



HAL
open science

La diffusion libre du savoir Accès libre et Archives ouvertes

Pierre Baruch

► **To cite this version:**

Pierre Baruch. La diffusion libre du savoir Accès libre et Archives ouvertes. L'Archicube, no. 3, dossier "L'édition et le Numérique", Association des anciens élèves, élèves et amis de l'École normale supérieure, 45 rue d'Ulm, 75005 Paris, www.archicubes.ens.fr, no. 3, p. 77-95, 2007. sic_00169330

HAL Id: sic_00169330

https://archivesic.ccsd.cnrs.fr/sic_00169330

Submitted on 3 Sep 2007

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.

La diffusion libre du savoir Accès libre et Archives ouvertes

Pierre Baruch*

Exegi monumentum aere perennius. (Horace)

"Puisse l'instruction générale s'avancer d'un pas si rapide que dans vingt ans d'ici il y ait à peine en mille de nos pages une seule ligne qui ne soit populaire ! C'est aux Maîtres du monde à hâter cette heureuse révolution." (Denis Diderot, avertissement du tome VIII de l'Encyclopédie)

Être lu, partout, toujours, l'ambition de tout auteur ! Le poète il y a deux millénaires, le philosophe, voici trois siècles, nous rejoignent dans ce souci d'éternité et d'universalité.

Mais Diderot évoque aussi l'aspect politique, la responsabilité des « maîtres du monde ». Plus près de nous –voici presque un siècle-, les dirigeants d'alors avaient délégué ce soin aux intellectuels, Bergson, Marie Curie, Einstein et autres, réunis dans la Commission pour la Coopération intellectuelle de la Société des Nations –l'oubliée SDN – .En 1922, son président, Henri Bergson proposait d'organiser « *les bibliothèques pour que toutes leurs ressources soient rigoureusement coordonnées et facilement accessibles,..... avec la constitution d'un catalogue commun*»

Projet alors utopique, les moyens techniques de l'époque ne le permettaient pas encore –la photocopie, le microfilm venaient d'apparaître -. C'est dans les années 90 que la réalisation en est devenue à notre portée: Internet et les outils associés ont montré que la diffusion universelle et instantanée du savoir n'était plus une fiction et, de plus, pouvait se faire à très faible coût.

La pérennité cependant est loin d'être assurée ; avec la dématérialisation, les « écrits » sont devenus presque aussi volatils que la parole, et les supports matériels eux-mêmes sont trop labiles, la pierre, le bronze ou même le papier restent plus sûrs que bandes magnétiques ou disques numériques !

Cette « révolution Internet » touche de nombreux domaines. La recherche scientifique est un de ceux qui tirent parti des immenses possibilités apportées par les réseaux mondiaux, au point de voir changer considérablement ses pratiques.

1- La publication scientifique, ses règles et ses supports

En effet, le savoir scientifique se bâtit par la communication entre ceux qui ont la charge de le faire progresser, les chercheurs : annonce de résultats nouveaux, d'idées nouvelles, mise à jour, évaluation, confrontation, validation ou réfutation, en un enchaînement permanent.

* Professeur émérite, Université Paris 7-Denis Diderot, pierrebaruch@orange.fr

Les nouvelles perspectives apportées par Internet vont-elles radicalement modifier le système de la communication scientifique, ou verrons-nous seulement une adaptation des pratiques existantes ? La question est à examiner dans le contexte des exigences propres à la recherche.

L'échange direct de correspondances a été le premier mode de communication et a perduré les siècles durant. Pour assurer une diffusion plus étendue à une communauté scientifique ne cessant de croître, apparurent des périodiques prenant en charge la publication de travaux originaux. Ce furent, dès le 17^{ème} siècle, sous l'égide des Académies et des Sociétés savantes, des publications comme les « *Philosophical Transactions of the Royal Society* », ¹, créés à Londres en 1665. Ces journaux spécialisés ont depuis foisonné, dépassant largement le cadre des Sociétés savantes et des Académies pour s'étendre aux éditeurs commerciaux. Ils sont complétés par les ouvrages d'enseignement et de recherche, des thèses, des présentations à des congrès, et par des écrits destinés au grand public.

Cette presse scientifique ² s'est créé un code de conduite spécifique, répondant aux critères exigés par la méthode scientifique. Son rôle est, non seulement d'assurer la diffusion des travaux, mais aussi de l'entourer d'un certain nombre de garanties : identification des auteurs, indexation, critères d'originalité, validation scientifique – effectuée couramment sous forme d'une évaluation par les pairs (*peer review*). Permanence et stabilité sont nécessaires pour la constitution des corpus de références bibliographiques ; la conservation reste assurée, à court terme par les éditeurs, dans le long terme par les bibliothèques,

Dans ce type de publication, les auteurs sont très rarement rémunérés, leur motivation primordiale étant la diffusion de leurs écrits ; ce caractère est souvent mal compris, certains éditeurs ont tendance à assimiler cette gratuité à un manque de valeur intellectuelle. L'auteur garde la propriété intellectuelle de son œuvre, n'en cède qu'un droit d'exploitation à l'éditeur. Cette situation, consacrée par l'usage tant que la diffusion ne se faisait que par l'imprimé, est devenue source de contradictions avec les nouvelles possibilités fournies par Internet : édition simplifiée, diffusion instantanée et universelle à coûts quasiment nuls.

On a donc vu proliférer sur la Toile des lieux contenant des textes scientifiques, sites de laboratoires ou d'établissements, pages personnelles, blogs, etc. Leur création et leur accès sont certes faciles, mais, pour la plupart il leur manque la permanence – une page web est volatile ; les moyens d'authentification, d'indexation ne sont pas présents en général.

Concilier ces garanties avec la rapidité et l'universalité que présente Internet est donc devenu urgent. Mais une autre demande se fait jour dans la communauté scientifique, pour l'accès libre (AL) ³ (« *Open Access* »), aux documents issus de la recherche, sans frais pour le lecteur.

Une profonde mutation des modes de communication est donc en train de se faire. D'un côté, la presse scientifique, qui offre les garanties décrites ci-dessus, a pris le tournant de la publication « en ligne », avec des modalités différentes suivant les cas. Parallèlement à ce système traditionnel de publications, imprimées ou électroniques, à accès payant ou gratuit, un nouveau mode de publication scientifique apparaît : la « communication scientifique directe » ⁴, fondée sur l'autoarchivage, par les chercheurs eux-mêmes, de leur production dans des « archives ouvertes » (AO), qui ajoutent à l'accès libre la notion de dépôt libre. La physique a été la première discipline à utiliser ce mode, avec le serveur d'articles en ligne ArXiv avec des bénéfices notables pour la facilité de communication.

Ces deux modes de communication ⁵ ne s'excluent pas, ils peuvent s'appliquer simultanément au même document, comme nous allons le voir.

Les journaux scientifiques et le libre accès

La plupart des journaux scientifiques a dès maintenant ajouté à la publication traditionnelle sur papier une édition électronique

La publication traditionnelle, quelle soit sur papier ou maintenant fréquemment accompagnée d'une mise en ligne, est aux frais du lecteur, qu'il s'agisse de consulter la collection de périodiques, ou d'acheter un article individuel en copie papier ou par accès payant au site web de l'éditeur, (*pay per view*).

En général ces coûts sont supportés par l'institution (bibliothèque universitaire, laboratoire, etc) et non par le chercheur lui-même, qui peut alors consulter de son bureau sur son ordinateur la collection complète des revues auxquelles son institution est abonnée. C'est une facilité fantastique, qui fait rêver ceux qui ont passé des heures et des jours à chercher et attendre un article. En fait, le système est devenu complètement transparent aux yeux de l'utilisateur, qui, de plus, n'a pas toujours conscience de son coût réel. Or celui-ci obère de plus en plus les budgets des établissements. La politique commerciale des grandes maisons d'édition internationales qui dominent le marché, la multiplication des journaux très spécialisés ont donné lieu à une inflation considérable. La figure 1, collationnée⁶ par l'association américaine des bibliothèques de recherche (ARL) montre l'évolution des dépenses de ces bibliothèques, comparée au coût de la vie (CPI : consumer price index)

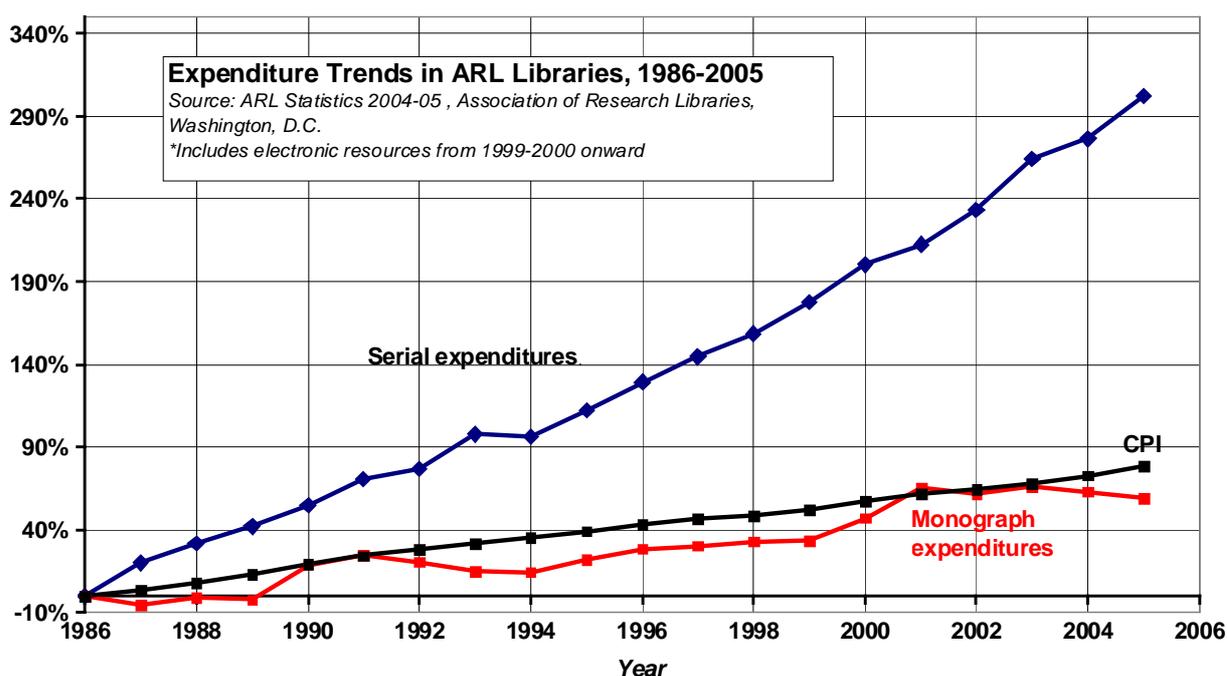


Figure 1 : Les dépenses des bibliothèques universitaires américaines ont cru dans les 20 dernières années beaucoup plus vite que l'indice des prix (« Serial expenditures » : dépenses en périodiques. Source : Association of Research Libraries)

Par réaction, certaines universités (Cornell) ont évoqué une résiliation massive de leurs abonnements ; des tractations commerciales conduisent à des regroupements d'établissements bénéficiant de conditions plus avantageuses (Consortium Couperin en France). Les éditeurs proposent aussi des « bouquets », regroupant un ensemble thématique de journaux, complétés par des aides à la recherche bibliographique (par exemple le programme Scopus d'Elsevier). A côté de ces puissantes maisons internationales existent quelques éditeurs de taille moyenne

ou petite, souvent sans but lucratif car adossés à des sociétés scientifiques ou à des institutions, universités, académies, organismes de recherche, mais fonctionnant sur le même mode payant pour le lecteur..

Cependant, ce modèle n'est pas figé. En effet, dans les dix dernières années on a assisté à la création de journaux purement électroniques, sans support papier, en accès totalement libre. Un tel mode de fonctionnement est possible financièrement en raison des coûts très réduits de l'édition et de la diffusion électronique et souvent de l'appui par l'institution.

Ces journaux en accès libre répondent à une demande formulée avec de plus de force par la communauté scientifique : pourquoi la connaissance, bien public, doit-elle subir la loi du marché, pourquoi les chercheurs, qui fournissent l'information sans rétribution spécifique, devraient ils payer pour la recevoir, pourquoi la puissance publique, qui subventionne la quasi totalité de la recherche, doit-elle de nouveau être mise à contribution pour accéder à ces résultats ?

L'édition commerciale, menacée par les réductions de crédits des bibliothèques, par la concurrence des journaux électroniques en libre accès et sensible aux demandes de cette clientèle assez particulière, a donc réagi.

La publication en libre accès : la « barrière mobile »

D'un côté, la plupart des journaux scientifiques admettent maintenant que les articles déjà publiés deviennent librement accessibles sur leur site web après un certain délai d'embargo (« barrière mobile »), allant d'un an ou moins pour les sciences exactes, à plusieurs années, parfois jusqu'à 5 ans pour certaines revues en sciences humaines et sociales (SHS). Ces différences sont à rapporter aux pratiques propres aux disciplines

Les sciences de la nature (y compris les mathématiques) publient essentiellement dans des journaux internationaux ; parfois très spécialisés, et, pour beaucoup, édités par des groupes privés multinationaux souvent en position dominante. L'évaluation, par un comité de lecture, des textes soumis est la règle. La rigueur de cette sélection devient un élément de hiérarchisation des journaux, avec le très discuté facteur d'impact. Il en résulte qu'être publié dans un journal très sélectif valorise son auteur. Les éditeurs des journaux classiques mettent en avant ce rôle, et le coût du processus de sélection, pour justifier les tarifs d'abonnement et la mise sous embargo temporaire. Par contre, dans ces disciplines la durée de vie utile des articles, est relativement courte, quelques années en moyenne; avec évidemment des écarts notables pour des articles fondateurs qui ne se périment pas. Ce trait autorise donc une durée d'embargo assez réduite

Pour les SHS, le paysage est sensiblement différent. Certes les revues internationales sélectives existent mais à côté, on trouve des nombreuses revues nationales, à diffusion limitée et parfois de périodicité incertaine. Le livre, ouvrage individuel ou collectif, est un support courant. On assiste là à une variété de pratiques en ce qui concerne les modalités de diffusion

Un exemple intéressant est fourni en France par des « portails » d'édition électronique en SHS, comme *revues.org*⁷ qui assure, pour près de 70 revues, l'édition électronique d'articles originaux. Celles qui ont à la fois les supports papier et électronique se conforment au modèle d'abonnement payant, alors que celles qui n'ont que l'édition électronique, sans support papier sont en accès libre, dès la parution sur le site de *revues.org*. Le portail est soutenu par le CNRS, qui, associé aux universités de Provence et d'Avignon et à l'EHESS, a créé à cet

effet l'unité de service *CLEO*, élément du TGE «*ADONIS*»⁸, visant à perfectionner les moyens de communication en SHS. De même, *Persée*⁹, réédite en ligne des numéros anciens d'une vingtaine de revues, pour la direction des bibliothèques et l'université Lyon 2. *CAIRN*¹⁰ publie en ligne un ensemble de revues, pour le compte d'un groupement d'éditeurs commerciaux, donc avec le régime d'abonnements payants

On reste ici dans un modèle classique où l'éditeur assume les coûts de fabrication et de diffusion et les répercute sur l'abonné ; l'accès libre intervient après amortissement. L'auteur cède ses droits, ne reçoit pas de rétribution, mais n'a pas à contribuer, sauf parfois des frais réduits («*page charges*» pour des clichés, des tirages à part, etc.). Les quelques journaux en accès libre dès parution échappent apparemment à ce modèle, mais en fait la part du lecteur y est prise en charge, sous forme de subvention, par l'institution.

La Publication en libre accès : « l'Auteur ou l'Institution paye »

Un autre concept s'est fait jour depuis quelques années, dans lequel c'est l'auteur qui assume tous ces coûts, avant même la parution de l'article – l'auteur ou.. l'institution de tutelle et le bailleur de fonds. L'idée a été lancée en 1999 par un groupe de scientifiques américains, dont H. Varmus, prix Nobel et a vu sa première mise pratique par la création de la « Public Library Of Science » (PLOS)¹¹, association sans but lucratif, soutenue par la fondation G. & H. Moore. PLOS publie depuis 2003 plusieurs journaux, en sciences de la vie (PLOS Biology), suivant ce modèle où il est demandé à l'auteur environ 2500\$ par article publié.(indépendamment des «*page charges*»). Le texte est édité suivant les règles classiques, en particulier soumis à un comité de lecture qui peut accepter ou refuser la publication. Seuls les articles acceptés donnent lieu à paiement. L'article publié est mis en ligne, en accès libre sur le site de PLOS, et simultanément apparaît dans l'édition papier du journal (payante, à un prix modique ou téléchargeable gratuitement).

Ce mode de publication (*Open Access Publishing - OAP*) s'est étendu depuis à divers journaux scientifiques, commerciaux ou liés à des sociétés savantes¹² et été prôné par diverses organisations (Budapest Open Access Initiative 2, BOAI-2)¹³. Le coût pour l'auteur reste du même ordre qu'avec PLOS, de 500 € à quelques milliers d'euros. D'importants organismes de recherche, dont récemment le CERN, incitent leurs chercheurs à publier dans des journaux affichant ce label, et ils en assument directement le coût.

Ce concept est séduisant par sa logique –il ne peut y avoir de recherche sans publication, et celle ci doit avoir la diffusion la plus vaste, donc il faut inclure le coût de cette diffusion dans le budget du projet de recherche.

Cependant, on peut se poser quelques questions. Implicitement, sa mise en œuvre suppose des laboratoires suffisamment dotés, pouvant mettre de côté une part de leurs ressources pour les publications. Le modèle s'applique mieux à des grosses organisations de recherche, dans une logique quasiment industrielle.. Comment le chercheur isolé, la petite unité de recherche, il en existe, même dans les sciences dures, pourra-t-il le faire ? Sans précaution, l'opération se ferait aux dépens du budget d'équipement, ce qui suscite l'inquiétude des laboratoires. Une solution verrait un transfert des crédits des bibliothèques vers les laboratoires –opération délicate administrativement, au moins en France, et évidemment peu acceptable par les bibliothèques. En fait, il s'agira de mettre en place un système de subvention des éditeurs, ce qui ne manquera pas d'accentuer leur pouvoir de contrôle du processus de publication. Aussi, les journaux qui prennent cette voie continuent à accepter des articles non payés d'avance, qui suivent la voie ordinaire, avec des délais plus importants, et ne sont donc pas librement accessibles sur Internet. Les bibliothèques devront elles garder les abonnements payants pour

accéder aux deux types d'articles, auquel cas la consultation de ce type de revues serait payée deux fois, en amont et en aval de la parution. ? Les tenants de ce système répondent qu'il ne s'agirait que d'une période transitoire, que des exemptions seront mises en place pour les cas délicats –en particulier pour les chercheurs de pays à faibles ressources.

On voit se confronter ici des conceptions différentes de la nature de la recherche scientifique. D'un côté un système structuré, dans lequel la publication joue un rôle similaire de celui de la publicité dans les entreprises : il s'agit de faire reconnaître la valeur des résultats scientifiques, éventuellement de les monnayer –subventions, brevets, licences, ... les dépenses de publication sont alors à incorporer explicitement dans les coûts de production, et le système OAP apparaît comme normal ; on peut y trouver une certaine analogie avec le budget de publicité d'une entreprise, incorporé dans le coût total du produit.. De l'autre, la conception idéaliste et romantique de la recherche, le chercheur travaillant pour le bien commun. Paradoxe de la science, l'auteur d'un article scientifique ne reçoit pas de rémunération de l'éditeur, il lui a cédé gratuitement ses droits d'auteur, implicitement ou non. L'éditeur se rembourse de ses frais de mise en forme et de diffusion, soit par les abonnements, soit en de les mettant à la charge de l'auteur.

Sur un autre plan, des inquiétudes se font jour sur le risque de limitation de la liberté du chercheur. L'organisme payeur ne sera-t-il pas tenté par une forme de censure préalable ? par un choix entre auteurs si, paradoxalement, la production scientifique d'une équipe est trop abondante ?

L'expérience manque pour répondre à ces interrogations, des ajustements seront évidemment nécessaires ; il ne faut pas conclure trop vite que « *Open Access Publishing* » est la seule réponse à la mise en place de l'accès libre à la connaissance scientifique, Les Archives Ouvertes le permettent aussi, avec d'autres modalités, comme nous le verrons. Y a-t-il une complémentarité, une compatibilité entre ces deux concepts ?

2-Les Archives Ouvertes

Le prototype en est le serveur ArXiv¹⁴, créé, voici une vingtaine d'années, sur l'impulsion des physiciens des hautes énergies, désireux de pouvoir rapidement échanger documents et données entre les grands centres mondiaux de cette discipline, dont le CERN où venait d'être mis au point le World Wide Web, ensemble de méthodes et de protocoles permettant de tels échanges via le naissant réseau mondial de l'Internet. Maintenant, ArXiv est devenu un mode de communication courant, non seulement pour la physique et les mathématiques, mais aussi pour l'informatique, la statistique, la biologie quantitative. Y sont déposés des textes dignes de publication (e-prints), mais non soumis, des projets d'articles en cours de soumission (pre-prints). Les pre-prints deviennent post-prints, après acceptation et parution. Certains éditeurs acceptent même que soit mis en ligne, comme post-print, un fac-simile (fichier pdf) de l'article tel qu'il a été publié.

Le mouvement en faveur des Archives Ouvertes s'est développé dans la communauté scientifique dans les années 2000, avec ses pionniers et ses prophètes, comme Stevan Harnad¹⁵ (Universités de Southampton et de Québec) qui s'intitule, non sans humour, « *archivangéliste* » et qui, avec ses collègues de Southampton, a développé des logiciels de dépôt et des protocoles d'indexation. Plusieurs manifestations ont permis d'organiser ce mouvement et d'y impliquer –parfois seulement de façon symbolique –les organismes de recherche (déclarations de Berlin¹⁶, initiatives¹⁷ de Budapest).

Depuis, d'autres « dépôts » (*repositories*) ont vu le jour¹⁸, « institutionnels » liés à un établissement (université, laboratoire), lequel maintient sur un serveur dédié les archives de la production scientifique de ses membres, toutes disciplines confondues, ou thématiques, comme ArXiv pour la physique ou PubMedCentral¹⁹ pour les sciences de la vie. Des annuaires²⁰ donnent la liste des dépôts d'archives existant.

HAL serveur d'archives multidisciplinaire pour la recherche française

A côté d'initiatives locales (quelques universités et laboratoires), le CNRS, sous l'impulsion de Franck Laloë, chercheur en physique à l'ENS (laboratoire Kastler-Brossel) a, dès 1998, envisagé l'intérêt d'inciter ses chercheurs à déposer leur production scientifique sur un serveur dédié, en libre accès, suivant l'exemple donné par ArXiv. Une unité de service, le Centre pour la Communication Scientifique Directe (CCSD), hébergée par le centre de calcul de l'IN2P3, a été créée pour en développer les outils. La base de documents HAL, pour *hyper article en ligne* (<http://hal.archivesouvertes.fr/>), avec un ensemble logiciel original, est entrée en service en 2000, pour le CNRS.

Le soutien public par l'Académie des Sciences (juin 2005) a conduit, en juillet 2006, à la signature²¹ d'un « *protocole associant l'ensemble des Universités et des Grandes Ecoles par l'intermédiaire de leurs Conférences, et les organismes de recherche que sont le CEMAGREF, le CIRAD, le CNRS, l'INRA, l'INRIA, l'INSERM, l'Institut PASTEUR et l'IRD autour de la mise en place d'une plate-forme commune de dépôt de publications et d'écrits scientifiques, (qui) constitue un engagement sans précédent de la part d'institutions nationales.* ».

Fallait-il, suivant le modèle britannique, concevoir une base d'archives par université, comme certains établissements l'avaient déjà fait ? Cependant, une large base de données, au niveau national, gérée par un organisme de recherche puissant, présenterait de nombreux avantages. C'est en particulier le cas en France, où la multiplicité des affiliations – un même chercheur peut relever simultanément de plusieurs organismes (université, CNRS, CEA, etc.) – rend difficiles la mise en place et, surtout, la coordination des archives propres aux divers établissements

Ainsi, la tâche en a été confiée au CCSD, pour mettre en œuvre HAL comme base d'archives commune. HAL est fortement inspiré d'ArXiv, mais plus élaboré et surtout bien plus large en termes de disciplines : on y trouve des apports en quantités croissantes, non seulement en « sciences dures » (mathématiques, physique, chimie, etc.), mais aussi en sciences humaines et sociales. HAL, qui contient maintenant (sept. 2007) près de 50000 documents, en reçoit quelques 2 000 par mois, chiffre en augmentation constante (*figure 2*). Les directions générales de l'INRIA, de l'INSERM et du CNRS ont incité par lettre tous les directeurs de laboratoires dépendant de leurs établissements à faire en sorte que la production scientifique de ces laboratoires soit déposée dans HAL..

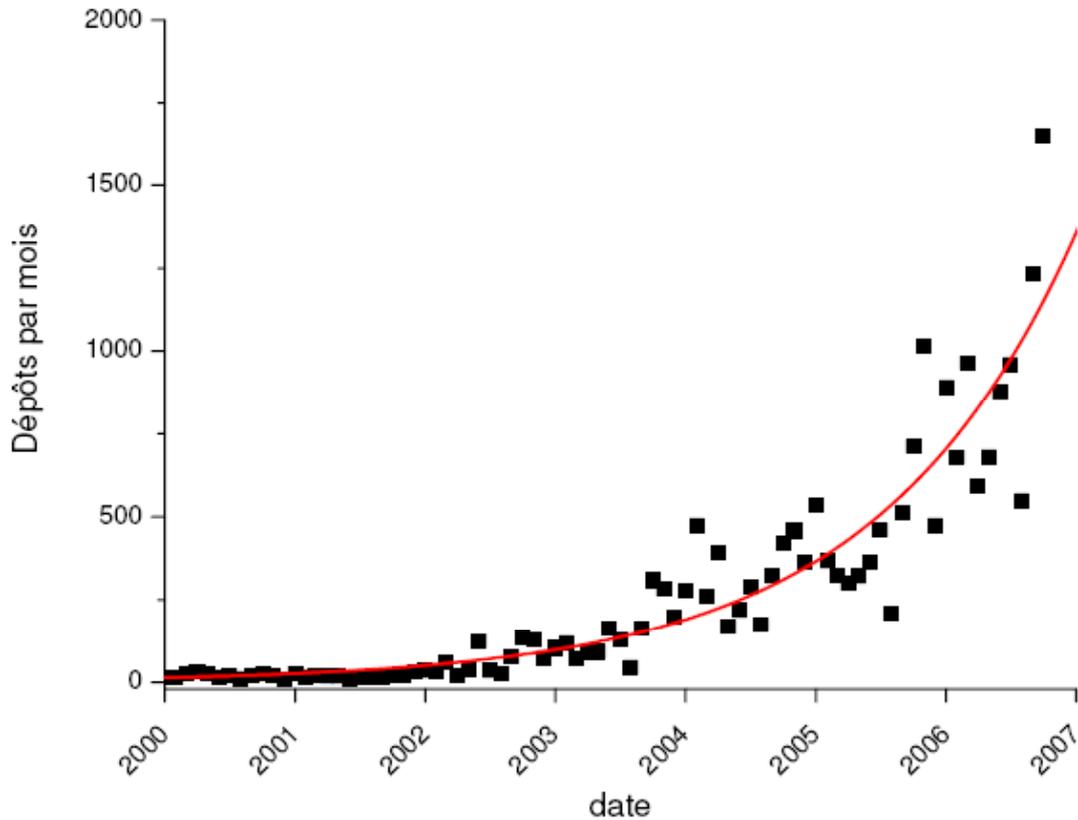


Figure 2: Le rythme de dépôt de documents dans l'archive ouverte française HAL s'accroît rapidement. Ce rythme est à comparer avec le total de la production scientifique française, qui est d'environ 10000 articles par mois (Source : CCSD-CNRS)

Que déposer dans ces bases de données ? Tout document que le chercheur souhaite porter à la connaissance de la communauté scientifique : manuscrit d'article soumis pour publication, contribution à un ouvrage scientifique, mémoire de thèse, rapport d'intérêt scientifique général ou article déjà publié par une revue traditionnelle (sous réserve de l'accord de celle-ci), ce qui les rend immédiatement et gratuitement disponibles à leurs collègues du monde entier. Contrairement aux revues scientifiques traditionnelles, aucun intermédiaire n'intervient, ni comité éditorial ni éditeur commercial. Certes, les responsables du système effectuent un examen élémentaire du contenu avant la mise en ligne, mais cet examen ne dure que quelques minutes par document ; il a juste pour but de vérifier l'identité de l'auteur, la pertinence scientifique, sans évaluation détaillée, et d'éliminer ainsi les dépôts évidemment inadéquats. Le serveur enregistre non seulement le texte, mais aussi les métadonnées, à savoir le nom de l'auteur, son affiliation, la date, les références bibliographiques, etc., ce qui facilite les recherches bibliographiques et l'établissement de listes de publications. Le document est daté –important pour établir les priorités– et ne peut plus être modifié, sinon par la mise en ligne d'une nouvelle version, tout en en laissant visibles les états précédents. Si l'auteur envisage la publication dans une revue classique, il le signale et, après acceptation et parution, met en ligne le texte final avec les références, sous certaines conditions d'accord avec l'éditeur (mise en ligne de la version auteur –fichier soumis au journal– ou de la version éditeur-fichier du texte mis en forme pour parution, respect d'un délai) ; les métadonnées doivent rester accessibles.

Pour retrouver un document, on utilise les moteurs de recherche et indexeurs, soit généraux, soit associés à la base d'archives, (voir l'encadré 1). HAL, comme les autres bases, est compatible avec le protocole d'interopérabilité OAI-PMH, qui permet d'extraire les métadonnées, mais son propre moteur de recherche est plus riche, permettant, entre autres, la recherche en plein texte

Encadré 1

Retrouver des documents, les moteurs de recherche

Les outils électroniques d'édition facilitent la documentation, de façon parfois surabondante. On a donc intérêt à utiliser des services spécialisés, plutôt qu'un attrape-tout comme Google.

Guides :

<http://www.searchengineguide.com/>

SHS <http://album.revues.org/>

Quelques moteurs généralistes :

Google Scholar <http://scholar.google.com/>,

The Virtual Library : <http://vlib.org/>

INIST-CNRS / <http://articlesciences.inist.fr/>

CrossRef <http://www.crossref.org/>

ISI Web of Science <http://scientific.thomson.com/products/wos/> (accès restreint)

Elsevier Science Direct: <http://info.sciencedirect.com/> (accès restreint)

Recherche limitée aux archives ouvertes <http://www.citebase.org/>

<http://www.oaister.org/>

Exemples de moteurs thématiques :

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

SHS <http://www.in-extenso.org/>

Physique <http://scitation.aip.org/>

Médecine, biologie : <http://medline.cos.com/>

etc

Les moteurs en accès restreint sont plus performants, mais nécessitent un abonnement. Certains ont une version limitée en accès libre. Se renseigner auprès des bibliothèques institutionnelles ou des centres de documentation pour en connaître les possibilités d'accès.

On remarquera que la recherche et l'échange de documents scientifiques par de telles méthodes sont voisins du *peer to peer*, ce système d'échanges directs entre internautes qui inquiète tant les industries audiovisuelles. Aussi, il existe un risque que les restrictions légales adoptées récemment pour combattre les chargements illicites de fichiers musicaux ou vidéo soient appliquées sans discernement à d'autres domaines, notamment celui des échanges scientifiques.

La Structure de HAL, les portails

Bien que très centralisée en apparence, la plate-forme de HAL permet une grande souplesse, car il est possible, à partir de l'ensemble des données qu'elle contient –HAL générique-, de créer des sous-ensembles qui apparaîtront à l'utilisateur –auteur ou lecteur –comme des portails autonomes, institutionnels ou thématiques, ou même comme des collections personnelles de documents.

Ainsi, on compte dès maintenant une quinzaine de portails, servant des institutions nationales : INRIA, INSERM, IN2P3 (en préparation INRA, CIRAD, CEMAGREF, IRD, etc), consacrés à des champs disciplinaires : Sciences de l'homme et de la société, éducation, sciences de l'univers, environnement, sciences de l'information et de la communication, langue et littérature basques, Le dépôt des thèses en ligne se fait sur le portail TEL, qui contient maintenant près de 8000 documents, un autre portail accueille les notes de cours. Les universités et les grandes écoles créent progressivement leurs propres archives, parfois par migration sur HAL de leurs dépôts antérieurs, cependant le nombre d'établissements impliqués est encore faible. Cette remarque ne concerne pas seulement le développement de HAL, car les établissements qui maintiennent d'autres systèmes d'archivage sont aussi une minorité. On trouvera toutes indications utiles dans la page d'accueil de HAL ainsi que sur le site <http://www.archives-ouvertes.fr/>.

Chaque portail détermine de façon autonome sa politique d'archivage, sous la réserve de respecter le niveau minimum de qualité défini pour HAL-générique : si, pour la physique, les documents mis en ligne ne sont pas tous destinés à la publication, HAL INSERM ne présente que des articles acceptés pour publication. En fait ces choix reflètent les us des disciplines concernées : la physique suit le modèle d'ArXiv et les documents déposés dans HAL peuvent, au choix de l'auteur être automatiquement transférés dans ArXiv, tandis qu'en médecine et sciences de la vie, la référence est PubMedCentral, qui n'accepte que des articles publiés ou en cours de publication ; l'INSERM a en fait passé un accord avec cet organisme pour des échanges réciproques.

La répartition entre disciplines (figure 3) montre certes la prédominance historique de la physique et des mathématiques, mais une part importante maintenant, et inespérée, des sciences humaines et sociales, malgré la nouveauté de ce mode de communication. Cependant, certaines disciplines restent peu représentées, comme la chimie. Pour la médecine et les sciences de la vie, la mise en place de HAL-INSERM est trop récente pour que les chiffres en soient significatifs. Les réticentes des chimistes envers l'autoarchivage peuvent être expliquées, en partie, par les liens historiques avec l'industrie, qui impose des règles strictes de protection de la propriété industrielle – l'American Chemical Society s'oppose à l'auto-archivage pour tous ses périodiques. La physique, pionnière de l'autoarchivage avec ArXiv, se trouve dans une situation paradoxale : la puissante American Physical Society ne met pas en accès libre les articles parus dans ses revues aussi prestigieuses que la Physical Review, quelle que soit leur date. Mais elle tolère, tacitement, que ses auteurs déposent leurs manuscrits dans ArXiv ; Qui plus est, elle a constaté qu'il n'y a pas de concurrence entre les deux modes de publication et n'a pas vu de baisse des abonnements. L'archive ouverte sert alors d'outil de travail, tandis que le revue garde ses fonctions de référence et de garantie d'excellence.

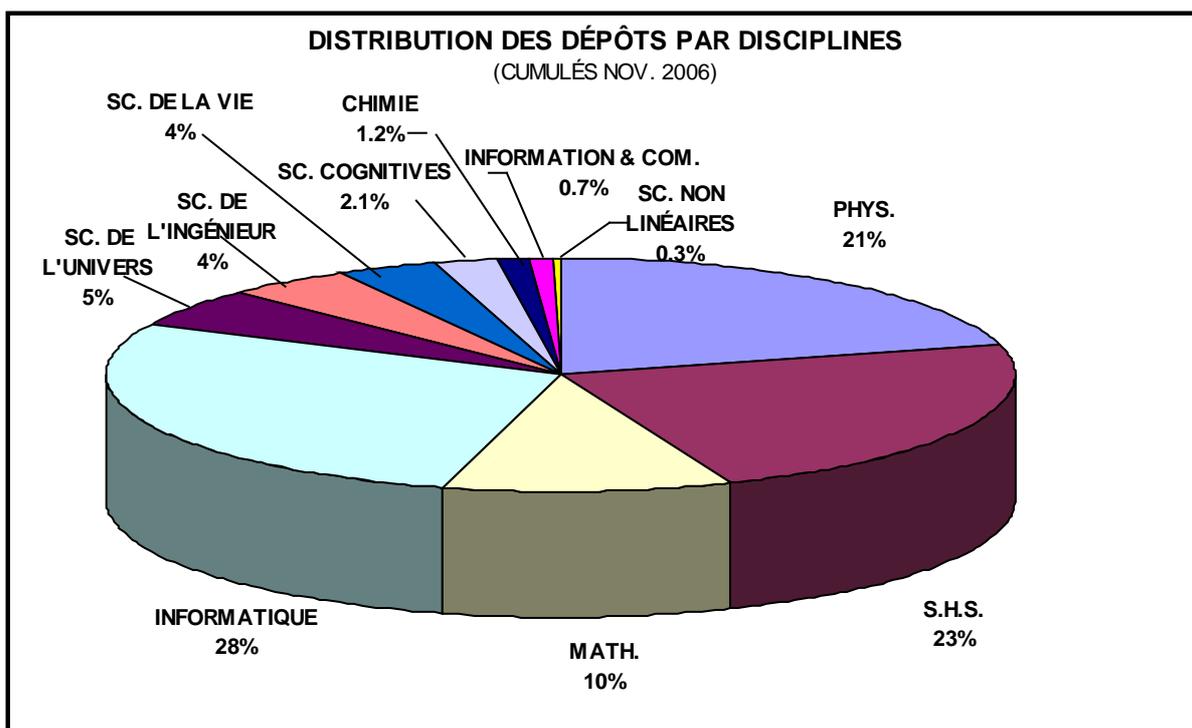


Figure 3 : répartition par discipline des dépôts dans l'archive ouverte HAL (Source : CCSD-CNRS)

Les autres dépôts d'auto-archivage en France, au nombre d'une vingtaine, contiennent quelques 10 000 documents. On en trouvera le détail dans les tableaux annexes²² mis en ligne sur @rchiveSIC, portail de HAL pour les sciences de l'information et de la communication.

La vocation de Hal ne se limite pas aux chercheurs français. Non seulement des dépôts venant de l'étranger sont possibles et existent, sans contraintes linguistiques, mais des organismes de recherche de pays européens envisagent de l'utiliser pour leurs propres besoins. Il est d'ailleurs intéressant de constater que les critiques de cette centralisation à la française s'éteignent alors que certains pays s'appliquent maintenant à fédérer leurs actions locales sur le plan national (DEPOT au Royaume Uni, DARE aux Pays-Bas, DIVA en Suède et Norvège). La démarche est inverse de celle de HAL, mais la rejoint dans la mesure où les archives locales gardent leur autonomie, mais peuvent apparaître à l'utilisateur comme issues d'un portail unique.

Qu'apportent les Archives Ouvertes ? Perspectives

A côté des nombreux avantages des archives ouvertes, quelques difficultés ou incertitudes se font jour, liées à leur mise en place et au rôle qu'elles joueront, à terme, dans le monde de l'édition et de la communication scientifiques.

La mise en route et le maintien d'archives ouvertes ont un coût négligeable par rapport aux sommes dépensées par les établissements de recherche pour leur documentation scientifique. Or il faut savoir que ces dépenses ont fortement augmenté au cours des 20 dernières années et préoccupent de très nombreux laboratoires et organismes de recherche (voir la figure 1). En France, les dépenses pour les abonnements aux revues scientifiques sont ainsi de plusieurs centaines de millions d'euros chaque année. Le coût des archives ouvertes est 500 à 1 000 fois plus faible, car il tient pour l'essentiel aux frais de personnel – quelques ingénieurs et informaticiens.

Assurer la pérennité des archives, essentielle pour les scientifiques, impose des contraintes techniques. Les adresses électroniques des documents doivent être parfaitement stables pour que ces documents servent de référence dans des travaux scientifiques ultérieurs. Les migrations logicielles et matérielles doivent être réalisées à temps pour que le corpus de documents accumulés reste disponible pendant au moins plusieurs décennies. Dans ces conditions, on peut espérer que la durée de vie des documents électroniques soit au moins aussi longue que celle des documents imprimés sur papier (généralement quelques siècles). Mais pour atteindre cet objectif, une certaine centralisation dans la gestion des archives sera nécessaire.

Certains avantages des archives ouvertes sont évidents. Il y a ainsi la rapidité de la communication scientifique, sachant que la publication d'un article dans une revue traditionnelle prend parfois plus d'un an. On a aussi une prise de date immédiate, puisque l'archive ouverte certifie une date de dépôt pour chaque version.

Un autre avantage est, pour le lecteur, un regroupement de l'information dans des réservoirs uniques, qui deviennent ainsi en quelque sorte la mémoire de l'humanité dans certains domaines scientifiques. Les recherches avancées sur les documents sont facilitées par l'utilisation d'un moteur de recherche opérant sur une base unique et homogène.

Les archives ouvertes offrent par ailleurs un accès gratuit à tous les scientifiques du monde entier, que ce soit pour soumettre des documents ou pour les consulter, d'où une plus grande visibilité (à la disparition de l'Union soviétique, bien des membres de la brillante école de physique théorique russe ont pu survivre intellectuellement dans leur discipline en accédant gratuitement, grâce au serveur *ArXiv*, aux travaux les plus récents). Qui plus est, l'accès libre et à distance à la lecture va dans le sens d'une démocratisation du savoir, dans la mesure où il facilite pour le grand public la consultation des articles scientifiques professionnels.

En outre, les archives ouvertes permettent de se libérer de la notion de document figé une fois pour toutes dans une publication. En effet, un auteur peut déposer dans la même «boîte» des versions successives – et datées – d'un même texte. On pourra ainsi trouver sur l'archive la version de l'auteur soumise à la revue traditionnelle, la version modifiée et acceptée (avec le lien vers l'article publié), mais aussi des versions ultérieures où, par exemple, des erreurs typographiques dans les données ou les équations auront été corrigées. C'est particulièrement utile pour les auteurs d'articles de synthèse, qui peuvent effectuer des mises à jour après publication, en tenant compte à la fois des réactions des lecteurs et des derniers développements scientifiques. Enfin, en cas d'erreur, l'auteur peut aisément déposer une version comportant un erratum, voire se rétracter en chargeant une nouvelle version vide précisant que le document a été retiré pour telle ou telle raison, tout en gardant les versions précédentes.

Une question fondamentale posée aux archives ouvertes et qui reste débattue est celle de la validation scientifique, à laquelle les chercheurs sont très attachés, pour la notoriété qu'elle leur apporte, et pour les conséquences sur leur carrière. Faire valider un article par un comité de lecture composé d'experts est un processus long et coûteux. Le service n'est assuré actuellement que dans le processus classique de publication dans des revues payantes – à coûts supportés par le lecteur ou par l'auteur. Il serait beaucoup trop lourd pour des archives ouvertes. Cet argument est d'ailleurs mis en avant par les éditeurs pour justifier l'accès payant comme l'embargo. Cependant, le dépôt d'un article dans une archive ouverte est maintenant souvent concomitant avec la soumission à une revue, avec l'acceptation du manuscrit ou avec

sa publication ; le déposant indique alors les références de l'article publié, qui constitue une caution scientifique.

Encadré 2 **Travaux pratiques**

Lecteur d'Archicube, vous êtes certainement aussi auteur :

Vous avez certainement dans vos œuvres un texte que vous souhaitez préserver pour la postérité, sinon gravé dans le bronze, tout au moins archivé sur un serveur sûr. Alors, déposez le sur HAL.

Vous disposez du manuscrit sous forme d'un fichier (Word, pdf, ...)

Vérifiez d'abord, avec le moteur de recherche interne de HAL, ou avec OAlster ou Citebase (voir encadré 1) qu'il n'a pas déjà été déposé, sur HAL ou une autre Archive Ouverte, par une main bien intentionnée.

Assurez vous, d'après votre contrat, que votre éditeur ne vous interdit pas formellement l'autoarchivage.

La suite est alors très simple : à partir du portail générique de HAL (<http://hal.archives-ouvertes.fr/>), ou de l'un de ses portails secondaires, thématique ou institutionnel (visibles dès la page d'accueil), identifiez vous, puis suivez les instructions jusqu'à la phase finale de transfert du document. Vous recevrez un accusé de réception, vous donnant les coordonnées (URL) de votre œuvre.

Si l'éditeur vous a fourni votre texte sous sa forme finale, tel qu'il a été publié, en fichier pdf par exemple, et s'il vous y autorise, déposez cette « version éditeur », de plus belle présentation que votre « fichier auteur » original.

Bien entendu, vous pouvez toujours déposer des textes originaux, non encore soumis à un journal, les faire évoluer en fonction des commentaires qu'ils vont susciter, avant de les présenter pour publication.

D'autres procédures ont été essayées, telles que la critique en ligne, non anonyme, du document, jusqu'à consensus. Mais elles n'ont pas encore débouché en pratique. La validation par les pairs est-elle indispensable ? L'attribution en 2006 de la prestigieuse médaille Fields au mathématicien russe récompense des travaux qui n'ont pas été publiés dans une revue, mais seulement mis en ligne dans ArXiv, sans passer par le filtre des experts. Le cas est encore trop singulier pour servir de modèle.

Les archives ouvertes et les revues cohabiteront

En liaison avec cette question de la validation scientifique, certaines critiques portent sur le risque de voir les archives submergées par des articles inintéressants, douteux ou faux. Des chercheurs médiocres ou en mal de reconnaissance, voire des amateurs un peu fous, ne pourront-ils pas se faire une publicité indue ? En théorie, le risque existe, mais l'expérience montre que les dispositifs de filtrage élémentaire déjà en place ont évité les dérives, et les erreurs sont minimales.

Un autre aspect qui peut soulever des difficultés est le rôle des archives ouvertes par rapport à celui des revues classiques à comité de lecture. Les archives ouvertes coupent-elles l'herbe sous le pied des éditeurs de ces revues, comme on aurait pu le craindre ? L'expérience des physiciens avec le serveur ArXiv, qui a aujourd'hui 15 ans, montre que non ; la circulation des grandes revues de physique n'a pas été affectée par l'usage intense qu'ils en font. Beaucoup de journaux maintenant acceptent qu'un manuscrit qui leur est soumis soit aussi mis en ligne dans une Archive Ouverte. Certains proposent même à l'auteur de se dispenser de la formalité d'envoi du manuscrit et de traiter directement le texte archivé.

On remarque même un effet coopératif entre le système des archives ouvertes et le fonctionnement des revues à comité de lecture. Certains auteurs, avant de soumettre leur document à une revue, commencent par le déposer sur une archive ouverte, afin de bénéficier des réactions spontanées des collègues ; ces réactions conduisent parfois à des améliorations substantielles du texte, qui donneront lieu à une seconde version – et c'est elle qui sera soumise à une revue.

D'un autre côté, il est fréquent que l'éditeur scientifique consulte sur l'archive ouverte l'ensemble de la production de l'auteur, pour se faire une idée de sa personnalité scientifique. Son choix de rapporteurs sera également guidé par l'archive ouverte, celle-ci lui permettant de voir si tel ou tel collègue est actuellement intéressé par le sujet, et est donc susceptible d'accepter d'écrire un rapport.

Par ailleurs, les lois concernant le droit d'auteur et son transfert éventuel à un éditeur doivent être respectées. Cependant, différentes situations sont possibles. Dans la plupart des cas, l'éditeur garde la propriété de la forme définitive du texte qu'il a publié. Sauf accords spéciaux, cela ne permet de mettre en ligne qu'une version auteur. La majorité des journaux scientifiques acceptent maintenant le dépôt d'une version auteur sur une archive ouverte institutionnelle, avec des modalités variables selon les cas. Certains éditeurs imposent toutefois un délai entre l'acceptation ou la publication de l'article et sa mise en ligne, embargo qui va de quelques mois à un an.

La pénétration des archives ouvertes dans la vie des laboratoires va croissante. Elle n'est pas uniforme suivant les disciplines. Les archives ouvertes suscitent parfois, sinon une opposition, du moins des réticences. Les chercheurs tiennent à juste titre aux revues classiques, en ligne ou sur papier, pour les garanties qu'elles apportent : évaluation par les pairs, prestige de la revue le cas échéant. Ils ne voient pas forcément la nécessité d'un système parallèle, ou n'ont pas conscience de la charge financière considérable que les abonnements aux revues représentent pour leurs institutions. D'où parfois une certaine indifférence, et une méconnaissance de l'intérêt de ces bases de données d'un nouveau genre. La pénétration est évaluée en moyenne aujourd'hui à 15% de la production scientifique, mais varie suivant les disciplines. Le présent article montre bien les limites actuelles de la couverture de la littérature spécialisée par les AO : sur la vingtaine de citations faites à des publications, classiques ou électroniques, seules 4 ou 5 renvoient à des documents déposés dans des AO ; cette proportion serait évidemment plus élevée pour un texte de physique ! Pour augmenter la visibilité des publications, l'obligation de dépôt en AO est prônée par les experts, mais se heurte à des réticences dans la communauté scientifique.

Il est clair qu'une archive ouverte n'est pas une nouvelle forme de revue traditionnelle, ou simplement l'agrégation dans un même portail électronique du contenu de nombreuses revues. Il s'agit d'un nouvel outil de travail, avec ses règles intellectuelles propres. Avec cet outil, la référence unique dans la communication entre scientifiques n'est plus l'article publié, et le document scientifique peut avoir une vie intéressante avant et après sa publication, parfois même sans publication du tout.

Les archives ouvertes nous font revenir à la vieille tradition de l'échange direct de correspondances scientifiques entre savants, qui a perduré des siècles durant, à côté du développement de l'édition scientifique telle que nous la connaissons aujourd'hui. Celle-ci, à son tour, ne disparaîtra pas, mais cohabitera avec les archives ouvertes. Comment concilier, la viabilité et la demande à l'accès universel ? Quel sera le point d'équilibre ?

Remerciements : Cet article est issu des discussions et réflexions menées depuis plusieurs années par Euroscience et son groupe de travail sur la publication scientifique. L'objectif en est de faire entendre, à côté des experts, nombreux et passionnés, la voix des principaux intéressés, les chercheurs. Je remercie Françoise Praderie et Siméon Anguelov, à l'origine de cette action, Hélène Bosc, qui a modéré les discussions, et les collègues participant. Les avis et conseils de Franck Laloë, pionnier des Archives Ouvertes, sont pour beaucoup dans la préparation de ce point de vue.

¹ le "Journal des Sçavans" en France, à la même époque, correspondait plus à nos revues de « vulgarisation », visant à faire connaître à un public non spécialisé les avancées de la science ; voir :Jean Claude Guédon, "In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing" 138th Association of Research Libraries Meeting, Toronto, Ontario, May 23-25, 2001 <http://www.arl.org/arl/proceedings/138/index.html> (version française : <http://doc-iep.univ-lyon2.fr/Edelec/>)

² Le terme « scientifique » sera utilisé ici pour couvrir aussi bien sciences « exactes » (ou sciences de la nature, suivant la terminologie anglo-saxonne) que sciences humaines et sociales (SHS). La suite de l'article ne concerne que les publications de recherche, en excluant les revues de vulgarisation, les manuels et autres ouvrages, qui suivent les lois du marché.

³ Terminologie et sigles ne sont pas fixés. Nous utiliserons ici AL pour Accès Libre, quand le lecteur peut accéder sans coût au document, et AO pour Archive Ouverte, quand l'auteur peut déposer sans frais son document, qui sera en accès libre. On évite ici l'ambiguïté anglo-saxonne, où le sigle OA, fréquemment rencontré, renvoie tantôt à Open Access (AL) tantôt à Open Archives (AO)

De même, les textes en anglais utilisent fréquemment les termes « gold journals » et « green journals » pour désigner respectivement les journaux en libre accès (payé par l'auteur) et ceux qui autorisent en plus le dépôt dans une archive ouverte de documents soumis à publication, ou publiés. Moins sujette à confusion est la dénomination « Open Access Journal » ou « OA Publishing », pour le premier type..

⁴ Baruch, P et Laloë, F, 2007. 'Les Archives ouvertes, quels atouts ?' *Pour la Science*, 352: 12-15

⁵ Baruch P. 2007, "Open Access Developments in France: the HAL Open Archives System", *Learned Publishing*, à paraître (Oct. 2007)

⁶ <http://www.arl.org/bm~doc/arlstat05.pdf>

⁷ <http://revues.org>

⁸ Voir le communiqué du CNRS à <http://www2.cnrs.fr/presse/journal/3441.htm> et la présentation d'ADONIS par Y. Maignien, directeur, à http://rpist.inist.fr/IMG/pdf/M._Maignien.pdf

⁹ <http://www.persee.fr>

¹⁰ <http://www.cairn.info>

¹¹ <http://www.publiclibraryofscience.org/>

¹² Le site « ROMEO » <http://www.sherpa.ac.uk/romeo.php>, ou <http://romeo.eprints.org> (Univ. de Southampton) recense la position des éditeurs sur le libre accès. L'Université de Lund (Suède) tient à jour une liste de journaux en libre accès : <http://www.doaj.org/>

¹³ voir référence 17

¹⁴ Initiée en 1991 à Los Alamos par le physicien américain Paul Ginsparg, tenue maintenant par Cornell University, (USA), la base de données ArXiv (<http://arxiv.org/>) a maintenant des ramifications internationales ; le miroir pour la France (<http://fr.arxiv.org/>) est géré par le CNRS. A ce jour (août. 2007), près de 450 000 articles y sont déposés

¹⁵ voir son blog : <http://openaccess.eprints.org/>

¹⁶ Organisée par l'Institut Max-Planck de Berlin en 2003, cette réunion a vu la signature par des universités et des organismes de recherche d'une déclaration d'intention en faveur des AO

<<http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>>. Cette manifestation s'est renouvelée en 2004 et 2005 (<http://www.eprints.org/berlin3/program.html>),

¹⁷ Les « Initiatives de Budapest » 1 et 2 (<http://www.soros.org/openaccess/fr/index.shtml>), fruits de réunions organisées par la Fondation Soros en 2001 et 2003 ont préconisé l'auto-archivage, BOAI-1 et la publication dans des journaux en libre accès (« auteur-institution paye »), BOAI-2. Ces réunions ont aussi élaboré les premières procédures d'interopérabilité des archives ouvertes : protocole OAI-PMH et ses successeurs,, qui permettent l'indexation et le « moissonnage » des documents déposés par les moteurs de recherche.

¹⁸ Voir bibliographie et liens dans :

- Chanier, T. *Archives ouvertes et publication scientifique*. Paris, L'Harmattan, 2004 ; http://archivesic.ccsd.cnrs.fr/sic_00001103

-
- Aubry, C.; Janik, J., eds. *Les archives ouvertes. Enjeux et pratiques. Guide à l'usage des professionnels de l'information*. Sciences et Techniques de l'Information. Paris : ADBS Editions, 2005.
 - L'Hostis, D et Aventurier, P. *Vers une obligation de dépôt . . . ?*, Paris, INRA, 2006 ;
http://archivesic.ccsd.cnrs.fr/sic_00115513

L'INIST-CNRS publie sur son site les actualités du Libre Accès et des Archives Ouvertes :

<http://openaccess.inist.fr/>

¹⁹ <http://www.PubMedCentral.nih.gov/>, archives des National Institutes of Health (NIH, USA), consacrée aux sciences biomédicales et de la vie

²⁰ Directory of Open Access repositories DOAR <http://www.opendoar.org/>, (Univ. de Nottingham) et Registry of Open Access repositories ROAR : <http://roar.eprints.org/> (Univ. de Southampton). A utiliser avec précaution, la compilation par un robot est source d'erreurs.

²¹ <http://www.ccsd.cnrs.fr/IMG/pdf/Communique-depresse11oct.pdf>.

²² Consulter <http://archivesic.ccsd.cnrs.fr/> (l'URL exacte sera fournie ultérieurement). La version anglaise est dans l'article cité en référence 5.

Voir aussi : *Archives ouvertes et archives institutionnelles des établissements d'enseignement supérieur et de recherche*. Paris, MENSUR-DGES, 2007. (<http://www.sup.adc.education.fr/Bib/Acti/ao/ao.htm>)