *NewsAgent Topics* et la *Review of Information Science* (RIS) réalisée par les concepteurs du dépôt, dont les initiales des prénoms forment le sigle JITA.
- Bib-Info est une archive institutionnelle, conçue et financée en interne pour l'archivage de conférences et de travaux n'ayant pas vocation à être publiés (« *unpublished* ») : le plan de classement interne reflète en ce sens des préoccupations strictement professionnelles liées à la bibliothéconomie ;
- Le dépôt institutionnel QUT, développé et financé par la *Queensland University Technology*, est un dépôt pluridisciplinaire de travaux scientifiques qui s'appuie sur un plan de classement généraliste national, l'*Australian and New Zealand Standard Research Classification*. Dans ce plan de classement, les LIS sont une sous-catégorie des *Information and Computing Sciences*⁸.

Cette étude préalable fait apparaître, avant même la mise en œuvre du moissonnage, des problématiques spécifiques au champ scientifique exploré, en termes de définition et de structuration. Des problématiques communes aux archives ouvertes sont également notables à ce stade : typologie et modalités organisationnelles des dépôts, qualité de l'indexation nécessaire à leur identification disciplinaire dans un répertoire international, instabilité des dépôts. Notons à propos

⁷ RCLIS : Research in Computing, Library and Informaion Science

⁸ <http://eprints.qut.edu.au/view/subjects/>

de ce dernier point que l'archive ouverte DLIST (*Digital Library of Information Science and Technology*), maintenue par l'université d'Arizona et comprenant 1481 documents à ce jour⁹, était inaccessible à la période de l'étude et n'a donc pu être inscrite dans l'échantillon.

3. Les problèmes identifiés pour le moissonnage d'archives ouvertes en sciences de l'information

Le test de moissonnage a été réalisé avec le logiciel PKP Harvester. Pour évaluer les conditions d'exploitation des items moissonnés, deux types d'observations ont été menées : d'une part chaque champ du formulaire de recherche proposé par le moissonneur a fait l'objet de requêtes tests, et d'autre part les résultats de requête ont été comparés à l'affichage (en liste simple, en notices pleines, et avec les notices des dépôts d'origine) puis analysés champ par champ.

Les difficultés identifiées à l'issue de ce test rencontrent d'autres retours d'expériences sur le moissonnage de données OAI-PMH (Lazoge, 2003 ; Coleman *et al.*, 2005 ; Merceur, 2006), les travaux d'évaluation de masse des usages du format Dublin Core non qualifié dans les archives ouvertes par les professionnels de l'information (Ward, 2004 ; Jordan, 2006 ; Efron, 2007 ; Foulonneau, 2007), et les retours d'usage d'une enquête qualitative menée auprès de 141 administrateurs de dépôts (OCLC : Childress *et al.*, 2009).

Sont constatés des freins à l'interopérabilité organisationnelle (i.o.), technique (i.t.) et sémantique (i.s)¹⁰ (Innocenti *et al.* 2010) à toutes les étapes :

- à l'étape de l'identification des dépôts par la requête *List ?verb=identifier* : les fiches d'identification des dépôts ne sont pas toujours complètes (i.o) ; les schémas xml sont parfois invalides (i.t.) (Lazoge, 2003 ; Efron, 2007) ;
- à l'étape du paramétrage du moissonneur, au moment de la sélection des sets : les sets varient d'un dépôt à l'autre, recouvrant des informations de nature et de granularité différentes (i.o) (Lazoge, 2003 ; Coleman *et al.*, 2005) ;
- à l'étape du moissonnage, certaines données ne sont pas collectées (i.t. entre le logiciel de moissonnage et les logiciels d'archives ouvertes) (Efron, 2007) ;
- à l'étape de l'évaluation des conditions d'exploitation des items moissonnés, on remarque : des doublons ou des fiches supprimées (i. o.) (Foulonneau, 2007 ; Merceur, 2007) ; une exploitation non homogène de la DTD Dublin Core non qualifiée (i. s) (Lazorge, 2003; Ward, 2004; Jordan, 2006 ; Efron, 2007) qui résulte

⁹ <http://dlist.sir.arizona.edu/arizona/handle/10150/105066> ; 1481 dépôt au 15 avril 2011.

¹⁰ Voir les préconisations de l'IDABC, *EIF - European Interoperability Framework for pan-European eGovernment services*, 2004.

notamment de l'interprétabilité des cadres conceptuels et sémantiques du Dublin Core non qualifié (Childress *et al.*, 2009). Le tableau 3 présente une typologie détaillée de ces différences. Enfin, nous notons également des limites induites par les politiques distinctes des dépôts en termes de typologies documentaires et de classification thématique, tel que signalé dans la première partie de l'article.

Les problèmes identifiés ici sont communs à l'ensemble des projets visant l'interconnexion d'archives ouvertes. La partie qui suit circonscrit les freins à une approche disciplinaire pour les sciences de l'information.

Type	Commentaire	Détail - Exemples
Exploitation ou non des champs (Obligation ou non de données)	Les études quantitatives permettent de déterminer le degré d'exploitation des champs (Ward, 2004 ; Jordan, 2006 ; Efron, 2007).	Les champs <i>creator</i> , <i>identifier</i> , <i>date</i> , <i>format</i> sont massivement complétés ; les champs <i>title</i> , <i>subject</i> , <i>description</i> , <i>format</i> , <i>type</i> , <i>language</i> , <i>publisher</i> , <i>contributor</i> viennent ensuite ; les champs <i>rights</i> , <i>source</i> , <i>relation</i> , <i>coverage</i> sont très peu voire pas du tout exploités.
Différence formelle de codification	Format des dates	Zulu ou Greenwich Mean time ; formats assigné par le logiciel ; formats assigné par l'utilisateur (AAAA, AAAA-MM ; AAAA-MM-JJ)
	Code langue (RFC 3066 ou ISO 639 préconisés par le DCMI)	Codes langues mais également texte libre : en, En, EN, english, English
	Règles d'écriture des noms propres	Pour le champ <i>creator</i> : NOM, Prénom ; Nom, Prénom ; nom, prénom ; Nom, P. ; etc.
	Formats des items (format MIME préconisé par le DCMI)	Format bien respecté suivant les études citées. Mais données non reconnues par PKP lors du test.
Type d'information attendue / exprimée pour un champ donné	Interprétabilité conceptuelle de la DTD DC non qualifiée	- <i>publisher</i> = hébergeur du dépôt, éditeur de la revue, institutions de l'auteur, etc. - <i>date</i> = date d'enregistrement ou date du document ?
Types de valeurs possibles	En fonction de la politique des dépôts, les types de valeur ne se recouvrent pas	- <i>type</i> : selon les politiques de contenu des dépôts (typologies documentaires) ; - <i>subject</i> : Classifications thématiques, indexation libre manuelle, indexation full text
Nombre de valeurs par champ	Champs qui peuvent être mono ou multivalués malgré les recommandations du DCMI pour des champs monovalués	- plusieurs auteurs dans un même champ <i>creator</i> - titres et résumés en plusieurs langues à la suite
Vocabulaire et langue d'expression des valeurs	Un même élément d'information, dans un même champ, peut être exprimé de plusieurs façons	Pour un même type de document par exemple : Article de revue à comité de lecture, <i>Journal article</i> (<i>On-line</i> , <i>Unpaginated</i>), <i>Article</i> (<i>Print/paginated</i>), <i>Research Paper</i> , <i>Paper</i> , <i>Article</i> .

Tableau 3. Typologie des différences d'exploitation de la DTD Dublin Core nq

Difficile convergence des archives ouvertes en SIC

4. Le difficile recouvrement sémantique des plans de classement en sciences de l'information

Nous présentons ici la diversité et le difficile recouvrement des types de valeurs possibles pour le champ *subject*, selon les classifications des dépôts (tab. 4).

ArchiveSic: Plan de classement interne (report de la terminologie anglaise des sets, équivalent du plan de classement interne francophone)	E-LIS: Catégories de niveau 1 du JITA	QUT: Sous-ensemble LIS du plan de classification générale australien	BibInfo: Plan de classement interne (traduction de l'allemand)
<ul style="list-style-type: none"> -Bibliometry, scientometry ; - Cinema, art, esthetic; -Scientific communication and information ; - Conflict, information strategy, intelligence ; - Information retrieval ; -Information/communication Law ; - Economy, cultural industry ; - Electronic publishing ; - Education, e-learning, training ; - Knowledge management; - History of information/communication; - Hypertext, hypermedia; - Information system engineering; - Mass media; - Muscology; - Organisation and communication; - Theory of information/communication; - Others 	<ul style="list-style-type: none"> -Theoretical and general aspects of libraries and information; -Information use and sociology of information ; - Users, literacy and reading ; - Libraries as physical collections ; - Publishing and legal issues ; - Management ; - Industry, profession and education ; - Information sources, supports, channels ; - Information treatment for information services ; - Technical services in libraries, archives and museums ; - Housing technologies ; Information technology and library technology 	<ul style="list-style-type: none"> -Health Informatics; -Human Information Behaviour; - Information Retrieval and Web Search ; - Librarianship ; - Library and Information Studies not elsewhere classified ; - Organisation of Information and Knowledge Resources; - Records and Information Management (excl. Business Records and Information Management); - Social and Community -Informatics 	<ul style="list-style-type: none"> -Acquisition; -Développement; -Les services de bibliothèque, maîtrise de l'information, e-tutoriels ; - Conservation, numérisation ; - Publications électroniques, édition électronique ; - Portails, bibliothèques virtuelles, OPAC ; - Gestion, contrôle opérationnel, bibliométrie ; - Bâtiments et équipement des bibliothèques ; - Droit, Politique ; - Promouvoir la lecture, la bibliothèque partenaires de l'éducation ; - Profession, éducation, formation ; - Certains types de bibliothèques et de collections ; - Bibliothèque interculturelle de travail ; - Autres

Tableau 4. Plans de classement thématiques pour 4 archives ouvertes en SIC-LIS

On observe notamment :

- des domaines d'étude différents: e.g. pas de catégories consacrées à la bibliothéconomie ou aux bibliothèques, aux usages et comportement, ni aux technologies de l'information sur ArchiveSIC alors que ce sont des thématiques de niveau 1 dans le JITA, que l'on retrouve également sur QUT et Bib-Info. *A contrario*, seule ArchiveSIC traite des arts, des industries et économie de la culture ou encore du Knowledge Management. Egalement : la Communication est incluse sur ArchiveSic alors qu'elle n'est jamais mentionnée dans les autres plans de classement ;

Difficile convergence des archives ouvertes en SIC #

- des approches différentes de la discipline: approches théorique et épistémologique (e.g. catégories Histoire et Théorie des SIC sur ArchiveSIC) vs approche technique et applicative (Bib-Info), en lien avec les métiers de la documentation et de la bibliothéconomie.

- un ancrage différent pour la discipline, en SHS et/ou en STM (e.g. *Health Informatics* sur QUT, où les LIS sont une sous-catégorie des *Computing and IS*).

Ces disparités se repèrent au-delà de l'échantillon et du type de dispositif étudié : classifications des grandes bases de données comme Francis, ou de l'éditeur scientifique Ebsco¹¹ par exemple. Elles font écho à l'histoire ou construction de ce champ scientifique, à son caractère inter et pluridisciplinaire, ainsi qu'aux objets d'étude envisagés et postures épistémologiques adoptées par les chercheurs du champ au niveau national et international (Buckland *et al.*, 1998 ; Fondin, 2001 ; Zins, 2005). Le manque induit d'un consensus autour d'une classification commune, telle qu'elle existe dans d'autres disciplines (code JEL en économie par exemple), ou tel qu'il a pu se faire pour des projets spécifiques comme le moissonneur pour les sciences de la mer Avano (Merceur, 2007), **sont** une limite à l'interopérabilité sémantique des archives ouvertes de la discipline (Coleman *et. al.*, 2005).

5. Conclusion, recommandation

L'étude montre les difficultés de l'interconnexion d'archives ouvertes et interroge sur les modes de convergence des Sciences de l'information à l'international. Plus largement, les enjeux de transversalité que laissait espérer l'interrogation commune de ces archives ouvertes disciplinaires se heurtent sur le terrain à une disparité de pratiques constatées. La convergence nécessiterait de repenser à la source les modes de production de ces dispositifs selon des procédures partagées au sein de la communauté scientifique et entre gestionnaires de ces archives comme l'ont mis en œuvre les économistes avec le service EconomistOnline par exemple (Hellemans, 2007)

6. Bibliographie

Buckland M., Liu Z., "History of Information Science", Buckland M., Hahn T. B., *Historical studies in Information Science*, Medford, Information-Today, ASIS, 1998.

Chartron G., Gallezot G., Noyer J.-M., Rossi C., "Conception d'une Archive ouverte en SIC : le sens de la technique," *Hypertextes hypermedias : créer du sens à l'ère du numérique, H2PTM'03*, Paris, Hermes, 2003.

¹¹ <http://www.ebscohost.com/academic/library-information-science-technology-abstracts-with-full-text>

Difficile convergence des archives ouvertes en SIC

- Childress E., Park J. R., Dublin Core metadata semantics: an analysis of the perspectives of information professionals, *Journal of Information science*, vol. 35, n°6, p. 725-739.
- Coleman A., Roback J., "Open Access Federation for Library and information Science. dLIST and DL-Harvest", *D-Lib Magazine*, vol. 11, n°12, 2005.
- De Robbio A., Coll I. S., "E-LIS: an international open archive toward building open digital libraries", *High Energy Physics Libraries Webzine*, vol. 11, 2005.
- Efron M., "Metadata Use in OAI-Compliant Institutional Repositories", *Journal of Digital Information*, vol. 8, n°2, 2007.
- Fondin H., « La Science de l'information : posture épistémologique et spécificité disciplinaire », *Documentaliste – Sciences de l'information*, vol. 38, n° 2, 2001, p. 112-122.
- Foulonneau M., "Information redundancy across metadata collections," *Information Processing & Management*, vol. 43, 2007, p. 740-751.
- Hellemans J., "Nereus-economists online : Les Archives ouvertes dans le domaine de l'économie", *Cahiers de la Documentation*, vol. 61, n°3, 2007, p. 5-14.
- Innocenti P., Ross S., Vullo G., *Interoperability for digital repositories: towards a policy and quality framework*, Open Repositories Conference 2010 (OR2010), Spanish Foundation for Science and Technology, 6-9 juillet 2010, Madrid (Espagne).
- Jordan M., « The CARL metadata harvester and search service », *Library Hi Tech*, vol. 24, n°2, 2006, p. 197-210.
- Lagoze C., Sompel H.V.D., "The making of the Open Archives Initiative Protocol for Metadata Harvesting," *Library Hi Tech*, vol. 21, jan. 2003, p. 118-128
- Bherwani, M., Hirwade, M.A., « Facilitating searches in Multiple Bibliographical Databases: Metadata Harvesting Service Providers », *Liber Quarterly*, vol. 9, n°2, 2009, p. 158.
- Merceur F., "Avano, assessment of one year management of a thematic OAI-PMH harvester", *33rd IAMSLIC Annual Conference: Changes of the Horizon*, Sarasota, 7-11 octobre 2007, Sarasota, 2008, IAMSLIC.
- Puplett D., "Subject Repositories: European Collaboration in the International Context", *Ariadne*, n°62, 2010.
- Ward J., "Unqualified Dublin Core usage in OAI-PMH data providers," *OCLC Systems & Services*, vol. 20, 2004, p. 40-47.
- Xia J., "Disciplinary repositories in the social sciences", *Aslib Proceedings: New Information Perspectives*, vol. 59, n° 6, 2007, p. 528-538.
- Zins C., "Conceptions of Information Science", *JASIST*, vol. 58, n°3, 2007, p. 335-350.