

Des archives ouvertes sur le futur

Nicolas Balacheff

directeur de recherche au CNRS

Laboratoire d'Informatique de Grenoble

Equipe Modèles et Technologies pour l'Apprentissage Humain

Lorsque je cherche un article ou un livre pour les besoins de mon travail, me rendre dans une bibliothèque n'est plus la seule solution. Je trouve une grande part de la documentation dont j'ai besoin grâce aux abonnements électroniques auxquels mon laboratoire souscrit, et - de façon plus incertaine mais souvent surprenante - grâce à Internet qui met à ma disposition un ensemble impressionnant de ressources y compris des ouvrages complets (depuis quelques temps avec [Google Print](#), et plus récemment avec [Europeana](#). Pourtant, bien des références qui me seraient utiles sont encore difficiles d'accès, soit qu'elles correspondent à des textes non encore « publiés » (les délais sont terriblement longs), ou que les modes de distribution des éditeurs ne soient pas assez performants (ce qui est le cas de bien des petits éditeurs pourtant par ailleurs très engagés et soutenant solidement notre activité). Il semble que la situation de mes collègues mathématiciens et physiciens soit bien meilleure ; surtout pour ces derniers qui font un usage intensif de publications mises systématiquement à disposition en ligne. Les « archives ouvertes ». Voyons de quoi il s'agit...

Le concept d'archive ouverte est né au début des années 90, au laboratoire de physique de Los Alamos, avec la création d'[ArXiv](#) pour rendre disponible, collectivement et de façon ouverte, les résultats des travaux les plus récents. Cette archive a connu un succès rapide ; elle réunit aujourd'hui plus de 300 000 documents et on relève environ 140 000 connexions par jour^[1]. L'initiative des physiciens s'est ensuite étendue à d'autres disciplines scientifiques : mathématiques, biologie, informatique. Plus important, une véritable révolution de la communication scientifique s'est imposée avec la création de l'[Open Archive Initiative](#) à laquelle adhère la plupart des disciplines scientifiques et des secteurs importants des sciences humaines et sociales (SHS). En France, les archives ouvertes se sont développées sous l'impulsion de Franck Laloë, fondateur du *Centre pour la Communication Scientifique Directe* ([CCSD](#)). La base de données *Hyper Article en Ligne* ([HAL](#)), hébergée par le CCSD, accueille les archives de toutes les disciplines, des mathématiques aux SHS. Mais avant de considérer les archives ouvertes du point de vue de notre domaine, je voudrais donner quelques précisions sur ce qu'elles sont en général et apporter quelques réponses à des questions souvent suscitées par cette nouvelle forme de communication.

L'idée de départ est simple : les chercheurs sont les producteurs des contenus des publications scientifiques, avec la banalisation de l'accès à des traitements de textes performants ils ont aussi en charge la préparation des documents avec une mise en page permettant l'impression, puis ils sont les premiers et presque les seuls acheteurs de ce qu'ils ont produit... Le coût est très élevé au regard du budget des équipes. Notons, de plus, que les citoyens qui financent la recherche, paient deux fois le coût de la diffusion de ses résultats ! Les progrès récents de la communication numérique, et

la généralisation du web, ont fait tomber les derniers obstacles à la communication scientifique directe. Chaque chercheur peut mettre à la disposition de ses pairs, et au-delà à la disposition de quiconque est intéressé, les résultats de ses travaux, pour un coût marginal dans la mesure où cela ne consomme qu'une très petite part de l'infrastructure et des ressources qu'il ou elle utilise professionnellement.

La recherche française est fortement engagée dans le mouvement de la communication scientifique directe, avec un soutien remarquable de l' [Académie des Sciences](#) et des grands établissements (CNRS, INRIA, INSERM, etc.) Le secteur des [SHS](#) est présent—mieux que cela n'est le cas dans d'autres pays—y compris les sciences de l'éducation qui constituent un sous-ensemble important de ces archives. Des archives ont été créées pour accueillir les publications de travaux sur les technologies pour l'éducation et la formation, à l'initiative des informaticiens ([EIAH\[2\]](#)) et des sciences de l'information et de la communication ([TEMATICE](#)). Deux archives pour un même domaine ? C'est effectivement le cas, mais les ancrages disciplinaires sont si forts, et les problématiques parfois tellement contrastées, qu'il ne faut pas trop s'en étonner. Le cas est intéressant pour remarquer que cela ne crée pas de difficulté dans la mesure où ces archives « savent » communiquer à propos de leurs contenus (grâce au protocole OAI-PMH[3], mais peu importe...). Les systèmes d'information modernes permettent de donner sans perdre, de déconstruire et reconstruire la diversité et la complexité ; ces atouts sont particulièrement stratégiques pour la recherche pluridisciplinaire.

Pour avoir participé activement à la création de l'archive EIAH, puis dans ce même domaine des technologies pour l'apprentissage, à la création de l'archive européenne [TeLearn](#), j'ai pu constater qu'après avoir souscrit à l'idée généreuse du partage, qui séduit aisément, bien des réticences se font jour. La première est celle de la propriété. En effet, les publications que nous souhaitons partager ont été ou vont être publiées, et la question du copyright se pose. La politique des éditeurs est très variée, elle va de l'accord pour la mise en ligne de versions initiales ou de pre-prints, jusqu'à l'embargo total. On trouve sur le site britannique [ROMEO/SHERPA](#), une information détaillée sur la position d'un grand nombre d'éditeurs. Il faut pour ces textes être prudent, vérifier ce que l'on a signé, et éventuellement confirmer auprès de l'éditeur. En revanche, il y a une quantité importante de textes qui sont « publiés » dans des ouvrages périssables sans propriété clairement identifiée - la fameuse littérature grise - et qui sont rapidement introuvables alors qu'ils sont effectivement utilisés. Ces documents devraient être archivés, et l'être ouvertement pour être accessibles facilement. A nouveau, la question de la propriété est posée. Mais cette fois, elle est d'une autre nature, il s'agit de la protection de l'auteur. On est un peu effrayé de mettre sur Internet un document dont on ne sait pas par qui il sera récupéré et ce qui en sera fait. On doit comprendre que, dans ce cas, les archives ouvertes, par leur caractère institutionnel, apportent une sécurité par le référencement qui atteste l'antériorité et la matérialité de la publication. Un bel exemple de document qui devrait systématiquement être diffusé de cette façon est la thèse, et le moyen existe : [Thèse en ligne](#).

Enfin, une objection souvent avancée est celle de la qualité. Les revues éditées classiquement apportent, avec plus ou moins de fiabilité, une garantie de qualité. Qu'en est-il des archives ouvertes ? La réponse est décevante : les archives n'apportent pas de garantie. La vérification des documents déposés est minimale, disons qu'elle consiste en la vérification du contenu qui doit correspondre au thème scientifique, et de la forme qui doit être au moins celle que l'on donne à un article que l'on soumet. Pour le reste... En fait, les « archives ouvertes » sont un moyen de diffusion gratuite et facile, l'attribution d'une marque de qualité reste à la charge des communautés scientifiques qui doivent s'organiser pour cela soit en créant des Journaux ouverts ou des collections (c'est par exemple le cas de la revue [STICEF\[4\]](#), soit en créant des « tampons » de qualité (celui

d'une institution comme le CNRS ou l'INSERM, ou d'une communauté organisée comme le « [Faculty of 1000](#) » en biologie). On peut pour conclure retenir l'image suivante : les archives ouvertes constituent une ressource en matière première, ce que l'on peut en faire est laissé à notre initiative. Nous n'en sommes qu'au commencement.

Bien que rapide, ce billet (ce *blog* ?) devrait donner à son lecteur une idée un peu plus précise de ce que sont les archives ouvertes et lui permettre de percevoir leur intérêt. Il répond aussi à quelques unes des questions que l'on se pose souvent à leur propos. Tout cela a un but : vous amener à considérer ce que nous pouvons faire dans le domaine qui nous occupe, celui de *la recherche sur l'enseignement et l'apprentissage des mathématiques*, notamment la didactique des mathématiques. Jusqu'ici, les chercheurs de notre secteur n'ont marqué que peu d'intérêt pour ce type de diffusion. Ce n'est probablement pas par manque d'attention à ce qu'Internet peut apporter. Des sites comme [Publimath](#) ou [MATHDI](#) donnent accès - pas toujours de façon ouverte - à de l'information bibliographique, mais il s'agit surtout de notices. Peu de sites offrent des ressources en ligne de façon systématique (une idée qui guida la création de [La lettre de la Preuve](#), il y a une dizaine d'années). Un tour sur les sites web des IREM[5] est décevant de ce point de vue, bien que l'on puisse trouver quelques textes. Du côté de l'INRP la mise en ligne récente de la [Revue Française de Pédagogie](#) représente un progrès significatif. On se prend à espérer que les IREM fassent de même. Tout cela prendra du temps, à la fois parce qu'il faut des moyens, de l'énergie, mais aussi parce qu'il faut changer les mentalités. Quoiqu'il en soit, chacun peut dès aujourd'hui contribuer à faire évoluer les choses. Sans avoir besoin de beaucoup plus qu'un clavier et un peu d'altruisme, il est possible pour tout chercheur, sur sa seule initiative, de télécharger les articles et communications de notre domaine dans l'entrepôt « *Histoire et perspective sur les mathématiques* » de HAL. J'encourage en particulier le dépôt systématique de pré-publications sur le portail documentaire de Mathdoc, en remarquant que « *mathematics education* », avec quelques sous-rubriques, fait partie des thèmes éligibles.

Cet engagement de chacun préparera et soutiendra les initiatives institutionnelles dont nous ne saurions nous passer. L'ARDM[6] devrait susciter un mouvement dans ce sens. Quant aux IREM, ils ont un trésor de publications qui devrait être mis à disposition sous la forme d'une archive ouverte. Enfin, l'INRP a un rôle tout particulier à jouer :

Dans le contexte actuel, il est difficile d'écrire ces lignes en ignorant le sort de *l'Institut National de la Recherche Pédagogique*. Aussi j'aimerais terminer ce billet par un « envoi », au risque de paraître trop solennel - mais je suis sûr que le lecteur saura faire la part des choses. A l'exemple des grands établissements de recherche, l'INRP doit prendre l'initiative d'organiser une communication scientifique moderne dans son secteur d'activité. Il y en a les moyens et les compétences. Par ce geste concret, il affirmera sa capacité à *rassembler la recherche française sur l'enseignement et l'apprentissage des savoirs académiques et professionnels*. La communication scientifique n'est pas qu'une affaire de gestion et d'infrastructure, c'est aussi et d'abord une affaire de science et de politique scientifique. Une action volontaire dans le domaine des archives ouvertes, en bonne coordination avec l'ensemble des forces universitaires, en construisant sur ces archives les moyens de leur exploitation, avec une perspective européenne affirmée, aura sans aucun doute un impact international pour la diffusion de nos valeurs scientifiques et éducatives.

L'INRP le peut. Un INRP acteur de la structuration de la recherche, fédérateur de nos efforts, interface entre le monde de la recherche et celui très vaste de ses utilisateurs : l'éducation, la formation et plus largement une économie dont on nous dit qu'elle est celle de la connaissance.

Notes :

[1] D'après C. Aubry et J. Janik (2005) *Les archives ouvertes, enjeux et pratiques*. Paris : ADBS éditions.

[2] Environnements informatiques pour l'apprentissage humain

[3] Open Archive Initiative's Protocol for Metadata Harvesting

[4] Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation

[5] Institut de Recherche sur l'Enseignement des Mathématiques

[6] Association pour la Recherche en Didactique des Mathématiques