

Laurent Romary

Inria et Université Humboldt de Berlin

COMMUNICATION SCIENTIFIQUE : POUR LE MEILLEUR ET LE PEER

Les fondements de la communication scientifique ont profondément évolué au cours des quinze dernières années. En réponse à la crise des revues (*serials crisis*) et conscients des nouvelles potentialités technologiques d'Internet, des chercheurs, bibliothécaires ou décideurs ont cherché à mettre en œuvre des moyens de diffuser largement et si possible librement l'ensemble du patrimoine des publications scientifiques. Ceci s'est traduit notamment par la prolifération au niveau international d'archives dites institutionnelles, permettant à tout chercheur de déposer ses productions pour une diffusion immédiate en ligne.

Cependant, la situation globale est loin d'être claire. Le caractère très fragmenté du paysage des archives électroniques¹ rend difficile la perception de l'usage effectif qu'en font les chercheurs, ainsi que l'impact que cet usage, potentiellement croissant, peut avoir sur l'édition scientifique traditionnelle. Une meilleure compréhension de ces phénomènes, à la fois sur le plan technique et sociologique, permettrait non seulement

d'améliorer la collaboration avec les éditeurs, mais aussi de définir des schémas directeurs pour les politiques d'information scientifique des organismes de recherche.

Fruit d'un débat très animé au niveau européen, le projet PEER (*Publishing and the Ecology of European Research*, <<http://www.peerproject.eu>>) vise à apporter des éléments de réponse dans le cadre d'un espace de collaboration entre éditeurs et chercheurs.

Émergence du projet PEER dans un contexte politique agité

Les réflexions ayant mené à la définition du projet PEER remontent au début de l'année 2007, quand, dans le contexte de l'annonce par la Commission européenne des *Recommandations concernant l'évolution technique et économique des marchés de l'édition scientifique en*

Europe (Dewatripont *et al.*, 2006), les communautés de chercheurs et d'éditeurs se sont lancées dans une bataille de tranchées à coups, respectivement, de pétitions (cf. <www.ec-petition.eu>) et de communiqués de presse (cf. <www.stm-assoc.org>), plaçant *de facto* la Commission européenne en position d'arbitre. De part et d'autre, la conscience est forte que des avancées politiques majeures sont en train de se préparer. Les chercheurs et les organisations dont ils font partie anticipent la possibilité d'une «obligation de dépôt» (*mandate*) pour les publications en Europe, ce que les éditeurs ne verraient pas d'un bon œil. Par ailleurs, les responsables gouvernementaux tant nationaux qu'euro-péens comprennent de mieux en mieux les enjeux des archives ouvertes de publications, mais cherchent malgré tout à préserver un équilibre pour garantir une communication scientifique performante, susceptible entre autres de fournir les moyens d'évaluation de la qualité des recherches qu'ils financent.

La Commission européenne, inspirée par les travaux du groupe d'experts sur les bibliothèques numériques, a donc réuni autour d'une table les représentants d'institutions scientifiques et les éditeurs, et a tenté de définir des actions communes visant notamment à offrir une transition contrôlée vers différents modèles de diffusion libre de l'information scientifique (Commission européenne, 2005). Très tôt émerge l'idée d'un «Observatoire de l'édition scientifique» dont la mission serait de suivre les évolutions du domaine et d'identifier, à l'aide de différents indicateurs (production scientifique, usages, modèles économiques sous-jacents) les éventuels changements dans ses grands équilibres. Le projet PEER, financé dans le cadre du programme eContent+, est la première étape pour développer conjointement un tel observatoire, en se focalisant sur les archives.

PEER : une plate-forme de communication entre les éditeurs et la communauté scientifique

Le projet PEER repose sur cinq partenaires : éditeurs scientifiques² regroupés dans le STM (association internationale des éditeurs des domaines scientifique, technique et médical); organismes de recherche (Max-Planck Society, Inria³); universités (bibliothèque de l'Université de Göttingen); agences de recherche (European Science Foundation). Le projet collabore par ailleurs étroitement avec la fondation SURF⁴, l'Université de Bielefeld et l'initiative européenne Driver⁵.

Son objectif principal est d'observer les effets d'un dépôt systématique et à grande échelle des manuscrits-auteurs après évaluation par les pairs, à savoir la phase 2 du cycle de vie d'un article scientifique (phase 1 : manuscrit soumis ; phase 3 : version finale de l'éditeur). Les questionnements envisagés tournent autour des aspects suivants :

- impact du dépôt systématique sur la viabilité des journaux (impact sur les abonnements et/ou les téléchargements sur le site de l'éditeur) ;
- impact sur la diffusion de l'article scientifique (accroissement cumulé des accès et/ou des citations) ;
- impact sur l'écologie globale des productions de recherche (par exemple, transfert de soumission d'un journal vers un autre) ;
- détermination des facteurs favorisant le dépôt par les auteurs et évaluation des coûts correspondants ;
- conceptions de modèles permettant de faire coexister l'édition scientifique «traditionnelle» avec l'auto-archivage.

Le dispositif PEER est conçu comme un espace de communication entre scientifiques et éditeurs. À un premier niveau, il s'agit de mieux comprendre les perspectives propres à chaque groupe vis-à-vis de l'*Open Access* en général et de la place des archives en particulier. À un deuxième niveau, des réponses communes sont envisagées de façon à réduire les craintes ou les défiances de part et d'autre. Que ce soit en termes de contribution des éditeurs au dépôt systématique ou d'identification des critères de confiance pour les archives (gestion des périodes mobiles d'embargo, fournitures de statistiques d'accès), le but est de mieux articuler les archives aux processus éditoriaux traditionnels. Enfin, le projet PEER est l'occasion de définir des directives précises sur l'échange tant des métadonnées que des textes intégraux (le contexte technique actuel est en effet très fragmenté, avec notamment une absence de stratégie claire de normalisation).

Vues sous un angle plus critique, les perspectives d'études sont le résultat de compromis difficiles :

– Les éditeurs visaient une évaluation de l'impact des archives ouvertes sur l'économie de l'édition. Leur approche a été contrebalancée par une autre permettant, du point de vue de l'auteur, d'observer l'impact sur la diffusion des contenus de la recherche (le nombre de citations a-t-il crû ? les chercheurs proposent-ils plus ou moins leurs articles aux journaux en ligne ? quels sont les critères pour choisir une revue scientifique ?...).

– Le choix d'articles en phase 2 est le résultat d'un compromis. Pour les éditeurs, il n'était pas question que le dépôt concerne des articles en version finale ; la version manuscrite (phase 1) n'aurait pas offert de valeur scientifique ; finalement, l'article en phase 2 a fait consensus, puisqu'il contient l'essentiel du contenu scientifique (la relecture par les pairs a été effectuée) et que les différences entre phase 2 et phase 3 sont en général peu pertinentes, sauf dans quelques domaines

(où la précision des formulations est cruciales, comme le biomédical ou le droit).

Corpus et méthode

Les travaux du projet PEER reposent sur un corpus de plus de 300 journaux scientifiques fournis par les éditeurs participants, et répondant aux quatre critères suivants : tout d'abord, 20 % du contenu est européen (déterminé par l'auteur-correspondant) ; ensuite, la qualité des revues, reflétée notamment par le « facteur d'impact », est bonne (ce critère, non absolu, est considéré au cas par cas, en fonction des journaux proposés par les éditeurs) ; de plus, autant que possible, les domaines concernés sont relativement larges ; enfin, la période d'embargo, définie par les éditeurs, est courte (12 mois ou moins) pour permettre des observations fiables au cours du projet (pour la plupart des journaux, un rétro-dépôt d'articles préalable à la période d'observation est envisagé, de façon à disposer très rapidement d'un nombre suffisant d'éléments observables).

Ce corpus de revues est divisé en deux : dans l'un, les éditeurs déposent directement le manuscrit-auteur et les métadonnées (affiliations, volume et quantième, pagination, ISSN, DOI, etc.) ; dans l'autre, les auteurs, incités par un courrier *ad hoc*, effectuent eux-mêmes le dépôt (dans ce dernier cas, les métadonnées complètes de publication sont fournies par les éditeurs malgré tout, ce qui garantit la cohérence des contenus entre eux). Le dépôt proprement dit, qu'il s'agisse d'un dépôt par l'éditeur ou d'un dépôt par l'auteur, s'effectue par le biais d'un portail (le « PEER Depot ») qui gère la normalisation des données et redirige les informations sur les archives de publication qui participent au projet.

Les domaines de recherche

Les activités de recherche ne sont pas conduites directement par les membres du projet, mais elles sont sous-traitées, après appel d'offre, à des équipes indépendantes⁶. Les axes de recherche, initiés depuis avril 2009, portent sur les aspects suivants :

– *Les comportements*. L'équipe du département des Sciences de l'information et du LISU à l'Université de Loughborough (UK) est chargée d'identifier les comportements des auteurs et des usagers d'archives de publication.

– *Les usages*. Le groupe CIBER de l'UCL (University College London) est chargée de définir les origines et les conditions d'usage des articles déposés dans les archives et de définir des indicateurs facilitant le suivi de tels usages.

– *Les modèles économiques*. L'appel d'offre lancé en septembre 2009, porte sur la détermination de l'ensemble des coûts (éditeurs et archives) associés aux différents modes de dépôt et d'accès envisagés, et la mise en perspective de ces coûts avec les coûts actuels de publication d'articles scientifiques.

La plate-forme technologique

Le projet PEER ne propose pas d'archivage propre ; les articles qui y sont déposés aboutissent dans des dépôts d'archives ou *trusted repositories* qui acceptent de recevoir les publications et se plient à des règles techniques précises (modèles de données, interface *Sword* de dépôt) et éditoriales (fourniture de statistiques communes normalisées, gestion fine des affiliations) garantissant un traitement uniforme du dépôt des manuscrits-auteurs. Ce réseau réduit d'archives est choisi en fonction de la

capacité à assurer un service fiable au-delà du projet, ainsi qu'une visibilité internationale.

De même, les éditeurs acceptent de se conformer à un certain nombre de contraintes liées notamment à la fourniture d'un jeu minimal de métadonnées et permettant le suivi précis des informations liées à chaque article. Par contre, il a été décidé d'accepter tous les formats (XML) des différents éditeurs, afin d'avoir une idée de la complexité des opérations de normalisation à réaliser.

Le «PEER Depot» joue un rôle particulier en ce qu'il reçoit toutes les informations primaires, en vérifie les contenus puis transfère une version normalisée vers les différentes archives. Il permet en fait aux éditeurs et aux archives d'avoir un interlocuteur technique unique qui assure la cohérence et la complétude des opérations de dépôt. L'une des activités importantes du «PEER Depot» a ainsi été de définir un format conforme aux directives de la TEI (*Text Encoding Initiative*, <www.tei-c.org>) permettant de reprendre le contenu intégral des articles ainsi que les métadonnées correspondantes (cf. Bijsterbosch *et al.*, 2009, pour les spécifications techniques). L'un des enjeux principaux est de définir un format non propriétaire qui garantisse l'interopérabilité entre archives (et les outils associés de recherche et de présentation), ainsi qu'un archivage pérenne des contenus.

Bilan à mi-parcours

Un premier bilan à mi-parcours montre que le projet PEER a conduit à une amélioration du dialogue entre éditeurs et chercheurs sur les attentes réciproques en matière d'archives. De nouveaux éditeurs ont ainsi souhaité rejoindre le projet en cours de route. Surtout, le climat de confiance, établi autour de la mise en place de la plate-forme de dépôt et des actions de recherche,

a permis de valider les processus de transfert d'information, qui n'étaient pas nécessairement acquis au moment du lancement du projet.

Sur le plan technologique, la normalisation des formats (cf. Holmes & Romary, 2009) nécessitera probablement des développements qui vont au delà du projet PEER lui-même. Il apparaît en effet nécessaire de définir un vrai cadre de standardisation pour les articles scientifiques, alors que cet aspect est peu considéré par les éditeurs privés. Des formats spécifiques, du type DTD de la *National Library of Medicine*, sans véritable objectif de qualité technique, de maintenance ou de documentation (ce qui caractérise un standard), et utilisés de façon très variable dans les milieux de l'édition, nuisent à la diffusion des publications scienti-

fiques. On a ainsi pu observer une grande hétérogénéité des usages des textes des articles en XML, empêchant une exploitation unifiée des contenus.

Enfin, les partenaires académiques du projet ne sont pas naïfs et voient que le milieu de l'édition cherche à maintenir un marché dont le rendement, au regard d'autres domaines d'activités du secteur de l'information, est élevé. Il reste que, dans un contexte de forte pression politique au niveau européen, avec la perspective d'une obligation intégrale de dépôt pour les publications associées aux projets du Huitième Programme-cadre, il est important de mieux connaître les modèles de collaboration possibles entre éditeurs et chercheurs, ainsi que les schémas viables de déploiement d'archives qui peuvent y être associés.

NOTES

1. Cf. Romary & Armbruster (2010). Chaque institution (université, organisme de recherche) a cherché à mettre en œuvre sa propre archive de publication, sans toujours anticiper sur les difficultés à maintenir de telles infrastructures de taille sous-critique. La France représente à cet égard une exception avec le consensus qui s'est créé autour d'une archive unifiée, HAL (<<http://hal.archives-ouvertes.fr>>).
2. Les éditeurs suivants sont directement impliqués dans le PEER: BMJ Publishing Group, Cambridge University Press, EDP Sciences, Elsevier, IOP Publishing, Nature Publishing

Group, Oxford University Press, Portland Press, Sage Publications, Springer, Taylor & Francis Group, Wiley-Blackwell.

3. <<http://www.inria.fr>>. Voir aussi HAL-Inria: <<http://hal.inria.fr>>.
4. <<http://www.surffoundation.nl/en/Pages/default.aspx>>.
5. <<http://www.driver-repository.eu>>.
6. Une partie non négligeable du management lié au projet consiste, comme on peut s'en douter, à gérer le plus finement possible les éventuels conflits d'intérêt.

RÉFÉRENCES BIBLIOGRAPHIQUES

BIJSTERBOSCH, M., BRÉTEL, F., BULATOVIC, N., PETERS, D., VANDERFEESTEN, M., WALLACE, J., PEER. D3.1. *Guidelines for Publishers and Repository Managers on Deposit, Assisted Deposit and Self-Archiving*, 2009. En ligne sur <http://www.peerproject.eu/fileadmin/media/reports/D3_1_Guidelines_v8.3_20090528.Final.pdf>.

COMMISSION EUROPÉENNE, *i2010: l'initiative "Bibliothèques numériques"*, 2005. En ligne sur <http://ec.europa.eu/information_society/activities/digital_libraries/index_fr.htm>.

DEWATRIPONT, M., GINSBURGH, V., LEGROS, P., *et alii*, *Study on the Economic and Technical Evolution of the Scientific Publication*

Laurent Romary

Markets in Europe, Bruxelles, Commission Européenne, 2006.
En ligne sur <http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf>.

HOLMES, M., ROMARY, L., «Encoding models for scholarly literature», in KAPIDAKIS, S. (dir.), *Publishing and Digital Libraries: Legal*

and Organizational Issues, 2009. En ligne sur <<http://hal.archives-ouvertes.fr/hal-00390966/fr>>.

ROMARY, L., ARMBRUSTER, C., «Beyond institutional repositories», à paraître dans *International Journal of Digital Library Systems*, 2010.
En ligne sur <<http://hal.archives-ouvertes.fr/hal-00399881/fr>>.