



HAL
open science

Pérenniser la publication scientifique: le mode “ Archives Ouvertes ”

Gabriel Gallezot

► To cite this version:

Gabriel Gallezot. Pérenniser la publication scientifique: le mode “ Archives Ouvertes ”. séminaire INRIA, 2-6 octobre 2006, Amboise. Sous la direction de Lisette Calderan, Bernard Hidoine et Jacques Millet, 2006, Amboise. sic_00105002

HAL Id: sic_00105002

https://archivesic.ccsd.cnrs.fr/sic_00105002

Submitted on 10 Oct 2006

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.

"Pérenniser le document numérique, séminaire INRIA, 2-6 octobre 2006, Amboise. Sous la direction de Lisette Calderan, Bernard Hidoine et Jacques Millet. Paris, ADBS Éditions, 2006".

Pérenniser la publication scientifique : le mode « Archives Ouvertes »

Gabriel Gallezot

Université de Nice - Urfist PacaC

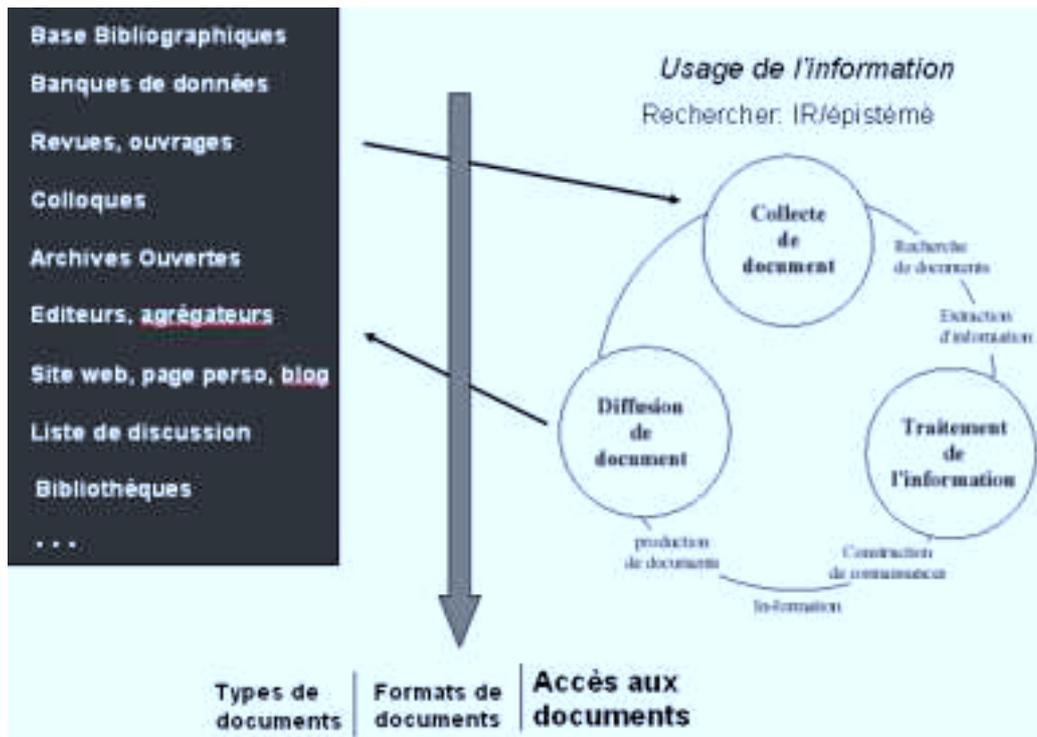
Pour bien appréhender la question de la pérennisation de la publication scientifique, il convient de rappeler brièvement le contexte de cette activité nodale du travail de la recherche.

1 La publication scientifique

1.1 Son inscription dans le cycle de l'IST

La publication d'un document scientifique est à la fois l'aboutissement et le commencement d'un processus de recherche. L'aboutissement consiste à publiciser les résultats d'une recherche, le commencement à entreprendre de trouver des documents déjà publiés pour engager une nouvelle recherche. Il y a une relation directe entre la recherche d'information (*information retrieval*, IR) et la recherche (*épistémé*). Ce lien, ou plus précisément ce liant, est introduit par la publication scientifique. Pour l'*épistémé*, c'est de découvrir, de produire de nouvelles connaissances. Pour la recherche d'information, il s'agit d'appréhender la complexité d'un phénomène par le repérage et la gestion d'un corpus de documents pertinent et le plus exhaustif possible afin d'extraire des informations et de les transformer ensuite en nouvelles connaissances [Ertzscheid & Gallezot, 2003]. Nous nommons ce processus le cycle de l'information scientifique et technique (IST), dont la figure 1 donne une représentation.

FIGURE 1 - LE CYCLE DE L'IST



L'activité informationnelle des chercheurs est circonscrite par ce modèle. Le « bouclage¹ » d'un cycle induit la construction de connaissances, d'informations et de documents. La collecte peut être réalisée à partir de banques de données, de sites web, d'expériences dans les laboratoires, de « butinage » (*browsing*) dans les rayonnages d'une bibliothèque... - autant de sources d'information. Le traitement correspond à l'activité cognitive des chercheurs ou à des manipulations par des outils informatiques. La diffusion est définie comme l'ensemble des opérations nécessaires à la propagation des connaissances. La transformation de l'information s'inscrit dans ce processus où la connaissance est la formation des idées, l'information est la mise en forme des connaissances (*in*-formation) et l'information inscrite sur un support constitue un document [Gallezot, 2002]. La pérennité de la publication scientifique doit donc s'appréhender en amont et en aval de ce processus : de la collecte des documents nécessaires à la production de textes et de données jusqu'à leur diffusion qui permettra une nouvelle collecte, etc.

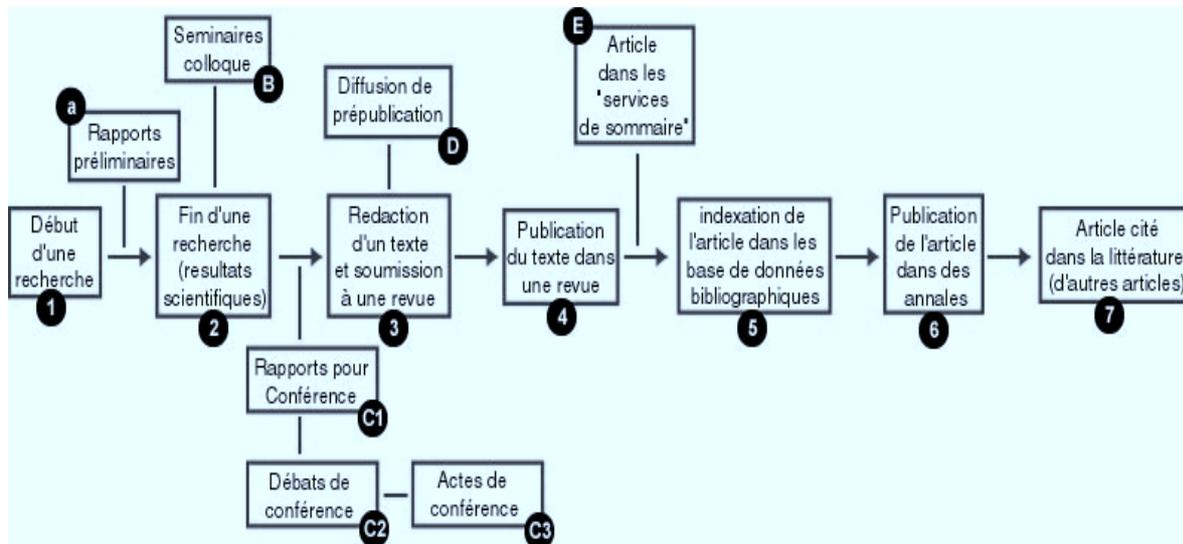
1.2 Le cycle de l'IST, sa place dans la communication scientifique

Le cycle de l'IST montre l'activité de recherche. Ce processus doit aussi être considéré dans la chaîne globale de la communication scientifique, notamment pour indiquer les différents statuts de documents possibles, les étapes de l'« écriture scientifique », les contingences d'ordre éditorial et, finalement, la vie même d'un texte scientifique ; et pour appréhender ainsi les « objets » à pérenniser. Le modèle de communication scientifique de Garvey et Griffith des années 1970², bien qu'il doive être remodelé (avec les documents numériques et selon les communautés scientifiques), présente l'avantage de décomposer de façon simple et linéaire le flux de l'IST (figure 2).

1 Le terme « bouclage » désigne la réalisation des différentes étapes du cycle, le passage de la collecte au traitement et du traitement à la diffusion.

2 Ce schéma est issu et traduit de l'ouvrage de Crawford, Hurd et Weller [1996, p. 9- 35].

FIGURE 2 - MODELE DE LA COMMUNICATION SCIENTIFIQUE SELON GARVEY/GRIFFITH



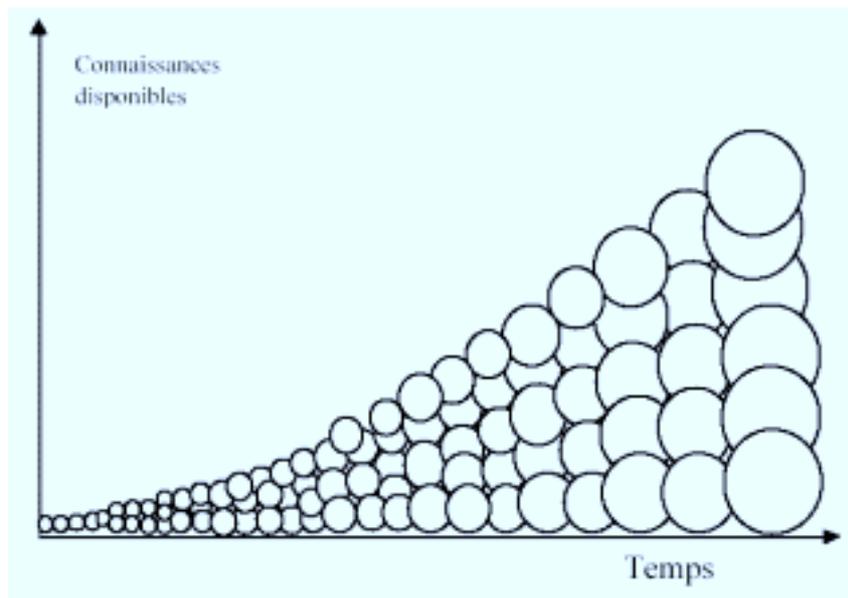
Ce schéma montre combien le processus de communication est lié à la pérennisation de la production scientifique. Chaque étape propose finalement un mode de pérennisation spécifique : l'inscription sur des carnets de laboratoire, dans des bases de données internes, des sites web, des archives ouvertes, des revues, des bases bibliographiques, etc. Ces modes de pérennisation sont associés aussi bien aux types de documents (texte, donnée primaire, etc.) qu'à leurs statuts (accès restreint, public, etc.). L'accès réseau généralisé, la production numérique des documents et leur stockage dans des entrepôts internes (accès restreint) ou publics modifient considérablement cette étape non seulement en terme de rapidité d'accès à l'information, mais aussi en termes d'élargissement du spectre de recherche documentaire ou encore de pérennisation élargie de la production scientifique. Par le biais d'un accès concentré et interopérable, les grandes banques de données internationales, les archives ouvertes et les outils de recherche associés offrent un plus grand partage de la production scientifique.

1.3 La publication scientifique, partitionnement de la science et accroissement de la quantité d'information

L'activité de publication scientifique trouve son origine dans la nécessité de publiciser les résultats de la recherche pour que d'autres recherches se fondent sur ces résultats et permettent de construire de nouvelles connaissances. Au passage, cette publicisation permet d'afficher la paternité d'une découverte et assure ou non une notoriété à l'auteur du texte. Ces deux aspects de la publication scientifique conduisent au partitionnement des savoirs. Le mythe du savant du siècle des Lumières, au savoir encyclopédique, s'est effacé pour laisser la place à un cloisonnement des spécialités. Un partitionnement des savoirs qui infère un cloisonnement des disciplines et impose finalement un fort accroissement de la publication scientifique.

La figure 3, empruntée à Swanson, illustre bien ce cloisonnement inhérent au processus de développement de connaissances nouvelles [Swanson, 1986]. Ajoutons à cette première raison une seconde qui consiste à créer « sa » revue pour marquer « son » champ de recherche et ainsi colliger encore un plus grand nombre d'articles sur un thème de recherche.

FIGURE 3 - LA COMPARTIMENTATION DES SAVOIRS



Partant de ce constat, tous les modèles de découverte de connaissance se proposent ainsi de rechercher des connexions latentes entre des disciplines devenues étanches. L'innovation ne consiste donc plus toujours à découvrir quelque chose qui n'existait pas, mais à transposer dans un autre domaine de la connaissance un phénomène déjà validé par ailleurs [Boutin, 2006]. La pérennisation de documents scientifiques est bien entendu une question sensible en découverte de connaissance, qu'il s'agisse de recherche sur des bases bibliographiques [Pierret, 2004], de corpus documentaires issus du web [Gordon, 1996] ou de bases de données primaires [Gallezot, 2002].

Le partitionnement des savoirs et l'accroissement de la quantité de textes scientifiques imposent plus que jamais la pérennisation de l'accès aux contenus de différents types de documents, non seulement dans le contexte de découverte de connaissance mais, plus généralement, dans le cadre d'un partage, voire d'un transfert de connaissances.

2 La question de la pérennisation de la publication scientifique

Nous avons vu dans la partie précédente sur la publication scientifique que la question de sa pérennisation se posait dans le contexte même de l'activité de recherche, mais aussi en terme patrimonial en raison du caractère public de cette publication. L'enjeu de la pérennisation est multiple et recouvre différentes facettes. Cette partie veut faire le point sur la démarche de pérennisation du document scientifique en mettant en évidence les raisons de cette démarche, ses acteurs, les documents concernés et les processus mis en œuvre.

2.1 Pourquoi pérennise-t-on ?

Nous venons de l'indiquer, la nécessité de pérenniser s'envisage sous différents aspects. Avec la question de l'accès aux documents et celle de la fonction patrimoniale, on pense à l'activité des bibliothèques. On pense aussi à l'effort de constitution des premières bases de données dans les années soixante, capables de capitaliser l'ensemble des références des publications scientifiques et autres brevets. On pense encore au web qui par les URL unifie l'accès aux documents, en attendant la généralisation d'URI, de DOI, d'identifiants OAI ou encore d'« ontologies métiers »

capables d'impulser la structuration de contenus pour accéder à l'information plus finement ou de manière granulaire, et avec une traçabilité.

Il s'agit donc de pérenniser l'accès aux documents dans leur format natif, dans leur granularité ou simplement par leur identification. De même il convient d'appréhender s'il s'agit de conserver le support et son contenu, uniquement le contenu, voire simplement la trace de son existence.

Bref, la pérennisation concerne les sédiments de la connaissance aussi bien en termes de conservation du patrimoine que de traitement de l'information. Elle sert ce que Fleck [1979] appelle le « collectif de pensée », une communauté de personnes échangeant mutuellement leurs idées et maintenant une interaction intellectuelle³.

2.2 Qui assure cette pérennisation ?

La question des acteurs de la pérennisation de la publication scientifique est vaste et sera traitée de manière plus restreinte dans la troisième partie de ce chapitre, sous l'angle des archives ouvertes. Rappelons toutefois ici les principales forces en présence.

Les scientifiques : ils produisent et, par leurs communautés ou leurs institutions, édictent des règles, des protocoles, des normes, des processus qui assurent aux données et aux textes une certaine stabilité et une certaine pérennité.

Les bibliothèques : elles donnent accès aux documents qu'elles conservent. Avec la documentation électronique, elles ont tendance à déléguer l'archivage aux éditeurs et à se concentrer sur la fourniture d'accès et la mise à disposition de lieux de consultation, en même temps qu'elles essaient de coordonner des projets d'entrepôts de documents dans nos universités.

Les éditeurs : la pérennité économique de leur activité dépend de la pérennité (accessibilité, archivage) des bases de données bibliographiques et des titres de périodiques qu'ils gèrent.

2.3 Que pérennise-t-on ?

Des documents bien entendu, mais de quels types ? Au regard de la première partie, aucun document scientifique ne doit être négligé. Les articles, les ouvrages occupent une place importante, mais le cycle de l'IST se nourrit de bien d'autres documents essentiels. La collection de données issues de la paillasse ou du terrain, les notes ou cahier de labo, les rapports intermédiaires sont autant de contenus inscrits susceptibles de produire des connaissances.

Un exemple prégnant du rapport entre collection de données primaires et nouvelle connaissance peut être apporté par la biologie moléculaire et la naissance de la bioinformatique. L'accumulation de données annotées issues du séquençage de génome a érigé le principe de double publication et ainsi permis l'accessibilité et le stockage pérenne de séquences nucléotidiques et ce dès le début des années quatre-vingt⁴. Cette démarche de pérennisation a donné naissance à une nouvelle discipline : la bioinformatique. Si elle s'attache à trouver ou développer des moyens de traitement automatique des données biologiques, cette discipline est soucieuse de la qualité de ces données et de leur pérennisation à travers la spécification de standards à vocation universelle pour les objets, pour les concepts biologiques et leurs relations ; spécification favorisées par l'apparition de nouvelles normes de médiation informatique ayant une portée

3 Thought collective et thought community.

4 En 1984, dans les « Instructions to Authors » du Journal of Biological Chemistry, il fut pour la première fois demandé aux auteurs de publier leurs séquences directement dans une banque de données comme GenBank ou EMBL Data Library. Un an plus tôt, la même revue informait les auteurs potentiels que leurs séquences ne seraient pas publiées dans la revue et qu'il n'était pas nécessaire de soumettre la séquence avec la proposition d'article.

générique, comme XML. On peut évoquer, par exemple, des travaux qui se sont très tôt penchés sur les ontologies d'objet biologiques pour rendre interopérables les bases de données⁵ [Schulze-Kremer, 1998].

Cette pérennisation des données primaires touche bien entendu toutes les disciplines dites des sciences exactes ou dures, où la nécessité de reproduire les expériences et la computation des données est primordiale. En SHS, toutefois, le développement des archives ouvertes pourrait permettre, au-delà du stockage du texte d'une publication, d'entreposer le matériel de recherche. Si, pour l'instant, les archives ouvertes sont utilisées comme une réplique électronique de la production papier, des appendices peuvent et devraient agrémenter les dépôts (données, images, annexes, films, liens, etc.). Pour des raisons techniques ou de coût, le support papier ne se prête pas à l'ajout de documents connexes à un article publié. Il s'agit ici de penser la pérennisation de l'association texte/donnée : un « document augmenté ». Si le stockage reste brut et non structuré, à l'inverse de la génomique, par exemple, un tel document présente au moins l'avantage de rendre possible l'accès aux données.

Nous avons évoqué, au paragraphe 1.3, le travail de Swanson à propos de la découverte de connaissance. Ce type de recherche est rendu possible par la constitution de bases bibliographiques et la pérennisation d'un travail de catalogage. Pour poursuivre avec les deux exemples cités (génomique et découverte de connaissance), la constitution de bases de données, au-delà d'un accès pérenne des documents, propose aussi de nouveaux modes de recherche d'information en liant par exemple une collection de séquences nucléotidiques à une collection de références bibliographiques (cf. Pubmed⁶) Ainsi, la pérennisation de documents dépasse la simple notion d'accès ou d'archivage patrimonial, pour entrer directement au cœur du processus de la recherche.

2.4 Modèles éditoriaux et enjeux d'une pérennisation de la production scientifique

Le mouvement du Libre Accès, les nouvelles dynamiques scientifiques à l'œuvre sur le web, les nouveaux habits des chercheurs, notamment en matière d'auto-archivage, les prises de position politiques et aussi parfois polémiques de certains d'entre eux (Lawrence Lessig notamment refusant de publier dans des revues « à exclusivité »), ainsi que l'arrivée de nouveaux entrants majeurs aux ambitions clairement affichées (Google, notamment, avec son service « Scholar⁷ ») - cet ensemble de paramètres bouscule pour le moins les modèles traditionnels de la publication et du même coup questionne la pérennisation de la production scientifique sous les aspects de l'accessibilité, de la visibilité, de l'évaluation et de la production de résultats.

À travers le modèle de communication scientifique établi par Garvey et Griffith (cf. *supra* § 1.2), modifié par l'accès réseau généralisé, la production numérique des documents et leur stockage dans des entrepôts internes (accès restreint) ou publics permettent, comme nous l'avons vu, un plus grand partage de la production scientifique. Entre le modèle éditorial traditionnel et celui introduit par l'auto-publication, quels sont les enjeux d'une pérennisation de la production scientifique ?

Considérons, le modèle suivant :

5 Par exemple : www.geneontology.org/.

6 www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed.

7 <http://scholar.google.com>.



Ce cycle correspond au processus traditionnel de publication où l'auteur produit un texte, qui est ensuite évalué. Un travail d'édition se met en place (corrections, mise en page, etc.), et le texte est alors diffusé. Ce texte est ensuite acquis par un lecteur (*via* une bibliothèque, une librairie, etc.) et puis le cycle peut recommencer avec la rédaction d'un nouveau texte à l'issue de plusieurs lectures.

L'auto-publication sur le web change ce modèle pour mettre directement en relation l'auteur et le lecteur, y compris dans le cadre d'un processus de revue par les pairs dans lequel des « non-pairs » disposent aussi d'un pouvoir de validation ou a minima de prise en considération de leurs commentaires. C'est alors la valeur d'usage qui devient la clé de voûte de la diffusion des connaissances.

Accès

Avec l'auto-publication de documents scientifiques, on peut observer plusieurs cas de figure : du chercheur qui, de façon isolée, héberge ses travaux sur sa page personnelle, au site web d'institution (laboratoire, université, société savante) qui expose certaines recherches de ses membres, jusqu'aux Archives Ouvertes (*Open Archives*) sur lesquelles sont déposées librement les productions (système appelé auto-archivage de *pre-prints* ou *post-prints*). Cette dispersion de l'auto-publication scientifique n'est pas sûre rendre durables l'accès (téléchargement et lecture) et la conservation d'un document. La réponse à apporter se trouve dans le développement d'entrepôts qui permettent leur interopérabilité *via* un standard comme OAI-PMH.

Visibilité

La publication des résultats de la recherche est un point nodal de l'activité du chercheur. Ses compétences sont évaluées par les textes qu'il publie et plus précisément par la visibilité de ses travaux, par le nombre de citations faites par ses pairs dans leurs propres publications. Ainsi l'accès durable (visibilité et stabilité) aux références bibliographiques, d'une part, puis l'accessibilité du plein texte, d'autre part, pour citer des travaux de recherche sont indispensables au mode d'évaluation de la recherche actuellement en vigueur. Les archives ouvertes permettent (notamment) d'optimiser la potentialité d'être cité.

Évaluation

Aujourd'hui les chercheurs ont le choix de rendre publique une pré-publication (*pre-print*) au sein d'une large communauté sans le contrôle d'un comité de lecture et/ou de perpétuer le schéma traditionnel de la publication scientifique où le contrôle des pairs conduit à une évaluation et une validation du contenu scientifique du document. Nombre d'éditeurs mettent par ailleurs en place des modèles mixtes permettant de garantir la libre circulation des résultats publics de la recherche et la valeur scientifique de ces documents, valeur pour laquelle l'usage seul ne saurait suffire. Les Archives Ouvertes ont développé des outils comme Opcit ou Citeseer⁸ qui offrent un autre mode d'évaluation de la recherche⁹ mais, tant que ces outils ne seront pas stabilisés,

8 <http://opcit.eprints.org/>, <http://citeseer.ist.psu.edu/>.

9 Si l'évaluation des chercheurs se réalise encore avec le rapport citation/lieu de publication (revue de rang n, facteur d'impact) ? Citeseer et OpCite sont des outils et des projets qui permettent de mesurer le nombre de fois

développés, utilisés et repérés par les instances d'évaluation¹⁰, les deux modèles et tout le spectre de leurs hybridations continueront de cohabiter.

2.5 Pérennité du modèle d'auto-publication

Le phénomène d'auto-publication [Gallezot, Rébillard et Vidal, 2004] n'est finalement pas nouveau si l'on pense au début de l'imprimerie et aux tracts. La ronéo et la photocopieuse, en leur temps, permettaient l'auto-publication de contenus. L'effet diligence décrit par Perriault [2002]¹¹ a simplement, là aussi, joué à plein. Il trouve aujourd'hui son apogée grâce d'une part au web, qui élargit la diversité des contenus et la dimension du public touché, et d'autre part aux outils de mise en forme des contenus à l'usage aisé et aux coûts d'équipement très abordables pour l'accès.

Si nous pouvons voir en Internet un projet politico-économique reconduit par une convergence technologique [Lacroix, Miège et Tremblay, 1994], nous pouvons appréhender le web comme l'émergence d'un espace d'expression et, pour ce qui nous concerne, d'une possibilité de diffusion de textes scientifiques.

Depuis l'origine scientifique de l'auto-publication sur le web jusqu'au Web 2.0, c'est le même souhait qui est formulé : la libre diffusion de contenus à des fins de visibilité, de simple expression directe ou d'échange. Par ce recours au lexique de la liberté et de l'ouverture, les initiateurs des expériences d'auto-publication sur le web teintent leurs actions d'un militantisme pour l'échange réciproque et la liberté de communication qui trouve ses origines dans l'idéal de l'internet scientifique des années quatre-vingt [Flichy, 2001].

À l'inverse, les tenants du processus classique de publication, sous couvert d'une nouveauté qu'il est nécessaire d'analyser, brandissent, comme seuls arguments pour mieux conserver le système d'évaluation et d'édition en place, la qualité des documents ou la légalité, alors même que de plus en plus d'études convergent pour attester que la qualité scientifique n'est aucunement altérée (c'est parfois même le contraire) par le jeu de systèmes d'évaluation ouverts, et ce indépendamment des champs dans lesquels ces études ont été menées [Lawrence, 2001 ; Kurtz, 2004 ; Odlyzko, 2002 ; Antelman, 2004]. Au final, ces différentes revendications soulignent l'enjeu de l'auto-publication comme modèle alternatif aux formes économico-éditoriales dominantes de la diffusion des savoirs.

Dans le cadre de l'édition scientifique, le modèle économique de l'auto-publication est celui de la publication à compte d'auteur, avec pour coût principal la mise en place des outils et infrastructures de diffusion. Ce modèle est même repris par des éditeurs scientifiques commerciaux, où c'est l'auteur du texte qui paie sa publication et non plus le lecteur (abonnement individuel, abonnement à des bouquets par bibliothèques, etc.). Et ce en « réponse » au

qu'un article ou qu'un auteur a été cité dans la bibliographie d'un texte déposé dans une archive ouverte. Il s'agit de rendre compte de l'importance des travaux d'un auteur non plus à travers des revues dont l'ISI (SCI) a déterminé le facteur d'impact, mais à travers des articles déposés dans une archive (CiteSeer) ou sur des plateformes respectant le protocole OAI (projet OpCite). Ainsi la revue n'est plus le centre du processus d'évaluation, elle n'est plus l'attracteur des « bons » papiers. Chaque article en ligne renferme potentiellement des travaux de grande importance que citeront d'autres chercheurs dans d'autres textes, et ce à une échelle qui, selon les projets, peut varier d'une archive à tous les articles en libre accès.

10 Il faut noter que la Grande Bretagne parse ses archives ouvertes (cf. IRRA : <http://irra.eprints.org/>) pour évaluer la recherche.

11 Une invention technique met un certain temps à s'acclimater pour devenir une innovation, au sens de Bertrand Gille, c'est-à-dire à être socialement acceptée. Pendant cette période d'acclimatation, des protocoles anciens sont appliqués aux techniques nouvelles. Les premiers wagons avaient la forme des diligences.

mouvement de l'Open Access qui prône notamment la liberté de l'auteur à disposer de sa production.

Les différentes expériences d'auto-publication sur le web se sont pour la plupart inscrites dans une dynamique non marchande, leurs initiateurs considérant bien souvent que l'inféodation au marché des formes traditionnelles et dominantes de publication était de nature à limiter la liberté de communication. Le développement de sites institutionnels pour l'édition scientifique et technique (recours aux financements publics), les pages personnelles (bénévolat), les réseaux *peer to peer* d'échange de fichiers constituent autant de traductions de cette revendication d'un accès libre aux contenus à partir d'échanges non marchands. Ce constat introduit la question de la prise en charge de la pérennisation. Les chercheurs doivent-ils, à travers leurs institutions, assumer la pérennisation de leur production scientifique ou peuvent-ils déléguer cette tâche aux éditeurs commerciaux ? La mobilisation des bénévoles et des organisations publiques, leurs énergies et leurs moyens perdureront-ils ?

L'histoire des technologies de communication, en particulier dans le domaine des médias, nous enseigne que les médias alternatifs comme les *fanzines* ou les radios libres sont bien souvent marginalisés sur le long terme au regard du poids du secteur marchand. L'auto-publication est-elle un mode alternatif voué à disparaître ou bien à coexister avec les modèles traditionnels ? On peut avancer qu'Internet n'est pas un média unidirectionnel, et qu'il se démarque ainsi de l'histoire de médias traditionnels. L'effet réseau ramifie, démultiplie, agrège selon la figure du rhizome... Le web a permis de donner au phénomène déjà ancien de l'auto-publication une consistance et une résonance particulières. Dans le cadre du développement socio-historique de l'internet, l'auto-publication constitue une modalité originale d'appropriation sociale du réseau, marquée par une dialectique mutualisation/marchandisation : des solutions techniques qui permettent l'accroissement de la production de contenus payants et gratuits, leur large diffusion ou leur mutualisation ; un mode de diffusion professionnel opposé à un mode de diffusion alternatif ; et une logique économique marchande confrontée à une logique non marchande.

Cette vision duale de l'auto-publication est-elle le reflet d'une phase intermédiaire de constitution ou d'un phénomène installé mais fluctuant au gré des évolutions techniques et des appropriations de l'Internet ? La persistance et la recomposition incessante de production/diffusion de contenu sur le web par des individus et des collectifs sont, à ce sujet, éclairantes. Des modèles parallèles continueront sans doute encore de coexister, selon une géométrie nécessairement variable.

3 Un mode de pérennisation : les archives ouvertes

Nous avons brièvement indiqué quelques enjeux de la pérennisation des documents scientifiques à travers leur inscription dans l'activité scientifique. Les aspects importants que nous avons soulignés ont trait à l'accès, au partage de connaissances. C'est là une démarche ancienne, mais régulièrement reconduite sous l'impulsion de politiques scientifiques, de politiques de l'IST et de l'émergence de technologies sans cesse renouvelées.

« Trois dynamiques se croisent aujourd'hui pour prôner le partage des connaissances scientifiques : le développement du Web et son architecture, la saturation des bibliothèques et les nouvelles relations avec les éditeurs, et enfin la mondialisation et les conférences au sommet qui l'accompagnent. Ces dynamiques sont à la source d'initiatives et de débats internationaux nourris. Elles pourraient donner naissance à une nouvelle organisation générale des connaissances, plus équitable, retrouvant ainsi les rêves des pères fondateurs des sciences de l'information, à condition de : 1) reconnaître la fonction patrimoniale des articles scientifiques, en leur conférant

le statut de bien public global quelques années après leur publication ; 2) favoriser l'échange direct entre chercheurs par une incitation aux archives ouvertes ; 3) reconnaître la plus value des revues et préserver le pluralisme par un marché équitable. [...]

» Ainsi, il serait naïf de considérer que les trois courants qui convergent en ce début de millénaire pour prôner un accès le plus immédiat possible aux connaissances scientifiques, reflètent des idées radicalement nouvelles. Bien au contraire, il s'agit d'un vieux rêve de l'humanité, dont d'ailleurs l'œuvre trop méconnue de Paul Otlet n'est qu'un des relais, plutôt récent. Néanmoins, à l'instar de ce dernier et, espérons-le, avec une réussite plus manifeste, nous nous trouvons dans une de ces périodes où l'Histoire hésite, où des bascules sont possibles. En effet, rarement sans doute d'aussi fortes pressions et opportunités ont coïncidé ; et, si les acteurs savent en trouver le chemin, il est peut-être possible d'imaginer de nouveaux compromis ouvrant très largement un accès généralisé aux connaissances scientifiques. [Salaün, 2003] »

C'est dans cette perspective que cette troisième partie se focalise sur les Archives Ouvertes comme un mode de pérennisation de la publication parmi d'autres ; nous observerons particulièrement le cas des Archives Ouvertes en France [Gallezot 2005].

3.1 Le Libre Accès dans le processus de communication scientifique

Sans remettre en cause la fonction organisatrice ni le travail des éditeurs, il faut aussi admettre que les connaissances inscrites sur les différents supports que représentent les revues ne circulent pas encore de façon optimale [Harnad 2001].

Quelques pionniers ont tracé de nouvelles voies en ce sens : Ginsparg a su valoriser la culture d'échanges de sa communauté en organisant une vaste banque internationale de données diffusant des pré-publications, devenue aujourd'hui l'une des principales sources d'information de son domaine [Ginsparg, 1994] ; le cogniticien Stevan Harnad milita pour le principe du *scholarly skywriting*, à savoir l'utilisation du réseau pour maximiser la diffusion des écrits scientifiques au-delà des supports existants [Harnad 1991].

Le mouvement engagé par Paul Ginsparg a permis le développement d'un questionnement sur les modèles éditoriaux en place. En dévoilant certaines des potentialités techniques et économiques offertes par le numérique, Ginsparg a amené les chercheurs à s'interroger sur leur lien de dépendance vis-à-vis des éditeurs traditionnels.

Plusieurs idées fortes organisent alors ce mouvement. La première, conformément au projet des membres fondateurs du web, consiste à exploiter les potentialités associatives dans la mise en commun des ressources. La deuxième consiste à exploiter le faible coût de fabrication eu égard à la capacité de dissémination des œuvres. La troisième consiste à rendre visibles, aux yeux d'un plus grand nombre, les différents états, à différentes étapes, de la production d'un document, et de permettre ainsi un plus large éventail de la critique. Ce point est important car il intervient en amont de la stabilisation et de la légitimation finale et traditionnelle par les pairs dont le système de filtrage est largement dominé par les éditeurs et un nombre relativement réduit et stable d'évaluateurs.

La quatrième idée est que l'accroissement des documents disponibles, la multiplicité des réseaux de liens en amont et en aval appellent le développement de nouvelles fonctions éditoriales permettant de travailler sur le processus d'écriture lui-même, sur les conditions de production et celles de citation des documents. Filtrage, navigation, gestion des points de vue peuvent rendre possible une meilleure appréhension de la vie du document ou bien de la communauté des

œuvres comme incomplétude en procès de production¹² pour reprendre ici la formule de A. N. Whitehead [1995].

3.2 Libre Accès, Archives Ouvertes et revues en ligne

Les Archives Ouvertes constituent un pan du mouvement du Libre Accès. L'accessibilité des documents, et plus particulièrement des textes scientifiques, au protocole OAI-PMH concerne aussi les revues et plus généralement tout types de documents.

L'*auto-archivage* (possibilité donnée à un auteur de déposer ses textes dans un entrepôt de documents) et la *pré-publication* (mise en ligne d'une version non évaluée d'un texte) sont des spécificités des Archives Ouvertes, mais toutes les formes de documents peuvent être concernées par ce mode de diffusion, à commencer par les *post-prints*, le matériel de recherche, les supports de présentation, etc.

Si le stockage de documents n'a pas besoin des Archives Ouvertes pour exister, le standard OAI-PMH lui donne une résonance importante. En effet, depuis l'avènement du web, un chercheur peut rendre accessibles ses publications sur son site personnel, ou sur le site de son laboratoire. Elles pourront alors être indexées par des moteurs de recherche spécialisés (Google Scholar, par exemple) ou généralistes. Déposées dans une Archive Ouverte, ces mêmes publications pourront néanmoins être repérées en tant que textes disponibles sur le web (si l'archive est indexée par les moteurs) mais aussi grâce à l'utilisation spécifique du protocole OAI-PMH qui fournit aux moissonneurs (OAIster, par exemple) les métadonnées descriptives qui vont favoriser la signalisation et la localisation de ces documents.

Auto-archivage et pré-publication mettent aussi en exergue un questionnement sur l'évaluation scientifique. Quand un chercheur dépose une pré-publication dans une Archive Ouverte, ses pairs accèdent directement au contenu, évaluent la qualité de ses propos selon leurs propres critères, et peuvent commenter le texte (*peer commentary*¹³, ou directement par courriel adressé à l'auteur). Libre au chercheur de corriger ou non son texte en fonction de la pertinence des commentaires. Dans le cas d'une revue à comité de lecture, le « libre au chercheur » se transforme en « nécessité de corriger ».

Le choix se résume-t-il donc à rendre public un *pre-print* au sein d'une large communauté ou de perpétuer le schéma traditionnel de la publication scientifique ? Sans autres modes d'évaluation de la recherche (cf. Opcit ou Citeseer) ces deux voies de publicisation des résultats de la recherche doivent être utilisées. Une part des revues concède désormais à l'auteur le droit d'auto-archiver son article. Le facteur d'impact (ou, pour de nombreuses disciplines de SHS, « *les-revues-qui-comptent* ») reste encore le critère d'évaluation majeur de la qualité scientifique des publications d'un chercheur. Il semble que le verrou psychologique de l'évaluation ait sauté, qui conduisait certains acteurs de la recherche française à camper sur leur position, et que cette possibilité de « double publier » permette désormais une expansion plus sereine de la libre diffusion des textes scientifiques.

¹² Plus prosaïquement il s'agit d'appréhender l'ensemble des corpus documentaires de la planète à travers le cycle de l'IST qui boucle à l'infini

¹³ Stevan Harnad, The invisible hand of peer review, www.nature.com/nature/webmatters/invisible/invisible.html.

3.3 Pré-publication, post-publication, publications électroniques et documents numérisés

Une pré-publication ou *pre-print* est un texte finalisé par son (ou ses) auteur(s), mais non encore publié (évalué, corrigé, etc.) par un éditeur. La pré-publication, sous forme papier, circulait généralement entre pairs et, depuis que les serveurs d'Archives Ouvertes existent, elle peut y être déposée.

Une post-publication ou *post-print* résulte du processus de publication géré par les éditeurs. Un chercheur envoie sa pré-publication, elle est évaluée par un comité scientifique, des corrections sont souvent demandées à l'auteur, qui renvoie alors le texte définitif (quelques allers et retours de ce type sont possibles), une mise en page est réalisée par l'éditeur (avec généralement l'aide de l'auteur qui respecte une feuille de style), la diffusion est à la charge de l'éditeur.

La publication électronique ou *e-print* est le terme employé pour indiquer la diffusion électronique d'un texte, qu'il s'agisse d'une pré- ou d'une post-publication.

On peut aussi distinguer la numérisation de documents qui, sans que l'on se préoccupe de savoir si ces textes ont été évalués ou non, rend accessibles électroniquement des textes existant initialement au format papier (articles de revues, extraits d'ouvrages, note d'auteur, etc.) et qui peuvent être déposés sur les serveurs d'Archives Ouvertes.

Le protocole OAI-PMH et les différents logiciels permettant l'archivage de documents rendent techniquement perméable la frontière entre pré-publication et post-publication, revue et archives. Non qu'il s'agisse d'un mélange des genres : c'est simplement que, aux yeux de l'utilisateur qui recherche l'information, le statut de ces items n'apparaît plus comme premier élément de sélection, du moins comme un élément constitutif de son choix.

Le contrat de lecture est tout autre, plus complexe et plus simple à la fois. Plus simple car c'est bien l'accès au contenu qui est essentiel, plus complexe justement dans un souci de traçabilité (origine et validité de l'information contenue dans le document). Les interfaces de recherche unifiées brouillent le statut du document... mais est-ce vraiment essentiel ? Où est la valeur du document ? Dans son contenu ou sur son « étiquette » ? Les questions restent identiques : qui publie, qui évalue, qui lit ?

3.4 Les documents concernés

Les articles scientifiques

Nous l'avons déjà évoqué plus haut, il peut s'agir de *pre-prints*, de *post-prints* et plus généralement de publications électroniques (*e-prints*). Et, dans le cadre de la définition stricte, il s'agit de publications électroniques auto-archivées, déposées par l'auteur ou un tiers à qui il peut avoir délégué le dépôt.

Si les archives se sont principalement constituées autour des *pre-prints*, elles contiennent aujourd'hui de plus en plus de *post-prints* (accord avec éditeurs, pas de cession de droit signée, *corrigenda*¹⁴).

14 Voir à ce sujet : RoMEO (www.lboro.ac.uk/departments/ls/disresearch/romeo/), SHERPA (www.sherpa.ac.uk/romeo.php?all=yes) et la « stratégie » Harnad/Oppenheim sur le contournement du copyright (www.ecs.soton.ac.uk/~harnad/Tp/2-Resolving-the-Anomaly/sld007.htm et www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm#Harnad/Oppenheim).

Documents numérisés

Qu'il s'agisse de Gallica¹⁵, Numdam¹⁶ ou Persée¹⁷, ces projets ont en commun la numérisation de documents. Ces projets visent à rendre publics sous forme électronique des documents existants sous une forme papier, difficilement consultable (accès et fragilité).

La « notice OAI » contenant les métadonnées (schématiquement : une indexation avec les descripteurs Dublin Core) associées au document numérisé permet une plus grande visibilité de ces archives de documents numérisés qui, pour certains, préexistaient à l'apparition du protocole OAI-PMH.

Les thèses et HDR

Ce type de documents constitue un autre point d'entrée important pour les archives ouvertes. Comme les articles scientifiques, les thèses présentent des résultats récents, mais pas (totalement¹⁸) publiés.

Ces thèses ou habilitations à diriger des recherches (HDR), issues du travail d'universitaires et de chercheurs appartenant à un laboratoire ou/et à une université, sont généralement stockées en bibliothèque, microfichées. Leur diffusion reste très faible.

Aussi leur publication en ligne au sein d'archives ouvertes permet non seulement une plus large diffusion des résultats scientifiques de l'auteur, mais elle valorise aussi les laboratoires et les universités d'où elles proviennent.

Adjuvants à la recherche

À l'instar de banques de données génomiques, les archives pourraient justement être des lieux où des informations complémentaires pourraient être ajoutées au document maître¹⁹.

Un schéma en couleur, une grille d'entretien, une photo de réaction, une illustration animée, des enquêtes, un carnet de laboratoire, des listes statistiques, etc. - bref, toutes sortes de résultats qu'il est difficile de publier sur papier, car ils sont trop volumineux ou trop coûteux à reproduire.

Ouvrages

Ce type de document n'est normalement pas concerné par les Archives Ouvertes. Néanmoins, un projet de numérisation comme Gallica propose des monographies qui seront accessibles sur un serveur OAI.

Les nouveaux projets de numérisation de documents reformuleront certainement l'orientation à propos de ce type de document. Le projet Print/book de Google, la « réponse » française autour de Gallica, ou encore le projet d'Internet Archive de Brewster Kahle (numérisation de plus d'un

15 <http://gallica.bnf.fr/>.

16 <http://www.numdam.org/>.

17 <http://www.persee.fr/>.

18 Selon l'usage des communautés scientifiques : une thèse peut être constituée de trois articles, déjà publiés dans des revues, auxquels on ajoute une introduction, une conclusion et une discussion (STM). Ce peut être aussi l'inverse : la thèse est découpée en trois parties pour faire l'objet d'autant de publications (SHS). Ou encore, notamment pour les habilitations à diriger des recherches (HDR), le document fait l'objet (après travail éditorial) d'une publication d'ouvrage.

19 Voir principe de « double publication » séquence génomique et article scientifique [Gallezot, 2002a].

million de livres provenant d'une douzaine de bibliothèques dans cinq pays²⁰), s'ils ne respectent pas nécessairement le protocole OAI-PMH, permettront néanmoins l'accessibilité aux documents (avec certaines limites selon les projets).

Documents de tous formats (vidéo, audio, etc.)

Nous avons surtout insisté sur le texte et particulièrement sur les articles scientifiques, mais d'autres types de documents peuvent être archivés. Des archives ouvertes qui ont trait aux langues proposent des séquences sonores, d'autres proposent des séquences vidéo²¹...

Notices bibliographiques

Si ce dernier type de documents semble évident, puisque le principe du protocole OAI-PMH est d'associer au document archivé une notice descriptive, cette seule notice ne devrait pas être identifiée comme document.

Plusieurs situations sont à observer. Un certain nombre d'archives ouvertes proposent uniquement cette notice de métadonnées, soit parce que le document n'est pas encore disponible (il peut ne pas être encore numérisé, un temps de latence peut être demandé par un éditeur avant auto-archivage ou l'accord des ayants droit peut ne pas avoir été obtenu), soit parce qu'il ne le sera jamais.

Cette opération permet toutefois, de constituer une base bibliographique autour d'un laboratoire, d'une thématique, d'une communauté, d'une institution, d'une production nationale²², etc.

Les objectifs des Archives Ouvertes sont ici détournés, néanmoins des documents peu ou pas référencés dans des banques de données bibliographiques pourront ainsi être signalés et retrouvés.

3.5 Typologies des acteurs

Les communautés scientifiques (chercheurs)

La publication des résultats de la recherche est un point nodal de l'activité du chercheur. Ses compétences sont évaluées par les textes qu'il publie et plus précisément par la visibilité de ses travaux, par le nombre de citations qui en sont faites par ses pairs dans leurs propres publications. Ainsi, en publiant ses travaux dans des archives ouvertes, le chercheur optimise la potentialité d'être cité et partage ses résultats.

Chaque communauté scientifique adopte des usages singuliers face à l'auto-archivage. Ces usages varient selon les disciplines, selon la place des revues par rapport aux livres (distinction à observer entre sciences humaines et sociales (SHS) et sciences, techniques et médecine (STM), par exemple), la taille de la communauté, la position des *gatekeepers*²³, les types d'archives

20 Peter Suber, Le gigantesque projet de bibliothèque de Google, SPARC Open Access Newsletter, January 2, 2005, issue 81, www.earlham.edu/~peters/fos/newsletter/01-02-05.htm#google. Traduction de M. Delhaye dans la revue de Biblioacid, février 2005, n° 2, <http://biblioacid.typepad.com/ba/pdf/BAv2n1.pdf>.

21 Les Archives ouvertes peuvent rendre disponibles toutes sortes de documents. Seules des limites de stockage et de transmission des données sont peut-être à observer, ainsi que la disponibilité des lecteurs de médias nécessaires pour pouvoir lire les documents.

22 Projet de portail SUDOC.

23 « Les gardiens des portes de la discipline, les sages qui veillent sur le développement de la discipline, ce sont les gardiens du paradigme dominant » Jean-Claude Guédon, in: Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing.

ouvertes proposés, l'éventuel accord des éditeurs du domaine... - bref, de l'histoire constituée de la communauté scientifique considérée et de la prise de position des institutions qui l'entourent.

Les bibliothèques (bibliothécaires/documentalistes)

Le mouvement des Archives Ouvertes doit aussi beaucoup aux bibliothèques, notamment aux bibliothécaires et documentalistes américains qui se sont très tôt élevés contre l'impossibilité de fournir à leurs usagers tous les documents qu'ils cherchaient. Le coût des abonnements aux revues ne cessant de croître²⁴, les bibliothèques ont l'obligation d'opérer des choix de bouquets de revues en ligne. Ces derniers ne contentent bien évidemment pas toutes les communautés scientifiques présentes sur les campus. En France, les acteurs de la documentation participent ou initient aussi des projets²⁵.

Les institutions (notion de service public)

Le récent protocole d'accord²⁶ entre de nombreuses institutions scientifiques et universitaires²⁷ indique le Centre pour la communication scientifique directe (CCSD) et sa plate-forme HAL comme moyen (direct ou indirect) pour réaliser l'identification, la diffusion, la valorisation, la promotion et le suivi de la production scientifique de leurs chercheurs et enseignants-chercheurs. Il précise encore, qu'il s'agit de « permettre à la fois aux chercheurs de communiquer directement leurs travaux à l'ensemble de la communauté internationale vers des archives ouvertes, aux institutions de recherche (universités, EPST et grandes écoles) de rassembler, d'afficher, de valoriser, de promouvoir et de pérenniser leur production scientifique et, de façon générale, à la recherche française d'améliorer sa visibilité au sein de la communauté scientifique internationale ».

Il faut remarquer que cet élan provient des acteurs de la recherche, sans qu'aucune position politique n'ait été indiquée sur ce sujet qui touche pourtant le bien public.

Les entreprises (éditeurs)

La fusion des groupes d'édition et l'internet ont mis en exergue un modèle économique qui ne semble pouvoir tenir. En effet, l'édition scientifique est passée de sociétés savantes à l'oligopole d'agrégateurs de contenus. Tenable sur support papier, cette situation ne l'est plus avec les moyens de diffusion offerts par Internet. Les chercheurs souhaitent pouvoir obtenir tous les contenus scientifiques dont ils ont besoin pour leur activité, mais ils veulent aussi que leurs articles soient visibles le plus rapidement possible. La diffusion directe des résultats scientifiques semble être efficace. L'industrie de l'édition scientifique se réorganise autour de modèles économiques différents (auteur/payeur, par exemple) avec des services ajoutés, et ils influencent encore l'évaluation de la production scientifique.

24 Source ARL : www.arl.org/stats/arlstat/graphs/2003/monser03.pdf.

25 On notera : Couperin qui mutualise les achats de bouquets pour les Universités et les EPST ; Persée, un portail de revues en SHS numérisée ; Sparte, un projet d'archivage de thèse ; STAR, une plate-forme de recensement de thèses archivées ; le portail SUDOC qui se propose finalement d'être un moissonneur national...

26 Protocole d'accord en vue d'une approche coordonnée, au niveau national, pour l'archivage ouvert de la production scientifique. www.couperin.org/article.php3?id_article=366.

27 CEMAGREF, CIRAD, CNRS, CPU, INRA, INRIA, INSERM, institut PASTEUR, IRD, Conférence des grandes écoles.

Le politique (une politique de l'IST)

En France, nous avons eu dans les années soixante-dix une politique de l'IST qui concernait la constitution et l'interrogation des bases de données. Face à ce renouveau de l'accès au contenu scientifique, un élan et une impulsion politique forts seraient souhaitables. Des avancées ont lieu : en plus des prises de position et des projets en faveur du Libre Accès de la part d'organismes de recherche scientifique et d'établissements universitaires²⁸ vus plus haut, on rappellera une déclaration des ministères de la recherche de trente-quatre pays, dont la France, qui affirmaient en janvier 2004 que le Libre Accès « contribue de façon décisive aux progrès de la recherche scientifique et à l'innovation ». Ces avancées constituent un substrat favorable à l'émergence d'une politique nationale.

Conclusion

La question de la pérennité de l'archivage des documents numériques et, à travers elle, celle de l'archive ouverte est donc à appréhender sous les aspects de l'accessibilité, de la visibilité, de l'évaluation et de la production de résultats scientifiques, avec quelques items sous-jacents comme :

- les aspects économiques de coût de développement et de maintenance des archives ;
- les aspects politiques de la recherche : bien public, accès public, évaluation de la recherche, visibilité institutionnelle, etc. ;
- le mode de structuration des dépôts, les formats natifs et leurs « lecteurs » *ad hoc*, l'« XMLisation » et sa multitude de mise en forme indépendante des « lecteurs », etc.

L'argument de l'accès à l'information scientifique et technique des années soixante - soixante-dix est reformulé de façon similaire pour les banques de données génomiques, puis de manière générale pour le Libre Accès, avec toutefois un changement de paradigme : de l'accès à la signalisation du document (la référence bibliographique) à l'accès au document lui-même, et plus profondément du document (objet) à son contenu (information).

Ce changement a été possible grâce à une prise de conscience des chercheurs et des responsables de bibliothèques devant les difficultés d'accès aux textes scientifiques. Des propositions concrètes pour auto-archiver des documents ont permis de contourner ces problèmes.

Si la France n'est pas en reste, avec un nombre de projets non négligeable et quelques institutions importantes qui revendiquent leur adhésion au Libre Accès, il semble néanmoins que le protocole d'accord en passe d'être signé peut donner une impulsion forte en faveur de ce mouvement et peut contribuer ainsi, de façon décisive, non seulement aux progrès de la recherche scientifique mais à l'innovation, à la pérennisation des sédiments cognitifs et à leur insertion durable dans les espaces ou environnements numériques de travail (ENT) en projet dans nos universités. On retrouve ici les rêves d'une « archithèque »²⁹ formulés par Jean-Michel Salaün : il s'agit bien de gérer le patrimoine informationnel numérique de la communauté qu'elle sert et qui lui fournit ses moyens.

28 Cf. les projets cités plus haut.

29 <http://grds04.ebsi.umontreal.ca/jms/index.php/2006/09/04/70-reves-d-architheque>. Voir aussi le chapitre de cet ouvrage rédigé par Jean-Michel Salaün.

Références

- ANTELMAN K. (2004). *Do Open-Access Articles Have a Greater Research Impact?*
www.lib.ncsu.edu/staff/kantelman/do_open_access_CRL.pdf
- BOUTIN E., GALLEZOT G., QUONIAM L. (2006). *Détecter l'innovant sur le web par des techniques non booléennes : méthode, outils, application.*
http://archivesic.ccsd.cnrs.fr/sic_00079095/en/
- CRAWFORD S. Y., HURD J. M., WELLER A. C. (1997). *From Print to Electronic; The Transformation of Scientific Communication.* Medford : Information today Inc ; Silver Spring : Asist monograph series
- ERTZSCHEID O., GALLEZOT G. (2003). *Chercher faux et trouver juste.* X^e colloque bilatéral franco-roumain, CIFSIC, Université de Bucarest, 28 juin - 3 juillet 2003, SFSIC - Université de Bucarest
- FLECK, L. (1935 ; 1979). *Genesis and Development of a Scientific Fact.* 2e éd. by Thaddeus J. Trenn and Robert K. Merton, transl. by Frederick Bradley, foreword by Thomas S. Kuhn. Chicago : Chicago University Press, 1979. - 222 p. (1^{ère} éd. en allemand, 1935)
- FLICHY P. (2001). *L'imaginaire d'Internet.* Paris : La Découverte
- GALLEZOT G. (2005). « Les Archives Ouvertes en France ». In Aubry C. et Janik J. (dir.), *Les Archives Ouvertes : enjeux et pratiques.*, Paris : ADBS Éditions/CNRS-MRCT
- GALLEZOT G. (2002a). « La recherche in silico ». In Chartron G. (dir.), *Les chercheurs et la documentation électronique : nouveaux services, nouveaux usages.* Paris : Édition du Cercle de la Librairie
- GALLEZOT G. (2002b). « Exploration informationnelle et construction des connaissances en génomique ». *Les Cahiers du numérique*, n° 3, p. 121-136
- GALLEZOT G., RÉBILLARD F., VIDAL G. (2004). « L'auto-publication ». In Chartron G. et Rebillard F, *Modèles de publication sur le web*, Rapport d'activités AS-CNRS 103.
http://archivesic.ccsd.cnrs.fr/sic_00001159
- GINSPARG P. (1994). « First Steps Towards Electronic Research Communication ». *Computers in Physics*, vol. 8, n° 4. Repris dans *Solaris*, n° 3,
www.info.unicaen.fr/bnum/jelec/Solaris/d03/3ginspar.html
- GORDON M. D., LINDSAY R. K (1996). « Toward discovery support systems: a replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil ». *Journal of the American Society for Information Science*, vol. 47, p. 116-128
- KURTZ M. J. (2004). *Restrictive access Policies Cut Readership of Electronic Research Journal Articles by a factor of Two.* Document de travail. <http://opcit.eprints.org/feb19oa/kurtz.pdf>
- LACROIX J.-G., MIÈGE B., TREMBLAY G. (dir.) (1994). *De la télématique aux autoroutes électroniques : le grand projet reconduit.* Presses de l'Université de Grenoble
- LAWRENCE S. (2001). « Online or Invisible ? » *Nature*. 411(6837):521,
www.neci.nec.com/~lawrence/papers/online-nature01
- HARNAD S. (2001). « For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving ». <http://cogprints.soton.ac.uk/documents/disk0/00/00/16/39/index.html>
- HARNAD S. (1991). « Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge ». *Public Access Computer Systems Review*, 2 (1), p. 39-53. Voir également l'ensemble des publications de Stevan Harnad à l'adresse <http://cogsci.soton.ac.uk/~harnad/intpub.html>

- ODLYZKO A. (2002). « The rapid evolution of scholarly communication ». *Learned Publishing*, 15:7-19
- PIERRET J.-D., BOUTIN E. (2004). « Découverte de connaissances dans les bases de données bibliographiques. Le travail de Don Swanson : de l'idée au modèle ». *ISDM*, n° 149
- PERRIAULT J. (2002) *L'accès au savoir en ligne*. Paris : Odile Jacob
- PEDAUQUE R. T. (2003). *Document : forme, signe et médium, les re-formulations du numérique*. RTP-DOC (CNRS 33) (collectif d'auteurs). http://archivesic.ccsd.cnrs.fr/sic_00000511.html
- PEDAUQUE R. T. (2005a). *Les déplacements documentaires - version annotée*. RTP-DOC (CNRS 33) (collectif d'auteurs). http://rtp-doc.enssib.fr/dw2_out.php?id=96
- PEDAUQUE R. T. (2005b). *Le texte en jeu : permanence et transformations du document*. RTP-DOC (CNRS 33) (collectif d'auteurs). http://archivesic.ccsd.cnrs.fr/sic_00001401.html
- SALAÛN, J.-M. (2003). *Publications scientifiques : Web, bibliothèques et bien public mondial*. http://archivesic.ccsd.cnrs.fr/sic_00000438/en/
- SCHULZE-KREMER S. (1998) « Ontologies for Molecular Biology ». In *PSB 98, On-Line Proceedings*
- SWANSON D. (1986). « Fish oil, Raynaud's syndrome, and undiscovered public knowledge ». *Perspect. Biol. Med.*, 30, 7-18
- WHITEHEAD A. N. (1995). *Procès et réalité : essai de cosmologie*. Paris : Gallimard