

Projet « Evaluation en ligne des archives ouvertes »

Jean-Pierre Demailly
Professeur à l'Université de Grenoble I (Institut Fourier)
Membre de l'Institut Universitaire de France
Rédacteur en Chef des Annales de l'Institut Fourier

Le but de ce document est de décrire un projet visant à mettre en place des comités éditoriaux supervisant l'évaluation en ligne des pré-publications scientifiques déposées sur les archives ouvertes. Une fois évaluées, corrigées ou amendées suivant les critères habituels des comités de rédaction des grandes revues scientifiques, les pré-publications électroniques prendraient alors le statut de véritables publications, certifiées par une caution scientifique.

Le projet est actuellement en cours d'élaboration, en collaboration avec le CCSD (Centre pour la Communication Scientifique directe, situé à Villeurbanne, soutenu par l'IN2P3), et les responsables de l'archive ouverte arXiv, dont la logistique est localisée à l'université Cornell aux Etats-Unis. Le démarrage effectif pourrait intervenir au printemps 2004.

Au terme de quelques années, si l'évaluation en ligne atteint une masse critique, cet outil pourrait permettre aux chercheurs de disposer d'un accès beaucoup plus rapide à l'information scientifique, avec des coûts de gestion très réduits. Il serait également beaucoup plus facile de développer des moteurs de recherche spécialisés pour indexer les publications et effectuer des recherches exhaustives plein texte.

Rationale

Les publications scientifiques et l'état de la technologie se situent aujourd'hui à une croisée de chemins. La majeure partie des publications scientifiques, des grandes revues internationales, sont encore aujourd'hui éditées sous la forme traditionnelle de volumes papier, diffusés par courrier ordinaire. L'acquisition de ces publications se fait essentiellement grâce à des abonnements annuels ou pluriannuels -- souvent par l'intermédiaire de consortiums d'achat, par exemple du fait de la législation sur les marchés publics en Europe.

Même lorsqu'une édition électronique existe, l'accès en est limité par le caractère propriétaire des documents, et le copyright restrictif qui s'y applique.

Bien qu'une proportion non négligeable des grandes revues soit éditée par des acteurs issus du monde académique (universités, grands instituts de recherche), une part de plus en plus importante de l'édition scientifique est aujourd'hui capturée par de grands éditeurs privés, via les regroupements d'entreprise et les phénomènes de concentration économique.

Tout ceci a pour conséquences de renforcer la situation de monopole des grands éditeurs, et d'entraîner un renchérissement notable du coût d'acquisition des revues par les bibliothèques universitaires.

Pourtant, depuis une vingtaine d'années environ, la plus grande partie du travail technique aboutissant à la rédaction d'un article de recherche est effectuée directement par le chercheur ou par son laboratoire de recherche. La plupart du temps, le chercheur utilise un traitement de texte et des logiciels de traitement de données spécialisés, et met en page lui-même son

Projet « Evaluation en ligne des archives ouvertes »

manuscrit dans un format électronique standard. Dans le même temps, des archives électroniques ouvertes de prépublications se sont constituées, permettant un accès centralisé à un grand nombre de publications.

Aujourd'hui, le seul rôle dévolu aux revues traditionnelles n'est plus que celui de l'évaluation et la validation des travaux par les pairs. Ceci se fait grâce à l'intervention des Comités de Rédaction des revues.

Le processus est en général le suivant : pour chaque manuscrit soumis à une revue, le comité de rédaction fait appel à un ou plusieurs experts extérieurs qui statuent sur l'intérêt et la validité de l'article, et suggèrent le cas échéant des corrections ou des améliorations.

Le Comité décide alors de publier ou non l'article, en fonction de la notoriété de la revue et du niveau supposé de l'article. En cas d'acceptation, le comité fait procéder aux corrections nécessaires. Dès lors, l'auteur cède le droit de diffusion de l'article à la revue. Le manuscrit, qui était le plus souvent librement disponible avant sa publication définitive, perd alors son caractère de publication ouverte.

Ce schéma bien établi (compte tenu des technologies qui avaient prévalu jusqu'à il y a quelques années) comporte à l'évidence de graves défauts :

- L'inconvénient le plus durement ressenti par la communauté scientifique est celui du coût d'équipement exorbitant des bibliothèques : avec l'explosion quasi-exponentielle du nombre des articles publiés, et le renchérissement des coûts d'édition et de publication, les budgets ne peuvent plus suivre. Ceci conduit régulièrement les bibliothèques à envisager des désabonnements massifs. Dans les pays en voie de développement, il n'existe pratiquement pas de centres universitaires capables de s'équiper, et ceci est un frein considérable pour les jeunes chercheurs de ces pays.
- Les chercheurs ne peuvent avoir accès aux volumes papier qu'avec un retard de plusieurs mois (sinon plusieurs années) par rapport aux versions électroniques diffusées par les archives ouvertes, lorsque celles-ci fonctionnent.
- La communauté scientifique, qui maîtrise toute la chaîne de l'élaboration des connaissances, en perd le contrôle au stade final de la publication.
- L'accès à l'information est plus difficile, les chercheurs ne peuvent pas aisément bénéficier des technologies modernes telles que les moteurs de recherche, sauf peut être via des moteurs de recherche propriétaires coûteux et partiels.
- Le chercheur qui souhaite publier se voit parfois refuser une soumission de bonne qualité, du fait d'une possible inadéquation du niveau de celle-ci avec celui de la revue à laquelle il s'est adressé. Le chercheur peut alors souffrir d'un retard de reconnaissance, surtout s'il s'agit d'un jeune chercheur.

Depuis de nombreuses années, il a été proposé de mettre en place des revues électroniques proposant un libre accès en ligne, ce qui permettrait de résoudre une grande partie des problèmes mentionnés ci-dessus. Pour cela il est indispensable que les chercheurs prennent conscience des enjeux et, notamment de l'importance des questions de propriété intellectuelle et de licences de publication. En particulier, les chercheurs devraient systématiquement se détourner des éditeurs exigeant des clauses de cessions de droits abusives ou trop restrictives.

Projet « Evaluation en ligne des archives ouvertes »

Evaluation en lignes des archives ouvertes

Je voudrais ici décrire les grandes lignes d'un projet qui va dans le même sens -- mais peut-être encore plus loin d'une certaine façon. Ce projet concerne la branche mathématique de l'archive ouverte arXiv, mais pourrait sans doute aussi s'adapter à d'autres archives. La partie technique sera prise en charge par le CCSD.

ArXiv a été fondée en 1991 par Paul Ginsparg avec le serveur hep-th (High Energy Physics). Elle est aujourd'hui dépositaire de très nombreux manuscrits dans différents domaines des Mathématiques et de la Physique. On peut estimer qu'arXiv a capté environ 5% de toutes les pré-publications de mathématiques. Dans certains secteurs comme la géométrie algébrique où l'archive a été fonctionnelle dès 1992, la proportion est beaucoup plus élevée; les chercheurs de ce domaine utilisent d'ores et déjà arXiv comme leur principale source d'information.

Dans ces conditions, puisque l'accès aux publications est centralisé, il apparaît souhaitable de mettre également en place un mécanisme d'évaluation en ligne des travaux, et un accès centralisé aux conclusions de cette évaluation. Voici les grandes lignes du fonctionnement que j'envisage. Ces points ont fait l'objet de discussions avec plusieurs personnes, notamment Greg Kuperberg, chairman de la branche mathématique d'arXiv, ainsi que Franck Laloë, Directeur du CCSD, et Daniel Charnay, Directeur Adjoint du CCSD.

1. Des comités de rédaction thématiques se mettent en place progressivement dans chacune des grands domaines des mathématiques recensés par la base de données arXiv. Ces comités seraient cependant complètement indépendants de l'archive sur le plan scientifique, et auraient pour mission et pour responsabilité de faire procéder à l'évaluation des travaux relatifs à leur

spécialité. Une taille réduite de 5 à 6 membres paraît appropriée, une règle souhaitable étant que les comités comportent des chercheurs de plusieurs pays, en provenance de zones géographiques variées. Bien entendu, se créent également aux interfaces des comités réunissant les compétences recouvrant les domaines en interaction.

2. Chaque comité de rédaction sollicite un ou plusieurs experts (referee(s)) chargés d'examiner les travaux dont il prend la responsabilité. Les experts se voient garantis leur anonymat s'ils le souhaitent. L'expert évalue l'article et rend ses conclusions suivant une grille standardisée :

A. L'article a pu être vérifié dans tous ses détails et peut être considéré comme valide en l'état des connaissances.

B. L'expertise laisse supposer que les résultats sont corrects

C. L'expertise fait apparaître que certains résultats sont douteux, incomplets ou insuffisants

D. L'article est partiellement ou totalement erroné

En outre, l'expert prépare une review complétant le résumé de l'auteur, en situant notamment l'article par rapport aux théories ou résultats déjà connus, ainsi que les perspectives ou retombées envisageables. Cette review pourrait avoir une forme voisine de ce que proposent actuellement les Maths Reviews ou le Zentralblatt. Après mise en forme éventuelle, ces

Projet « Evaluation en ligne des archives ouvertes »

informations sont mises en ligne sur le serveur du CCSD sous la responsabilité des comités éditoriaux, si possible grâce à une plate-forme technique unifiée permettant l'accès au contenu de l'article et l'accès à son évaluation. Le comité de rédaction s'engage sur les conclusions A,B,C,D ci-dessus, en faisant appel à plusieurs referees si nécessaire. Le comité éditorial s'engage également sur la partie "review" en essayant de situer les perspectives de l'article.

Dans tous les cas, et surtout dans celui où l'évaluation fait apparaître des problèmes, l'auteur se voit proposer un délai d'au moins un mois pour corriger les problèmes constatés ou pour interagir avec le(s) referee(s) via le comité de rédaction, avant la mise en ligne des conclusions.

3. Les comités éditoriaux peuvent être sollicités par les auteurs, mais peuvent également prendre l'initiative d'évaluer toute soumission qui relève de leur domaine de compétence. L'archive ouverte maintient un fichier des articles en cours d'évaluation, et des comités procédant à celle-ci. Ces informations sont disponibles à l'ensemble des chercheurs consultant l'archive, et donc en particulier aux membres des comités de rédaction.

4. Les auteurs peuvent éventuellement solliciter plusieurs comités s'ils le souhaitent, notamment en cas de contestation. Les bases de données d'arXiv et du CCSD maintiennent l'historique complet de toutes les versions successives des articles soumis, ainsi que de leurs évaluations successives par les comités de rédaction. Toute information peut-être révisée ou complétée à tout moment en fonction de l'évolution des connaissances.

5. Les auteurs peuvent continuer à soumettre leurs travaux aux revues traditionnelles, même lorsque ceux-ci ont déjà été évalués par l'un des comités de rédaction associés à arXiv. Cependant, le but à terme (d'ici plusieurs années) est de faire en sorte que les revues traditionnelles deviennent progressivement inutiles. L'objectif est en effet que la valeur des travaux présentés par les auteurs puisse être mesurée de façon fiable et précise à partir des conclusions et des reviews disponibles en ligne, ainsi que par les statistiques de consultation des travaux.

6. Une plate-forme technique commune au travail de l'ensemble des comités de rédaction est mise en place. Le CCSD prend en charge ce développement. Des listes de discussion privées sont mises automatiquement à disposition de chacun des comités. Le CCSD développe en liaison avec arXiv une interface standardisée pour déposer les conclusions et les reviews des comités dans les bases de données et leurs miroirs. Toutes ces données sont mises à la disposition de la communauté scientifique en libre accès.

7. Un bureau comprenant des représentants des mathématiciens des divers champs disciplinaires, des diverses zones géographiques ainsi que des représentants d'arXiv et du CCSD est créé. Sa mission est de superviser la mise en place, le renouvellement et la coordination du travail des comités de rédaction.

8. Dans toute la mesure du possible, les technologies utilisées pour l'archivage, la mise en ligne des données et les formats de documents utilisent des standards ouverts et des logiciels libres. Outre les aspects éthiques, le but est de garantir la pérennité du traitement des données sur de longues périodes. Les comités de rédaction prêtent attention à la qualité des documents électroniques soumis, et demandent aux auteurs de procéder aux améliorations nécessaires sur ces points.

Projet « Evaluation en ligne des archives ouvertes »

9. Le travail des comités de rédaction et des referees est considéré comme entrant dans les missions naturelles des enseignants-chercheurs, et ne nécessite donc aucun financement (il faut signaler que c'est la situation très majoritaire qui prévaut aujourd'hui pour les comités éditoriaux des revues de mathématiques). Les frais d'infrastructure des serveurs, de leur fonctionnement et de leur maintenance sont pris en charge par des subventions récurrentes des institutions hôtes, des associations scientifiques et des états. On peut imaginer que ces sommes seront très faibles par rapport aux gains réalisés en terme de coût d'équipement des bibliothèques universitaires, lorsqu'une proportion croissante de publications scientifiques sera en libre accès.