

INFORMATIONS

Documentation mathématique Enjeux et éléments de réponse

Pierre Bérard (*Cellule MathDoc et Institut Fourier, Grenoble*)

Enjeux

La documentation joue un rôle primordial dans tous les domaines scientifiques et tout particulièrement en mathématiques. Trois raisons au moins pour ce qui nous concerne.

- Une publication mathématique contient l'intégralité d'un résultat qui est ainsi immédiatement utilisable par la communauté scientifique.
- Un résultat mathématique jouit, a priori, d'un caractère pérenne et l'utilisation d'articles anciens est fréquente. Ainsi, les œuvres de H. Poincaré sont toujours d'actualité aujourd'hui.
- Il y a un caractère "imprévisible" dans les mathématiques. Certaines notions, certains résultats ont, parfois contre toute attente, un impact important sur les sciences (c'est la "unreasonable effectiveness of mathematics in the natural sciences" évoquée par le physicien E. Wigner), sur l'ingénierie ou sur les mathématiques elles-mêmes, des années après leur découverte.

La documentation est, pour un mathématicien, une source essentielle d'inspiration, de concepts et d'outils. C'est dans les exemples, parfois du passé (surfaces minimales, équations différentielles, *etc.*), qu'il forge son intuition. C'est dans le corpus des résultats acquis qu'il trouve ses outils. C'est souvent en faisant des rapprochements inattendus entre champs mathématiques a priori éloignés qu'il fait progresser sa science (ce qui explique, au moins en partie, l'importance des revues mathématiques généralistes, et elles sont majoritaires). La documentation constitue donc en mathématiques une ressource "stratégique" à laquelle la communauté consacre beaucoup de soins et une part importante de son budget de recherche. L'avènement de l'ère électronique n'a pas changé fondamentalement la relation des mathématiciens à la documentation, mais elle a renforcé l'importance de certaines ressources documentaires¹.

- Importance des bases de données pour retrouver les informations pertinentes dans une littérature cumulative, à croissance exponentielle (moins de 1 000 articles par an vers la fin des années 1800, plus de 70 000 aujourd'hui).
- Importance de la littérature grise (en particulier par l'intermédiaire de serveurs de prépublications, dont certains commencent à jouer le rôle d'archives numériques).

¹ La communauté mathématique n'est bien sûr pas la seule à se préoccuper de ces questions. Mentionnons, également celles des physiciens et des chercheurs en sciences de la vie qui sont très actives dans ce domaine.

- Importance de l'accès électronique aux revues spécialisées (versions en ligne et archives numérisées) et des liens croisés entre ressources numériques.
- Importance de l'archivage à très long terme (en particulier celui des documents numériques).

La concentration dans le monde de l'édition, la création de services à valeur ajoutée par les éditeurs commerciaux et la pression budgétaire sur les bibliothèques (augmentations des tarifs d'abonnement, sur-coûts pour les accès électroniques) le montrent, la documentation est devenue un *enjeu commercial* extrêmement important².

Tous les aspects de la documentation (ainsi que les modalités de sa diffusion) sont concernés, depuis les prépublications jusqu'aux revues, en passant par les bases de données, la numérisation des fonds anciens (et même les logiciels de gestion documentaire ou, aujourd'hui, de gestion électronique de documents). De grands groupes commerciaux, plus préoccupés par les gains des actionnaires que par la diffusion de la science, tentent aujourd'hui d'occuper l'ensemble du terrain : celui des revues et des services associés bien sûr mais aussi, et c'est nouveau, celui des prépublications, *etc.* En accaparant, par le biais de certains accords de consortium, les moyens budgétaires dont nous disposons, ils mettent également en péril les éditeurs de taille modeste et en particulier les éditeurs à caractère académique.

La documentation est pour nous, producteurs et utilisateurs de contenu, un enjeu scientifique et politique. Serons-nous demain libres d'accéder à la documentation dont nous avons besoin, à quel prix et dans quelles conditions ?

Concernant les revues, et au delà des questions techniques (contraintes budgétaires, marchés, *etc.*) et des stratégies des établissements ou des organismes de recherche, il me semble primordial de veiller à plusieurs points fondamentaux :

- Choix des titres selon des critères scientifiques (rapport qualité/prix : voir à ce sujet l'analyse faite par U. Rehmann à Bielefeld³ et l'article K. Frazier⁴), en adéquation aux priorités des laboratoires.
- Garanties sur la pérennité des documents et sur les conditions d'accès (à très long terme) aux collections (numériques ou imprimées).
- Soutien à l'édition académique, souvent de grande qualité, qui est un facteur de modération des coûts d'abonnements (voir l'analyse de U. Rehmann).
- Implication forte de la communauté scientifique (qui alimente les revues, qui fait partie des comités éditoriaux, qui assure le contrôle scientifique des publications et qui, parfois, se bat très vivement contre les hausses excessives des tarifs⁵) dans la gestion des moyens documentaires, à tous les échelons.

Le réseau national des bibliothèques de mathématiques (RNBM) et la Cellule MathDoc mènent des actions concertées dans cette direction. Ces actions ont permis, à ce jour, la conclusion d'accords pour faciliter l'accès aux bases de données Zentralblatt MATH et Mathematical Reviews, aux versions électroniques

² On pourra à titre d'exemple, lire l'article :

www.nature.com/nature/debates/e-access/Articles/butler.html

³ www.mathematik.uni-bielefeld.de/~rehmann/BIB/

⁴ www.dlib.org/dlib/march01/frazier/03frazier.html

⁵ Certains comités éditoriaux de revues qui pratiquaient des tarifs excessifs n'ont pas hésité à démissionner en bloc, puis à créer une revue concurrente.

de revues de mathématiques du service LINK, la mise en place d'un service de sommaires portant sur un ensemble de près de 900 revues, en mathématiques et domaines connexes. Deux autres actions sont actuellement menées par la Cellule MathDoc, en collaboration avec les laboratoires et les bibliothèques de mathématiques. Ces actions participent d'un effort international visant à donner à la communauté scientifique les moyens de jouer pleinement son rôle dans la gestion des ressources documentaires.

Les programmes de numérisation rétro-active

Différents programmes sont consacrés à la numérisation systématique de fonds anciens dans le monde. Le programme JSTOR⁶ est pluri-disciplinaire et il permet en particulier d'accéder aux fonds anciens de la vingtaine de revues mathématiques américaines. La société américaine de physique a numérisé l'intégralité du fonds des Physical Reviews (1893 – 1997)⁷. Les bases de données Mathematical Reviews et Zentralblatt MATH ont également mené de tels programmes et la numérisation du Jahrbuch über die Fortschritte der Mathematik est en cours⁸.

Il existe d'autres programmes de numérisation, comme ceux menés à l'université de Göttingen, à l'Université Cornell ou à la Bibliothèque nationale de France⁹. Ces programmes peuvent répondre à plusieurs objectifs : souci patrimonial, mise à disposition des fonds anciens auprès d'une communauté de lecteurs élargie, création d'un ensemble scientifique cohérent par insertion de liens croisés entre les fonds anciens numérisés et les fonds nativement numériques ou entre fonds numérisés et bases de données.

Parce qu'ils valorisent l'ensemble d'un corpus documentaire, les programmes de numérisation représentent des enjeux scientifiques importants, mais aussi des enjeux financiers pour les éditeurs commerciaux. S'il est difficile aujourd'hui de concevoir l'avenir d'une revue sans version électronique, il sera difficile demain de concevoir celui d'une revue dont le fonds ancien n'aura pas été numérisé. C'est pour que les revues françaises puissent répondre à cette exigence de demain, que le programme de NUMérisation de Documents Anciens Mathématiques (NUMDAM)¹⁰ a été lancé. Soutenu par la Direction de la Recherche et par le Centre National de la Recherche Scientifique, il est piloté par la Cellule MathDoc.

Une première phase, significative et financée par le CNRS, permettra de numériser de l'ordre de 220 000 pages (environ 8 000 articles). Cinq revues généralistes et une série d'actes de colloques sont concernées par cette phase¹¹.

Le travail préliminaire à la première phase du programme NUMDAM (dépouillement des collections en vue de la numérisation, création et alimentation de la

⁶ www.jstor.org

⁷ prola.aps.org

⁸ www-irma.u-strasbg.fr/EMIS/projects/JFM/

⁹ gallica.bnf.fr

¹⁰ www-mathdoc.ujf-grenoble.fr/NUMDAM/

¹¹ Annales de l'Institut Fourier, Bulletin et Mémoires de la Société Mathématique de France, Publications Mathématiques de l'Institut des Hautes Études Scientifiques, Journées Équations aux Dérivées Partielles, ainsi qu'une autre revue importante pour laquelle des négociations sont en cours.

base de données des articles, rédaction du cahier des charges techniques) s'est déroulé de septembre 2000 à mai 2001. Un appel d'offre pour les opérations de numérisation a été émis par le CNRS fin juin 2001. Le choix de l'opérateur et la notification du marché devaient intervenir avant la fin de l'année 2001. La première phase du programme NUMDAM devrait s'achever au printemps 2003. Les principes généraux suivants ont été retenus.

- Le programme NUMDAM est conçu pour s'insérer dans un effort global de la communauté scientifique internationale et il collaborera, autant que faire se peut, avec les institutions qui développent des programmes similaires (en mathématiques ou plus largement).
- L'interface donnant accès aux documents numérisés sera orientée "utilisateurs" (moteur de recherche, qualité de l'affichage et de l'impression).
- La base de données NUMDAM (notices bibliographiques des articles, avec résumé quand il existe et plein texte caché) sera librement accessible sur la Toile. Des liens croisés seront introduits, entre les articles des fonds numérisés et les notices des bases de données mathématiques : Jahrbuch über die Fortschritte der Mathematik, Mathematical Reviews et Zentralblatt MATH.
- Les textes des articles (mode image) seront aussi librement accessibles que possible. Leur intégration dans le flux des articles nativement numériques est prévue. Ces deux points prendront en compte les contraintes économiques des revues, par exemple par l'instauration d'un "créneau mobile".

Les phases ultérieures du programme NUMDAM devraient être rapidement programmées. Elles permettront de traiter d'autres revues mathématiques publiées en France, ainsi que des documents importants (séminaires, cours, *etc*). Dresser la liste des documents à numériser, localiser et réunir les originaux en vue de leur traitement demandera un travail important auquel seront associés mathématiciens, historiens des sciences et bibliothécaires.

N'hésitez pas à adresser à la Cellule MathDoc vos suggestions concernant les documents à numériser.

Serveurs de prépublications et de thèses

Les prépublications, les thèses et les habilitations jouent, on le sait, un rôle très important. Elles permettent la communication des résultats de la recherche bien avant leur publication dans une revue spécialisée. Ces documents sont de plus en plus souvent mis en accès libre sur la Toile : pages personnelles, serveurs des laboratoires, ou archives de prépublications telles que arXiv¹².

En mettant en place des *index nationaux de littérature grise* en 1997, la Cellule MathDoc se donnait pour objectif d'améliorer la visibilité des prépublications, des thèses et des habilitations de mathématiques publiées en France, ainsi que celle des laboratoires qui les mettent en ligne et qui assurent ainsi la première validation scientifique¹³. Cet objectif a été en partie rempli avec la montée en puissance des index et la collaboration avec l'index international Math-Net.preprints/MPRESS.

Ces index reposent sur le ramassage, dans différents sites en France, de fichiers

¹² www.arxiv.org

¹³ www-mathdoc.ujf-grenoble.fr/prepub.html

contenant des méta données – auteur(s), titre, date, mots clés, codes de classification mathématique – pour permettre une indexation efficace, les textes eux-mêmes restant sur les serveurs locaux.

Ce système de collecte d'informations réparties sur des sites variés présente des avantages (mise en place relativement aisée, plus grande responsabilité laissée aux laboratoires) mais également des inconvénients importants (les méta données ne sont pas toujours conformes, les pages personnelles sont parfois volatiles, le système est difficile à modifier parce qu'il repose sur un réseau d'interlocuteurs ...). Une évolution est aujourd'hui indispensable.

En vue d'améliorer le système des index nationaux, la Cellule MathDoc collabore actuellement avec le *Centre pour la Communication Scientifique Directe* (CCSD, UPS 2275 du CNRS)¹⁴.

L'objectif principal est de participer à une internationalisation de l'archive **arXiv**, en particulier par la création de miroirs *actifs* capables de recevoir des soumissions d'articles et de diffuser les textes (les miroirs actuels sont tous passifs : les soumissions ne peuvent se faire que sur le serveur maître, les miroirs n'assurent que la diffusion).

Une première réalisation dans cette direction sera bientôt finalisée avec la mise en service d'un *serveur pour les thèses*, comportant une section *thèses de mathématiques*.

La mise en place d'un miroir actif de l'archive **arXiv** pour les prépublications prendra sans doute quelques mois. Le cahier des charges est en cours de discussion. Il est souhaitable que le système permette le dépôt d'une prépublication par un individu, ainsi que le dépôt systématique par un laboratoire (qui sera alors dûment identifié).

Des informations sur la mise en place de ces nouveaux dispositifs sont disponibles à l'adresse :

www-mathdoc.ujf-grenoble.fr/CCSD-CMD/

Les laboratoires et les écoles doctorales peuvent dès à présent prendre contact avec la Cellule MathDoc pour organiser leur participation à ces serveurs de thèses et prépublications.

¹⁴ ccsd.cnrs.fr