

## Numérisation/conservation partagée : la collaboration BnF/Cellule MathDoc ; pistes de travail pour de nouveaux services.

Table ronde, Journées des pôles associés, le 1<sup>er</sup> Juillet 2004, BnF, Paris

Elizabeth CHERHAL, Cellule MathDoc, UMS5638, Université Joseph Fourier, Grenoble  
(Pôle associé pour la numérisation concertée en mathématiques)

### Résumé

Nous présentons les actions de la Cellule MathDoc en matière de numérisation partagée notamment la collaboration avec la BnF. La Cellule MathDoc est présentée comme producteur de documents numérisés et comme développeur de services agrégatifs fournissant un accès unique à plusieurs corpus de documents numériques (exemples : les programmes LINUM et mini-DML).

### Présentation de la Cellule MathDoc

La Cellule MathDoc, unité mixte de service CNRS/Université Joseph Fourier, est une équipe formée de mathématiciens, informaticiens et documentalistes. Elle a pour mission de rendre accessible la documentation mathématique recherche. Ses principaux services au public sont : le portail documentaire mathématique,<sup>1</sup> où l'on trouve de nombreux services tels que les catalogues fusionnés des bibliothèques de mathématiques, les index des thèses et prépublications de mathématiques, etc..., le serveur de documents numérisés NUMDAM<sup>2</sup> (Numérisation de Documents Anciens Mathématiques) et très prochainement un nouveau serveur pour les revues académiques de mathématiques.

### Bilan des actions concernant la documentation numérisée.

Pour faire suite à la communication de Laurent Guillopé dans ces mêmes journées en 2003, il convient de donner quelques informations rapides sur la poursuite des actions qu'il y a évoquées :

Le programme NUMDAM est arrivé au terme de sa première phase. En un an deux nouvelles collections très importantes ont été ajoutées : les *Publications Mathématiques de l'IHÉS*<sup>3</sup> (1959→) et les *Annales scientifiques de l'École normale supérieure* (1864→). NUMDAM entre dans sa deuxième phase avec la mise en ligne prochaine des *Annales Blaise Pascal*, et une partie des *Annales de l'Institut Henri Poincaré*. L'interface de consultation a été améliorée, en particulier grâce à l'ajout de liens croisés. Enfin, NUMDAM dispose maintenant d'un serveur OAI<sup>4</sup>.

Le « *répertoire bibliographique des sciences mathématiques* » est maintenant numérisé dans son intégralité en mode image par la BnF. Cette numérisation vient compléter la numérisation en mode texte (avec production de métadonnées et insertion de liens) qui se poursuit en collaboration avec le laboratoire de philosophie et d'histoire des sciences<sup>5</sup>.

Le « *journal des mathématiques pures et appliquées* » a été numérisé jusqu'en l'an 1940 par la BnF, et la production des métadonnées des articles postérieurs à 1881 suivra prochainement.

### Producteur et agrégateur

Le travail de la Cellule MathDoc en matière de documents numériques est à la fois celui de producteur et d'agrégateur. Production pour numériser et mettre en ligne les documents, agrégation pour rendre accessibles, à travers un guichet unique, des documents numérisés en provenance de plusieurs sources.

#### La production :

Le service NUMDAM propose au public un serveur web comportant le feuilletage et la recherche. Les métadonnées (ou notices) accompagnant les documents eux-mêmes contiennent, en plus des données bibliographiques habituelles, des résumés et des listes de références. Cela permet

<sup>1</sup> <http://math-doc.ujf-grenoble.fr>

<sup>2</sup> <http://www.numdam.org>

<sup>3</sup> Institut de hautes études supérieures

<sup>4</sup> Open Archives Initiative <http://www.openarchives.org/>

<sup>5</sup> Unité mixte de recherche Université Nancy 2 et CNRS

l'inclusion de nombreux liens croisés : du document vers les documents qui le citent, d'une citation vers l'original, etc...

Le serveur NUMDAM, aussi intéressant qu'il soit, ne forme qu'une petite pierre dans l'édifice de la numérisation de la littérature mathématique mondiale. C'est pourquoi les métadonnées sont fournies également à travers un serveur OAI<sup>6</sup>, afin de faciliter le partage et le développement de services plus importants.

### L'agrégation :

Par agrégation nous entendons le fait de regrouper en un seul endroit les métadonnées concernant plusieurs sources de documents, et faciliter l'accès à ces documents à travers une interface unique.

- Le programme LINUM<sup>7</sup> (livres numérisés de mathématiques) est né en 2002. Il a pour ambition de proposer un guichet unique pour **tous** les livres numérisés de mathématiques consultables sur la Toile. LINUM permet actuellement l'accès à des livres numérisés disponibles sur une dizaine de sites : les principaux sont à Cornell, U.Michigan, Göttingen, la Biblioteka Wirtualna Matematyki à Varsovie et enfin la BnF. Les métadonnées sont parfois fournies gracieusement par les producteurs de ces services, et parfois catalogués par nos soins. Il est certain que l'implantation de serveurs OAI devra faciliter grandement cet échange, qui jusqu'à présent se fait sur la base du volontariat (envoi de fichiers par exemple).
- Le projet « mini-DML » se situe dans le contexte du projet mondial « WDML<sup>8</sup> » (World Digital Math Library) dont le but ambitieux est d'offrir une version numérique de toute la littérature mathématique. La mini-DML est un prototype de guichet d'accès à cette littérature numérique.

Le principe de la mini-DML est de collecter des métadonnées de base (auteur, titre, date de publication, références, identifiant unique, URL) pour tout document mathématique numérique en ligne (article de journal, congrès ou séminaire, prépublication, livre, etc...), et les rendre accessibles à travers une interface simple. Le prototype mini-DML contient des métadonnées en provenance des collections suivantes : NUMDAM, BnF, Euclid<sup>9</sup>, ArXiv<sup>10</sup>, ces deux derniers services comportant un serveur OAI.

La mini-DML permet de trouver par exemple des documents écrits par un auteur au siècle dernier, numérisés d'un côté par la BnF et de l'autre par NUMDAM, ou des articles d'un auteur moderne se trouvant dans différents journaux actuels, ainsi que dans ArXiv. L'exemple du journal « *Annals of Mathematics* », dont les articles peuvent se trouver sur quatre sites différents, montre à quel point une initiative agrégative est utile pour le chercheur d'aujourd'hui.

### En conclusion (numérisation et partage)

En lançant le programme NUMDAM, nous avons résolu ne de pas numériser un document se trouvant déjà numérisé par un autre service dont l'accès est libre. La préparation de documents numériques et leur installation sur des serveurs est chose suffisamment coûteuse pour ne pas s'y attaquer plusieurs fois. Nous sommes donc de fervents partisans du partage en matière de numérisation. Rendre disponible les documents numérisés est déjà un immense pas en avant, mais rendre disponible les métadonnées accompagnant ces documents est tout aussi important. C'est pourquoi nous avons tenu à installer un serveur OAI et nous nous réjouissons de la prochaine installation d'OAI à la BnF.

---

<sup>6</sup> <http://www.numdam.org/oai>

<sup>7</sup> <http://math-doc.ujf-grenoble.fr/LiNuM/>

<sup>8</sup> <http://www.ceic.math.ca/WDML/dml/index.shtml>

<sup>9</sup> <http://projecteuclid.org/>

<sup>10</sup> <http://front.math.ucdavis.edu/>