

Recherche en sciences de l'information

LES SERVICES POUR LES ARCHIVES OUVERTES : DE LA RÉFÉRENCE À L'EXPERTISE

Emma Bester

A.D.B.S. | « Documentaliste-Sciences de l'Information »

2010/4 Vol. 47 | pages 4 à 15

ISSN 0012-4508

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-documentaliste-sciences-de-l-information-2010-4-page-4.htm>

Distribution électronique Cairn.info pour A.D.B.S..

© A.D.B.S.. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

[**étude**] Le développement des archives ouvertes a été récemment marqué par l'apparition de services associés – des services qui ne se limitent pas aux seuls documents ou références et qui, recourant à des technologies avancées, prennent en compte les besoins de l'utilisateur final. Après en avoir défini les enjeux proprement scientifiques, **Emma Bester** brosse dans cet article un tableau de l'offre de services actuellement disponibles ; cet état de l'art s'appuie sur une grille d'analyse fonctionnelle appliquée à sept archives ou dispositifs serviciels adossés à des archives ouvertes. L'auteure constate que ces services à valeur ajoutée visent à mettre l'auteur au cœur des données d'information, et s'interroge sur l'exploitation possible de ces dispositifs pour élargir les réseaux scientifiques et renouveler les circuits de repérage et de sollicitation des experts et évaluateurs d'un domaine.

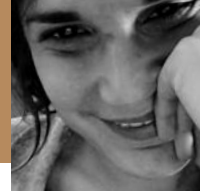
Les services pour les archives ouvertes : de la référence à l'expertise

EN VINGT ANS, LES ARCHIVES OUVERTES SONT DEVENUES des dispositifs significatifs de la communication scientifique dans de nombreux domaines [1]. Entendues au sens strict comme des dépôts de pré et post-publication d'articles scientifiques suivant le principe de l'auto-archivage¹, et imaginées comme supports de communication scientifique directe à coût minimal, les archives ouvertes s'appuient sur des logiciels conçus comme des interfaces rudimentaires de dépôt, de collecte et d'accès aux documents [2].

Les professionnels de l'information scientifique et technique s'engagent désormais à dépasser ce stade infrastructurel et à proposer des services associés de seconde génération, orientés utilisateur final en fonction de besoins identifiés, s'appuyant notamment sur des technologies avancées telles que le *text* et le *data mining* [3] [4] [5] [6]. La question des services à valeur ajoutée a été l'un des enjeux remarqué du congrès Open Repositories 2010 (OR2010) [7].

Si ces réflexions sont également pensées pour les archives ouvertes institutionnelles, les dépôts thématiques, nationaux ou inscrits dans des projets européens et internationaux, sont pressentis pour supporter prioritairement le développement de services : ils s'adressent en effet à des communautés importantes en nombre et bien identifiées, ils bénéficient de soutiens financiers et ils sont le plus à même de permettre des économies d'échelle [8] [9].

Dans cette étude, nous nous proposons, après une présentation des principaux enjeux associés au développement de services pour les archives ouvertes, de dresser un état des avancées et tendances des offres de services actuellement disponibles pour les archives ouvertes. Nous avons étudié les dispositifs sélectionnés, répondant à des critères de fiabilité, de masse critique et de couverture géographique, typologique et disciplinaire, au travers d'une grille d'analyse fonctionnelle. Ce travail s'enrichit également des projets



Diplômée de l'INTD et doctorante en sciences de l'information sous la direction de Ghislaine Chartron, au sein de l'équipe CNAM-DICEN (Dispositifs d'information et de communication à l'ère numérique, EA 4420, <http://dicen.cnam.fr>). Emma Bester s'intéresse particulièrement à l'information scientifique et technique, à la communication scientifique et aux problématiques des archives ouvertes et du libre accès. Elle a récemment publié, avec Pierre Mounier, une communication sur les « Usages des ressources en libre accès dans les bibliothèques universitaires et services communs de documentation. Le cas de Revues.org » dans les actes du colloque *L'information scientifique et technique dans l'univers numérique : mesures et usages* (sous la direction de Chérifa Boukacem-Zeghmouri, ADBS Éditions, 2010).

bester.emma@gmail.com

présentés lors du congrès OR2010 qui s'est tenu en juillet 2010 à Madrid sous l'égide de la FECYT² [10].

Partant du constat que les services associés aux archives ouvertes se déplacent peu à peu des seules références et/ou documents pour mettre l'auteur au cœur des données d'information, la conclusion s'arrête sur l'opportunité d'exploiter ces dispositifs pour renouveler les circuits de mise en visibilité et d'appel à contribution des évaluateurs, rapporteurs ou experts d'un domaine scientifique.

1 DES SERVICES À VALEUR AJOUTÉE POUR LES ARCHIVES OUVERTES : QUELS ENJEUX ?

Ce sont les enjeux proprement internes aux dynamiques scientifiques qui nous intéressent ici. L'enjeu premier de l'implémentation de services avancés réside à notre sens dans la fidélisation des usagers et l'encouragement au dépôt : il s'agit de mieux atteindre l'objectif originel de libre accès aux savoirs scientifiques porté par le mouvement de l'Open Access. La qualité de service de l'offre commerciale (ScienceDirect, SpringerLink ou le Web of Science, par exemple) d'une part et, d'autre part, la multiplication des dispositifs de communication scientifique et les difficultés corrélées d'appropriation par les chercheurs³ constituent des contraintes fortes pour l'évolution des archives ouvertes.

Idee selon laquelle ces archives peuvent s'articuler avec les modalités actuelles d'évaluation de la

¹ Voir la définition proposée sur le site de l'Inist (<http://openaccess.inist.fr/spip.php?rubrique9>). À noter cependant que les archives ouvertes intègrent de plus en plus de documents de nature diverse, notamment dans le cadre institutionnel où elles tendent à regrouper l'ensemble de la production institutionnelle.

² Fundación Española para la Ciencia y la Tecnología

recherche, voire la bousculer, mobilise également la réflexion portée sur les services [11]. Dans l'optique institutionnelle, elle suppose alors l'exhaustivité des dépôts comme condition de réussite⁴. Ces dispositifs sont aussi pressentis pour assurer avec succès les fonctions de collaboration, et plus largement de facilitation de la communication spécialisée et de co-production de la science [12]. Enfin, les archives ouvertes peuvent être envisagées comme supports de visualisation et d'analyse des tendances de la recherche en train de se faire, grâce à l'analyse des contenus et par la mise en visibilité des réseaux d'acteurs [13].

Corrélativement à ces objectifs, il nous semble que, parce qu'elles sont par nature des « équipements d'authentification » [14] et qu'elles supportent les savoirs experts contributifs autant qu'« interactionnels » (i.e. professionnels) [15], les archives ouvertes peuvent concourir à renouveler les modalités de repérage d'un expert. La question mérite d'être abordée

/////

TABLEAU 1 – LES SEPT DISPOSITIFS OBJETS DE CETTE ÉTUDE

	Archives ouvertes				Moissonneurs		Réseau social
	E-LIS	HAL-SHS	PubMed Central	RePEc	Avano	Driver	Research GATE
Date de création	2003	2005	2000	1997	2006	2006	2008
Porteurs institutionnels	AePIC, CILEA (Italie) Ministère de la Culture (Espagne)	CNRS, CCSD (France)	NCBI, NIH, NLM (États-Unis)	Réseau international d'institutions de recherche (70 pays)	Ifremer (France)	Confédération européenne d'institutions de recherche	Initiative individuelle de trois chercheurs*
Nombre de documents déposés **	10 766	24 567	19 millions	925 000	221 960	2,5 millions	400 000 (nombre de membres)
Discipline	Sciences des bibliothèques et de l'information	Sciences humaines et sociales	Sciences biomédicales	Économie	Sciences maritimes et aquatiques	Multi-disciplinaire	Multi disciplinaire
Adresse électronique	http://eprints.rclis.org	http://halshs.archives-ouvertes.fr	www.ncbi.nlm.nih.gov/pmc	http://repec.org	www.ifremer.fr/avano	www.driver-community.eu	www.researchgate.net

*I. Madisch (Harvard University), H. Fickenscher, S. Hofmayer (Medical School of Hannover)

** à la date du 28 avril 2010

///// dans la mesure où les procédures de sélection des experts dans les processus d'évaluation des articles ou des projets de recherche restent aujourd'hui limitées et controversées [11] [16]. Les enjeux de relégitimation des scientifiques comme figures expertes, en ce sein mais également dans leur interaction avec le reste de la société, sont également à prendre en considération [16] [17].

Il s'agit à présent d'examiner comment et dans quelle mesure les services installés sur les archives ouvertes répondent déjà à ces objectifs.

2 MÉTHODOLOGIE

• **Échantillon.** Cette étude, effectuée en février-mars 2010, porte sur une sélection de quatre archives ouvertes et deux méta-archives ouvertes répondant à des critères de fiabilité (longévité, porteurs), de masse critique (nombre de dépôts et de déposants), et de couverture géographique, typologique et disciplinaire (dépôts nationaux ou internationaux, thématiques ou pluridisciplinaires, STM ou SHS). Ont été sélectionnées les archives ouvertes E-LIS⁵, HAL-SHS⁶, PubMed Central et RePEc⁷, et les moissonneurs Avano et Driver⁸ [tableau 1]. Le réseau social ResearchGATE a été adjoindé à cet échantillon pour comparaison. Ce dernier, ainsi que les moissonneurs, sont des exemples de dispositifs serviciels s'adossant à des archives ouvertes ; ils donnent à voir les types de services pensés autour des archives ouvertes.

Grille d'analyse. C'est une approche fonctionnelle qui a été choisie pour analyser ces objets hétérogènes, comme la plus à même d'établir de façon transversale des familles de fonctions et les fonctionnalités servicielles associées.

Tableau des résultats. Le tableau 2 présente, pour chacun des dispositifs étudiés, les familles de fonctions identifiées et les fonctionnalités observées.

3 RÉSULTATS DE L'ÉTUDE

Le développement qui suit détaille les fonctionnalités identifiées [tableau 2]. D'abord les fonctionnalités structurelles d'alimentation, de validation, d'identification, de consultation et d'accession. Puis les fonctionnalités innovantes de personnalisation, de publication, de contextualisation des références, de communication et de collaboration entre usagers. L'étude s'enrichit également, suivant les fonctionnalités, d'une mise en perspective avec des projets présentés lors de la rencontre internationale OR2010, révélateurs des enjeux serviciels mobilisant la communauté des archives ouvertes.

³ Voir à ce propos l'étude du Research Information Network : « If you build it, will they come ? How researchers perceive and use web 2.0 » (2010) : www.rin.ac.uk/our-work/communicating-research/using-relevance-web-20-researchers

⁴ Voir par exemple le workshop des 16 et 17 novembre 2010 organisé en Suède autour de la question : « Les archives institutionnelles peuvent-elles être un support de l'évaluation ? » : www.slu.se/en/library/about/projects/oaworkshop

⁵ E-prints in Library and Information Science

⁶ Hyper Article en Ligne - Sciences de l'Homme et de la Société

⁷ Research Papers in Economics

⁸ Digital Repository Infrastructure Vision for European Research

⁹ Voir par exemple la page « Manuscript Submission System » du NIH pour PubMed Central : www.nihms.nih.gov/help/#slideshow

¹⁰ Munich Personal RePEc Archive <<http://mpra.repec.org/>>

¹¹ International Registry for Authors : www.iralis.org

¹² RePEc Author Short-ID Lookup : <http://ideas.repec.org/cgi-bin/shortid.cgi>

¹³ Distributed Open Access Reference Citations : www.isn-

Alimenter, valider, identifier

À l'étape de l'alimentation, plusieurs types d'acteurs peuvent intervenir : soit usagers directs (lecteurs, auteurs-dépôts, administrateurs alimentant et/ou validant), soit collaborateurs facilitateurs (partenaires éditoriaux, de recherche). Tous peuvent s'appuyer sur divers services d'aide au versement : tutoriels d'utilisation des *workflows* et/ou masques de saisie avec possibilité de contacter les administrateurs⁹, outils d'aide à la mise en conformité des contenus avec les exigences techniques de l'édition électronique (validateurs XML, SGML, etc.), référentiels sur la politique Open Access des éditeurs (lien direct vers Sherpa-Romeo sur HAL-SHS, par exemple) ou des fonds de financement de la recherche (Grant Lookup de PubMed UK). Notons également la médiation institutionnelle, comme dans le cas de RePEc où le MRPA¹⁰ assure cette fonction pour les auteurs dont l'institution n'est pas affiliée à ce projet.

L'archivage nécessite inscription et identification : la gestion des noms des auteurs et de leur affiliation est à cette étape un élément clé, pour ne pas dire problématique (graphies multiples, homonymies, évolutions patronymiques, affiliations plurielles) [18]. Sur HAL-SHS, des index automatiques permettent à l'utilisateur d'éviter les graphies multiples ou la saisie répétée. Sur E-LIS, il est conseillé de s'inscrire dans le répertoire international IRALIS¹¹, à partir duquel est attribué un identifiant à chaque auteur.

Dans le cas de l'archive décentralisée RePEc, il ne s'agit pas de gérer ces problématiques en amont, mais de permettre une harmonisation et une validation en aval, par l'auteur : le module AuthorClaim propose l'ensemble des graphies possibles du nom de l'auteur, que ce dernier peut valider ou invalider. Une requête est ensuite lancée dans le(s) dépôt(s) d'origine de l'auteur, en fonction de l'affiliation qu'il indique, et il peut alors valider ou invalider les items relevés comme étant ou non de son fait. Les références validées alimentent en retour la base de données centralisée du projet RePEc. Un identifiant auteur est attribué par ce biais, auquel on accède en requêtant un nom dans le module RePEc Author Short-ID Lookup¹².

Deux des sessions générales du congrès OR2010 étaient consacrées à ces problématiques. Elles reflètent des stratégies diversifiées selon les projets et les acteurs impliqués. Dans le cas du projet allemand DOARC¹³, visant une interface commune d'accès à des références et citations provenant de dépôts institutionnels éclatés, c'est une identification et une validation en aval par les auteurs qui ont été choisies : le dispositif s'appuiera sur le module AuthorClaim, développé à l'origine pour RePEc et présenté ci-dessus [10a]. Le projet ORCID¹⁴, qui réunit des partenaires commerciaux (Thomson Reuter, Elsevier, Nature, entre autres) et des partenaires institutionnels de type archives ouvertes (arXiv, INSPIRE, ADS, SSRN),



TABLEAU 2 – GRILLE D'ANALYSE FONCTIONNELLE : ÉTUDE COMPARÉE DE SEPT DISPOSITIFS

Les croix de grande taille (X) indiquent des fonctionnalités avancées,
celles de petite taille (x) indiquent des fonctionnalités minimales

Fonctions	Fonctionnalités => Services- Outils	HAL- SHS	E-LIS	RePEc	PubMed Central	DRIVER	AVANO	Research GATE
Identification	S'inscrire	X	X		X	X		X
	S'identifier	X	X	X	X	X		X
	Recevoir un identifiant auteur		X	X				
Alimentation	Déposer ses e-prints et/ou gérer ses réf/doc	X	X		X	X		
	Stocker en ligne d'autres documents							X
	Se faire aider (tutoriels, répertoires autorités, Q/R)	X	X		X			
Validation	Identifier/Valider ses publications			X	X			X
Consultation	Consulter (par catégories, auteurs, documents)	X	X	X	X	X	x	X
	Rechercher (recherche simple, avancée)	X	X	X	X	X	X	X
	Rechercher dans un module spécifique			X	X			
	Rechercher via navigateur (plugin)			X	X			
	Aide à la recherche (SOC)	JEL	JITA	JEL	MeSH			
Accession	Accéder à la référence et/ou au document	X	X	X	X	X	X	X
	Accéder au(x) profil(s) auteur(s) liés à la référence			X	X			X
	Accéder aux références citées			X	X			
	Accéder aux références citant			X	X			
	Accéder aux autres références du/des auteur/s	X		X	X	X		

(Suite du tableau 2 page suivante)

///// vise à constituer un répertoire centralisé d'auteurs auxquels seront attribués des identifiants uniques sur le même principe que les DOI (*digital object identifiers*) et en liaison avec la base de données gérée par Crossref [10b]. Enfin, une troisième stratégie, d'ordre national, était présentée : le répertoire national japonais des chercheurs, le KAKENHI, servira de base pour la constitution d'une base harmonisée et l'attribution

d'un identifiant unique [10c]. Ces projets rejoignent les diverses initiatives, institutionnelles ou privées, locales, nationales ou internationales, déjà prises en ce sens¹⁵.

Parcourir, rechercher

Les tableaux 3 et 4 relèvent les différentes modalités d'accès aux références et/ou documents selon les

SUITE TABLEAU 2 – GRILLE D'ANALYSE FONCTIONNELLE : ÉTUDE COMPARÉE DE SEPT DISPOSITIFS

Les croix de grande taille (X) indiquent des fonctionnalités avancées, celles de petite taille (x) indiquent des fonctionnalités minimales

Fonctions	Fonctionnalités => Services-Outils	HAL-SHS	E-LIS	RePEc	PubMed Central	DRIVER	AVANO	Research GATE
Accession	Accéder à des ressources/ données liées dans des réservoirs externes				X			
Personnalisation	Créer, gérer, alimenter son espace personnel	X	X		X	X		X
	Créer, gérer, alimenter son profil	x		X	X	x		X
	Gérer ses requêtes : sauvegardes, alertes, filtres, historique	X	X		X	X		
	Gérer ses résultats : tri, sélection, conservation, export	X	X	x	X	X	x	X
	Créer des collections de ref/doc				X	X		X
	Créer des collections de profils							X
Publicisation	Publier sa bibliographie	x		X	X	X		X
	Publier son profil			X				X
Communication, collaboration	Contacteur un auteur/membre			X	X			X
	Partager un espace de travail					X		X
	Donner son avis (commentaire, vote)		X					X
	Envoyer vers un autre dispositif (« Share », mail)				X	X		X
Contextualisation	Profil	x		x				
	Statistiques	x		X				
	Ranking			X	X			X
	Network graph							X

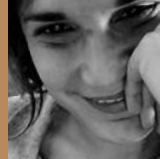


TABLEAU 3 – MODALITÉS DE CONSULTATION PRÉSENTES SUR LES ARCHIVES OUVERTES

	HAL-SHS	E-LIS	RePEc	PubMed Central	DRIVER	AVANO	Research GATE
Domaines, catégories, sujets ; vocabulaires contrôlés	X (HAL-SHS + JEL/ASiC)	X (JITA)	X (JEL)	X (MeSH)			X (RG)
Auteur	X	X	X				X (membre)
Pays, langue		X			X		
Type de documents	X		X		X		
Date/année de dépôt, publication	X	X					
Laboratoire, institution, affiliation							
Publication, revue, série - Éditeurs			X	X			
Collection, communauté	X		X	X	X		X
Ranking			X				X

plateformes. Celles-ci proposent des systèmes plus ou moins complexes d'aide à l'accès à l'information – combinant un nombre plus ou moins élevé de modalités de recherche –, systèmes qui reposent sur des choix documentaires, techniques et technologiques différents.

Les accès par consultation [tableau 3], valorisés sur HAL-SHS, E-LIS et Driver, dépendent essentiellement de la structuration des bases de données, des champs moissonnés et/ou alimentés, et d'une requête préétablie pour consultation. Ce sont ces champs de consultation qui composent les formulaires de recherche avancée [tableau 4]. *A contrario*, c'est la technologie du moteur de recherche qui est privilégiée sur PubMed et ResearchGATE. Ce dernier vante ainsi la technologie des algorithmes sémantiques : « *Are you working on an abstract? Paste a snippet of an abstract and find similar publications through our semantic algorithms.* »

La consultation par domaine ou sujet dépend de classements définis en amont, qui peuvent être propres au dispositif (comme les catégories proposées par HAL-SHS ou ResearchGATE) ou dépendre de classifications partagées par une communauté disciplinaire (code JEL en économie, par exemple). L'accès aux publications en série dépend quant à lui des partenariats contractés avec les éditeurs, tel que cela apparaît dans PubMed Central et RePEc (partenariats annoncés avec Elsevier et Springer, entre autres).

oldenburg.de/
projects/doarc2
14 The Open Researcher and Contributor ID : www.orcid.org
15 Voir la page Author Identification du wiki International Repositories Infrastructure : <http://repinf.pbworks.com/Author-identification>
16 New Economics Papers : <http://nep.repec.org>

Personnaliser : veille, gestion documentaire

À ces modalités d'accès par recherche ou consultation s'ajoutent des fonctionnalités propres à la veille, qui peuvent être intégrées au dispositif ou faire l'objet de modules complémentaires [tableau 5]. On peut ainsi, dans HAL-SHS, créer par flux RSS ou courriel des alertes relatives à des domaines, des auteurs et/ou des mots-clés du titre des *e-prints*. Dans Driver et PubMed, il est possible de conserver ses requêtes personnalisées et de les transformer en alertes. Dans le cadre du projet RePEc, le module NEP¹⁶ offre de s'abonner à des flux thématiques prédéterminés, alimentés des nouveaux *working papers* par des experts. Dans le cas de ResearchGATE, la souscription à des notifications par courriel va concerner essentiellement l'aspect social, vie du/de mon réseau (nouveaux membres dans un groupe, nouveaux posts dans un groupe ou d'un membre en particulier, etc.). Mais certaines notifications concernent les publications : nouvelles publications dans une catégorie, un groupe, un profil – et, le plus intéressant peut-être : la possibilité d'être informé lorsqu'une référence demandée devient disponible en texte intégral.

Des fonctionnalités de gestion documentaire, plus ou moins avancées selon les dispositifs, complètent encore ces offres de service. Hormis l'exportation de références bibliographiques et le téléchargement de documents, dont la valeur ajoutée augmente avec la

/////

///// diversité des formats et environnements supportés, les fonctionnalités avancées requièrent la création d'un espace personnel [tableau 5].

Non envisagée sur Avano, l'offre de service associée y est de fait réduite au minimum. Les fonctionnalités de gestion documentaire sont respectivement nulle sur RePEc, où la création d'un espace personnel (AuthorPage) est destinée à l'authentification du chercheur et de ses publications, et faible sur E-LIS, où l'espace Your Page sert uniquement à la gestion des *e-prints* déposés. De même sur Hal-SHS, où il est possible de configurer des alertes (veille), mais pas de gérer ses propres résultats de requêtes.

Des options de conservation et d'organisation en collections sont par contre présentes dans les modules My NCBI (PubMed) et My Driver (Driver). La phase II du projet Driver prévoit plus largement de déployer l'outil ScholarLink [10d], qui est un gestionnaire de références bibliographiques orienté à la fois *desktop* et Web, pensé pour faciliter la constitution de collections partagées. Plus proche encore du travail personnel du chercheur, le *plugin* MePrint, présenté à OR2010, proposera bientôt des fonctions d'analyse des données bibliographiques (occurrence des mots-clés) [10e]. Enfin, s'il n'est pas proposé sur ResearchGATE d'organiser ses références en collections, il y est par contre possible d'ajouter d'autres références bibliographiques à celles collectées via le dispositif et donc de centraliser ses références. Le service emprunte ici aux agrégateurs participatifs en ligne de références bibliographiques type CiteULike, sans en offrir les fonctionnalités de *tagging*, propres à la qualification et à l'organisation induites des références.

Publiciser

Ces options documentaires peuvent, le cas échéant, enrichir les pages publiques des membres inscrits aux dispositifs [tableau 5]. La publicisation, ou mise en

visibilité publique volontaire, peut concerner des données biographiques et/ou bibliographiques. Elle peut être interne au dispositif ou externalisée et/ou intégrée à un autre dispositif médiateur grâce à l'attribution d'une URL pérenne (*permalink*).

Dans le cas de RePEc, les données biographiques, minimales et de nature institutionnelle, sont complétées par la liste des publications du chercheur, collectée via le module AuthorClaim déjà présenté ; une adresse URL pérenne est alors associée à chaque page auteur. Sur PubMed et Driver, un *permalien* est attribué pour chacune des collections constituées, mais il ne s'agit pas ici de pages publiques internes au dispositif, personnelles et personnalisables. Dans ResearchGATE, données biographiques libres, bibliographie auteur (My Publications) et bibliographie générale (My Library) sont au contraire internes et constitutives du dispositif ; seuls les membres inscrits peuvent y accéder directement ou après en avoir fait la demande, selon le paramétrage des droits d'accès au profil. Il est également possible, sur HAL-SHS, de constituer une page publique via une liste des références déposées et des *permalien*s.

Accéder à des documents et/ou des références, des ressources bibliographiques liées et des données primaires liées

Outre l'accès à des références et/ou des documents, certaines plateformes proposent des accès à des données bibliographiques liées à la référence consultée.

Il est ainsi possible sur HAL-SHS et Driver, contrairement à E-LIS ou Avano, d'accéder à l'ensemble des publications d'un auteur ou de son institution d'affiliation contenues dans le dépôt. Le service EconomistsOnline, adossé aux données de RePEc et enrichi des références d'institutions académiques, donne accès, à partir d'une référence, à toutes les publications d'un auteur, aux articles de la revue source, ainsi

Références

[1] Hélène BOSCH. « Archives ouvertes : quinze ans d'histoire ». In : C. Aubry, J. Janik (dir.). *Les archives ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information*. Paris : ADBS Éditions, 2005. P. 27-54

[2] Thierry CHANIER. *Archives ouvertes et publication scientifique : comment mettre en place l'accès libre aux résultats de la recherche ?* Paris : L'Harmattan, 2004

[3] Ju JINGJIE. « Leveraging levels of information services and developing knowledge services. The trend of information services in libraries ». *Library Management*, 2006, vol. 27, n° 6/7, p. 354-361

[4] Robert CHAVEZ, Gregory CRANE, Anne SAUER, Alison BABEU, Adrian PACKEL, Gabriel WEAVER. « Services make the repository ». *Journal of Digital Information*, 2007, vol. 8, n° 2.

<https://journals.tdl.org/jodi/article/viewArticle/195/179>

[5] ASSOCIATION OF RESEARCH LIBRARIES. *The research library's rôle in digital repository services*. Washington : ARL, 2009. 52 p.

www.arl.org/bm-doc/repository-services-report.pdf

[6] Les CARR, Tim BRODY. « Citation Services (Overview) ». In : International Workshop on Co-operation on Citation Services

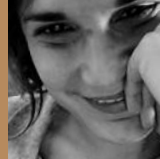


TABLEAU 4 – MODALITÉS DE RECHERCHE PRÉSENTES SUR LES DISPOSITIFS

	HAL-SHS	E-LIS	RePEc	PubMed Central	DRIVER	AVANO	Research GATE
Formulaire de recherche simple (mot-clé)	X	X	X	X	X	X	X
Formulaire de recherche avancée (équations booléennes)	X	X	X		X		
Affinage de la recherche par facette				X	X*		
Recherche dans l'historique				X	X	X	
Équations complexes, langage de recherche avancé			X	X		X	
Aide à la recherche multilingue (traduction des requêtes)			Economists Online		X		
Recherche sémantique d'un ensemble textuel							X
Modules de recherche spécifique externes			X**	X***			
Recherche intégrée au navigateur (plugin)			X	X			

* La recherche à facette n'est pas proposée pour la recherche fédérée mais lorsqu'on effectue une recherche dans un dépôt spécifique ; les facettes s'appuient sur la structuration des bases de données des dépôts d'origine.

** EDIRC : répertoire d'institutions de recherche

*** Citation Matcher : recherche de références bibliographiques ; Authority, ScholarUniverse : recherche d'experts

and Repositories, 14th June 2010, Utrecht, Netherlands.
<http://eprints.ecs.soton.ac.uk/21260>
 [7] Philip HUNTER, Robin TAYLOR. « Report on Open Repository 2010 ». Ariadne, 2010, n° 64. www.ariadne.ac.uk/issue64/or-10-rpt
 [8] Dave PUPPLET. « Subject repositories ». Ariadne, 2010, n° 62. www.ariadne.ac.uk/issue62/bl-subject-repos-rpt
 [9] Chris ARMBUSTER, Laurent ROMARY. *Comparing repository types*.

Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. 2009 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1506905
 [10] FUNDACION ESPANOLA PARA LA CIENCIA Y LA TECNOLOGIA (FECYT), MINISTERIO DE CIENCIA Y INNOVACION. OR2010. The 5th International Conference on Open Repositories.

Madrid. 6-9 juillet 2010. <http://or2010.fecyt.es>
 [10a] Michael MAUNE, Eberhard R. HILF. « Open Access Reference Citations Service with author-identification ». *Open Repository 2010, general session 1B: Citation & Bibliography*
 [10b] Simeon WARNER. « Author identifiers ». *Open Repository 2010, general session 4A: Data integration and desambiguation*

qu'aux articles indexés avec un même code de la classification JEL. EconPaper (RePEc) indique en outre les références citées dans l'article sélectionné ainsi que les articles le citant [figure 1].

L'ensemble de ces fonctionnalités est proposé dans PubMed Central. Au « *link to* » s'ajoute ici le « *link out* », les références renvoyant également à l'ensemble plus vaste PubMed.gov et étant mises en liaison avec les données du portail Entrez Database¹⁷ : *datasets* (bases de données primaires) multiples (GenBank, Nucléotides, etc.), bases de données bibliographiques, sites de référence comme Medline, sites d'autorités (thésaurus, taxonomies). De la même manière, le projet NEEO se propose, à terme, d'adjoindre des accès à des données liées pour les références indexées dans EconomistsOnline¹⁸.

Accéder à des profils et/ou des « experts »

Si les références bibliographiques et les documents sont au cœur des enjeux documentaires traditionnels, les offres de services et fonctionnalités visent de plus en plus l'individu, auteur d'un article, expert d'un domaine de recherche et/ou membre actif d'un réseau social. Ce sont des approches servicielles exploitées sur trois des dispositifs étudiés : PubMed, RePEc et ResearchGATE [tableau 2].

Le cas de ResearchGATE [figure 2] est particulier car, en raison de sa nature de réseau social, son fonctionnement repose, pour moitié, sur les données des profils des membres ; le second pilier réside dans les actions des utilisateurs, qui viennent généralement nourrir en retour les profils (le sien, celui des autres) [tableau 5]. L'accès à des profils est ici la raison même du dispositif : les membres ou *researchers* sont considérés comme des ressources à part entière, que l'on peut « *bookmark* » (collectionner), annoter (tags, commentaires), contacter. Pour faciliter le push personnalisé via la « *technologie des algorithmes*

sémantiques » – visant la mise en relation avec des membres, des groupes et des références bibliographiques –, les usagers sont invités à exprimer leurs centres d'intérêt en alimentant massivement leur profil.

Dans RePEc, des profils auteurs sont accessibles sur IDEAS et EconPaper. À partir des informations fournies par les auteurs eux-mêmes via le module d'enregistrement Author Service, sont présentés les listes de leurs publications et de leurs affiliations ainsi que leurs contacts possibles. Le profil est complété par les codes JEL d'indexation des publications et par les données statistiques de téléchargement et citation obtenues via CitEc et LogEc [figure 1]. Le profil est donc constitué à partir d'informations institutionnelles, des données prélevées sur les références bibliographiques et des usages quantifiés associés. Même si le module d'identification Author Service contient un champ de texte libre, les informations formelles dominent largement ; elles offrent une vue sur la présence de l'auteur dans le dispositif, mais l'accès aux références bibliographiques reste l'enjeu premier.

Deux modules de l'ensemble PubMed, Authority [figure 3] et Scholar Universe, sont quant à eux explicitement présentés comme des banques de données d'experts :

- « *The information provided on Authority.com web site is intended to help visitors to find experts in the specific areas of life-sciences [...]. Authority.com mission is to identify the experts and help visitors to find and evaluate them.*¹⁹ »

- « *Scholar Universe delivers reliable expert search results when searching for qualified experts based on their expertise, location, background, and more.*²⁰ »

Reposant sur la technique du datamining, des profils automatiques sont constitués à partir des informations disponibles dans une ou plusieurs bases de données, agrégeant liste des publications, mentions d'affiliation, contacts possibles (téléphone, adresse

¹⁷ Voir le sitemap du NCBI : www.ncbi.nlm.nih.gov/Sitemap
¹⁸ <http://dvn.iq.harvard.edu/dvn/dv/NEEO>

Références

[10c] Kei KURAKAWA, Hideaki TAKEDA, Akiko AIZAWA, Masao TAKAKU. « Researcher Name Resolver: A framework for researcher identification in Japan ». *Open Repository 2010, general session 4A : Data integration and desambiguation*
[10d] Gabriella KAZAI, Paolo MANGHI, Antonis LEMPESIS. « ScholarLink: A Reading List Metaphor for Collaborative Research ». *Open Repository*

2010, poster session 28
[10e] Patrick MCSWEENEY. « The Semantics of Reading Lists ». *Open Repository 2010, eprints Users Group : session 3A*
[11] Muriel LEFEBVRE. « L'évaluation des savoirs scientifiques : modalités et enjeux ». In : J. Schöpfel (dir.). *La publication scientifique : analyses et perspectives*. Paris : Hermès, 2008. P. 299-316

[12] Évelyne BROUDOUX, Ghislaine CHARTRON. « La communication scientifique face au Web2.0 Premiers constats et analyse ». *H2PTM'09 - Rétrospective et perspective - 1989-2009. H2PTM, France (2009)*
[13] Katy BÖRNER, Weixia HUANG, Micah LINNEMEIER, Russell J. DUHON, Patrick PHILIPPS, Nianli MA, Angela M. ZOISS, Han-ning GUO, Mark A. PRICE. « Rete-netzwerk-red : analyzing



TABLEAU 5 – FONCTIONNALITÉS DE GESTION DES DONNÉES BIOGRAPHIQUES ET BIBLIOGRAPHIQUES SUR LES DISPOSITIFS

	HAL-SHS	E-LIS	RePEc	PubMed Central	DRIVER	AVANO	Research GATE
Gestion personnalisée							
- Créer un espace personnel	Mon espace	Your pages	Author Page	My NCBI	My Driver		Home > My Profil
Gestion du profil							
- Publier des éléments biographiques	Nom-prénom, affiliation	Nom-prénom, affiliation	Nom-prénom, affiliation	Username	Username		Exhaustif, avec champs libres (texte)
- Publier des éléments bibliographiques : ses publications	Page personnelle (permalink)		via AuthorClaim (permalink)	My bibliography (permalink)	via collection (permalink)		My Publications
autres références				My bibliography	via collection (permalink)		My Library
Gestion documentaire							
- Gérer ses références et ses documents :							
déposer, gérer ses dépôts	X	X					
collecter (rechercher) ses publications	X		X AuthorClaim	X	X		X
stocker / uploader d'autres références							X
organiser en collections				X	X		
- Gérer ses résultats de recherche (références / documents) :							
conserver				X	X Dropbox		X
exporter / télécharger (formats)	X	X	X	X	X	X	X

and visualizing scholarly networks using the Network Workbench Tool ». *Scientometrics*, Budapest, 2010. DOI 10.1007/s11192-009-0149-0

[14] Yves TREPOS. *La sociologie de l'expertise*. Paris, PUF : 1996

[15] Harry COLLINS, Robert EVANS. *Rethinking Expertise*. Chicago and London: The University of Chicago Press, 2007

[16] Catherine VILKAS. « Des pairs aux experts : l'émergence d'un "nouveau management" de la

recherche scientifique ? » *Cahiers internationaux de sociologie*, 2009, vol. 1, n° 126, p. 61-79

[17] Isabelle SABELLE BERREBI-HOFFMANN, MICHEL LALLEMENT. « À quoi servent les experts ? ». *Cahiers internationaux de sociologie*, 2009, vol. 1, n° 126, p. 5-12

[18] Dorothea SALO. « Name authority control in institutional repositories ». *Cataloging and Classification Quarterly*, 2009, vol. 47, n° 34, p. 249-261

[19] Ghislaine CHARTRON. *Les chercheurs et la documentation numérique : nouveaux services et usages*. Paris: Éditions du Cercle de la librairie, 2002

[20] Gabriel GALLEZOT, Olivier LE DEUFF. « Chercheurs 2.0 ? » *Les Cahiers du numérique*, 2010

FIGURE 1 – ACCÉDER À DES DONNÉES BIBLIOGRAPHIQUES, BIOGRAPHIQUES ET STATISTIQUES SUR ECONPAPER (RePEc)

1 Cliquer sur le nom de l'auteur pour accéder à ses contacts, mentions d'affiliation et la liste de ses articles et working papers
 2 Voir les références de la base RePEc qui citent cet article (module CitEc)
 3 Cliquer pour voir les références citées dans l'article et présentes dans RePEc
 4 Voir les statistiques de consultation de cet article (nombre de fois où le résumé vu, nombre de téléchargements) (module LogEc)

FIGURE 2 – AUTHORATORY (BANQUE D'EXPERTS) : CONSTITUTION DE PROFILS AUTEURS AUTOMATIQUES (EXPLOITATION DES DONNÉES BIBLIOGRAPHIQUES DE PUBMED CENTRAL PAR DES TECHNIQUES DE DATA MINING)

électronique), co-auteurs, statistiques de téléchargements et de citations, domaines d'expertise via analyse de texte intégral et des termes d'indexation. Les profils Scholar Universe sont signalés, quand ils existent, dans les notices de PubMed par un « link to ». Cependant, la base fonctionne aussi parallèlement de manière autonome. Il ne s'agit plus, ici, de trouver des références bibliographiques mais des auteurs-experts en tant que tels. Si le service Authoratory est entièrement libre de droit pour les organisations à but non lucratif, l'accès aux données complètes de ScholarUniverse, généré par la société COS-Refworks, est payant. D'autres sociétés privées se positionnent également sur ce qui apparaît comme un « marché de l'expertise ». C'est le cas, par exemple, de Expernova²¹, qui moissonne les entrepôts OAI-PMH afin de constituer et vendre aux entreprises des banques de données d'experts.

Contextualisation vs évaluation

Ainsi les références, de plus en plus, sont présentées environnées d'éléments qualitatifs et quantitatifs, en offrant une vue contextualisée : profil plus ou moins exhaustif de l'auteur et des co-auteurs (notamment les affiliations), références citées ou citant le document et statistiques de consultation ont été mentionnés dans les cas de RePEc (EconPaper), PubMed et ResearchGATE. S'ajoutent dans ce dernier des options de visualisation du réseau de l'utilisateur membre, via un *network graph* ou carte où sont matérialisées les liaisons entre un membre et ses publications et/ou bibliographies, ses contacts et ses groupes. Des indications de *ranking* (classement) sont aussi proposées, fondées sur les votes des utilisateurs (*rating*).

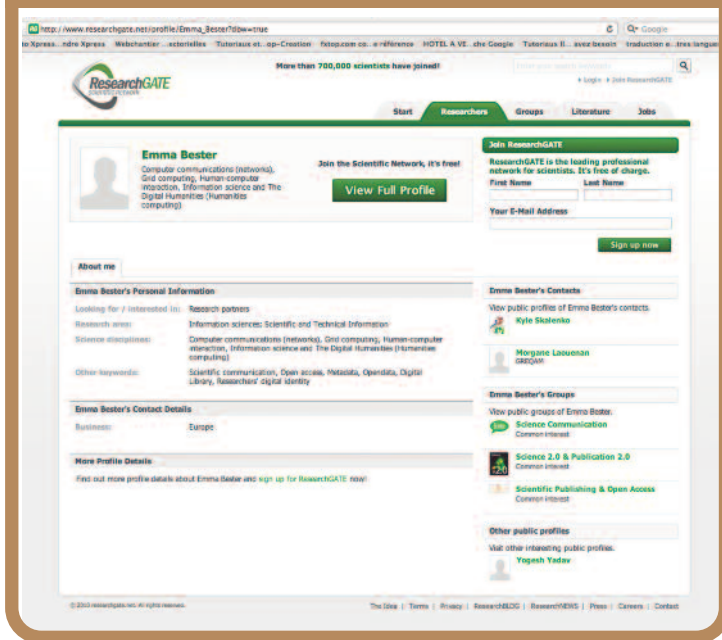
La base IDEAS (RePEc) propose également un accès dédié par *ranking* des auteurs et institutions et des documents les plus cités ou téléchargés²². Il procède d'une combinaison entre les données obtenues par les modules LogEc (statistiques de consultation et de téléchargement) et CitEc (taux de citation). Dans ce cas, on accède aux ressources directement à partir du classement : le facteur d'impact l'emporte sur la requête. C'est une logique distincte, qui s'est établie pour concurrencer, ou tout au moins équilibrer, les pratiques controversées d'évaluation s'appuyant principalement sur les données bibliométriques de corpus limités de revues [11].

Communiquer, collaborer

Une dernière famille de fonctionnalités concerne les fonctions de communication et de collaboration, entre auteurs et lecteurs d'une part et entre membres du dispositif d'autre part. Il est ainsi possible d'entrer directement en contact avec un auteur dans RePEc (IDEAS et EconPaper) et sur PubMed où les adresses électroniques sont notifiées respectivement sur les



**FIGURE 3 – RESEARCHGATE (RÉSEAU SOCIAL) :
CONSTITUTION DE PROFILS PERSONNALISÉS
PAR LES MEMBRES DU RÉSEAU**



tations servicielles actuelles se déplacent de plus en plus vers l'accès à des profils auteurs plus ou moins personnalisables et/ou proposent une vue contextualisée des références bibliographiques, en ce qu'elles sont augmentées de données qualitatives et quantitatives. Portées à l'international pour un champ scientifique donné, ces logiques combinées peuvent, de façon parallèle et complémentaire aux procédures et circuits de l'édition traditionnelle, concourir à l'amélioration de l'expertise scientifique [12]. Il sera intéressant d'observer sous cet angle les modalités d'usage des profils auteurs et des bases de données externalisées d'experts, telles qu'elles sont respectivement proposées sur RePEc et PubMed.

On peut également imaginer des approches servicielles tierces de type *mashup* , par exemple, donnant davantage la main aux chercheurs pour connecter l'ensemble de leurs activités de recherche et valoriser leur identité numérique [20]. Quelles que soient les approches choisies, il convient de s'intéresser aux archives ouvertes comme support de mise en visibilité des scientifiques et de leur(s) domaine(s) d'expertise, à même d'élargir les réseaux établis des habitudes de travail, d'enrichir les comités de lecture et d'aider à la mise en relation d'experts au niveau international. ●

Novembre 2010

pages profils ou sur les notices des références consultées. Dans Driver et ResearchGATE, les fonctions de communication et de collaboration sont internes au dispositif, elles permettent aux membres d'entrer en contact, de constituer des groupes et de mettre en commun un espace de travail (stockage de fichiers, collections partagées, calendrier partagé).

4 CONCLUSION ET PERSPECTIVES

Cette étude comparée a permis de repérer certaines avancées et tendances pensées pour les archives ouvertes et autour d'elles en terme d'offre de services à valeur ajoutée. Elles dépendent de la nature ou des objectifs premiers des dispositifs (archives ouvertes *stricto sensu*, moissonneurs, réseaux sociaux), des techniques documentaires et technologies déployées (qualité des métadonnées, référentiels d'indexation, *data* ou *text mining*), de la volonté et de la capacité portées de développement (équipe dédiée d'E-Print, développements de Thomas Krichel et Christian Zimmermann pour RePEc, soutien du JISC à PubMed UK, etc.), ou encore des partenariats engagés (avec les éditeurs privés, par exemple).

Elle montre plus spécifiquement que, au-delà de la double fonctionnalité structurelle de recherche et d'accès à des références bibliographiques, au-delà des enjeux d'intégration interne et externe [19], les orien-

- 17 Voir le sitemap du NCBI : www.ncbi.nlm.nih.gov/Sitemap
- 18 <http://dvn.iq.harvard.edu/dvn/dv/NEEO>
- 19 www.authoria-tory.com/about/about.htm
- 20 www.scholaruniverse.com/productinfo.jsp
- 21 www.expertnova.com
- 22 <http://ideas.repec.org/top>