



HAL
open science

Les archives ouvertes en France. Un potentiel documentaire pour la formation à distance

Joachim Schöpfel, Christiane Stock

► To cite this version:

Joachim Schöpfel, Christiane Stock. Les archives ouvertes en France. Un potentiel documentaire pour la formation à distance. *Distances et savoirs*, 2009, 7 (3), pp.443-456. 10.3166/ds.7.443-456 . sic_00481597

HAL Id: sic_00481597

https://archivesic.ccsd.cnrs.fr/sic_00481597

Submitted on 6 May 2010

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.



Distributed under a Creative Commons Attribution| 4.0 International License

Les archives ouvertes en France

Un potentiel documentaire pour la formation à distance

Joachim Schöpfel* — **Christiane Stock****

** Université Charles de Gaulle Lille 3
UFR IDIST - Laboratoire Geriico
BP 60149
F-59653 Villeneuve d'Ascq cedex
joachim.schopfel@univ-lille3.fr*

*** INIST-CNRS
2, allée du Parc de Brabois
F-54519 Vandœuvre-lès-Nancy cedex
christiane.stock@inist.fr*

RÉSUMÉ. Issues du mouvement vers l'accès libre à l'information scientifique, les archives ouvertes constituent une formidable ressource documentaire pour la formation et l'autoformation à distance, en particulier dans l'enseignement supérieur. L'article présente les résultats d'une étude empirique sur le contenu de 56 archives ouvertes en France en 2008. Parmi les différents aspects abordés : typologie des sites et organismes, domaines scientifiques, nombre des dépôts, types de documents, qualité des contenus, avec une attention particulière pour les contenus non commerciaux (littérature grise).

ABSTRACT. Being part of the open access movement to scientific information, the open archives are a significant documentary resource for distant and open learning, especially in Higher Education. Our article presents the results of an empirical survey on the content of 56 open archives in France in 2008. Information is provided on typology and hosts of archives, scientific subjects, deposit number, typology of documents, content quality etc. Specific attention is paid to non-commercial content (grey literature).

MOTS-CLÉS : archives ouvertes, archives institutionnelles, information scientifique, formation à distance, France.

KEYWORDS: open archives, institutional repositories, scientific information, distant learning, France.

Introduction

Les ressources en ligne constituent un formidable réservoir potentiel de savoir pour les communautés virtuelles d'apprentissage de la formation à distance dans le cadre de l'enseignement supérieur, aussi bien pour l'enseignant-formateur que pour l'apprenti qui élabore ses connaissances dans le contexte de son environnement numérique (cf. Drissi *et al.*, 2006).

Le contenu des archives ouvertes¹ s'ajoute en complémentarité, potentiellement et virtuellement, à l'offre des espaces documentaires et campus numériques de la formation à distance. En même temps, il favorise l'autoformation (Alava, 2000 et Albéro, 2003) grâce à sa technologie (interopérabilité), au référencement par les moteurs de recherche et au principe de l'accès non restrictif.

Malgré leur nombre encore limité à ce jour, le contenu des archives ouvertes en France évolue rapidement. Nous présentons ici les résultats d'une étude empirique et exhaustive menée en 2008, une sorte de photographie du paysage qui permet de mieux cerner et utiliser la richesse et le potentiel de cette nouvelle offre d'information scientifique. Une attention particulière est accordée au contenu non *marchand* communément appelé « littérature grise », souvent non accessible *via* les grands portails et bibliothèques numériques (littérature grise, cf. Schöpfel, 2006).

« De nouvelles possibilités de diffusion de la connaissance (...) en s'appuyant sur le paradigme du libre accès *via* l'internet, doivent être soutenues. (...) Une version complète de l'œuvre (...) est déposée (...) auprès d'au moins une archive en ligne, (...) gérée et entretenue par une institution académique, une société savante, une administration publique ou un organisme établi ayant pour but d'assurer le libre accès, la distribution non restrictive, l'interopérabilité et l'archivage à long terme. »²

Le 22 octobre 2003, dix-neuf organismes européens dont le CNRS signèrent cette Déclaration de Berlin sur le libre accès à la connaissance scientifique. En janvier 2006, la Commission Européenne publia une étude sur l'évolution du marché de l'édition scientifique avec des recommandations pour une politique en faveur des archives ouvertes (Dewatripont *et al.*, 2006). Lancée au début 2007, une pétition européenne pour l'accès libre aux résultats de la recherche publique a été signée par des centaines d'organisations et plus de 27 000 chercheurs³.

En France, les communautés scientifiques et les professionnels de l'information ont suivi ce mouvement vers l'accès libre à l'information scientifique⁴ avec intérêt et

1. « (...) un réservoir où sont déposées des données issues de la recherche scientifique et de l'enseignement et dont l'accès se veut ouvert, c'est-à-dire sans barrière » Cf. <http://openaccess.inist.fr/spip.php?mot12>. Cf. aussi la présentation par Schöpfel et Stock, 2008.

2. Cf. le texte original à <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

3. <http://www.ec-petition.eu/>

4. « Le mouvement du libre accès désigne l'ensemble des initiatives prises pour une mise à disposition des résultats de la recherche au plus grand nombre, sans restriction d'accès, que ce

engagement⁵. En juillet 2006, universités et établissements publics de recherche ont signé un premier accord pour le développement d'une infrastructure commune d'archives ouvertes pour le dépôt des publications des chercheurs et enseignants-chercheurs. Leur objectif : faciliter la communication directe et l'évaluation, augmenter la visibilité et notoriété de leur établissement sur le web, maîtriser les dépenses documentaires.

Au cœur du dispositif français, le Centre de Communication Scientifique Directe du CNRS à Lyon⁶ et le consortium COUPERIN qui a lancé son propre site dédié aux archives institutionnelles en novembre 2008⁷. Le projet européen DRIVER⁸ a évalué la France en 2007 comme un « pays avancé » en matière d'archives ouvertes (Van de Graaf et Van Eijndhoven, 2007).

Méthodologie et échantillonnage

Le référencement des archives ouvertes a fait de grands progrès depuis le lancement des premiers sites dans les années 1990, mais il n'existe pas à ce jour de répertoire exhaustif ni en France, ni ailleurs. La meilleure synthèse disponible date de 2007 et présente 40 sites français (Baruch, 2007).

En recoupant les informations de huit répertoires internationaux de référence⁹, nous avons identifié et sélectionné 56 sites. Pour chaque site nous avons vérifié s'il était localisé en France et s'il avait des dépôts récents.

Ces 56 archives ouvertes représentent un échantillon quasi-exhaustif au moment de l'étude (printemps 2008). Quasi-exhaustif veut dire : des sites référencés et répertoriés, avec un rayonnement et une visibilité internationale. Nous sommes conscients qu'il existait, au moment même de l'étude, d'autres archives ouvertes en France. Mais comme elles ne sont pas référencées, il est difficile de les identifier et d'en estimer le nombre.

Chacun des 56 sites de notre sélection a été caractérisé selon 35 critères. La collecte de données a été réalisée entre mars et mai 2008 avec une stratégie « open source », à partir d'informations publiées et/ou librement accessibles sur internet. Les éléments obtenus ont été intégrés dans une base de données, vérifiés, validés et le cas échéant, modifiés et/ou complétés avant l'analyse statistique ou qualitative.

soit par l'autoarchivage ou par des revues en libre accès. » Cf. <http://openaccess.inist.fr/spip.php?mot5>

5. Cf. l'étude de Bosc H., 2008 et le site de veille du CNRS <http://openaccess.inist.fr/>

6. <http://ccsd.cnrs.fr/>

7. <http://www.couperin.org/archivesouvertes/>

8. *Digital Repository Infrastructure Vision for European Research*, cf. <http://www.driver-community.eu/>

9. L'échantillonnage a été effectué entre mars et mai 2008. Cf. la liste des répertoires à la fin des références.

Trente sites, c'est-à-dire plus de la moitié, correspondent à des archives institutionnelles, mises en place par des universités telles que Lyon 2, Strasbourg ou Toulouse, par des écoles (Paristech, Télécommunications) ou par des organismes de recherche comme l'INRA, l'IRD ou l'IFREMER, afin de gérer et diffuser les ressources numériques créées par l'établissement ou par les membres (= chercheurs, enseignants, doctorants etc.) de sa communauté (cf. Lynch, 2003, adapté par Foulonneau *et al.*, 2007).

Treize archives sont de nature thématique, c'est-à-dire contiennent des dépôts d'un domaine scientifique particulier comme par exemple les mathématiques, la cristallographie ou les nouvelles technologies de l'information et de la communication.

D'autres sites sont réservés à un type de document spécifique (rapport, mémoire, thèse ou revue) ou ont une vocation patrimoniale (ex. IRIS de l'USTL) ou nationale (ex. HAL du CCSD)¹⁰.

Les premières archives ont été mises en place à partir des années 90 mais la plupart sont récentes et ont moins de cinq ans. Plus des deux tiers fonctionnent avec des logiciels développés par et pour le mouvement vers l'accès libre (DSpace, EPrints, HAL, LODEL), compatibles avec le protocole international OAI-PMH ; le reste s'appuie sur des développements spécifiques ou sur des systèmes de gestion de bibliothèque. Pour l'utilisateur, une plate-forme « standard » comme DSpace ou EPrints facilite la navigation au-delà de la recherche simple sur le site¹¹.

La volonté d'une visibilité internationale est évidente pour la moitié des archives et se reflète dans la langue des interfaces d'information, de consultation et de recherche. 54 sites ont des interfaces en français. Mais 25 sites proposent également une interface en anglais, au moins partiellement ; deux autres en plus en allemand et en espagnol. Deux archives sont entièrement publiées en anglais¹².

Domaines scientifiques

Les archives ouvertes françaises couvrent la plupart des domaines et disciplines scientifiques mais elles le font d'une manière inégale (cf. tableau 1).

Les archives multidisciplinaires totalisent un tiers des dépôts. Il s'agit surtout des sites institutionnels (grandes universités multidisciplinaires etc.) et des archives à

10. Cf. aussi <http://www.couperin.org/archivesouvertes/spip.php?article92>

11. Le paysage est appelé à changer dans les prochains mois, après le lancement en 2008 du nouveau système ORI-OAI, une solution spécifique dédiée aux archives institutionnelles de l'enseignement supérieur.

12. En fait, il s'agit de deux archives de données primaires factuelles avec des résultats scientifiques en cristallographie et en astronomie.

vocation nationale ou pour un type de document particulier (thèses etc.), sans lien avec une communauté scientifique particulière.

Domaine scientifique	Nombre d'archives	Nombre total de dépôts
Multidisciplinaire	19	252 093
Sciences humaines et sociales	20	223 368
Sciences appliquées	12	186 704
Sciences, médecine, sport	5	42 413

Tableau 1. *Domaines scientifiques des archives ouvertes*

Parmi les archives créées pour des disciplines spécifiques, les sciences humaines et sociales et les sciences appliquées tiennent le haut du pavé avec notamment la linguistique (3 sites), les sciences de l'information et de la communication (4 sites), l'ingénierie civile (5 sites), l'informatique (3 sites) et l'agronomie (3 sites).

Un tiers des dépôts se trouvent dans les archives SHS, un quart dans celles réservées aux sciences appliquées.

En comparaison, le nombre d'archives en sciences exactes est limité, et le nombre de leurs dépôts, assez faible (6 %). La véritable particularité du paysage français est l'absence d'une grande archive ouverte en sciences de la vie pour les publications en médecine et biologie. La raison est probablement l'option des deux grands établissements de recherche CNRS et INSERM en faveur d'une archive d'articles nationale et multidisciplinaire (= HAL) et la réputation du site américain PubMed Central qui draine les publications des chercheurs français¹³.

Contenu des archives et nombre de dépôts

Le contenu des archives présente une grande variété. Parmi les dépôts, nous avons trouvé plus de 20 types différents dont des livres dont certains avec une valeur patrimoniale (livres rares), des manuscrits, des cartes et images, des revues et articles de revues, des actes de congrès (communications), des cours, posters, brevets, thèses et mémoires, des rapports, des pré-publications et autres documents non publiés.

Mais l'utilisateur y trouvera aussi des échantillons de parole (enregistrements sonores) avec transcription, des données factuelles (résultats de recherche), des vidéos, logiciels et sites web.

13. Fin février 2009, les chercheurs du CNRS et/ou de l'INSERM avaient publié plus de 20 000 documents sur le site de PubMed Central <http://www.pubmedcentral.nih.gov/>, par rapport à moins de 5 000 dépôts en sciences du vivant sur le site HAL du CNRS.

Cinquante-trois archives contiennent des documents textuels (écrits) dont 7 intègrent d'autres types de documents (données factuelles, images, cartes, etc.). Seuls 3 sites ne contiennent aucun document texte.

Quatre archives, toutes produites par le CNRS, sont dédiées à une seule catégorie de documents, provenant de multiples institutions. Le CNRS a ainsi créé une archive nationale pour des revues en libre accès des domaines sciences humaines et sociales (revues.org hébergé par CLEO). Les trois autres archives sont dédiées à la littérature grise : LARA (INIST) propose le libre accès aux rapports scientifiques et techniques ; TEL (CCSD) contient des thèses en ligne ; MémSIC (CCSD), des mémoires en Sciences de l'Information et de la Communication (niveau licence et master).

Le type de documents le plus important, en termes du nombre de dépôts, est de loin l'article scientifique. A eux seuls, les cinq sites PERSEE, HAL, INRA, IRD et Revues.org totalisent ensemble plus de 360 000 articles scientifiques, surtout en SHS (plus de 200 000) mais aussi en physique, mathématiques, agronomie et d'autres disciplines.

La taille des archives oscille entre 16 et 172 215¹⁴ dépôts, avec une moyenne de 12 500 dépôts. Cependant 32 archives possèdent moins de 1 000 entrées ; ensemble, elles représentent 57 % des sites mais seulement 2 % du nombre de dépôts (cf. figure).

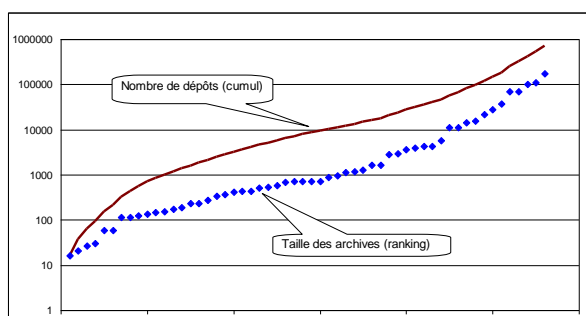


Figure 1. Taille des archives et nombre de dépôts (échelle logarithmique)

La liste des douze archives ouvertes les plus importantes en termes de dépôts souligne le rôle central du CNRS, l'importance de certains domaines scientifiques et l'absence d'un grand site d'information en sciences du vivant (cf. tableau 2).

14. Il s'agit de PERSEE, l'archive des revues en SHS.

Nom	Serveur	Organisme	Domaine
PERSEE	Lyon 2	Université Lyon 2	SHS
HAL	CCSD	CNRS	Multidisciplinaire, surtout
ProdINRA	INRA	INRA	Agronomie
Crystallography		Université Maine	Multidisciplinaire, sciences
Horizon Pleins		IRD	Multidisciplinaire
GALLICA ¹⁵		BNF	Patrimonial
NUMDAM	MathDoc	CNRS	Mathématiques
Revue.org	CLEO	CNRS	SHS
HAL SHS	CCSD	CNRS	SHS
I-Revues	INIST	CNRS	Multidisciplinaire
HAL INRIA	CCSD	CNRS	Informatique appliquée
TEL	CCSD	CNRS	Multidisciplinaire

Tableau 2. Les douze archives les plus importantes en termes de dépôts

Le contenu non commercial

Les articles scientifiques déposés dans les archives ouvertes sont également publiés dans des revues scientifiques. D'autres documents déposés par contre échappent plus ou moins aux circuits habituels d'accès et de diffusion. Il s'agit de documents qui ne sont pas commercialisés et dont l'utilisation n'est pas toujours facile. Pour cette « littérature grise »¹⁶ les archives ouvertes constituent un vecteur de communication intéressant. L'intérêt de la littérature grise : souvent, ces documents rendent accessibles des résultats de recherche plusieurs mois voire années avant leur publication dans les revues ou livres, et parfois ces résultats ne seront jamais publiés ailleurs (cf. Schöpfel, 2008).

Quatre archives sur cinq (80 %) contiennent au moins un type de littérature grise, c'est-à-dire notamment des thèses, rapports, conférences, cours ou documents de travail non publiés. Plus intéressant encore, 100 % des archives institutionnelles donnent accès à cette catégorie de documents, proposant ainsi un accès libre à une production scientifique riche en information et unique en son genre.

15. Le cas de la bibliothèque numérique GALLICA est problématique ; elle figure dans notre étude dans la mesure où certains répertoires internationaux la considèrent comme une archive ouverte à caractère patrimonial. *Stricto sensu*, elle ne devrait pas y être, même si les limites entre archives ouvertes et bibliothèques numérique parfois sont fluctuantes.

16. « [La littérature grise est] ce qui est produit par toutes les instances du gouvernement, de l'enseignement et la recherche publique, du commerce et de l'industrie, sous un format papier ou numérique, et qui n'est pas contrôlé par l'édition commerciale » (Schöpfel, 2006).

Type de document	Nombre de dépôts
Thèses et mémoires	37 014
Communications	23 765
Rapports	17 725
Cours	276
Travaux non publiés	179
Non spécifié	31 329
Total	110 288

Tableau 3. *La littérature grise dans les archives ouvertes*

Ces documents, dont le nombre total dépasse 100 000, correspondent à 18 % des dépôts textuels dans les archives ouvertes en France. Néanmoins, la répartition est inégale. Certains sites importants contiennent un nombre relativement élevé de documents gris (HAL, Horizon IRD, ProdINRA). D'autres archives de taille plus modeste se consacrent entièrement à cette catégorie comme TEL (thèses) ou LARA (rapports). Mais un tiers des archives ne contiennent aucun document gris (comme PERSEE) ou très peu (<10 %).

Le pourcentage élevé de documents « non spécifiés » dans le tableau s'explique par un problème récurrent de la littérature grise, le manque de contrôle bibliographique. Seule une partie des archives identifient clairement ces documents et ajoutent des métadonnées descriptives spécifiques, comme par exemple le lieu et la date d'une communication, le type et la discipline d'une thèse ou encore le numéro d'un rapport. Ce manque de description bibliographique ne facilite pas l'identification et la recherche.

Politique d'archivage et qualité des dépôts

La qualité d'une archive ouverte est liée à son contenu mais s'exprime aussi *via* son mode de gestion et communication. Avec le répertoire international OpenDOAR, nous pouvons distinguer cinq aspects d'une politique explicite pour la gestion d'une archive ouverte¹⁷ :

- conditions pour l'utilisation des métadonnées
- conditions pour l'utilisation des dépôts
- information sur les contenus (dépôts)
- information sur la soumission de publications (obligation de dépôt etc.)
- information sur la politique de conservation

17. <http://www.openoar.org> Au même sujet, voir aussi Jeffery et Assersson, 2007.

D'après notre étude, seulement la moitié des archives communiquent explicitement sur leurs approches, conditions et politiques. Ceci étant, nous avons trouvé un seul site¹⁸ qui propose une information complète sur l'ensemble des aspects mentionnés. C'est regrettable dans la mesure où cette information est nécessaire pour orienter l'utilisateur final. En plus, dans le contexte futur d'interopérabilité et d'échange entre sites l'expression claire des politiques faciliterait la création de services dérivés (cf. plus loin).

Concernant la qualité des contenus, il est difficile de faire la synthèse de l'ensemble dont l'effectif dépasse 700 000 dépôts au moment de l'étude. Mais il est tout à fait possible de fournir quelques éléments.

Articles scientifiques : d'après les informations sur les sites, voire dans les notices, la plus grande partie des articles déposés dans les archives ouvertes ont été publiés dans des revues scientifiques. Leur qualité dépend donc pour l'essentiel de la politique éditoriale (évaluation, sélection) des revues.

Contenu non-commercial : au moins une partie des dépôts gris a été explicitement évaluée. Il s'agit des travaux universitaires (thèses, mémoires). Mais les communications aussi ont été l'objet d'un « contrôle de qualité », d'une sélection par les comités scientifiques des journées d'études, colloques et conférences.

Résultats de recherche : il est par contre plus difficile de se prononcer sur le dépôt des données d'observation ou d'autres résultats de recherche. D'une certaine façon, leur qualité se révélera lors de leur ré-utilisation (vérification).

Modération (« label qualité ») : quelques sites mentionnent des administrateurs ou modérateurs qui peuvent agir comme éditeur scientifique pour la sélection des documents à déposer, sans toutefois agir sur le contenu (révision). D'autres archives n'acceptent que les publications revues par des pairs, donc en majorité les publications commerciales (dans la forme eprint ou postprint), le contrôle qualité étant reporté vers l'éditeur commercial.

Une estimation très prudente du pourcentage des dépôts ayant été soumis à une procédure d'évaluation aboutit à 60-65 % mais la part des documents évalués et/ou avec un label qualité est probablement bien supérieure.

Discussion

Notre étude décrit une nouvelle offre documentaire riche, variée et en évolution rapide. Nous avons déjà évoqué quelques problèmes qui sont liés surtout à la jeunesse de ce mouvement vers l'accès libre à l'information, comme par exemple le manque d'information précise sur les conditions d'utilisation ou l'absence de contrôle bibliographique dans le domaine de la littérature grise. D'autres problèmes,

18. OATAO, l'archive institutionnelle du PRES de l'université de Toulouse.

comme l'absence d'une couverture satisfaisante dans le domaine des sciences du vivant ou le nombre limité d'archives institutionnelles, renvoient à des choix politiques en matière d'information scientifique qui semblent en cours de révision, notamment dans le cadre de l'université numérique.

Cependant trois autres aspects ont attiré notre attention : l'accès aux documents, la fiabilité des informations et le suivi de l'utilisation des archives.

Le problème d'un accès restreint

En principe, les archives ouvertes offrent un accès sans barrière aux documents déposés. Or, la réalité n'est pas aussi simple. Selon nos chiffres, seulement 48 % des sites correspondent à ce principe. En fait, nous observons deux problèmes différents.

(1) La présence de références bibliographiques sans accès au texte intégral. Plusieurs organismes acceptent le dépôt ou l'import de notices de catalogage sans lien vers le document signalé. Sont concernés par exemple plusieurs sites du CCSD (HAL, HAL-SHS, HAL-INRIA). D'autres comme l'INRA ou l'IRD ont décidé d'enrichir progressivement bases de données ou catalogues bibliographiques avec le texte intégral. Nous avons trouvé ce mélange de notices et de documents dans 16 % des archives¹⁹. Pour l'accès à l'information, c'est plutôt gênant ; de l'autre côté, les liens vers le texte intégral devraient être progressivement ajoutés, et l'utilisateur peut dès à présent exclure les notices de sa recherche documentaire *via* les interfaces des sites.

(2) Limitation d'accès. Certaines archives protègent une partie du contenu, soit *via* l'intranet de l'établissement, soit *via* un mot de passe réservé à un groupe d'utilisateurs. D'autres sites comme I-Revues ou Revues.org agencent libre accès aux documents plus anciens et embargo temporaire aux documents plus récents, en interdisant aux non-abonnés par exemple l'accès aux deux dernières années. Ils existent des solutions techniques (couleurs, icônes) pour en informer l'utilisateur dès l'affichage des résultats de requête, mais celles-ci sont rarement utilisées.

Le problème des doublons

Le nombre précis des dépôts à un moment donné est difficile à cerner. Les chiffres varient d'un répertoire à l'autre, en fonction de la méthode de comptage (tous les dépôts y compris notices et données primaires, seulement dépôts avec texte intégral, seulement dépôts sans restriction d'accès etc.). Notre étude a privilégié le comptage de tous les dépôts mais il est clair qu'il y aura toujours des divergences (parfois significatives) avec d'autres approches et études.

19. Paradoxalement, l'une des raisons est l'utilisation de ces sites par leur organisme pour l'évaluation des chercheurs et les bilans – le signalement exhaustif des publications est donc plus important que la présence des documents eux-mêmes.

Ceci aussi pour une autre raison. Personne n'empêchera un auteur de déposer son texte dans plusieurs archives afin d'en assurer une plus grande visibilité. Le même document sera donc compté plusieurs fois. A cela s'ajoute une particularité de la plate-forme HAL. Le CCSD a décidé d'y verser systématiquement les dépôts d'autres archives. Ainsi une thèse déposée dans PASTEL ou TEL se retrouvera dans HAL. Un article déposé dans ArchiveSIC sera versé dans HAL-SHS et HAL. Ces doublons gonflent artificiellement le volume de HAL.

On rencontre une autre situation à Toulouse où plusieurs universités et grandes écoles ont mis en ligne des archives de thèses (INPT, INSA, UPS). Un site spécifique « Toulouse thèses » moissonne ces dépôts et intègre en plus les thèses de l'école vétérinaire locale. Récemment, le PRES Toulouse a mis en place une nouvelle archive institutionnelle OATAO (Open Archive Toulouse Archive Ouverte) qui recevra surtout les publications scientifiques des chercheurs (articles) mais également des thèses. Une archive unique pour toutes les publications et tous les établissements de l'enseignement supérieur de Toulouse est annoncée pour bientôt. En attendant, il faut vivre (et compter) avec des doublons.

Le problème des statistiques d'utilisation

Actuellement il est quasiment impossible de trouver une information précise sur l'utilisation des archives ouvertes. On ne sait pas comment les chercheurs, enseignants, étudiants etc. utilisent les dépôts. Seul un rapport de l'IFREMER (Merceur, 2008) fournit quelques éléments sur l'usage de l'archive institutionnelle ARCHIMER. Il peut y avoir plusieurs explications à cette absence d'information : des problèmes techniques liés au logiciel, des problèmes conceptuels avec les standards, l'absence de moyens pour la collecte et l'interprétation des données. Parfois, comme pour les archives du CCSD, le chercheur a accès aux statistiques de ses propres dépôts. Mais aucune information n'est fournie sur l'intégralité des dépôts ni non plus sur l'évolution des chiffres.

Il faut espérer que cette situation évoluera rapidement afin de pouvoir évaluer le comportement des utilisateurs, mesurer le retour sur investissement et développer de nouveaux services à partir d'une analyse réaliste.

Plaidoyer pour une synergie nécessaire

« L'apport des Nouvelles Technologies dans l'enseignement réside essentiellement dans (...) l'accessibilité facilitée aux connaissances et au savoir, le rendant en quelque sorte universel » (Crétin et Laïd, 2004). L'accessibilité aux connaissances et l'universalité du savoir sont aussi les maîtres-mots du mouvement vers l'accès libre à l'information scientifique.

La création, la mise en œuvre et le développement des archives ouvertes en France possèdent un potentiel significatif pour la formation à distance, aussi bien en

métropole que dans la coopération avec les pays francophones en voie de développement. « Openness in education involves (...) open access » (Guiton *et al.*, 2005). Pour l'enseignement à distance dans ces derniers pays, la gratuité d'accès et la qualité des contenus présentent une alternative aux grandes bases commerciales dont le prix peut être prohibitif.

Il s'agit de ressources numériques à caractère scientifique ou technique, accessibles 7/24, de partout et sans restriction. Une bonne partie de ces ressources, en particulier la littérature grise, sont des documents uniques et rares, riches en information (ex. rapports), ou excellent par leur fraîcheur (ex. preprints). D'autres ont le mérite d'être plus proches des préoccupations des étudiants (ex. mémoires, thèses).

Bien que des évolutions récentes (limitations d'accès, absence de texte intégral) aillent parfois à contre-courant du concept initial d'archives ouvertes, ces ressources dans leur ensemble constituent un réservoir formidable de savoir et de techniques.

L'intérêt de cette « économie des biens publics » pour la formation à distance, en particulier dans le cadre universitaire, a été identifié assez tôt (cf. Chartron et Minon 2004). Il est donc permis de s'étonner de l'apparente faible synergie entre les deux dispositifs, aussi bien par rapport aux technologies que par rapport aux contenus.

L'intérêt technologique du mouvement vers l'accès libre à l'information scientifique réside dans l'interopérabilité des systèmes à partir d'un protocole commun pour le stockage et le moissonnage des métadonnées des dépôts (OAI-PMH), « widely used globally in projects that aim at open information sharing in a broad spectrum of domains including scholarly communication, virtual museums, distant learning, cultural heritage etc. » (Van de Sompel, 2003).

Mais où sont aujourd'hui les concepts et réalisations qui connectent les deux dispositifs et systèmes ? Peut-être faudra-t-il attendre la réalisation de « l'université numérique » (Isaac, 2008), le bilan des projets précurseurs et la mise en place d'un réseau performant d'archives institutionnelles universitaires, éventuellement même au détriment des archives à vocation nationale, pour que la jonction sur le terrain se fasse entre « l'essor de l'enseignement à distance » et le développement d'une « approche éditoriale du patrimoine numérique de chaque établissement ».

Il n'en reste pas moins que le choix des contenus nécessite une forte implication et une bonne connaissance de la part du formateur. Pour cela, une information précise sur la politique d'archivage, de validation et sélection et sur le contenu est indispensable. Tous les sites des BU et SCD contiennent ce genre d'information sur leurs collections, et toutes les bibliothèques veillent sur la qualité et la précision des notices de leur catalogue pour que le lecteur s'y retrouve. Pourquoi pas les sites des archives ouvertes ? Pourquoi un signalement souvent rudimentaire ? Pourquoi des fonctionnalités de recherche parfois assez pauvres ? Il faudra replacer les archives ouvertes dans un contexte documentaire et de service et appliquer à leur gestion la même rigueur professionnelle qui assure l'excellence des bibliothèques scientifiques.

La synergie entre les archives ouvertes et la formation à distance se fera à partir de la connexion et/ou de l'interopérabilité des systèmes et avec une information précise sur les contenus. Elle passera par une double prise de conscience : la prise de conscience par les établissements qu'il y a un nouveau public pour leurs archives ouvertes et de nouveaux services à développer, et la prise de conscience par les professionnels de la formation à distance (et par les apprentis-étudiants eux-mêmes) de cette offre d'information scientifique à portée de main, une offre qui représente déjà aujourd'hui 10-20 % de la production scientifique de certains pays et qui prendra encore de l'ampleur dans les années à venir grâce à la politique des établissements d'enseignement supérieur et de la recherche, des agences de moyens et des administrations au niveau national et européen.

Bibliographie

- Alava S. (dir.), *Cyberespace et formations ouvertes, Vers une mutation des pratiques de formation ?*, Bruxelles, De Boeck, 2000.
- Albéro B. (dir.), *Autoformation et enseignement supérieur*, Paris, Lavoisier, 2003.
- Baruch P., "Open Access Developments in France: The HAL Open Archive System", *Learned Publishing* 2007, vol. 20, n° 4, p. 267-282. <http://hal.archives-ouvertes.fr/hal-00176428/fr/>
- Bosc H., « L'auto-archivage en France : deux exemples de politiques différentes et leurs résultats », *Liinc em Revista* 2008, vol. 4, n° 2, p. 196-217. http://archivesic.ccsd.cnrs.fr/sic_00340885/fr/
- Chartron G., Minon M., « Quelle offre éditoriale numérique pour l'utilisateur 'virtuel' des universités françaises ? », *Pré-Conférence Satellite IFLA 2004, Management and Marketing Section*, São Paulo, 18-20 août 2004. http://archivesic.ccsd.cnrs.fr/sic_00001048/fr/
- Crétin R., Laïd B., « L'enseignement à distance : proposition de trois démarches », *Colloque TICE Méditerranée 2004*, Nice, 26-27 novembre 2004. <http://isdms.univ-tln.fr/PDF/isdms18/20-cretin-bouzidi.pdf>
- Dewatripont M. *et al.*, *Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe*, Bruxelles, Commission Européenne, 2006. http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf
- Drissi M.M., Talbi M., Kabbaj M., « La formation à distance : un système complexe et compliqué », *Revue de l'EPI (Enseignement Public & Informatique)*, septembre 2006. <http://www.epi.asso.fr/revue/articles/a0609b.htm>
- Foulonneau M. *et al.*, « Réseaux d'archives institutionnelles en Europe: logiques de développement et convergences », Non publié, 2007. http://archivesic.ccsd.cnrs.fr/sic_00205049/fr/>
- Guiton P. *et al.*, "The management of systems in open and distant learning", Hope A., Guiton P. (dir.), *Strategies for sustainable open and distant learning. World Review of Distant and Open Learning*, vol. 6. London, Routledge, 2005.

- Isaac H., *L'université numérique. Rapport à Madame Valérie Pécresse, Ministre de l'Enseignement Supérieur et de la Recherche*, Paris, 2008. http://www.nouvelleuniversite.gouv.fr/IMG/pdf/Rapport_univ_num.pdf
- Jeffery K., Asserson A., « Greyscape », *GL9 Conference Proceedings. Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape*, Anvers, 10-11 décembre 2007.
- Lynch C.A., "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age", *ARL Bimonthly Report* 2003, n° 226 (février), p. 1-7. <http://www.arl.org/newsltr/226/ir.html>
- Merceur F., *Fonctionnement et usages d'une archive institutionnelle*, Issy-les-Moulineaux, IFREMER, 2008. <http://www.ifremer.fr/docelec/doc/2008/rapport-4632.pdf>
- Schöpfel J., "Grey literature on bilingualism in Belgium", *Cahiers de la Documentation/Bladen voor Documentatie*, 2008, vol. 62, n° 2, p. 22-30. http://archivesic.ccsd.cnrs.fr/sic_00196598/fr/>
- Schöpfel J., « Le devenir de la littérature grise. Quelques observations », *Perspectives Documentaires en Education*, 2006, n° 62, p. 63-72. http://archivesic.ccsd.cnrs.fr/sic_00136829/fr/
- Schöpfel J., Stock C. "Grey literature in French Digital Repositories: A Survey", *GL10 Conference Proceedings. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society*, Amsterdam, 8-9 décembre 2008.
- Van de Graaf M., Van Eijndhoven K., *The European Repository Landscape: Inventory Study into the Present Type and Level of OAI-Compliant Digital Repository Activities in the EU*, Amsterdam, AUP, 2007. <http://dare.uva.nl/document/93725>
- Van de Sompel H., "Yesterday's Thoughts on Tomorrow's Scholarly Communication System", *Conference on Open Access to Knowledge in the Sciences and Humanities*. Berlin, 20-22 octobre 2003. http://www.cbib.ucl.ac.be/HVdS_EN.pdf
- BASE Bielefeld Academic Search Engine: <http://base.ub.uni-bielefeld.de/index.html>
- DSpace Repositories using DSpace – Alphabetical. http://www.dspace.org/index.php?option=com_content&task=view&id=596&Itemid=182
- EPrints Sites Powered by EPrints: <http://www.eprints.org/software/archives/>
- OpenDOAR Directory for Open Access Repositories: <http://www.opendoar.org/>
- ROAR Registry of Open Access Repositories: <http://roar.eprints.org/>
- Scientific Commons Register URL: <http://en.scientificcommons.org/register-repository>
- University of Illinois OAI-PMH Data Provider Registry: <http://gita.grainger.uiuc.edu/registry/>
- Webometrics Ranking Web of World Repositories: <http://repositories.webometrics.info/>