



Introduction au concept d'archive ouverte

Marie-Françoise Fily

► **To cite this version:**

| Marie-Françoise Fily. Introduction au concept d'archive ouverte. 2005. sic_00001523

HAL Id: sic_00001523

https://archivesic.ccsd.cnrs.fr/sic_00001523

Preprint submitted on 19 Jul 2005

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.

Introduction au concept d'archive ouverte

Marie-Françoise Fily
 Cirad-Dic, TA383/05 Avenue d'Agropolis, 34398 Montpellier cedex, France
 juin 2005

Résumé

Ce document a pour objectif d'informer les scientifiques d'établissements publics de recherche sur le concept d'archive ouverte et de les sensibiliser à l'utilisation d'archives ouvertes pour le dépôt, la diffusion et la valorisation de leurs propres publications.

Après un bref historique du mouvement en faveur du libre accès et une définition des archives ouvertes, sont discutés l'intérêt et l'utilisation des archives ouvertes dans le processus de publication scientifique, de recherche d'information et d'évaluation de la recherche. Les questions juridiques liées au dépôt sont également soulevées.

La notion d'archive ouverte est enfin illustrée par la présentation de moissonneurs, de plates-formes de gestion d'archives ouvertes et de projets en cours en France et en Europe.

Sommaire

1.	Le libre accès aux résultats de la recherche	3
1.1.	L'initiative de Budapest pour l'accès ouvert (BOAI) (14 février 2002)	3
1.2.	La déclaration de Bethesda pour l'édition en libre accès (11 avril 2003)	3
1.3.	La déclaration de Berlin sur le libre accès à la connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales (22 octobre 2003)	3
2.	Que sont les archives ouvertes ?	4
2.1.	Définition	4
2.2.	Les différents types d'archives ouvertes	4
2.3.	Les acteurs	5
3.	Archives ouvertes et activités de recherche : quel intérêt ?	6
3.1.	Dans le processus de publication : le chercheur en tant qu'auteur déposant	6
3.2.	Dans le processus de recherche d'information : le chercheur en tant qu'utilisateur	7
3.3.	Les archives ouvertes et l'évaluation de la recherche	8
4.	Aspects juridiques	8
4.1.	Les règles de base du droit d'auteur	8
4.2.	Les archives ouvertes et le droit	9
4.3.	Le projet Romeo	10
4.4.	Les licences Creative Commons	10
5.	Aspect technique	11
5.1.	Le protocole OAI	11
5.2.	Quelques moissonneurs	11
5.3.	Les plates-formes de création/gestion d'archive ouverte	12
5.4.	Quelques exemples de projets en France et en Europe	14
6.	Bibliographie indicative	16

Préambule

Ce document a été rédigé dans le cadre d'une démarche de sensibilisation de chercheurs en Agronomie tropicale au concept d'archives ouvertes.

Il ne doit donc pas être considéré comme un état de l'art sur la question.

1. Le libre accès aux résultats de la recherche

Le mouvement en faveur du libre accès aux résultats de la recherche revendique le droit à ce que la production scientifique écrite, activité inhérente au processus de la recherche, soit portée à la connaissance de l'ensemble de la communauté scientifique internationale et que ses contenus soient accessibles à tous, gratuitement et de façon pérenne.

L'histoire de ce mouvement est marquée par trois déclarations :

- la déclaration de Budapest décrit les deux stratégies du libre accès que sont l'auto-archivage et les revues en libre accès ;
- la déclaration de Bethesda précise ce qu'est une publication en libre accès ;
- la déclaration de Berlin étend la notion de publication en libre-accès à l'ensemble des résultats de la recherche.

1.1. L'initiative de Budapest pour l'accès ouvert (BOAI)¹ (14 février 2002)

Deux stratégies complémentaires sont définies par la déclaration de Budapest :

- l'auto-archivage : il s'agit du dépôt par son auteur d'un document en texte intégral sur un site web en accès public. Ce dépôt s'accompagne d'une saisie d'informations (métadonnées) permettant une recherche ultérieure (par nom, discipline, mot-clé, organisme, ...) ;
- les revues en libre accès : ce sont des revues diffusant gratuitement des articles à l'ensemble de la communauté scientifique, les frais de publication étant généralement pris en charge par les auteurs eux-mêmes ou leur institution.

1.2. La déclaration de Bethesda pour l'édition en libre accès² (11 avril 2003)

Cette déclaration définit ce qu'est une publication en libre accès :

- l'auteur et les titulaires du droit d'auteur accordent à tous les utilisateurs un « droit d'accès gratuit, irrévocable, mondial et perpétuel et leur concèdent une licence leur permettant de copier, utiliser, distribuer, transmettre et visualiser publiquement l'œuvre... à condition d'en indiquer correctement l'auteur ». Ils accordent également aux utilisateurs le droit de faire un petit nombre de copies papier de l'œuvre pour leur usage personnel ;
- la version complète de l'œuvre est déposée dès sa publication dans au moins un réservoir en ligne subventionné par un organisme reconnu « oeuvrant pour le libre accès, la diffusion sans restriction, l'interopérabilité et l'archivage à long terme ».

1.3. La déclaration de Berlin³ sur le libre accès à la connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales (22 octobre 2003)

Cette déclaration reprend la définition du libre accès de la déclaration de Bethesda et étend ce modèle à l'ensemble des résultats de la recherche et au patrimoine culturel.

Contrairement aux déclarations antérieures, les signataires en sont des institutions et non des individus ; ce qui a permis au libre accès d'acquiescer une validité institutionnelle.

Elle a été signée en France par le CNRS et l'Inserm en 2003, par l'Inra et l'INRIA en juillet 2004.

Une réunion dite de « Berlin 3⁴ » a eu lieu à l'université de Southampton le 1^{er} mars 2005. Elle recommande aux signataires de Berlin 1 :

- d'exiger de leurs chercheurs qu'ils déposent une copie de tous leurs articles dans une archive ouverte ;

¹ <http://www.soros.org/openaccess/fr/read.shtml> (version française)

² <http://biomedcentral.com/openaccess/bethesda/>

³ http://www.zim.mpg.de/openaccess-berlin/BerlinDeclaration_fr.pdf (version française)

⁴ <http://www.eprints.org/berlin3/>

- d'**encourager** leurs chercheurs à publier dans des revues en libre accès en leur accordant les **subventions** nécessaires.

2. Que sont les archives ouvertes ?

2.1. Définition

La notion d'archive ouverte et d'auto-archivage naît avec la création du premier serveur de publications en physique par Paul Ginsparg à l'université de Los Alamos en 1991 (ArXiv).

En 1999, la convention de Santa Fé lance le mouvement Open Archive Initiative qui a pour objectifs :

- de développer et de promouvoir des standards interopérables pour les bases d'articles scientifiques ;
- de définir un ensemble de protocoles techniques pour la description des données et leur interrogation ;
- d'oeuvrer pour que la constitution d'archives électroniques devienne un processus établi dans la communication scientifique.

De fait, l'objectif des archives ouvertes est de développer l'accès aux résultats de la recherche scientifique publiés et par là d'accroître l'impact des publications et la notoriété des auteurs, grâce à l'archivage institutionnel ou thématique de publications évaluées (post-publications) ou non évaluées (pré-publications).

Tout en continuant à soumettre leurs écrits à des revues ou à des colloques, les auteurs sont incités à les déposer sous forme de pré-publications (avant évaluation scientifique) ou de post-publications.

D'un point de vue technique, une archive ouverte est un entrepôt d'informations contenant des documents numériques quelle que soit leur antériorité, accessibles par une interface de type moteur de recherche. Il s'agit aussi de mettre en place un système d'archivage électronique à long terme assurant la pérennité des œuvres et de leur accès.

Suivant les archives, la typologie des documents acceptés peut varier : articles publiés et/ou acceptés dans les revues scientifiques, actes et communications de congrès, thèses, mémoires, HDR, posters, rapports, supports de cours, ouvrages et chapitres d'ouvrages, brevets, descriptions de projets, notices bibliographiques, images, sons, ...

Quel que soit le type de document déposé, le niveau scientifique attendu est celui d'une publication soumise dans une revue à comité de lecture.

Les dépôts sont effectués par les auteurs eux-mêmes. Souvent, les professionnels de la documentation interviennent pour enrichir les métadonnées associées aux documents.

Ces métadonnées reprennent au minimum les éléments de base du Dublin Core⁵. D'autres métadonnées peuvent être ajoutées (organisme, unités de recherche, thématique, ...).

La création de métadonnées lors du dépôt du document permet d'alimenter les outils de signalement, de faciliter la recherche ultérieure et l'extraction des informations bibliométriques du corpus.

Une fois déposés, les documents sont consultables sur Internet presque immédiatement dans le monde entier.

2.2. Les différents types d'archives ouvertes

2.2.1. Auto-archivage

L'auto-archivage est le processus qui consiste pour les auteurs à soumettre eux-mêmes leurs travaux aux sites web ou aux archives ouvertes de leur choix.

⁵ Le Dublin Core est défini par une norme ISO et comporte 15 éléments : title (titre), subject (sujet), description, coverage (couverture), source, relation, type, format, date, language (langue), identifier (identifiant), creator (auteur/créateur), contributor (contributeur), publisher (éditeur), rights (droits).

2.2.2. Archives thématiques

Les archives thématiques sont organisées autour d'une discipline, souvent au niveau international. Elles contiennent principalement des pré-publications (preprints) qui peuvent être éventuellement commentées et évaluées.

Exemples : ArXiv (physique) : <http://fr.arxiv.org/> (site miroir français)
 Research Papers in Economics (RePEc) (économie) : <http://repec.org/>
 CogPrint (sciences cognitives) : <http://cogprints.org/>
 Organic Prints (Agronomie) : <http://orgprints.org/>
 ArchiveSIC (Information et communication) : <http://archivesic.ccsd.cnrs.fr/>

Selon les domaines couverts, les archives thématiques sont plus ou moins reconnues, alimentées et utilisées par la communauté scientifique.

2.2.3. Archives institutionnelles

Les archives institutionnelles ont pour objectif de contenir l'ensemble de la production scientifique électronique d'une institution et d'assurer validité, visibilité et pérennité des écrits produits. Elles peuvent constituer des réservoirs pour tous les résultats de la recherche, y compris les données brutes et factuelles non incluses dans les publications.

Elles peuvent aussi être le moyen de valoriser une base documentaire en la rendant compatible avec le protocole OAI.

Le dépôt dans une archive ouverte institutionnelle n'empêche pas le dépôt éventuel dans une autre archive ouverte, par exemple thématique.

2.3. Les acteurs

2.3.1. L'administrateur de l'archive

L'administrateur de l'archive met au point la classification thématique et la typologie des documents acceptés, ainsi que le modèle de métadonnées de l'archive.

Il modère les dépôts : modération technique (fichiers aux formats acceptés), modération éditoriale (documents correspondant à la typologie définie).

Il peut modifier, supprimer et sauvegarder le contenu de l'archive.

2.3.2. L'auteur

L'auteur dépose ses propres publications après accord éventuel de ses co-auteurs.

Pour ce faire, il s'identifie auprès de l'archive. Il peut alors déposer un nouveau document ou consulter les commentaires de l'administrateur sur un dépôt déjà effectué.

L'auteur a l'obligation de respecter le contrat de publication qu'il a signé avec l'éditeur.

Par ailleurs, par l'acte de dépôt, il accorde aux utilisateurs un accès libre, gratuit, irrévocable aux publications déposées et les autorise à réutiliser ses publications (avec éventuellement certaines restrictions) sous réserve de reconnaissance de sa paternité (citation).

2.3.3. Le lecteur

Le lecteur navigue ou effectue des recherches dans l'archive.

Il lit en ligne ou télécharge les documents qui l'intéressent.

Il peut créer une alerte sur les thématiques qui l'intéressent.

2.3.4. Liens entre les acteurs

Des liens contractuels unissent ces trois acteurs.

En déposant son texte dans l'archive, l'auteur en cède à l'administrateur les droits de reproduction et de représentation sur support numérique, mais aux seules fins d'enseignement et de recherche.

L'administrateur s'engage à respecter l'intégrité du texte et à y faire figurer systématiquement le nom de l'auteur.

Le lecteur s'engage à n'utiliser le texte qu'à des fins d'enseignement et de recherche, en mentionnant toujours ses références bibliographiques et les droits d'utilisation qui lui sont attachés.

3. Archives ouvertes et activités de recherche : quel intérêt ?

La plupart des internautes n'utilisent qu'un outil de recherche, généralement un moteur de recherche de type Google. N'afficher ses publications qu'au travers d'un catalogue de bibliothèque (OPAC) interrogeable sur le seul site de son institution revient à ne pas être visible : les moteurs de recherche ne peuvent l'interroger.

Au contraire le protocole OAI-PMH qui permet aux archives ouvertes d'être interopérables est de plus en plus pris en compte par les moteurs de recherche.

Au-delà de l'amélioration de la communication scientifique directe, la mise en place d'une archive ouverte institutionnelle peut avoir de nombreux intérêts pour un chercheur :

3.1. Dans le processus de publication : le chercheur en tant qu'auteur déposant

3.1.1. Permettre la communication scientifique directe et rapide

La majorité des chercheurs accède aujourd'hui à Internet, l'utilise pour ses recherches d'information et veut s'en servir pour **diffuser ses travaux** : le chercheur qui produit déjà directement ses publications en format électronique, souhaite les rendre accessibles à l'ensemble de la communauté scientifique.

Mais cette diffusion est anarchique (sites personnels, sites d'équipes, sites thématiques). De plus, ces sites constituent une source d'information sans structure homogène et labile. Leur durée de vie est souvent liée à celle d'un engouement, d'un projet ou d'une équipe. Ils ne peuvent donc assurer la pérennité des publications scientifiques ni leur repérage exhaustif par les moteurs de recherche et leur visibilité sur Internet.

Une archive ouverte permet d'**organiser cette visibilité**, en tenant compte des problématiques de confidentialité, de droits d'auteurs et de reproduction qui ne sont actuellement pas toujours respectées.

Par ailleurs, le dépôt de publications dans une archive ouverte permet aux auteurs de **recueillir des commentaires d'autres chercheurs**, en particulier si l'archive propose un système d'alerte thématique sur les nouveaux documents déposés.

De plus, les résultats de recherche déposés sous forme de preprints sont immédiatement datés et **leur paternité n'est plus contestable**.

3.1.2. Améliorer son audience et sa visibilité face à ses pairs, à ses partenaires et à ses sources de financement

En permettant l'accès à sa production scientifique à partir d'un « réservoir » validé institutionnellement et scientifiquement, le chercheur **augmente sa reconnaissance par ses pairs**.

L'archive ouverte de son institution peut être visible

- Dans des portails documentaires
- À travers archives ouvertes thématiques, régionales, ou nationales
- À travers les moteurs de recherche tels Google ou Yahoo!

et acquiert ainsi une visibilité internationale.

Cette visibilité facilite la communication du chercheur avec ses partenaires et lui permet d'atteindre de nouveaux lecteurs, partenaires potentiels. Cela peut aussi avoir comme impact des collaborations internationales et des possibilités de financement.

De plus, en étant davantage visibles, les publications du chercheur sont davantage citées, ce qui augmente leur impact.

Par ailleurs, la mise en place d'une archive ouverte n'est pas incompatible avec la visibilité sur un site personnel ou d'équipe ou au travers d'un OPAC, mais l'information est conservée en un endroit unique (ce qui permet notamment d'assurer plus facilement sa pérennité).

3.1.3. Alimenter facilement les bases administratives servant à l'évaluation (transfert automatique des références)

Pour peu que les documents déposés soient correctement structurés, il est possible d'envisager l'extraction automatique des métadonnées et le transfert automatique des références bibliographiques vers les bases servant à l'évaluation.

Cela nécessite de développer, par type de document, un cadre normalisé de mise en forme qui permette une exploitation de la structure des documents. Il peut s'agir de simples feuilles de style word.

3.1.4. Permettre la conservation à long terme de ses fichiers sources (sans avoir la responsabilité de cette gestion)

La mise en place d'une archive ouverte permet d'éviter la multiplication des lieux de dépôt et de stockage des publications numériques d'une institution (catalogue de la bibliothèque accessible via un OPAC, sites d'équipes, sites thématiques, ...) tout en offrant la possibilité de consultation du texte intégral à partir de ces outils.

Elle offre ainsi aux chercheurs l'assurance que toutes leurs publications sont répertoriées et sauvegardées en un lieu unique et pérenne (non dépendant de l'organisation structurelle de l'établissement).

3.1.5. N' y consacrer que peu de temps

Le dépôt d'un document dans une archive ouverte ne demande pas beaucoup de temps. Il faut compter dix minutes par publication.

Par ailleurs, seul le premier auteur archive.

3.2. Dans le processus de recherche d'information : le chercheur en tant qu'utilisateur

3.2.1. Avoir facilement et gratuitement accès à des publications scientifiques validées

3.2.2. Disposer d'une interface de recherche unique dans la totalité des archives ouvertes

La création de telles interfaces est rendue possible grâce au protocole OAI-PMH qui permet l'interopérabilité entre les différentes archives ouvertes.

3.2.3. Éditer facilement sa liste de publications ou celle de son unité de recherche

Cette fonctionnalité est de plus en plus souvent offerte par les interfaces de gestion d'archives institutionnelles.

Le chercheur est ainsi assuré d'éditer des listes de publications toujours à jour.

3.3. Les archives ouvertes et l'évaluation de la recherche

La qualité scientifique d'un article se juge à son contenu, mais aussi à la notoriété de la revue dans laquelle il paraît et au nombre et au renom des auteurs qui le citent : c'est ce que l'on appelle l'impact de l'article.

De nombreuses études montrent que la disponibilité en libre accès d'un article augmente son impact scientifique⁶.

Par ailleurs, les concepteurs d'archives ont développé des outils bibliométriques destinés aux calculs de citations à partir des données bibliographiques des documents déposés, ainsi que des outils mesurant les corrélations entre le nombre de citations d'un article et le nombre de téléchargements (donc de lectures supposées) qui en sont faits.

Des études montrent que la courbe des téléchargements est identique à celle des citations, mais que les secondes se produisent plus tard que les premiers. **Le nombre de téléchargements permet donc d'anticiper le taux de citations à venir, ce qui permet notamment d'évaluer de jeunes chercheurs sur des publications récentes (donc non encore citées).**

Par ailleurs, sous réserve que les métadonnées saisies soient suffisantes, une archive ouverte permet de **mesurer la production scientifique** d'un organisme et aide à l'évaluation des chercheurs et des équipes de recherche.

Une archive ouverte pouvant contenir des informations non librement accessibles, il est possible de saisir et de « cacher » certaines données nécessaires à l'évaluation mais non diffusables aux moissonneurs de métadonnées.

3.3.1. Citebase

Citebase offre des outils de calcul des corrélations entre le nombre de citations d'articles contenus dans Citebase et le nombre d'accès à ces articles.

<http://citebase.eprints.org/cgi-bin/search>

3.3.2. Citeseer

Citeseer a développé un système de citations croisées, proche de celui du Science Citation Index de l'ISI.

<http://citeseer.ist.psu.edu/>

4. Aspects juridiques

4.1. Les règles de base du droit d'auteur

4.1.1. Notions d'originalité et de forme

Une œuvre est protégée dès sa création, sans nécessité d'aucune formalité de dépôt.

C'est uniquement la forme de l'œuvre qui est protégée. Les idées sont dites « de libre parcours ».

=> Tout document au stade de prépublication (avant signature d'un copyright) appartient à son auteur et est donc diffusable sous sa responsabilité. L'auteur est libre de déposer son œuvre dans une archive.

4.1.2. Les droits moraux et patrimoniaux

Les **droits moraux** liés à une œuvre sont incessibles, perpétuels et imprescriptibles.

Il s'agit :

- du droit au respect du nom (droit de paternité) ;
- du droit au respect de l'œuvre (l'œuvre ne doit pas être modifiée ni présentée dans un contexte risquant de la dénaturer) ;

⁶ Voir à ce sujet : « The effect of open access and downloads (« hits ») on citation impact : a bibliography of studies » <http://opcit.eprints.org/oacitation-biblio.html>

- du droit de divulgation (seul l'auteur peut décider de porter son œuvre à la connaissance du public) ;
- du droit au repentir et au retrait de l'œuvre.

Les **droits patrimoniaux** sont cessibles totalement ou partiellement.

Il s'agit :

- du droit de reproduction :
un changement de support (par exemple, la numérisation) est une nouvelle reproduction ;
- du droit de représentation (communication au public) :
la mise à disposition d'un document numérique sur un site relève du droit de représentation.

4.1.3. Les différents types d'œuvres

- L'œuvre à auteur unique :
l'auteur dispose de l'ensemble des droits sur son œuvre.
- L'œuvre de collaboration :
chacun des co-auteurs dispose de l'ensemble des droits sur l'ensemble de l'œuvre.
D'où la nécessité d'obtenir l'accord écrit de l'ensemble des co-auteurs pour pouvoir mettre à disposition un document numérique sur un site ou dans une archive ouverte.
- L'œuvre collective :
le coordinateur de l'œuvre dispose des droits sur l'ensemble de l'œuvre.

4.1.4. Les titulaires des droits

L'auteur est toujours une personne physique, excepté dans le cas d'une œuvre collective.

Les personnes morales ne peuvent être que cessionnaires des droits.

L'auteur salarié reste titulaire de ses droits d'auteur.

Il peut les céder à son employeur au fur et à mesure de la réalisation de ses œuvres, mais cette cession doit faire l'objet d'une description précise par écrit : modes de reproduction et de communication visés, durée et étendue de la cession.

La cession peut toutefois être recueillie par courrier électronique, par exemple au moment du dépôt institutionnel de la publication.

La cession globale des œuvres futures est nulle.

L'auteur agent public cède les droits des œuvres créées dans le cadre de sa mission à son administration de tutelle.

4.2. Les archives ouvertes et le droit

La mise en place d'une archive ouverte soulève des questions d'ordre juridique sur la propriété intellectuelle des documents, la propriété des résultats décrits et les clauses de confidentialité éventuelles qui y sont liées.

Un ensemble de liens juridiques et/ou contractuels lie les auteurs/déposants, les lecteurs et les administrateurs de l'archive ouverte. Ils ont pour objectif de garantir l'accès libre et gratuit aux documents, leur non-utilisation commerciale et le respect du droit des auteurs.

Les textes des licences de l'archive doivent figurer en ligne. Le contrat est accepté automatiquement par les utilisateurs lors des actes de dépôt ou de téléchargement.

Il est conseillé aux auteurs de déposer leur publication en deux temps :

- à titre de prépublication (preprint) pour en marquer l'antériorité ;
- lors de la publication pour déposer la dernière version scientifiquement validée du document.

La plupart des éditeurs considèrent le dépôt dans une archive ouverte comme un acte de diffusion à part : ils reconnaissent que le dépôt d'une pré-publication dans une archive ouverte ne constitue pas une publication antérieure.

Par ailleurs, l'auteur conserve le droit de diffuser une œuvre déposée dans une archive ouverte dans les conditions qui lui conviennent, sous réserve que les autres contrats de diffusion ne remettent pas en cause le dépôt.

Un document publié ne peut être déposé dans une archive ouverte que si l'auteur et les co-auteurs éventuels sont titulaires des droits de copyright.

De fait, le droit d'auto-archiver dépend du contrat signé par les auteurs au moment de l'acceptation de leur publication.

Plusieurs cas peuvent se présenter :

- l'auteur n'a pas signé de contrat d'édition lors de la publication de son document : il reste donc détenteur des droits sur son document et peut le déposer dans une archive ouverte ;
- il a signé un contrat d'édition ne prévoyant pas l'exploitation du document sur support électronique : il a le droit de le déposer dans une archive ouverte ;
- il a signé un contrat d'édition prévoyant explicitement l'exploitation du document sur support électronique : il doit demander l'autorisation de l'éditeur pour le déposer dans une archive ouverte.

Mais aujourd'hui, le copyright évolue et de nombreux éditeurs importants (dont Elsevier) permettent l'auto-archivage, parfois avec certaines restrictions de forme.

Par ailleurs, au moment de la signature du contrat de publication, il est souvent possible de rayer les clauses abusives ou d'insérer une clause autorisant l'auto-archivage.

4.3. Le projet Romeo

De plus en plus d'éditeurs acceptent le dépôt des articles publiés dans les archives ouvertes. Mais des nuances existent sur le format de la version finale de l'article qui peut être déposé. A titre d'exemple, depuis juin 2004, Elsevier autorise le chercheur à déposer son article dans l'archive ouverte **de son institution**, avec des restrictions de forme ; le dépôt de la version finale de l'auteur après acceptation de l'éditeur est autorisé, mais pas celui du tiré à part de l'éditeur au format pdf. De plus, un lien vers la page d'accueil de la revue doit être effectué.

Le projet Romeo référence les positions de 112⁷ éditeurs scientifiques sur l'auto-archivage des pré ou post-publications.

A chaque revue est attribuée une couleur en fonction de sa position sur l'auto-archivage :

- les revues blanches ne donnent pas explicitement le droit d'auto-archivage : ce droit doit être négocié au coup par coup ;
- les revues jaunes permettent d'archiver les pré-publications ;
- les revues bleues permettent d'archiver les post-publications ;
- les revues vertes permettent d'archiver les pré-publications, les post-publications ou les deux.

Sur 112 éditeurs éditant 8457 périodiques :

- 92 % autorisent l'archivage, soit des pré, soit de post-publications ;
- 8 % interdisent l'archivage (revues blanches).

<http://www.sherpa.ac.uk/romeo.php>

<http://romeo.eprints.org/publishers.html>

4.4. Les licences Creative Commons

Le projet Romeo recommande l'utilisation des licences Creative Commons qui permettent de préciser les droits des utilisateurs pour une œuvre en libre accès.

Ces contrats permettent de préciser à l'avance les droits d'utilisation d'une œuvre (texte, photo, musique, vidéo, site web, ...) disponible en libre accès, selon les conditions exprimées par l'auteur.

Les usages autorisés ou interdits sont au nombre de quatre et définis par une série d'icônes :

- paternité : l'auteur peut autoriser ou non l'usage libre de son œuvre à condition que sa paternité lui soit clairement reconnue et que son nom soit cité ;
- modification autorisée ou non ;
- utilisation commerciale autorisée ou non ;
- réutilisation autorisée ou non sous les mêmes modalités que le contrat d'origine.

⁷

Chiffres mai 2005

Le choix parmi ces quatre usages possibles donne lieu à six combinaisons de licences.

Un moteur de licence est proposé sur le site de Creative Commons. Il permet à l'auteur d'obtenir automatiquement un code HTML/RDF à insérer dans sa page. Ce code reproduit le logo Creative Commons avec un lien vers le type de contrat sélectionné.

En mettant son œuvre sous licence Creative Commons, l'auteur s'engage à ce qu'elle ne soit pas une contrefaçon.

Les licences Creative Commons sont disponibles en France depuis le 19 novembre 2004.

<http://fr.creativecommons.org>

5. Aspect technique

5.1. Le protocole OAI

(Open Archives Initiative Protocol for Metadata Harvesting ou OAI-PMH)

Ce protocole définit les normes à supporter par les archives ouvertes pour qu'elles puissent être interrogées simultanément, en une seule requête, quelle que soit leur localisation, sans duplication ni téléchargement des documents, ce qui les rend interopérables.

Il définit le langage par lequel communiquent le fournisseur de données (entrepôt ou data provider) et le fournisseur de services (service provider) qui rassemble des données collectées par un **moissonneur**.

La version 2.0 du protocole est utilisée depuis juin 2002 et est considérée comme la version définitive⁸.

Le fournisseur de données a la responsabilité de définir ce qu'il souhaite ou ne souhaite pas rendre disponible. Il conserve le contrôle total sur les données et les formate pour l'usage spécifique d'un ou de plusieurs moissonneurs. Il peut organiser son entrepôt en ensembles et sous-ensembles d'éléments et donc ne pas permettre à certains moissonneurs de collecter l'intégralité des métadonnées d'un entrepôt mais seulement un ou plusieurs sous-ensembles.

L'utilisateur final interroge uniquement le réservoir de notices du fournisseur de services, qui lui retourne en réponse la liste des notices pertinentes.

Ces notices proposent un lien hypertexte vers le document primaire, accessible sur le serveur du fournisseur de données. Ce serveur n'est en fait sollicité que pour la fourniture du document primaire.

5.2. Quelques moissonneurs

5.2.1. OAIster

Projet de l'université du Michigan

Il permet l'interrogation, par une recherche simple ou avancée, du contenu de plus de 400 archives ouvertes (5 387 890 enregistrements de 484 archives au 30 mai 2005).

Un accès par feuilletage présente brièvement chacune des ressources.

<http://oaister.umdl.umich.edu/o/oaister/>

5.2.2. ARC

<http://arc.cs.odu.edu/>

5.2.3. Citebase Search

Il permet l'interrogation de ArXiv, Cogprint et Biomed central

<http://citebase.eprints.org/cgi-bin/search>

⁸ The Open Archives Initiative Protocol for Metadata Harvesting. Protocol version 2.0 of 2002-06-14.
<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

5.2.4. Errol

Projet de OCLC Research

Il a pour objectif de rendre des entrepôts OAI navigables et propose une interface graphique pour parcourir les entrepôts.

<http://errol.oclc.org>

5.2.5. Scirus

Créé par Elsevier Science

<http://www.scirus.com/srsapp/>

5.2.6. Google et Yahoo!

Ils commencent à prendre en compte les archives ouvertes.

Dans une future version de Google Scholar, les articles en libre accès seront indiqués par une flèche.

5.3. Les plates-formes de création/gestion d'archive ouverte

5.3.1. Choix de la plateforme

Les logiciels permettant la création d'archives accessibles en ligne sont disponibles à travers une licence « open-source » et se conforment au protocole OAI-PMH.

Une liste des logiciels d'auto-archivage est maintenue par l'OAFForum :

http://www.oaforum.org/oaforum_db/list_db/list_software.php

Il est préférable de choisir une plate-forme déjà éprouvée, comme celle du CCSD (CNRS/INRIA).

5.3.2. Quelques plateformes

5.3.2.1. Archimede

Développé par l'université de Laval (Québec).

C'est un logiciel libre de gestion d'archives ouvertes, sous licence GPL.

Le texte de l'interface du système est indépendant du code. Cette approche facilite le développement d'une interface dans une langue spécifique sans avoir à modifier le code de l'application.

L'interface permet de faire des recherches dans les métadonnées aussi bien que dans le texte intégral des documents. Elle utilise LIUS (Lucene index update and search) : LIUS permet l'indexation de plusieurs formats de documents : XML, html, PDF, rtf, word, excel, txt, JavaBeans.

5.3.2.2. Castor

Développé par l'Ecole des Mines de Nantes, l'IRISA (INRIA Rennes) et Xerox (collaboration entre chercheurs, ingénieurs et documentalistes).

Cette plate-forme permet de stocker, annoter et décrire les documents produits par les étudiants, les enseignants, les chercheurs (articles, rapports, ..), et de les classer dans des collections privées ou publiques.

La plateforme gère les principaux formats de documents (MS Office, Open Office, ...) et convertit systématiquement les dépôts dans un format de documents XML structuré unique (schéma basé sur DocBook) : le but est d'assurer une pérennité maximale et une exploitation uniforme du fonds documentaire (recherche plein texte, classement, ...).

<http://www.emn.fr/castor>

5.3.2.3. CDSware

Développé par le CERN.

Il est utilisé par l'Ecole polytechnique fédérale de Lausanne (cf. [portail Infoscience](#)).

L'utilisateur a la possibilité d'accéder à l'historique de ses recherches, de se créer des alertes.

<http://cdsware.cern.ch:8000> (démonstration en ligne)

5.3.2.4. D-Space

Développé par le Massachusetts Institute of Technology (MIT) et HP.

Il est utilisé par la MSH-Alpes, l'Inserm, l'université de Lille 1 (Grisemine).

Cette plate-forme peut gérer toute la production numérique y compris les documents multimédia.

Le document est stocké à un seul endroit mais on peut le « voir » à partir de plusieurs collections/communautés.

Chaque communauté est constituée de 0, une ou plusieurs sous-communautés et/ou d'une ou plusieurs collections. Chaque sous-communauté est constituée d'une ou plusieurs collections.

Le nombre de communautés, de sous-communautés, de collections, de documents est illimité.

Chaque communauté ou sous-communauté possède sa propre page d'accueil.

Chaque collection peut définir son propre workflow, peut posséder son/ses administrateur(s) délégué(s) et posséder sa propre page d'accueil.

D-Space utilise le système Handle qui permet d'attribuer des identifiants pérennes aux ressources déposées de manière à ce qu'elles puissent être citées dans les publications.

D-Space intègre le moteur de recherche Lucene Jakarta.

<http://dspace.mit.edu> (démonstration en ligne)

5.3.2.5. Eprints

Développé par le département d'électronique et d'informatique de l'Université de Southampton.

Il est destiné au départ à archiver des articles scientifiques.

Il est possible d'établir le modèle de métadonnées en fonction du type de publication (ce modèle sert à créer le formulaire de dépôt (champs obligatoires et facultatifs)), d'organiser les documents dans des ensembles thématiques, de gérer plusieurs versions et plusieurs formats pour une même publication.

E-Prints dispose d'une interface en français

<http://demoprints.eprints.org> (démonstration en ligne)

5.3.3. La plate-forme créée par le CNRS : Hal (Hyper Articles en Ligne)

Cette plate-forme a été créée par le Ccsd (Centre pour la Communication Scientifique Directe), unité de service du CNRS, à partir du logiciel E-prints.

Créé en octobre 2000, le Ccsd a pour mission unique de « mettre à disposition des chercheurs de toutes disciplines, dans le monde entier, un circuit de communication directe leur permettant à la fois de charger leurs propres résultats de recherche et de consulter ceux des autres chercheurs », de façon gratuite.

Le Ccsd envisage de mettre son dispositif d'archives ouvertes « sous contrôle international » afin d'assurer sa pérennité à très long terme, de garantir le suivi des migrations technologiques et le suivi des standards.

Une institution peut demander au Ccsd l'ouverture de son archive propre, ce qui lui évite d'avoir à gérer le serveur correspondant, les sauvegardes, et ce qui facilite l'interopérabilité avec les autres archives. C'est l'option retenue par l'INRIA.

Hal est à la fois :

- un frontal pour tous les domaines scientifiques couverts par ArXiv : physique, mathématiques, une partie de la biologie (tous les documents déposés sur Hal dans ces domaines le sont simultanément sur ArXiv). Cela leur assure une meilleure visibilité internationale
- une archive multidisciplinaire avec vues personnalisées, pour les autres domaines scientifiques

La structure de Hal permet l'indexation de son contenu par les moteurs de recherche type Google ou Google Scholar.

Hal permet l'indication de l'affiliation des auteurs, y compris pour les chercheurs étrangers participant à la publication.

Une collaboration INRIA/Ccsd est en cours (accord cadre sur le co-développement de Hal) :

- sur la mise à disposition du logiciel Hal sous licence GPL à partir de 2005
- sur la mise au point d'un système d'affiliation électronique au niveau de l'archive

Des collaborations sont prévues avec d'autres EPST : Inserm, Inra, Ifremer, IRD.

L'ordre des étapes de dépôt peut-être paramétré au niveau de l'interface

On peut décider par exemple de commencer par le téléchargement du document afin d'en extraire automatiquement les auteurs, leurs affiliations, la référence bibliographique, ...

Ces possibilités d'automatisation dépendent du niveau de structuration du document source.

Hal permet la gestion de collections fondées sur un système de tampons :

- un ou plusieurs tampons sont attribués à toute instance qui en fait la demande (organisme, UR, ...)
- à l'aide des tampons, il est possible de distinguer, manuellement ou automatiquement, un ensemble de publications
- des vues sélectives peuvent ainsi être proposées affichant seulement les documents « tamponnés », sous forme de mini-site autorisant la navigation, la recherche et la consultation au sein de la collection de documents.

L'utilisateur a la possibilité de définir un profil personnel adapté à ses intérêts et de recevoir un message automatique à chaque fois qu'un nouvel article correspondant aux critères choisis est déposé.

Hal peut être décliné en versions multiples, appelées « environnement de dépôt » ou « environnement de consultation », avec texte propre, présentation graphique propre et éventuellement métadonnées supplémentaires (au-delà des métadonnées indispensables au dépôt).

<http://ccsd.cnrs.fr>

5.3.3.1. Fonctionnalités particulières de Hal

- Transfert des références bibliographiques vers la base de données servant de référentiel des publications
- Extraction automatique de listes de publications pour alimenter des sites d'unités de recherche, avec lien vers le texte intégral des documents
Ces listes peuvent être personnalisées avec un logo et/ou du texte libre
- Génération de listes de publications dans tous formats (XML, doc, pdf, ..) avec des possibilités de personnalisation de la présentation (ordre des données, présence ou non d'un titre, ...)
Des critères de sélection, tel le type de publication (par exemple, revues à comité de lecture) peuvent être spécifiés de façon à limiter l'extraction.
- Soumission d'un manuscrit à un périodique scientifique à comité de lecture
A la fin du dépôt d'un document, l'auteur peut le soumettre automatiquement pour publication à certaines revues scientifiques à comité de lecture.

5.4. Quelques exemples de projets en France et en Europe

Le nombre total d'archives ouvertes n'est pas connu précisément.

Leur recensement repose sur une auto-déclaration par leurs concepteurs sur un des sites suivants :

<http://www.openarchives.org> ou <http://www.oaforum.org> , de façon à ce qu'elles soient moissonnées par les principaux moissonneurs de services généralistes ou les principaux moissonneurs du domaine.

En mars 2005, 749 serveurs OAI étaient répertoriés sur l'un ou l'autre de ces sites.

5.4.1. Laboratoire IN2P3 du CNRS (Democrate-Hal)

L'objectif du projet du laboratoire IN2P3 est de valoriser la production scientifique de l'équipe en lui assurant visibilité extérieure, accessibilité, exhaustivité, pérennité.

L'outil utilisé est Hal.

Ce projet est à la fois la réponse à la demande de visibilité des chercheurs (interconnexion avec arXiv) et à une demande institutionnelle de référencement de la production.

Il implique une collaboration entre le chercheur qui dépose ses documents avec les métadonnées correspondantes et la documentaliste qui vérifie la pertinence des métadonnées existantes et intègre les métadonnées institutionnelles.

Un export automatique se fait vers le référentiel des publications du CNRS, [publi@cnrs](mailto:publi@cnrs.fr).

Les responsables du projet sont un chercheur et une documentaliste.

Site web : <http://democrite.ccsd.cnrs.fr>

5.4.2. Archive institutionnelle de l'INRIA

Le but de l'archive ouverte de l'INRIA est d'assurer la pérennité de la production scientifique de l'INRIA et de son accès et d'en assurer une meilleure visibilité, en minimisant les contraintes pour les chercheurs.

Elle doit aussi servir à recenser et connaître la production scientifique de l'institut (tableaux de bord, indicateurs, édition de bibliographies).

L'INRIA a fait le choix d'une collaboration avec le Ccsd pour utiliser la plate-forme Hal et mutualiser les moyens entre le CNRS et l'INRIA dans le développement de Hal.

Cette collaboration passe notamment par le détachement d'un ingénieur INRIA pour passer Hal en open source, le but final étant de créer une archive ouverte pour toute la communauté scientifique française.

L'INRIA a ouvert son archive le 27 avril 2005.

Les documents pouvant être déposés sont des articles soumis ou acceptés, des chapitres d'ouvrages, des rapports, des cours, des travaux universitaires et des communications de congrès.

Remarque : L'hébergement des données se fait au sein d'un centre de calcul CNRS.

Site web : <http://hal.inria.fr>

5.4.3. Centre Inra d'Avignon

Le but du projet du centre Inra d'Avignon est de constituer la mémoire de tous les types d'écrits du centre.

L'outil utilisé est D-Space.

Cet outil propose une structure de l'archive en communautés (par exemple par équipe de recherche) et collections (postprints, littérature grise, ...).

Le dépôt peut-être soumis à un système de validation (workflow).

Remarque : A l'Inra, les centres de recherche mènent pour l'instant leur réflexion sur les archives ouvertes de manière autonome. Le centre de Tours, par exemple, a créé une archive ouverte en physiologie animale avec le logiciel E-prints.

5.4.4. Infoscience (Ecole polytechnique fédérale de Lausanne)

Le projet infoscience a pour but de faciliter l'accès aux ressources scientifiques produites à l'EPFL : publications, preprints, rapports de recherches, projets, thèses, travaux d'étudiants, cours, etc.

L'outil utilisé est CDSware, logiciel libre développé par le CERN.

Au niveau de chaque notice, il est possible :

- d'obtenir le document en format PDF ;
- de transformer le document en format Postscript ;
- d'ajouter la notice dans son répertoire personnel ;
- d'accéder à la version imprimable ;
- d'envoyer la notice par messagerie électronique

Un répertoire personnel permet de gérer :

- ses propres alertes : enregistrer un thème, une recherche, un groupe de publications, etc. Si des modifications sont apportées aux contenus, l'utilisateur est averti par message électronique ;
- ses recherches : créer ses propres combinaisons de recherches, les stocker pour les retrouver plus tard.

6. Bibliographie indicative

Publications générales sur la problématique du libre accès

Battisti M. 2004. Libre accès aux résultats de la recherche : une politique pour un renouveau de la publication scientifique ? Journées d'étude Inist Documentaliste-Sciences de l'information, 2004, 41 (4-5) : 262-267

Chanier T. 2004. Archives ouvertes et publication scientifique : comment mettre en place l'accès libre et aux résultats de la recherche ? Paris : L'Harmattan, 2004. 186 p.

Salaun J.M. 2004. Libre accès aux ressources scientifiques et place des bibliothèques. [On-line]. [2004/12/15].

Bulletin des bibliothèques de France, 2004, 49 (6) : 20-30

<URL :

http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/11/41/sic_00001141_00/sic_00001141.pdf >

Le libre accès aux résultats de la recherche. [On-line]. [2004/12/15].

Vandoeuvre les Nancy, INIST, 2004

<URL: http://www.inist.fr/openaccess/article.php3?id_article=64 >

Publications techniques

Aubry C., Janik J. (dir.). 2005. Les archives ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information. Paris: ADBS, 332 p. (Sciences et techniques de l'information).

Aubry S. 2004. Archives ouvertes : nouveaux modèles de publication et de diffusion et nouveaux outils

In : INRIA. - Publier sur Internet : Séminaire INRIA , 27 septembre-1^{er} octobre 2004, Aix-les Bains

Paris : ADBS, 2004 : p. 209-237

L'auto-archivage : foire aux questions. [On-line]. [2004/12/15].

<URL: http://sophia.univ-lyon2.fr/boai/self-faq_fr.html >

Battisti M. 2005. Libre accès à l'information et droit d'auteur

In : Les archives ouvertes : actualités, enjeux, perspectives. Journée d'étude URFIST, Bordeaux, 31 mars 2005.]

<URL: http://urfist.u-bordeaux4.fr/battisti_fichiers/slide0001.htm >

Foulonneau M. 2005. Collaborer pour de nouveaux services culturels en ligne. Le protocole OAI, protocole de collecte de métadonnées de l'initiative des archives ouvertes.

Paris : Ministère de la culture et de la communication, 2005. 59 p.

<URL: http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/guide_oai.pdf >

Fournier D. 2004. Etude de la mise en place d'une archive ouverte institutionnelle, valorisant les publications scientifiques et universitaires des établissements d'Agropolis. Contribution au projet Canist

Marseille : Université d'Aix-Marseille, 2004. 72 p.

Mémoire (DESS Gestion des systèmes documentaires en IST)

ManuHal

<URL: http://ccsd.cnrs.fr/article.php3?id_article=15>

Nawrocki F. 2005. Le protocole OAI et ses usages en bibliothèque. [On-line]. [2004/02/15].

Paris, Ministère de la culture et de la communication, 2005

<URL: <http://www.culture.gouv.fr/culture/dli/OAI-PMH.htm>>