

Archives ouvertes, archives institutionnelles et protocole français

« **L**es archives ouvertes, lit-on dans la plaquette de présentation diffusée en 2006 et disponible sur le site Archives ouvertes¹, reposent sur la mise en place de serveurs où sont déposés et sauvegardés les documents issus de la recherche (projets d'articles, textes soumis pour publication, articles publiés...). La possibilité d'enregistrer les versions successives d'un texte enrichit la notion de document scientifique et peut donner naissance à de nouvelles pratiques d'écriture de chercheurs. Les archives ouvertes permettent d'accélérer la communication scientifique et de l'ouvrir gratuitement à tous les chercheurs qui ont accès à l'Internet. Cet accès libre et immédiat accroît la visibilité et souvent l'impact des résultats de la recherche. »

Rachel Creppy

Sous-direction des bibliothèques
et de l'information scientifique
rachel.creppy@education.gouv.fr

Un protocole français a été mis en place « pour une approche coordonnée, au niveau national, pour l'archivage ouvert de la production scientifique² ».

Le contexte européen

OpenDOAR³, le répertoire d'archives ouvertes universitaires maintenu par l'université de Southampton recense au 15 octobre 2007 un millier d'entrepôts, soit environ 15 % de la production scientifique courante mondiale. La moitié de ces entrepôts sont en Europe. C'est la marque

de l'engagement européen dans le mouvement international des archives ouvertes et de la communication scientifique directe, initié dans les années 1990 par des communautés américaines de chercheurs et de professionnels de l'information - ArXiv⁴ (1991) en physique et PubMed Central⁵ (2000) en sciences de la vie. Le mouvement collectif engagé en 2001 par l'initiative de Budapest (Inar, ABF, INPT...) a été solennisé en 2003 par la Déclaration de Berlin dont les premiers signataires français furent le CNRS, l'Inserm, l'Inria, l'Institut Pasteur, l'EPHE et l'université Lyon-II. C'est aussi en 2003 que l'Ifla et le Sommet mondial de la société de l'information ont pris position en faveur du libre accès. On peut mesurer l'engagement des acteurs institutionnels sur le site Juliet⁶, répertoire mondial

1. www.archives-ouvertes.fr/IMG/pdf/ArchivesOuvertes07.pdf

2. Cet article, qui fait le point sur son état d'avancement, est une mise à jour de la présentation faite par Danièle Duclos-Faure lors de la réunion annuelle des directeurs de bibliothèques universitaires les 22 et 23 janvier 2007. www.sup.adc.education.fr/Bib/intro/reuDir/2007/ao.pdf

3. www.opendoar.org

4. www.arxiv.org

5. www.pubmedcentral.nih.gov

6. www.sherpa.ac.uk/juliet

Titulaire d'un DEA en sciences de l'information et de la communication, des Capes en documentation et en histoire géographique et du DCB, **Rachel Creppy** est chargée des réseaux et projets coopératifs à la Sous-direction des bibliothèques et de l'information scientifique. Elle a auparavant été conservateur aux SCD des universités du Littoral-Côte-d'Opale et Paris-I, puis à l'Abes.

de la politique d'archives ouvertes des institutions de recherche, maintenu par Sherpa, le consortium britannique pour les archives ouvertes.

Leur effort est relayé par la Commission européenne, qui, en janvier 2006, dans une étude sur l'évolution économique et technique du marché des publications scientifiques en Europe, recommande l'accès libre et rapide aux résultats de la recherche publique, dans le respect des droits d'auteurs et droits voisins. De même Eurab (European Research Advisory Board), le conseil consultatif européen en matière de choix scientifiques, recommande, dans son rapport de 2006, un embargo de six mois au plus pour la mise à disposition ouverte des post-publications. Cet avis est repris par l'European Research Council-Scientific Council en septembre 2007. L'OCDE, elle aussi, émet en 2007 des recommandations favorables à l'accès ouvert aux données de la recherche financée sur fonds publics.

Enfin, l'Union européenne est à l'origine du projet Driver, Digital Repository Infrastructure Vision for European Research⁷, qui réunit sept universités, la fondation Surf (Pays-Bas), le Consiglio nazionale delle ricerche (Italie) et le CNRS (France) pour construire une infrastructure d'entrepôts ouverts d'articles et résultats de la recherche, ainsi que pour proposer une interface de consultation de ce réseau. Après une enquête, à laquelle trop peu d'établissements français ont pu répondre mais qui a rendu visible la diversité des modèles nationaux d'organisation des archives ouvertes, le projet Driver s'est engagé

7. www.driver-repository.eu

Archive ouverte ou institutionnelle ?

Le terme archive ne doit pas être considéré selon l'acception usuelle: ce néologisme est traduit de l'anglais « *archive* » que le mot « *repository* » tend à remplacer, et on peut y substituer dépôt, réservoir, entrepôt ou, dans certains cas, bibliothèque numérique.

Une archive institutionnelle est le lieu où une institution regroupe l'ensemble de sa production (de recherche, patrimoniale, pédagogique, administrative...) dans des espaces privés ou ouverts, comme le fait le Cern, Centre européen de recherche nucléaire¹.

Les archives institutionnelles des établissements d'enseignement supérieur et de recherche français sont consultables à partir de leur site web².

Une archive ouverte met à disposition sur internet des publications en texte intégral et en libre accès (*Open Access*).

1. <http://cdsweb.cern.ch>
2. www.education.gouv.fr/pid305/adresses-utiles.html

Le modèle national néerlandais : DARE (Digital Academic Repositories) ¹

L'archive ouverte nationale, DAREnet, contient 135 875 publications scientifiques des universités et organismes de recherche néerlandais.

Un réseau à l'architecture répartie fédère les dépôts institutionnels dans DAREnet, programme de la fondation Surf². Il regroupe toutes les universités, la bibliothèque nationale, l'académie royale des arts et des sciences (KNAW) et l'organisation néerlandaise pour la recherche scientifique (NWO).

Une approche qualitative: la base met en exergue deux sélections, Cream of Science, 48 021 publications de 200 scientifiques d'excellence et Promise of Science, 15 735 thèses (résultats au 15 octobre 2007)

La conservation pérenne est assurée par la Bibliothèque nationale des Pays-Bas.

L'architecture de DARE sert de modèle au projet Driver.

1. www.darenet.nl/nl/page/language.view/search.page
2. www.surf.nl/en/Pages/home.aspx

dans la phase de test des interfaces et des services dont la livraison est prévue fin 2007.

Le dispositif français en 2007

Le 6 juillet 2006, les principaux organismes de recherche et les établissements d'enseignement supérieur et de recherche représentés par la conférence des présidents d'université (CPU) et la conférence des grandes écoles (CGE) ont signé pour deux ans un Protocole d'accord en vue d'une approche coordonnée, au niveau national, pour l'archivage ouvert de la production scientifique. Aux signataires initiaux⁸ sont venus s'ajouter au printemps 2007 quatre nouveaux participants⁹. L'approche française est originale: alors que les autres pays ont le plus souvent adopté une approche fédérant des archives institutionnelles distribuées, archives-ouvertes.fr prévoit une plate-forme partagée, héritière de Hal (Hyper article en ligne), l'outil d'archive ouverte développé par le Centre pour la communication scientifique directe du CNRS, le CCSD. Cette plate-forme copilotée par les signataires du protocole sera interopérable avec les systèmes d'information locaux des établissements.

Les objectifs du protocole sont les suivants:

- pour les institutions: identifier, diffuser, valoriser, promouvoir, suivre

8. Cemagref, Centre national du machinisme agricole, du génie rural et des eaux et forêts; Cirad, Centre de coopération internationale en recherche agronomique; CNRS, Centre national de la recherche scientifique; CPU, Conférence des présidents d'université; Inra, Institut national de la recherche agronomique; Inria, Institut national de recherche en informatique et en automatique; Inserm, Institut national de la santé et de la recherche médicale; Institut Pasteur; IRD, Institut de recherche pour le développement; Conférence des grandes écoles.

9. CEA, Commissariat à l'énergie atomique; Ineris, Institut national de l'environnement industriel et des risques; Ifremer, Institut français de recherche pour l'exploitation de la mer; Inrets, Institut national de recherche sur les transports et leur sécurité.

la production scientifique des chercheurs et enseignants-chercheurs ;

- pour les chercheurs : communiquer leurs travaux à la communauté internationale, accélérer les échanges scientifiques, renforcer les facteurs d'impact ;

- améliorer la visibilité de la recherche française.

L'évaluation de la recherche n'est pas un objectif direct du projet. Il est convenu de :

- développer en commun une plateforme partagée de dépôt et la copiloter ;

- mettre en place un dépôt direct ou un dépôt indirect via le système d'information des établissements.

Avant 2009, le cadre juridique de la plateforme partagée et du CCSD devra être défini. Le Comité stratégique des archives ouvertes (Costrao) propose un groupement d'intérêt public (GIP), afin que toutes les par-

ties prenantes à la recherche puissent être représentées.

Les travaux des signataires du protocole

Le Comité stratégique des archives ouvertes définit les axes stratégiques, établit un plan de travail et désigne les représentants du comité scientifique et technique (Cost). Il suit directement deux chantiers, la rédaction d'un guide juridique pour les déposants et la mise en place d'une structure administrative pour la plateforme partagée.

De nombreux travaux sont en cours au sein du Constrao, et de son instance opérationnelle, le Cost, composé de chercheurs, d'informaticiens, de professionnels de la documentation des différents organismes signataires ou membres du Constrao. Un bureau a été créé en décembre 2006 pour assurer le suivi des dossiers et produire un document d'orientation générale.

Les travaux s'ordonnent selon plusieurs axes.

La communication

Elle s'adresse tant aux chercheurs, pour les sensibiliser à la nécessité de déposer, qu'aux décideurs et financeurs de la recherche, et aux éditeurs pour les convaincre de la complémentarité des archives ouvertes et du modèle traditionnel de publication scientifique. La sensibilisation des instances décisionnelles des universités a été jugée prioritaire, afin d'obtenir leur concours pour alimenter l'archive. Le 21 septembre 2007, Hal a passé le cap de 50 000 publications déposées. Mais le taux de dépôt spontané reste bas, il peut être amélioré par l'investissement des personnels de la documentation, le volontarisme des organismes, comme cela a été observé au Cemagref ou à l'Ineris, et enfin par des incitations financières.

La France en 2006 selon Driver

20 réponses pour la France : Ifremer, CNRS, universités Lyon-II et Reims, INP de Toulouse, Paristech, Pastel, École normale supérieure Lettres et sciences humaines...

86 % contiennent des thèses et 71 % des articles. 71 % des articles sont des *postprints*, 28,8 % des articles publiés.

Toutes les archives ouvertes recensées offrent des documents ouverts, quelques-unes (28 %) offrent également des documents accessibles sur le campus uniquement.

37 % sont des documents en sciences humaines et sociales (29 % en Europe), 38 % en sciences de l'ingénieur (21 % en Europe).

La structure des métadonnées

Les informations les plus sensibles sont celles qui décrivent les affiliations des chercheurs. Les universités étant soucieuses d'une bonne identification des publications des unités mixtes, ces données font l'objet d'une concertation organisée par le ministère de l'Enseignement supérieur et de la Recherche (Service de la recherche universitaire).

Autre information difficile, la nomenclature des disciplines qui ne fait pas encore l'objet d'un consensus. Des sous-groupes ont été constitués pour élaborer un glossaire, pour lister les types de publications et les référentiels des disciplines qui seront utilisés. L'intervention des professionnels de la documentation devrait permettre que les informations minimales exigées du chercheur lors du dépôt soient enrichies, structurées et validées selon des référentiels stabilisés pour garantir leur diffusion internationale et leur pérennisation. La production des référentiels eux-mêmes n'entre pas dans le périmètre des groupes de travail.

Le portail anglais : Intute *

Intute a choisi d'offrir des archives institutionnelles associant tous les types de documents utiles à l'enseignement supérieur et à la recherche, du blog à la base de données, en passant par le manuel électronique et les résultats bruts de la recherche : recherche et pédagogie sont associées pour placer l'utilisateur plutôt que les multiples producteurs au centre du dispositif. Une interrogation limitée aux archives ouvertes reste possible pour les 157 000 publications scientifiques provenant de 81 entrepôts académiques.

Le JISC, Joint Information Systems Committee, a pour stratégie de soutenir et de coordonner le développement et l'interopérabilité de dépôts institutionnels dans chaque université dans le cadre du Repositories Support Project. Un service central, The Depot, est proposé aux chercheurs pour servir de ramasse-miettes, pour les orienter vers le dépôt de leur institution de rattachement, ou en attendant que celui-ci soit créé.

La réflexion sur les métadonnées permettant l'interopérabilité des archives est très avancée (Eprints Application Profile, UKOLN, JISC).

* www.intute.ac.uk

Les relations entre les systèmes d'information des établissements d'enseignement supérieur et la plate-forme partagée

La plate-forme partagée échangera des informations avec les systèmes d'information des établissements, géant les données documentaires, les données pédagogiques et celles qui servent à la gestion et à l'évaluation de la recherche.

Une identification des bases avec lesquelles une connexion est souhaitable, outre les systèmes d'information des établissements, doit être faite. La connexion déjà réalisée d'Hal et d'ArXiv a été complétée par l'export des documents déposés dans la base de l'Inserm vers PubMed Central.

L'archivage pérenne

Ce point est traité par le Cines en collaboration avec le CCSD. Il s'agit non seulement de prévoir plusieurs sites répartis, afin de préserver les documents et leurs métadonnées, mais aussi de se donner les moyens de conserver durablement leur intelligibilité en palliant le vieillissement physique des médias de stockage, l'obsolescence technologique des solutions de stockage, la péremption logique des formats de données et de leur contexte, en assurant le renouvellement sur le long terme des architectures et des applicatifs.

L'implication des établissements universitaires

En 2004, le consortium Couperin avait publié un manifeste en faveur des archives ouvertes. Son département « Études et prospectives » a créé à la signature du protocole un groupe technique sur les archives ouvertes GTAQ, associé au Cost afin d'instruire les dossiers spécifiques à ses membres, de favoriser la mise en place des archives ouvertes et institutionnelles dans les établissements, d'alimenter la réflexion du Cost par les expériences des établissements.

Les choix des universités

Dépôt dans Hal : Rennes-II, Grenoble-I, Saint-Étienne, Nice Sophia Antipolis. Dépôt dans le système d'information de l'établissement : Besançon, Lyon-II. À Strasbourg, dépôt dans Hal via le système d'information de la Misha (Maison interuniversitaire des sciences de l'homme – Alsace) pour les publications régionales en sciences humaines.

La plate-forme Hal

Hal est un outil de communication scientifique directe entre chercheurs créé en 2000 par le CNRS et porté par une unité de service, le CCSD (UPS 2275)¹.

Sa vocation est internationale (interconnexion avec ArXiv et PubMed) et pluridisciplinaire.

Ouvert aux établissements d'enseignement supérieur en 2002, Hal regroupe les publications grâce à des tampons (certification souple, non limitée) pour offrir des vues institutionnelles ou par composantes.

Reposant sur le principe de l'auto-archivage, Hal est conçu comme un lieu de dépôt, direct ou indirect, et ne se propose pas de moissonner des archives institutionnelles. L'interface de dépôt peut être personnalisée à la demande d'un établissement.

Une évolution nécessaire est en cours pour s'adapter au changement d'échelle et pour gérer la double fonction actuelle d'archive institutionnelle du CNRS et d'archives ouvertes des établissements.

Un « Guide du dépôt et du bon usage de Hal » rappelle les règles de la propriété intellectuelle : tous les auteurs doivent approuver le dépôt (coauteur ou auteur cité), le document peut être rendu public, le contrat avec l'éditeur l'autorise (lien au site Romeo Sherpa)². L'auteur s'engage à ne pas retirer un document une fois qu'il est déposé.

1. www.ccsd.cnrs.fr
2. www.sherpa.ac.uk/romeo.php

On peut noter, parmi ses travaux, l'étude de l'interconnexion des systèmes locaux et de la plate-forme partagée et une enquête sur l'état d'avancement des archives ouvertes parmi ses membres, dont les premiers résultats ont été rendus publics lors d'une journée d'étude, le 21 mai 2007. La publication de cette enquête

permettra de connaître les choix et l'état d'avancement d'un échantillon significatif d'établissements (un tiers des adhérents Couperin). En effet, la participation à la plate-forme nationale n'est pas exclusive et les établissements sont encouragés à valoriser leur production sur de multiples médias. Ainsi, ils peuvent exposer leurs publications à travers la plate-forme partagée, leur archive propre ou régionale, et des archives thématiques comme le réseau Nereus dont les membres français sont Sciences Po, Toulouse-I et Paris-Dauphine (programme NEEQ, Network of European Economists Online).

Un modèle à définir

De nombreux chantiers restent à instruire : le modèle juridique des archives ouvertes, le modèle économique équilibrant les archives ouvertes, les revues en libre accès et les publications traditionnelles, l'accès ouvert aux résultats bruts, etc. L'intégration dans les projets européens conduira à se rapprocher des normes et standards qui se mettent en place pour les modèles de données et de métadonnées. Des travaux de standardisation sont aussi nécessaires pour rendre partageables les métadonnées propres aux plates-formes institutionnelles. Enfin, le dossier des archives ouvertes constitue un champ d'expérience permettant de tester les modes de la nécessaire coopération entre les établissements universitaires et les organismes de recherche en matière d'information scientifique.

Ce sera aussi pour les professionnels de l'information, bibliothécaires et documentalistes, l'occasion de redéfinir leur rôle dans la chaîne éditoriale, la diffusion de l'information, la formation et l'information des étudiants, des enseignants et des chercheurs.

Octobre 2007

Pour plus d'information : www.sup.adc.education.fr/bib/Acti/ao/ao.htm et www.archives-ouvertes.fr