



Université Joseph Fourier — C.N.R.S.

Cellule MathDoc - UMS 5638

BP 74 – 38402 Saint-Martin d'Hères cedex

www-mathdoc.ujf-grenoble.fr

RAPPORT d'ACTIVITÉ

Cellule MathDoc – UMS 5638

2002 – 2005

Cellule de Coordination Documentaire Nationale pour les Mathématiques
UJF – CNRS

Cellule MathDoc – UMS 5638
Université Joseph Fourier
B.P. 74
38402 Saint-Martin d'Hères cedex
(France)

Fax : 04 76 63 56 11

accueil@mathdoc.ujf-grenoble.fr

www-mathdoc.ujf-grenoble.fr/

SOMMAIRE

INTRODUCTION	5
RESUME	7
A. HISTORIQUE	7
B. PERSONNEL	7
C. BILAN ET PERSPECTIVES	8
1. <i>Documentation</i>	8
2. <i>Coopération Zentralblatt</i>	8
3. <i>Développement de logiciels</i>	9
4. <i>Numérisation</i>	9
5. <i>Édition</i>	10
LES ACTIONS DE LA CELLULE MATHDOC	10
A. LA DOCUMENTATION	10
1. <i>Le Catalogue Fusionné des Périodiques</i>	10
2. <i>Catalogue Fusionné des Ouvrages</i>	11
3. <i>Le service Simple de sommaires</i>	12
4. <i>Index nationaux des prépublications et thèses de mathématiques</i>	12
5. <i>Gallica-Math</i>	12
6. <i>LiNuM et mini-DML</i>	13
7. <i>Archives Bourbaki</i>	14
8. <i>Serveurs web d'information de la Cellule MathDoc</i>	14
9. <i>La lettre d'information de la Cellule MathDoc</i>	14
10. <i>Édition de séminaires et conférences</i>	15
B. LA COOPÉRATION AVEC LE ZENTRALBLATT MATH	15
1. <i>Développement logiciel</i>	15
2. <i>Fourniture de données bibliographiques de revues françaises</i>	16
3. <i>Gestion des abonnés et serveurs français, promotion</i>	16
4. <i>Contrats européens</i>	16
C. DÉVELOPPEMENT DE LOGICIELS	17
1. <i>European Database Manager for Mathematics (EDBM)</i>	17
2. <i>Outils de Gestion pour le programme NUMDAM</i>	17
3. <i>Logiciel de Gestion Documentaire pour les bibliothèques de Mathématiques</i>	19
4. <i>RUCHE - Outil de gestion rédactionnelle</i>	21
D. NUMÉRISATION	22
1. <i>Introduction</i>	22
2. <i>Présentation générale du programme NUMDAM</i>	23
3. <i>Bilan des réalisations (années 2001-2005)</i>	27
4. <i>Le processus de mise en ligne des revues</i>	29
5. <i>Développements spécifiques</i>	33
6. <i>Perspectives immédiates et futures</i>	34
E. ÉDITION : LE PROJET CEDRAM	35
1. <i>Enjeux : évolution de l'édition scientifique</i>	35
2. <i>Le projet : créer un pôle pour les revues de mathématiques</i>	37
3. <i>Avancement du projet</i>	38
4. <i>Futur du projet CEDRAM</i>	43
ANNEXES	45
A. MISSIONS	45
B. PUBLICATIONS	49
C. BUDGET	51
GLOSSAIRE PAR ORDRE ALPHABETIQUE	60

Introduction

Ce rapport couvre une période de quatre ans 2002-2005. Entre-temps un comité de pilotage a fait le point en septembre 2003. Au cours de ces quatre années, la Cellule MathDoc a beaucoup évolué tant au niveau de ses missions que de sa composition.

L'unité a été créée en 1995 par Pierre Bérard et Laurent Guillopé. Appelés à d'autres responsabilités, ils ont quitté la direction en 2002 et 2004 respectivement. Tandis que Thierry Bouche assurait la responsabilité des deux grands projets que sont la numérisation et le pôle d'édition, j'ai pris la succession de Pierre Bérard à la direction. Au printemps 2004, l'unité a déménagé pour intégrer de nouveaux bureaux dans la bibliothèque rénovée de l'institut Fourier.

Les deux missions initiales de la Cellule MathDoc étaient l'assistance aux bibliothèques de mathématiques pour ce qui concerne la documentation électronique et la coopération avec le Zentralblatt Math. En 2000, elle s'est vue confier la tâche de numériser les revues de mathématiques françaises et en 2005 celle de mettre en place un pôle d'édition et de diffusion pour ces revues. Pour accomplir ces nouvelles missions, son effectif permanent a été porté de 3 à 6 personnes (7 pendant 18 mois) et elle a employé en permanence 1 ou 2 contractuels.

Le projet de numérisation a pris une grande importance, mobilisant une grande part des capacités de l'unité. À l'issue de la deuxième phase, fin 2006, les principales revues françaises de mathématiques auront été numérisées à quelques exceptions près. Le travail devra se poursuivre dans plusieurs directions. D'une part nous sommes sollicités pour des travaux de numérisation assez divers comme les archives Bourbaki, les œuvres complètes de Schwarz et de Lichnerowicz, des revues ne contenant qu'une part de mathématiques comme le Bulletin Astronomique, des revues européennes. D'autre part, beaucoup de possibilités sur les modes d'accès à des ouvrages numérisés sont à explorer et à mettre en place, nous pourrions ainsi en faire un mode de consultation très différent de celui des ouvrages originaux. Enfin, les exigences de qualité définies dès le début de l'opération en font maintenant un modèle de référence international et nous amènent à avoir un rôle de conseil pour d'autres projets.

De même que le projet de numérisation a répondu à un besoin de mise en valeur des revues existantes, la transformation très rapide des modes d'édition nous a conduit à créer un pôle d'édition et de diffusion pour les revues de mathématiques. Son but est d'aider ces revues à passer le cap difficile qui verra l'édition papier peu à peu remplacée par l'édition électronique. Un certain nombre de revues ont répondu à cette offre, il reste à convaincre les autres de la nécessité d'évoluer assez rapidement.

Je voudrais terminer en remerciant toute l'équipe de MathDoc pour le travail accompli.

Yves Laurent, Directeur de la Cellule MathDoc
Décembre 2005

Résumé

A. HISTORIQUE

1995 : création de la Cellule MathDoc, Unité Mixte de Service 5638 sous la co-tutelle du Centre National de la Recherche Scientifique et de l'Université Joseph Fourier.

- Directeur : Pierre Bérard, directeur adjoint : Laurent Guillopé
- Personnel permanent : 2 ingénieurs UJF, 1 secrétaire CNRS

La Cellule MathDoc est créée avec pour missions :

- Apporter un soutien technique aux bibliothèques de mathématiques et aux laboratoires pour tout ce qui concerne la documentation électronique.
- Conduire la coopération Franco-Allemande sur Zentralblatt avec pour perspective de le transformer en un outil européen performant.

1999 : renouvellement pour quatre ans de l'unité. Direction inchangée.

2000 : début de l'opération de Numérisation des Documents Anciens (NUMDAM¹). Thierry Bouche rejoint l'équipe.

2001-2002 : recrutement de deux ingénieurs CNRS. Phase préparatoire de NUMDAM.

2003 : renouvellement pour quatre ans de l'unité. Nouveau Directeur : Yves Laurent.

2003-2004 : phase active de NUMDAM I. Préparation de NUMDAM II.

2005 : recrutement d'un ingénieur CNRS et remplacement de la gestionnaire. Début du programme CEDRAM (Centre d'Édition et Diffusion pour les Revues Académiques de Mathématiques). Phase active de NUMDAM II.

B. PERSONNEL

Personnel permanent :

- Claude Goutorbe, ingénieur (IR) Université Joseph Fourier, depuis octobre 1995.
- Elizabeth Cherhal, ingénieur (IE) Université Joseph Fourier, depuis novembre 1995.
- Monique Marchand, assistante (AI) CNRS, depuis juin 1996 (départ juin 2006).
- Catherine Barbe-Zoppis, ingénieur (IR) CNRS, depuis novembre 2001.
- Hélène Falavard, ingénieur (IE) CNRS, depuis août 2002. (CDD de janvier à juillet)
- Myriam Charles, assistante (TCE) CNRS, depuis janvier 2005.
- Laure Heigéas, ingénieur (IE) CNRS depuis février 2005.

¹ Pour tous les sigles, se reporter au Glossaire en dernière page du rapport.

Contractuels :

- Isabelle Costerg, depuis avril 2004 pour les projets NUMDAM et BNF ;
- Philippe Jacquier-Roux, depuis juillet 2005 pour le programme Ruche.

Direction :

- Yves Laurent, directeur de recherches CNRS, est directeur de l'unité depuis octobre 2002. Demande son renouvellement pour la période du prochain quadriennal 2007-2010.
- Thierry Bouche, Maître de Conférences à l'université Joseph Fourier, responsable du projet NUMDAM depuis octobre 2000, nommé responsable du programme CEDRAM par le comité de pilotage de ce projet en octobre 2005.

Suite à la démission de Laurent Guillopé qui était directeur adjoint depuis octobre 1995, l'unité demande la nomination de Thierry Bouche au poste de directeur adjoint de l'UMS 5638.

C. BILAN ET PERSPECTIVES

Nous diviserons l'activité de la Cellule MathDoc en cinq grands thèmes : documentation, coopération avec Zentralblatt Math, développement de logiciels, numérisation, édition. Les deux premiers sont à l'origine de la création de la Cellule MathDoc. Le projet de numérisation, plus récent a pris une importance grandissante et tous les membres de l'unité y sont impliqués. Le projet d'édition est en phase de démarrage. Nous lirons un rapport détaillé sur ces activités dans le chapitre suivant, nous allons noter ici les points les plus marquants.

1. Documentation

Seule ou en coopération avec le RNBM², la Cellule MathDoc a créé et entretient plusieurs catalogues (périodiques, ouvrages, thèses et prépublications) qui sont accessibles sur son serveur Web et permettent un accès centralisé aux données réparties dans les bibliothèques et les laboratoires.

Dans le domaine des thèses et prépublications, après avoir mis en place un système d'indexation des collections existantes, la Cellule MathDoc a collaboré avec le CCSD, créé en 2001, pour la mise au point de leur service spécialement en mathématiques.

Dans le domaine des documents numérisés, nous essayons de donner un accès facilité aux collections que l'on peut trouver à différents endroits. C'est ainsi que nous avons lancé le projet mini-DML en collaboration avec des partenaires étrangers (SUB Göttingen, Euclid, Springer). Avec la BNF, nous avons mis en place une interface, que nous avons appelée « Gallica-Math » et qui donne un meilleur accès aux collections mathématiques se trouvant dans la bibliothèque numérique Gallica.

2. Coopération Zentralblatt

Commencée dès 1995, la coopération avec le Zentralblatt Math s'est poursuivie dans ses axes principaux, développement logiciel, fourniture de données bibliographiques des revues

² <http://www.rnbnm.org>

françaises, gestion des abonnés français et promotion. Dans un premier temps cette coopération a bénéficié du soutien européen dans le projet LIMES avec la participation de plusieurs centres européens. Après la fin de ce projet et le rejet du projet suivant, la coopération a continué de manière bilatérale entre MathDoc et Zentralblatt.

Le contrat qui avait permis aux bibliothèques françaises de conserver un tarif figé pendant quatre ans jusque 2004 a été renouvelé dans le cadre d'une négociation globale du RNBM avec Springer.

Suite à une réunion qui a eu lieu au printemps à la Direction de la Recherche, un message a été envoyé pour demander une réflexion approfondie sur les possibilités d'amélioration du niveau scientifique de la revue. Un comité scientifique a été créé en 2005 ; présidé par Jean-Pierre Bourguignon, il a rendu un premier rapport qui devrait être examiné au prochain comité de coordination du Zentralblatt.

3. Développement de logiciels

Le logiciel de recherche et d'indexation EDBM est maintenant très largement utilisé aussi bien dans les applications internes à la Cellule MathDoc que dans de nombreux projets extérieurs (voir ci-dessous). Son développement s'est poursuivi dans le cadre du contrat européen LIMES. Un gros effort a été fait pour rendre faciles son installation et son adaptation, mais il faudrait encore développer la documentation et la flexibilité pour en faire véritablement un logiciel libre, c'est-à-dire utilisable par tous.

Le logiciel de gestion documentaire (LGD), destiné aux bibliothèques, a été développé et maintenu en coopération étroite avec le RNBM (plus spécifiquement la bibliothèque d'Orsay) et le réseau Mathrice. Il semble cependant que les bibliothèques utilisatrices, au moins les plus importantes, envisagent de se tourner vers des solutions commerciales.

De nombreux scripts et logiciels sont développés en interne pour différents usages, en particulier dans le cadre des programmes NUMDAM et CEDRAM, ils seront détaillés plus bas.

Enfin, en collaboration avec les *Annales de l'institut Fourier*, un logiciel de gestion de secrétariat de rédaction, RUCHE, a été acquis et est en cours d'adaptation.

4. Numérisation

L'opération de numérisation NUMDAM se déroule en deux phases correspondant à deux marchés séparés. La première phase, avec comme prestataire de service la société AIS, s'est terminée en septembre 2003. La seconde phase commencée immédiatement après avec la société Jouve se terminera début 2007.

Le bilan de la première phase est de 6 revues entièrement numérisées pour un total d'environ 200 000 pages (*Annales de l'institut Fourier*, *Journées ÉDP*, *Publications Mathématiques de l'IHÉS*, *Bulletin de la SMF*, *Mémoires de la SMF*, *Annales Scientifiques de l'ÉNS*). La deuxième phase a vu la numérisation du Séminaire de Probabilités de Strasbourg, des Annales de l'IHP et de celles de l'Université de Toulouse, des Annales de Gergonne et d'un certain nombre de grands séminaires, en particulier, le séminaire Bourbaki.

L'utilisateur peut immédiatement se rendre compte de l'intérêt qu'il y a à regrouper sur un même serveur le plus grand nombre possible de revues, aussi bien dans l'utilisation du moteur de recherches que dans l'utilisation des liens hypertextes entre les bibliographies et les articles correspondants. Comme on peut le voir dans le rapport ci-dessous, l'opération NUMDAM est menée

dans un grand souci de qualité technique. La qualité de la numérisation proprement dite (600 dpi) doit permettre d'obtenir un résultat durable tout en permettant dans le futur de tirer profit des améliorations des matériels et logiciels de consultation.

La qualité du travail accompli et les compétences développées au sein de la cellule dans un domaine encore peu exploré font que nous sommes sollicités de plusieurs côtés pour des conseils ou pour réaliser un travail de numérisation. Le crédit spécifique alloué en 2000 à la Cellule MathDoc pour mener ce projet, auquel se sont ajoutées d'autres subventions, sera épuisé à la fin de 2006. Pour pouvoir le continuer, il faudra donc trouver de nouveaux financements.

Un important travail reste à faire pour harmoniser les différents projets de numérisation dans le monde. Nous participons à un projet européen DML-Eu proposé dans le cadre de l'offre e-ContentPlus. Son but est la création d'un portail commun donnant accès aux différentes collections européennes. Le projet a été déposé en décembre 2005 et nous attendons une réponse en 2006.

5. Édition

Après plusieurs réunions préparatoires en 2003 et 2004 avec les responsables des principales revues de mathématiques françaises, un pôle d'édition a été mis en place au sein de la Cellule MathDoc, le projet CEDRAM. Le but du pôle est d'offrir aux revues une aide à plusieurs niveaux selon leurs besoins, dans les domaines de l'édition, de la publication électronique et papier, de la diffusion et de l'archivage. Des réunions régulières entre les responsables des revues françaises, qu'elles participent au pôle ou non, devraient permettre de réfléchir à l'avenir de l'édition qui sera sans aucun doute bouleversé par les nouvelles technologies.

Un comité de pilotage s'est réuni en octobre, il a nommé un directeur pour le pôle. Un ingénieur a été affecté au projet; après un temps de préparation en collaboration avec les *Annales de l'institut Fourier*, le travail d'édition commencera véritablement en janvier 2006 avec de nouveaux formats et la mise en ligne de la revue. Les autres revues suivront rapidement.

L'enjeu est maintenant de créer une structure véritablement professionnelle capable de gérer complètement les revues qui, comme les *Annales de l'ÉNS*, souhaitent une prise en charge globale.

Les actions de la Cellule MathDoc

A. LA DOCUMENTATION

1. Le Catalogue Fusionné des Périodiques

Le catalogue fusionné des périodiques contient environ 8 400 journaux de mathématiques morts et vivants, y compris des journaux électroniques d'accès libre. Pour chaque titre, les informations suivantes sont mentionnées :

- une notice du journal, provenant de la base de données de l'ISSN, ou d'un catalogage direct, dans le cas de journaux très récents ou purement électroniques ;
- les localisations de ce journal dans les différentes bibliothèques du RNBM, et les états de collections précis. Ces données sont obtenues par le retraitement informatique de

différents fichiers envoyés par les bibliothèques participantes (toutes les bibliothèques de mathématiques importantes y participent) ;

- différents URL associés aux titres : services de sommaires, site de l'éditeur, site de service de numérisation. Des listes de tous les journaux en ligne sont extraites dynamiquement de la base.

La base de données est régulièrement tenue à jour par :

- le contact avec les bibliothèques pour récupérer les nouvelles versions de leurs catalogues, réécriture parfois des programmes de retraitement des ces fichiers ;
- la mise à jour des notices de journaux et des url associés.

2. Catalogue Fusionné des Ouvrages

Forte de l'expérience, à la fois du projet européen EULER et du catalogue fusionné des périodiques, la Cellule MathDoc a entrepris une collaboration avec le RNBM et plus spécifiquement avec Raymond Douet, informaticien à la bibliothèque Jacques Hadamard d'Orsay. Cette base a vu le jour en 2002 et contient les notices de la plupart des bibliothèques de mathématiques. Le modèle « centralisé » (toutes les notices rassemblées a posteriori dans une même base) a montré son efficacité par rapport au modèle « décentralisé » antérieur (un accès réparti en temps réel aux différentes bases de données).

L'apport de la Cellule MathDoc dans cette application a essentiellement porté sur :

- l'analyse préalable ;
- la définition de la structure de la base ;
- la fourniture des programmes de lecture de données à la fois du projet EULER et du catalogue des périodiques ;
- le programme d'élimination des doublons dans les notices bibliographiques ;
- dans certains cas, une assistance aux bibliothèques : programme d'extraction des données du logiciel LGD, ainsi que d'autres explications préalables quant aux formats de fichier à fournir ;
- dernièrement, une analyse du contenu de la base, afin d'établir des spécifications pour son amélioration.

Il est à noter que les choix techniques (MySQL pour la base de données, perl pour son alimentation et php pour son interface web) sont simples et tout à fait adaptés, malgré un volume relativement important de données à traiter.

Suite au départ de Raymond Douet, la maintenance a été interrompue pendant près d'un an. Nous avons fait une analyse détaillée du contenu de la base et établit des spécifications pour son amélioration. Parmi celles-ci se trouve la possibilité d'importer des notices au format UNIMARC et ainsi étendre le catalogue à des bibliothèques utilisant ce format, en particulier dans le cadre du SUDOC³.

³ <http://corail.sudoc.abes.fr>

3. *Le service Simple de sommaires*

Ce service propose les sommaires de 696 périodiques de mathématiques et disciplines connexes, à partir de l'année 1997, avec possibilité de recherche par auteur, titre, nom de journal. La mise à jour est hebdomadaire.

Une quarantaine de laboratoires ou institutions ont demandé l'accès à ce service. Un service d'alerte est proposé, permettant aux utilisateurs qui le désirent de recevoir chaque semaine par courrier électronique les nouveaux sommaires des périodiques qui les intéressent. 94 utilisateurs sont abonnés (à une quinzaine de périodiques en moyenne) pour un envoi hebdomadaire de 200 à 300 sommaires.

Autrefois basée sur le système WAIS, l'application a été entièrement reprise et utilise maintenant le moteur EDBM (cf. ci-dessous). Sa réécriture a permis d'obtenir un système beaucoup plus fiable, qui ne nécessite que peu d'interventions, et dont l'évolutivité est grandement accrue.

La question de son maintien s'est posée en 2005. Un sondage a été effectué auprès des bibliothécaires du RNBM et de la communauté des mathématiciens pour répondre à cette interrogation. Ce sondage a montré que ce service était apprécié par ses utilisateurs. Il a donc été décidé de le maintenir pour les années à venir malgré le prix relativement élevé de l'abonnement à SWETS.

4. *Index nationaux des prépublications et thèses de mathématiques*

Les index nationaux des prépublications, thèses et habilitations de mathématiques sont basés sur le principe de la récolte de métadonnées sur différents serveurs et leur concaténation au sein d'une même base de données.

La récolte de ces métadonnées se fait de deux manières différentes :

- via le moteur de recherche « Harvest » sur les sites des instituts de mathématiques qui participent au projet ;
- via le protocole OAI-PMH, sur les sites du CCSD, HAL⁴ et TEL⁵.

Les serveurs du CCSD sont de plus en plus utilisés par les laboratoires de mathématiques. La Cellule MathDoc les y encourage et apporte une aide ponctuelle aux laboratoires désirant alimenter la base HAL avec des données déjà existantes. Progressivement, le système initial sera abandonné pour celui du CCSD.

5. *Gallica-Math*

La bibliothèque numérique Gallica de la Bibliothèque nationale de France, contient de nombreux textes mathématiques fondamentaux. Le moteur de recherche de Gallica permet de localiser les volumes. Mais l'accès aux volumes n'est pas le plus pertinent pour certains types de contenus, en particulier les collections d'articles (volumes de périodiques ou oeuvres complètes).

Dans le cadre du pôle associé BNF-Cellule MathDoc, un nouveau service « Gallica-Math » a été mis en place pour offrir un accès efficace à un choix de collections mathématiques numérisées par la BNF, au niveau article. L'application spécifique permet un feuilletage de la table des matières

⁴ <http://hal.ccsd.cnrs.fr>

⁵ <http://tel.ccsd.cnrs.fr>

de chaque volume et un système de recherche complet est mis en place au travers de la mini-DML (voir paragraphe suivant).

Les collections seront complétées au fur et à mesure.

Sont actuellement consultables sur ce serveur :

- les œuvres complètes de nombreux mathématiciens,
- le JMPA, *Journal de mathématiques pures et appliquées* : articles catalogués de 1836 à 1881, articles numérisés de 1836 à 1932,
- le RBSM, Répertoire bibliographique des sciences mathématiques (1894-1912), une des premières bibliographies dédiées aux mathématiques.

Le travail fourni par la Cellule MathDoc a consisté, d'une part à cataloguer les collections (manuellement ou par récupération semi automatique des tables de matières existantes), puis à développer des applications pour leur affichage sur le web. Une mention spécifique doit être faite pour le RBSM, dont la saisie est le résultat d'une collaboration avec le laboratoire de philosophie et d'histoire des sciences (Archives H. Poincaré, CNRS/UN2, Nancy).

6. LiNuM et mini-DML

La dispersion des ressources numérisées en mathématiques pose de sérieux problèmes aux utilisateurs et il serait extrêmement utile de disposer de moyens unifiés pour y accéder. Les deux grandes bases de données mathématiques MathSciNet et Zentralblatt mettent de plus en plus de liens vers les documents numérisés dans les notices existantes. De plus, la numérisation et la mise en ligne du Jahrbuch⁶ (1868-1942) par le Zentralblatt a permis de remonter dans le temps. Mais ces accès restent très incomplets.

La Cellule MathDoc propose d'une part la base LiNuM⁷ qui permet un accès unifié à d'importantes collections d'ouvrages de mathématiques numérisés dans le monde : Gallica, Cornell University Library (Digital Math Books Collection), Michigan University Library (Making of America et Historical Math Collection), Göttinger Digitalisierungszentrum (Mathematica), ICM Varsovie (Biblioteka Wirtualna Matematyki), notamment. Ce service est alimenté par l'envoi de catalogues par les institutions concernées, il demande un travail de maintenance relativement léger, mais assez manuel.

Par ailleurs, les grands projets fédérateurs de type Digital Mathematical Library (DML puis WDML et sa version européenne DML-eu) tardant à se mettre en place, la Cellule MathDoc a proposé un système léger, le projet mini-DML⁸. Pour éviter les écueils des grands projets, nous privilégions les relations bilatérales et nous proposons d'utiliser la technologie OAI-PMH afin de récolter de façon automatique toutes les données qui peuvent l'être. Des discussions sont en cours avec la bibliothèque de Göttingen et le projet Euclid d'une part et avec l'éditeur Springer d'autre part pour mettre en place les serveurs OAI avec des standards de métadonnées simplifiés.

⁶ <http://www.emis.de/MATH/JFM/>

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

⁸ <http://www.numdam.org/minidml/>, voir <http://www.numdam.org/minidml/litterature/article-minidml.pdf>.

À l'heure actuelle, la version de la mini-DML installée à Grenoble permet de chercher des collections vivantes dont le catalogue est diffusé par OAI (project Euclid, arXiv, NUMDAM...) et d'autres collections pour lesquelles la Cellule MathDoc a, elle-même, produit un catalogue exploitable (Gallica-math, *Annals of Math.* sur JSTOR, *Fundamenta Mathematicae* sur ICM Varsovie...).

7. Archives Bourbaki

Nous avons été chargés de mettre en ligne les archives Bourbaki, c'est-à-dire l'ensemble des documents conservés par le groupe et qui contiennent en particulier les différentes versions préliminaires de leur traité. Elles avaient été confiées au CNRS et étudiées par une unité regroupant mathématiciens et historiens dirigée par Christian Houzel qui a été dissoute par la suite. Nous avons récupéré les résultats d'une première numérisation et complété celle-ci pour 600 pages qui avaient été omises. Après une évaluation qui doit permettre de décider s'il faut recommencer la première numérisation, nous allons créer un site web pour les publier courant 2006.

8. Serveurs web d'information de la Cellule MathDoc

Deux serveurs d'information sont opérationnels :

(1) Un serveur institutionnel : la Cellule MathDoc y apporte des informations sur ses activités et projets, ses publications, son personnel, etc. Ce serveur a subi des modifications dans le courant de l'année 2005 par le changement de son aspect visuel, par la refonte de certaines rubriques.

(2) Un serveur pour le public : le Portail documentaire mathématique :

ce serveur donne accès à la documentation mathématique, que ce soit les applications de la Cellule MathDoc ou d'autres sources d'informations connues. Il a été mis en place fin 2002, et est apprécié par les utilisateurs, bibliothécaires comme mathématiciens.

La mise en place du service « OpenResolver », basé sur le standard « OpenURL » permet à l'utilisateur de poser une question et d'avoir en retour un ensemble de liens cliquables « précodés » vers les bases où il est susceptible de trouver une réponse. Ce service lui évite de consulter successivement 4 ou 5 rubriques du serveur et de remplir 4 ou 5 formulaires différents. De plus, il a l'avantage d'informer l'utilisateur sur les possibilités de trouver l'information recherchée.

Outre les services documentaires proprement dits, la Cellule MathDoc maintient également un annuaire des laboratoires et bibliothèques de mathématiques (coordonnées, responsables, site web...) et un annuaire des écoles doctorales et masters de mathématiques.

Les applications disponibles via ce portail sont régulièrement mises à jour. La mise en place d'une rubrique « actualités » permet aux utilisateurs de suivre la vie du portail.

9. La lettre d'information de la Cellule MathDoc

La lettre d'information de la Cellule MathDoc a été mise en place en janvier 2003 avec l'objectif d'informer largement la communauté des mathématiciens sur les services de la Cellule MathDoc. Les premiers numéros envoyés nous ont permis d'avoir un retour positif de la part des utilisateurs sur les services que nous proposons. Cette lettre est envoyée conjointement à une liste d'abonnés (environ 150 personnes actuellement, mathématiciens et bibliothécaires) et à des listes de diffusion telles que celle du RNBM ou celle du réseau Mathrice.

10. Édition de séminaires et conférences

Deux actions, index du séminaire de probabilités de Strasbourg initié en 1997 et soutien à l'édition des Journées équations aux dérivées partielles (JÉDP) commencée en 1998 se sont poursuivies.

Leur insertion dans NUMDAM avec la numérisation rétrospective (réalisée pour les JÉDP pour les années 1975-2000 au cours de sa première phase, programmée pour les volumes 1967-2002 du séminaire de probabilités au début de sa seconde phase démarrant en 2003) marque une nouvelle étape.

Pour les JÉDP, les années récentes (*i.e.* post 2001) ont été insérées dans NUMDAM. Un cédérom, contenant un extrait statique de NUMDAM, a été réalisé à l'automne 2003 (voir la partie sur NUMDAM pour le détail).

B. LA COOPÉRATION AVEC LE ZENTRALBLATT MATH

La coopération avec le Zentralblatt Math (ZM), une des missions confiées à la Cellule MathDoc à sa création, a continué dans ses axes principaux :

- développement logiciel,
- fourniture de données bibliographiques de revues françaises,
- gestion des abonnés français et promotion.

Laurent Guillopé, puis Yves Laurent ont représenté la Cellule MathDoc au comité semestriel de coordination du Zentralblatt Math. Pour le compte de la Direction de la recherche, Pierre Bérard a réalisé avec le Directeur chargé des mathématiques une enquête sur le contenu scientifique du Zentralblatt Math : la Cellule MathDoc a apporté son soutien technique pour certaines phases de cette enquête.

La Cellule MathDoc a été un des partenaires du contrat européen LIMES (2000-2004).

1. Développement logiciel

Développé par Claude Goutorbe, l'ensemble logiciel EDBM l'a été à l'origine pour le serveur de la base Zentralblatt Math (ZM). La version 1.0 a été finalisée en septembre 2001, avec installation sur l'ensemble des 11 miroirs internationaux en janvier 2002, des modules spécifiques (traitement des diacritiques, mise à jour incrémentale, indexation) ont été achevés et livrés avec leur documentation. Cependant, les mises à jour de la base de données ZM continuent à être effectuées par envoi de cédéroms, sans que le module d'indexation ne soit utilisé : le renouvellement du processus de production à Berlin tarde à prendre corps.

La version 1.0 du ZM-lookup a été achevée en janvier 2002 : cet outil logiciel permet d'associer à une référence bibliographique l'identifiant de la notice ZM si elle existe. Cette version a été utilisée par le programme NUMDAM : 58 529 références bibliographiques traitées, avec 44 302 identifiants obtenus (37 845 identifiants rendus par l'outil MR-lookup pour les Mathematical Reviews). L'outil, disponible sur le réseau⁹, traite des lots de références bibliographiques légèrement balisées (auteur, titre, année de publication). Une extension pour traiter des références

⁹ <http://www-mathdoc.ujf-grenoble.fr/ZMATH/zlink.html>

non balisées existe à l'état de prototype (elle a été testée avec grand succès sur un ensemble d'environ 17 000 références provenant du Michigan J. Math.)

La version 2.0 du logiciel a été fournie au FIZ en septembre 2004 ; certaines applications ont d'ores et déjà été adaptées à cette version (base de données des périodiques pris en compte par Zentralblatt), mais la base principale n'est cependant pas encore disponible dans cette nouvelle version.

2. Fourniture de données bibliographiques de revues françaises

Les revues concernées sont les *Annales de l'institut Fourier* et toutes les publications (revues et monographies) de la SMF. Les métadonnées sont régulièrement fournies au Zentralblatt dès la publication et parfois avant. Les données des revues participant à CEDRAM pourront prochainement être régulièrement et automatiquement ajoutées à cette fourniture de données.

Lors de la rencontre avec la direction de l'INIST en mars 2003, il a été discuté de l'alimentation du ZM en données bibliographiques saisies par l'INIST pour la base Pascal, avec réciproquement le déchargement de données du ZM dans Pascal. Le FIZ Karlsruhe et l'INIST sont intéressées mais le projet n'a guère avancé à ce jour.

3. Gestion des abonnés et serveurs français, promotion

La Cellule MathDoc tient le rôle d'intermédiaire entre les abonnés français et Springer-Verlag (Paris et Heidelberg) : la vérification des factures (pour les types d'abonnement et la bonne prise en compte des accords de consortium) et le suivi des règlements a requis beaucoup de temps (nouvelles règles de TVA, erreurs,...) à Monique Marchand. L'accord de 1999, qui organisait les divers consortiums et stabilisait les tarifs d'abonnement au niveau de 1999 pour quatre ans, a profité aux abonnés français (par ex. l'abonnement pour la version en ligne est de 3 977 Euros, alors que le prix pour public pour 2004 est 5 300 Euros -non compris les 15 % de réduction pour les abonnés de la SME). La renégociation de cet accord a été menée par le RNBM et a abouti au début de l'année 2005.

Les serveurs du réseau français (Grenoble, Marseille et Orsay) ont eu leurs données régulièrement mises à jour, ainsi que le serveur du réseau international à Strasbourg. L'interface d'accès du réseau français devra être mise à jour. Des analyses statistiques fines sur les consultations (des abonnés ou en mode restreint) du ZM à travers les serveurs français et les serveurs de Berlin et Strasbourg (réseau international) ont été effectuées sur trois semestres à partir de janvier 2001.

4. Contrats européens

De 2000 à 2004, la Cellule MathDoc a été un des partenaires du contrat européen LIMES (Large Infrastructure in Mathematics Enhanced Services, dans l'action Accès aux infrastructures de la recherche). Intervenant principal de la Cellule MathDoc, Claude Goutorbe a travaillé à 70 % de son activité sur ce contrat. Outre les développements logiciels mentionnés ci-dessus, il faut mentionner la tenue d'une des réunions semestrielles du projet à Nantes en octobre 2002. Une partie des revenus du contrat LIMES a pu être réinvestie dans des actions autres de la Cellule MathDoc. En continuation de ce contrat, la Cellule MathDoc a participé à la définition du projet CITIZEMS (participation compatible avec les autres projets européens de la Cellule MathDoc). Ce

projet n'a malheureusement pas été retenu ce qui va limiter sérieusement le processus d'europanisation, si difficile à réaliser concrètement.

C. DÉVELOPPEMENT DE LOGICIELS

1. *European Database Manager for Mathematics (EDBM)*

Développé à l'origine pour l'interface Web de la base de données Zentralblatt-MATH, le logiciel EDBM est maintenant utilisé par de nombreuses autres applications, développées par la Cellule MathDoc comme par d'autres institutions.

Citons, à Berlin: ERAM (Jahrbuch), MATH-DI (didactique des Mathématiques), CompuScience (Informatique); dans un cadre européen: EDBM est utilisé comme moteur de recherche et d'indexation par le projet EULER.

Plus récemment, l'INRIA a basé son service de sommaires et de fournitures de documents Eurydice sur la nouvelle version du moteur EDBM.

Parmi les applications développées par le Cellule MathDoc, citons les interfaces Web d'un certain nombre de bases de données bibliographiques: SemProba (*Séminaire de probabilités de Strasbourg*), LiNum (catalogue de Livres Numérisés Mathématiques), RBSM (Répertoire Bibliographique des Sciences Mathématiques), JMPA (accès à la numérisation par la BNF du Journal de Mathématiques Pures et Appliquées), le serveur de sommaires de la Cellule MathDoc, et bien sûr le plus important le serveur du projet de numérisation NUMDAM.

Le développement de cette dernière application a fourni l'occasion de nouvelles améliorations importantes du moteur EDBM, nécessitées par la plus grande complexité de ce service sous certains aspects.

Au cours des quatre dernières années, les développements ont principalement portés sur les aspects suivants :

- le moteur de recherche, utilisant des bibliothèques "open source", a été largement réécrit pour être adapté aux versions les plus récentes de ces bibliothèques, gagnant au passage de nombreuses fonctionnalités utiles ;
- un nouveau module d'indexation a été développé. Ce module a l'avantage de permettre un contrôle plus fin du processus d'indexation. Il est également notablement plus efficace ;
- un certain nombre d'outils de plus haut niveau, utilisables dans le cadre de l'écriture des applications elles-mêmes, ont été mis en place, constituant un squelette largement réutilisable et configurable. Grâce à ces outils, la création d'une application de complexité moyenne ne demande que quelques jours.

2. *Outils de Gestion pour le programme NUMDAM*

Le programme NUMDAM a nécessité des développements logiciels spécifiques et très ciblés. Ils concernent les points suivants : le contrôle qualité et la gestion des métadonnées. Ces développements ont été réalisés dès le lancement de la phase de production du premier marché et depuis, tout au long de la vie du programme.

2.1. Procédures de contrôle qualité de NUMDAM

Après réception des données transmises par notre prestataire de service, une première phase de travail concerne le contrôle de la qualité de l'ensemble des données. C'est une application qui a été développée au sein de l'unité (langage de programmation perl), et qui, depuis le début du programme NUMDAM, ne cesse d'évoluer pour intégrer des vérifications de plus en plus poussées des données.

Nous travaillons à la fois sur les fichiers images des articles numérisés et à la fois sur les fichiers de métadonnées décrivant chaque volume et chaque article numérisé. Nous cherchons à extraire le maximum d'information de chacun pour les recouper entre elles et ainsi s'assurer de leur justesse.

Une première partie concerne tout d'abord l'intégrité et la validité des données. Les fichiers sont-ils au bon format, à la bonne résolution et bien lisibles par les outils standards ? Il peut en effet se produire des incidents au cours de leur production chez le prestataire, comme au cours de leur transfert entre les différents supports informatiques jusqu'à notre unité. Un comptage des répertoires et fichiers physiques reçus est effectué pour comparaison avec les métadonnées d'une part et les données attendues d'autre part (dépouillement préalable effectué pour chaque revue). Nous vérifions aussi que tout est bien présent dans le fichier article, avec le comptage du nombre de page, le contrôle de la présence du texte caché par exemple dans le cas des fichiers pdf et djvu. Nous regardons aussi les métadonnées, plus précisément, les fichiers XML qui décrivent chaque volume et chaque article: sont ils bien conformes à la dtd de NUMDAM ? Chaque champ contient-il le bon type de données attendu, nombre, chaîne de caractères, nom de fichier; ces fichiers sont-ils bien présents dans les répertoires livrés ?

Un deuxième travail concerne ensuite la cohérence globale de l'ensemble des données entre elles. Nous analysons la nomenclature des différents répertoires et fichiers articles pour nous assurer qu'elle correspond bien aux identifiants du volume et de ses articles décrits dans les métadonnées. Nous recoupons différentes sources d'informations pour contrôler que celles-ci sont bien exactes : identifiants articles et informations issues de la description du volume, nombre de pages des articles et fichiers correspondants, langue des titres, langue des articles corroborées avec l'analyse du texte issu des images des articles (appelé « OCR »), qui fournit le texte caché des fichiers pdf et djvu, nombre de références bibliographiques présentes dans les métadonnées avec celui déduit du texte caché, ...

Au cours de ces différentes phases, nous pouvons être amenés à demander au prestataire une correction des données transmises ou une nouvelle numérisation si nous constatons des erreurs graves. Nous pouvons aussi effectuer nous même des corrections moins importantes ou isolées.

À l'issue de ce travail de contrôle exhaustif des données et de celui du contrôle par échantillon, le lot est validé auprès du prestataire de service et pourra être importé dans la base de données NUMDAM afin de préparer la mise en ligne de la revue.

2.2 « Gestion globale des données » de NUMDAM

2.2.1 Gestion des métadonnées

Une fois le contrôle qualité de la revue effectué et l'étape validée, l'ensemble des métadonnées recueilli auprès du prestataire de service est géré par une base de données

relationnelle (MySQL). Cette structure de données offre une organisation synthétique des données, qui permet un accès suivant des « angles » très différents aux informations, des observations plus fines des contenus, des corrections globales si nécessaire ainsi que la possibilité d'extraire des données statistiques globales.

Cette gestion des métadonnées par une base de données intermédiaire constitue en outre une source supplémentaire de contrôle et d'amélioration qualitative des données. En effet, ce changement de format des données se révèle intéressant à plusieurs titres.

D'une part, la comparaison entre données avant import et après export de la base peut mettre en évidence des écarts liés à des erreurs initiales de format ou de structure.

D'autre part, nous avons ajouté à cette étape de nouvelles procédures de contrôle nous permettant d'améliorer encore la qualité des données. Ainsi toute nouvelle revue fait l'objet d'un travail d'analyse plus poussé au cours de son import dans la base de données afin de réduire encore l'écart possible avec l'oeuvre initiale. Pour les titres des articles par exemple, nous avons développé des procédures d'analyse de l'orthographe (élaboration d'un index de termes mathématiques) et d'uniformisation de la syntaxe (espaces, guillemets, points, majuscules...), tout en veillant à respecter au plus près l'article original. De même, nous travaillons sur la syntaxe des noms d'auteur et des bibliographies. Nous recherchons aussi d'éventuels oublis de relations entre articles.

Suite à cela, les métadonnées sont extraites de la base de données sous forme de catalogues XML décrivant chaque volume et vont pouvoir être utilisés pour l'affichage en ligne de la revue sur NUMDAM.

2.2.2 Gestion des auteurs

La gestion de l'ensemble des auteurs soulève différentes questions : recherche de doublons, correction de noms ou prénoms erronés, liens entre différentes graphies d'un même auteur, recherche des coordonnées des auteurs récents pour les problèmes de droits de diffusion...

Nous avons, à cette fin, mis en place une base de données de référence de l'ensemble des auteurs de NUMDAM et développé une interface facilitant l'accès à toutes ces données. Cette base nous permet d'identifier précisément chaque auteur, de rechercher des formes proches, de repérer et corriger des noms mal orthographiés suite à des erreurs de catalogage, et dans le cas d'abréviations ou d'erreurs présentes dans la publication initiale, de lier une forme incorrecte ou alternative à sa référence. Elle nous permet en outre d'automatiser des procédures de contact des auteurs (formulaire envoyés par mël), lorsque nous avons besoin de les contacter au sujet des droits de diffusion des articles.

3. Logiciel de Gestion Documentaire pour les bibliothèques de Mathématiques

Le logiciel de gestion documentaire est un projet à l'initiative de la Cellule MathDoc en collaboration avec l'institut Fourier et le Réseau national des bibliothèques de mathématiques. Il trouvait sa motivation dans le contexte particulier des bibliothèques de mathématiques, confrontées à des problèmes d'informatisation de leur fonds. Les systèmes intégrés de gestion de bibliothèque disponibles sur le marché se révélaient en effet des ensembles complexes et coûteux à l'achat comme en maintenance, et basés sur des systèmes propriétaires (codage "privé" des données du fonds documentaire, import-export au bon vouloir des propriétaires du logiciel). Partant de cette analyse et pour répondre à leurs besoins précis, la Cellule MathDoc a entrepris le développement d'un prototype de logiciel de gestion documentaire.

Le travail conjoint des ingénieurs de la Cellule MathDoc et de bibliothécaires du RNBm a permis l'établissement d'un cahier des charges précis du LGD. La première maquette, développée par Claude Goutorbe de la Cellule MathDoc, a été disponible en mars 2000. Réalisée à partir de logiciels libres (Python - MySql - GTK), elle limite les fonctionnalités du logiciel à l'essentiel et a pour double objectif de fournir une plate-forme de test pour les bibliothécaires et d'évaluer la faisabilité du développement d'un tel logiciel. Ce prototype permet la gestion du catalogage, des lecteurs et des opérations de prêt. Il fournit aussi d'autres fonctionnalités utiles comme l'import de bases existantes de type "texto", l'export des données, le « nettoyage » des index, l'édition de codes barres, etc. Fonctionnel, il a rapidement été installé et mis en service à la bibliothèque de l'institut Fourier. Depuis, diverses installations ont été réalisées. Actuellement, il est en usage dans les bibliothèques de l'institut Fourier, du laboratoire de mathématiques de Montpellier ainsi qu'à celle d'Orsay.

La Cellule MathDoc considère aujourd'hui que le LGD a su montrer sa faisabilité et pleinement rempli sa mission. Il a aussi révélé un certain nombre de difficultés et de réserves rencontrées lors de sa diffusion, à travers des questions d'ordre technique et structurelle :

- problème de la conversion/récupération des données du fonds initial,
- manque de fonctions secondaires ou annexes telles que le bulletinage, la gestion des commandes,...
- question des moyens humains nécessaires pour assurer: l'installation, l'assistance technique « hot-line » aux utilisateurs, le besoin régulier d'adaptation aux « nouvelles technologies », le développement des fonctionnalités...

En outre, l'adoption du LGD comme logiciel documentaire nécessite une prise en charge informatique locale et une volonté commune de la part de la direction, des bibliothécaires et des informaticiens.

Depuis son installation au sein de l'institut Fourier, des évolutions ont été apportées au LGD. Mickael Marchand, ingénieur informatique de l'institut Fourier a ainsi travaillé sur la correction et l'adaptation de certaines fonctionnalités (interrogation publique du catalogue via Internet, historique des prêts par lecteur, outil de récolement pour les bibliothécaires...), permettant de maintenir son utilisation dans de bonnes conditions et avec des librairies informatiques à jour. En fonctionnement dans cette bibliothèque depuis 1999, le LGD a pu démontrer ses qualités et sa robustesse.

D'autres adaptations ont été effectuées, comme à la bibliothèque d'Orsay, où Raymond Douet a apporté des améliorations et de nouvelles fonctionnalités (gestion de la base via une base temporaire, journalisation des opérations, développement d'un module internet pour la consultation du catalogue, affichage détaillé et impression des notices ouvrages, prêts rétroactifs, historique, gestion de la validité des comptes lecteurs, etc.).

L'initiative de 2002, sous l'égide du Réseau national des bibliothèques de mathématiques pour tenter d'apporter des améliorations au logiciel actuel et surtout de lancer de nouveaux développements n'a pas abouti.

Le LGD a ainsi été distribué aux différentes bibliothèques en ayant fait la demande. La qualité de cet outil et la fiabilité dont il fait preuve dans les bibliothèques où il est en usage ne sont plus à démontrer. Aujourd'hui, la Cellule MathDoc se désengage progressivement du projet alors que, par ailleurs, d'autres solutions « libres » ont récemment émergé. Que ce soient des outils comme « KOHA » ou bien « PMB », ils sont le fruit de projets internationaux et s'avèrent des

logiciels de gestion documentaires de qualité. Traduits en différentes langues, ils évoluent régulièrement grâce à des équipes importantes de développeurs, pour inclure les nouveaux besoins et les nouvelles possibilités techniques. Encadrés par des mailing-list actives, ils sont aussi accompagnés de « petites » sociétés de service pouvant localement assurer l'installation et le paramétrage du logiciel, mais aussi une « hot line » ou des services plus ponctuels. Afin de faciliter les transitions éventuelles avec d'autres logiciels ou de permettre l'échange de données avec d'autres fonds bibliothécaires, la Cellule MathDoc travaille actuellement au développement de « passerelles » vers le format Unimarc des données bibliographiques issues du LGD.

4. RUCHE - Outil de gestion rédactionnelle

4.1 Origine et objectifs du projet

Courant 2004, RUCHE est né de la volonté d'accélérer et de faciliter la gestion rédactionnelle de la revue *Traitement du Signal* (<http://www.traitementdusignal.fr>). Les objectifs de RUCHE sont les suivants :

- accélérer les délais de publication et de fabrication des numéros,
- assurer aux acteurs de la revue un accès relationnel, transparent et immédiat aux données,
- automatiser et fiabiliser les procédures, les courriers électroniques, les gestions de formats,
- s'adapter aux nouvelles technologies de la communication et de l'édition.

La réalisation de ce projet s'est faite grâce à la collaboration étroite entre la revue *Traitement du Signal* et « nb notabene », société de conseil en communication et de design graphique. La revue *Traitement du Signal* et le GRETSI ont joué le rôle prédominant de maître d'œuvre et ont écrit le cahier des charges. D'autre part, à l'aide de l'analyse et de la réflexion menées par « nb notabene », Philippe Jacquier-Roux, embauché par la revue *Traitement du Signal* pendant plus d'un an, a développé le logiciel jusqu'à sa mise en place.

4.2 Technologies et fonctionnalités de RUCHE

RUCHE est développé à l'aide des technologies classiques utilisées pour les applications internet : PHP/MySQL. D'autre part les fichiers PDF sont générés avec la classe Fpdf.

RUCHE se décompose en trois accès :

- accès "public" qui permet aux auteurs de faire une demande de soumission d'article et de s'identifier ;
- accès sécurisé "public identifié" qui permet de :
 - accéder à ses informations personnelles et les modifier,
 - déposer un article,
 - consulter les analyses sur un article et y répondre (auteurs),
 - récupérer et lire des articles (comité de lecture), les annoter, émettre un avis ;

- accès sécurisé "réservé" au secrétariat de rédaction qui permet de :
 - envoyer des courriers automatiquement,
 - gérer le transfert de fichiers,
 - suivre et modifier l'état d'avancement des articles,
 - gérer les numéros, archiver.

4.3 Acquisition et adaptation de RUCHE

RUCHE a été acquis en 2005 par l'Association des Annales de l'institut Fourier auprès de la revue *Traitement du Signal* via la licence CeCILL, licence française de logiciel libre élaborée par le CEA, le CNRS et l'INRIA.

Philippe Jacquier-Roux, embauché par la Cellule MathDoc pour une période de 9 mois, est en charge d'adapter et de mettre en ligne le logiciel RUCHE pour les *Annales de l'institut Fourier*. Sa mise en place effective est prévue pour début février 2006.

Par la suite la Cellule MathDoc se chargera de la maintenance de RUCHE et de son adaptation aux autres revues intéressées.

D. NUMÉRISATION

1. Introduction

Nous vivons une époque charnière quant à la diffusion du savoir et des résultats de la recherche. L'imprimé aura été pendant 500 ans le vecteur, et dans une certaine mesure le moteur, de la culture et de l'érudition. Cette singularité est aujourd'hui largement remise en cause par les nouveaux modes de production, de conservation et d'accès aux textes. S'il est impossible aujourd'hui de prédire le futur des publications scientifiques, on ne peut ignorer que les recherches de documentation se font désormais avant tout sur Internet, et que l'accès aux textes de référence en un clic est un atout immense pour leur rayonnement dans le monde entier.

De nombreux programmes de numérisation sont en cours dans le monde. En Europe, les programmes généralistes de la Bibliothèque nationale de France (Gallica) ou de la bibliothèque universitaire de Göttingen offrent des collections très riches en accès libre. Le projet le plus ambitieux dédié spécifiquement aux revues savantes est probablement l'américain JSTOR¹⁰ (de l'ordre de 600 titres anglophones, dont 30 en mathématiques, pour 3 millions d'articles et 20 millions de pages) qui a d'une certaine façon défini la qualité et l'ergonomie minimale attendue d'une telle opération. Son succès est tel qu'un assez grand nombre de bibliothèques universitaires américaines sont en train de retirer les archives papier des revues concernées pour les archiver en lieu sûr, mais loin des utilisateurs !

Le programme NUMDAM, lancé en l'an 2000 grâce à des fonds provenant de la direction de la Recherche, a eu pour première mission de définir les fonctions, attendues par les chercheurs, d'une publication électronique des archives de revues de recherche en mathématiques, donc la nature et le format des données capturées lors de l'opération de numérisation, puis de concevoir une interface puissante et ergonomique pour l'exploitation de ces données. La réussite de ce travail se mesure à l'impact qu'il a eu, NUMDAM est actuellement considéré au niveau mondial comme

¹⁰ JSTOR : www.jstor.org.

une référence et un certain nombre de projets déjà très avancés envisagent de retraiter leurs collections pour les mettre à niveau (citons par exemple le projet soumis à la fondation Moore par la communauté des mathématiciens américains sous le chapeau de l'AMS et du MSRI de Berkeley, qui va jusqu'à envisager de numériser à nouveau des journaux déjà disponibles sur JSTOR ; le programme français PERSEE¹¹ dédié aux revues de sciences humaines s'est aussi très largement inspiré de NUMDAM).

Mais la qualité première de NUMDAM est celle des collections rendues accessibles, elle revient avant tout à la longue tradition d'exigence de l'école mathématique française. Le programme a permis la mise en ligne de revues prestigieuses et cette bonne image a conduit le reste des revues indépendantes françaises à nous solliciter, ainsi que plusieurs revues européennes.

On peut résumer les objectifs du programme par quelques mots clefs :

- *Préserver le patrimoine* mathématique « français » en assurant un archivage numérique fidèle pris en charge par une institution publique.
- Augmenter sa *visibilité* et permettre à tous d'y *accéder* grâce à la puissance des outils de navigation et à l'ergonomie générale du site.
- Donner de ce fait un *outil* performant à la *recherche* contemporaine (et future).
- Soutenir les publications académiques en exposant la richesse de leur production.

Soulignons également quelques enjeux spécifiques qui font de la numérisation de revues mathématiques une opération attendue et délicate :

- La documentation mathématique ne se périme pas : le recours aux textes originaux est toujours riche d'enseignement, pas seulement d'un point de vue historique.
- Elle est utile pour d'autres sciences, de façon *asynchrone* : les résultats de la recherche mathématique sont la *lingua franca* de la science, mais les chimistes, physiciens, économistes et autre cybernéticiens ont souvent recours à des résultats anciens (parfois à des branches entières des mathématiques qui ont été abandonnées depuis par les mathématiciens). L'un des défis d'une numérisation d'articles de mathématiques est de rendre des services aux *utilisateurs* des mathématiques.
- Elle est valide comme un *tout*, qui forme un réseau (international) de références. Il est donc crucial de permettre aux lecteurs de remonter dans l'histoire d'un résultat, de proposer le maximum de liens significatifs.

2. Présentation générale du programme NUMDAM

2.1 Moyens mis en œuvre

2.1.1 Moyens humains

La conduite du programme NUMDAM requiert une partie importante du personnel de la Cellule MathDoc. Citons les principales fonctions :

- Thierry Bouche (maître de conférences à l'institut Fourier) : conseiller scientifique responsable du programme ;

¹¹ <http://www.persee.fr>

- Hélène Falavard (ingénieur documentaliste) : chef de projet chargée des relations avec les différents prestataires, de la coordination de l'équipe au sein de la Cellule MathDoc, de la création et de la maintenance du serveur Web, du contrôle qualité et de l'encadrement du personnel vacataire ;
- Isabelle Costerg (documentaliste vacataire) : chargée du contrôle qualité sur échantillon, de la préparation des collections (dépouillement, organisation des envois des collections au prestataire, ...) et du catalogage fin des articles (ajout des formules de mathématiques au format LaTeX) ;
- Catherine Barbe-Zoppis (ingénieur développement) : chargée de la mise en place des outils de contrôle des données (script de contrôle exhaustif des données, base de données des auteurs, import-export des données avant mise en ligne, ...), de l'installation et de la maintenance des serveurs de développement et Web ;
- Claude Goutorbe (ingénieur développement) : chargé du développement de l'interface de recherche et de feuilletage, de l'établissement des liens externes vers les Mathreviews, le Zentralblatt Math. et le Jahrbuch ;
- Elizabeth Cherhal (ingénieur documentaliste) : a contribué à la mise au point de la structure XML des données.

2.1.2 Moyens financiers et matériels

2.1.2.1 Soutiens financiers

Le programme a été lancé sur crédits CNRS grâce à l'attribution d'une subvention de la Direction de la Recherche en l'an 2000. Cette subvention a permis de réaliser non seulement la première phase du programme mais permettra également de réaliser la totalité de la deuxième phase (soit au total 500 000 pages numérisées).

Il a bénéficié d'aides supplémentaires :

La région Rhône-Alpes nous a attribué une subvention dans le cadre du programme BRAIN¹². Grâce à ce soutien financier, nous avons pu faire réaliser la numérisation des grands séminaires mathématiques qui se sont tenus à l'IHP, dont le séminaire Bourbaki et les cahiers du séminaire d'histoire des mathématiques. Au total, cette opération spécifique, qui a porté sur 32 000 pages, a été financée pour moitié par cette subvention.

L'université Paul Sabatier de Toulouse a financé pour partie la numérisation des *Annales de la faculté des sciences de Toulouse*. Le montant global qui nous a été alloué en 2005 a permis de financer en totalité le travail de numérisation et de création des articles par le prestataire. Rappelons que cette collection représente environ 36 000 pages et plus de 1 000 articles.

2.1.2.2 Soutiens matériels

D'un point de vue logistique, nous avons bénéficié de l'aide de nombreuses bibliothèques du RNBM qui nous ont confié leurs collections afin de les numériser :

- la bibliothèque de l'institut Henri Poincaré : les collections complètes des *Annales de l'IHP* (A, B et C), des séminaires de l'IHP ainsi que l'édition originale des *Cahiers du séminaire d'histoire des mathématiques* ;
- la bibliothèque de l'ÉNS : la collection complète des *Annales scientifiques de l'ÉNS* ;

¹² Bibliothèque Rhône-Alpes d'Informations Numériques : programme régional destiné à financer la numérisation du patrimoine local.

- la bibliothèque de l'IRMA à Strasbourg : la collection complète du *Séminaire de probabilités de Strasbourg* ;
- la bibliothèque de l'institut Fourier nous prête régulièrement des volumes pour compléter les collections reçues d'autres bibliothèques ;
- la bibliothèque du laboratoire Jean Leray (Nantes) nous a prêté une collection complète des actes de Journées ÉDP.

Nous avons également conclu un accord de partenariat avec la bibliothèque municipale Carré d'art de Nîmes qui nous a confié le temps de la numérisation la collection complète des *Annales de Gergonne*.

Nous avons également pu compter sur le soutien matériel de la Société mathématique de France, de l'IHÉS et des Annales de l'institut Fourier qui nous ont prêté une collection complète de leurs publications.

2.2 Choix techniques et stratégiques

2.2.1 Les spécifications

Les choix techniques que nous avons effectués au début du projet sont restés les mêmes, évoluant uniquement au niveau du catalogage des bibliographies. Nous avons donc choisi :

- une qualité de capture des images de 600 dpi monochrome, désormais considéré comme un standard international ;
- des formats de fichiers destinés à l'archivage (TIFF) et à la diffusion sur le Web (PDF et DjVu) ;
- un format de catalogage ouvert avec un niveau de détail élevé et intègre (XML/Unicode).

Avec l'expérience du premier marché, nous avons en effet choisi de faire évoluer le niveau de catalogage des bibliographies, passant d'un catalogage de bas niveau (seules les informations sur les auteurs, les années de publications et le titre étaient balisées) à un catalogage plus fin qui prend en compte tous les types d'information (éditeurs, pagination, revue, tome, fascicule, ...). Ce choix a été dicté par l'amélioration du travail de liaison vers les bases de données MathSciNet, Zentralblatt et Jahrbuch.

2.2.2 Le créneau mobile

La notion de créneau mobile pour régir la diffusion du texte intégral des articles a été mise en place. Le principe du créneau mobile est de définir un délai durant lequel les articles récents ne sont pas librement accessibles sur le serveur www.numdam.org. Ce délai est défini par et pour chacune des revues et varie entre 0 et 10 ans. Actuellement, sa valeur est majoritairement de 5 ans.

Revue	Créneau mobile	Disponibilité des données bibliographiques	Disponibilité du texte intégral
Annales de la faculté des sciences de Toulouse	5 ans	Années 1887 à 2002 incluse	Années 1887 à 1999 incluse
Annales de l'institut Fourier	5 ans	Années 1949 à 2000 incluse	Années 1949 à 1999 incluse
Annales de l'institut Henri Poincaré		Années 1930 à 1964 incluse	Années 1930 à 1964 incluse
Annales de l'institut Henri Poincaré, section B, Probabilités et Statistiques	5 ans	Années 1964 à 2000 incluse	Années 1964 à 1999 incluse
Annales de l'institut Henri Poincaré, section C, Analyse non linéaire	5 ans	Années 1984 à 2000 incluse	Mai 1984 à 1999 incluse
Annales mathématiques Blaise Pascal	0	Années 1994 à 2002 incluse	Années 1994 à 2002 incluse
Annales scientifiques de l'École normale supérieure	5 ans	Années 1864 à 2000 incluse	Années 1864 à 1999 incluse
Annales de l'université de Grenoble		Années 1945 à 1948 incluse	Années 1945 à 1948 incluse
Bulletin de la SMF	5 ans	Années 1872 à 2000 incluse	Années 1872 à 1999 incluse
Journées Équations aux dérivées partielles	0	Années 1974 à 2000 incluse	Années 1974 à 2003 incluse
Mémoires de la SMF	10 ans	Années 1964 à 2000 incluse	Années 1964 à 1994 incluse
Publications mathématiques de l'IHÉS	5 ans	Années 1959 à 2000 incluse	Années 1959 à 1999 incluse
Séminaire de Probabilités de Strasbourg	5 ans	Années 1967 à 2002 incluse	Années 1967 à 1999 incluse

2.3. Problèmes juridiques

Dès l'origine du programme NUMDAM, les aspects juridiques liés à la numérisation de revues ont été une préoccupation centrale. Nous avons d'ailleurs pris conseil auprès d'un expert juridique spécialisé dans le droit d'auteur qui nous a proposé différents modèles de contrats de cession de droits. Rappelons que le droit d'auteur est protégé par des conventions internationales, européennes et par le droit français. Un auteur est titulaire des droits (patrimoniaux) d'exploitation de son œuvre, dès l'instant où il l'a produite, pour une période courant jusqu'à 70 ans après sa mort. En règle générale, il ne peut céder ses droits qu'explicitement ; en conséquence, un auteur des années 1910 (ou ses héritiers) dispose des droits pour toute exploitation relevant des technologies de l'information.

Dans le cas des revues toujours vivantes, nous avons opté pour la solution suivante : nous demandons aux auteurs de céder les droits liés à la publication électronique à leur éditeur et concluons un accord de diffusion avec l'éditeur. La Cellule MathDoc sert de lien entre les auteurs et l'éditeur de la revue. Le cas des revues dites mortes n'est pas à ce jour réglé.

Pour faciliter la tâche des auteurs et des éditeurs, nous avons mis au point une opération de mailing avec formulaire pré-imprimé de cession de droits à retourner, signé, à la revue.

2.4. Actions de promotion

Depuis les débuts du programme NUMDAM, nous avons cherché à le faire connaître auprès des différentes communautés concernées, que ce soient les chercheurs en mathématiques, en histoire des sciences, les documentalistes. Nous nous sommes également concertés dans la mesure de nos moyens avec d'autres projets de numérisation à l'échelon international, notamment dans le domaine des mathématiques (DML¹³, WDML¹⁴, DML-EU¹⁵).

¹³ *Digital Mathematical library*, planning project NSF n° DUE-0206640, cf. <http://www.library.cornell.edu/dmlib/>.

¹⁴ *World digital library, projet de l'Union mathématique internationale*, cf. <http://www.wdml.org/>.

¹⁵ Portail européen pour une bibliothèque numérique universelle de mathématiques : projet soumis au FP6 en 2003 (bien évalué mais non retenu), nouveau projet soumis en octobre 2005 pour le programme eContentPlus.

Nous avons donc participé à de nombreuses conférences ou réunions.

Plus spécifiquement, nous avons organisé quelques actions de promotion :

1. École du RNBM (octobre 2004) : à l'occasion de cette semaine de formation des documentalistes des bibliothèques de mathématique, nous avons réalisé un dépliant destiné à faire connaître aux utilisateurs des bibliothèques de mathématiques les différentes revues disponibles en ligne à cette date. Ce carton a été distribué à l'ensemble des participants.
2. Newsletter de la Société européenne de mathématiques (juillet 2004) : une plaquette de présentation du programme NUMDAM a été réalisée, en anglais, pour insertion dans le numéro 53 de la newsletter de la SME. Un tirage de 3 000 exemplaires a été effectué pour cette occasion.
3. Conférence « Linking and searching collections of retrodigitised mathematical journals » (novembre 2003) : un atelier a été organisée par la Cellule MathDoc en marge de la conférence DocEng 2003 (Symposium on Document Engineering). À cette occasion, différentes présentations du programme NUMDAM sur le plan technique ont été effectuées.
4. Les journées de Nantes (janvier 2003) : à l'occasion de l'ouverture au public du serveur NUMDAM, une conférence « JEDP : un quart de siècle numérisé » a été organisé à Nantes par Laurent Guillopé en collaboration avec le laboratoire Jean Leray (UMR 6629 UN/CNRS) et le GDR Analyse des équations aux dérivées partielles (GDR 2434 CNRS).

Outre les actions décrites ci-dessus, nous informons régulièrement la communauté des mathématiciens et les bibliothécaires du RNBM de l'évolution et de l'avancement de NUMDAM grâce à l'envoi trois fois par an d'une lettre d'informations et grâce à l'envoi de mails aux listes de diffusion bibli@rnbm.org et tous@rnbm.org.

3. Bilan des réalisations (années 2001-2005)

3.1 Chronologie du programme

De longs mois de test et de préparation ont été nécessaires avant la première mise en ligne de données sur le serveur www.numdam.org, que ce soit pour le premier marché contracté avec la société AIS, ou le second avec la société Jouve. La chronologie ci-dessous permet de situer les grandes étapes de vie de ce programme depuis son lancement en juin 2000.

- de juin 2000 à décembre 2001 : rédaction du cahier des charges, préparation des collections, choix du prestataire ;
- de janvier 2002 à décembre 2002 : écriture des spécifications techniques, réalisation du banc d'essai, lancement de la production ;
- de janvier 2003 à septembre 2003 :
 - o suite et fin du premier marché : poursuite et finalisation de la production, affichage successif des six premières collections sur le serveur www.numdam.org ;
 - o démarrage du second marché : modification du cahier des charges, choix du nouveau prestataire, préparation des premières collections ;
- de octobre 2003 à avril 2004 : établissement d'un planning de production, réalisation de lots tests, finalisation de la chaîne de production ;

- de mai 2004 à décembre 2004 : lancement de la nouvelle production, premier affichage de nouvelles revues, rédaction du cahier des charges pour l'opération BRAIN (Bibliothèque Rhône-Alpes d'informations numériques, hors marché pour cause de source de financement différente), choix du prestataire ;
- de janvier 2005 à décembre 2005 : poursuite de la production, affichage successif des collections sur le serveur www.numdam.org.

3.2 La première phase : 2001-2003

Le bilan de la première phase est de six revues entièrement numérisées et mises en ligne sur le serveur Web (Annales de l'institut Fourier, Journées ÉDP, Publications mathématiques de l'IHÉS, Bulletin de la SMF, Mémoires de la SMF, Annales Scientifiques de l'ÉNS) pour un total d'environ 200 000 pages et 8 000 articles.

Titre	Volumes	Articles	Pages
<i>Annales de l'institut Fourier</i> (créée en 1949 - Association des AIF)	156	1 810	51 054
<i>Annales scientifiques de l'ÉNS</i> (créée en 1864 - ENS sous contrat avec Elsevier)	295	1 864	68 442
<i>Bulletin de la SMF</i> (créée en 1872 - Société Mathématique de France)	167	2 608	45 774
<i>Journées ÉDP</i> (créée en 1974 - CNRS)	28	469	5 580
<i>Mémoires de la SMF</i> (créée en 1964 - Société Mathématique de France)	134	396	18 118
<i>Publications mathématiques de l'IHÉS</i> (créée en 1959 - IHÉS sous contrat avec Springer)	92	344	17 424
Total de la première phase	872	7 491	206 392

3.3 La seconde phase : 2003-2006

Le bilan de la deuxième phase se décline en 3 parties distinctes : (1) les revues numérisées et mises en ligne, (2) les revues numérisées et en cours de traitement pour mise en ligne, et (3) les revues dont la numérisation a été planifiée pour les prochains mois.

(1) À ce jour, nous avons procédé à la mise en ligne de sept revues supplémentaires (*Annales de la faculté des sciences de Toulouse*, *Annales mathématiques Blaise Pascal*, *Annales de l'institut Henri Poincaré*, *Annales de l'institut Henri Poincaré, section (B) Probabilités et statistiques et section (C)*, *Analyse non linéaire*, *Annales de l'université de Grenoble et Séminaire de probabilités de Strasbourg*) pour un total d'environ 93 000 pages et 3 900 articles.

Titre	Volumes	Articles	Pages
<i>Annales de la faculté des sciences de Toulouse</i> (créée en 1887 - Université de Toulouse)	207	1 034	36 051
<i>Annales mathématiques Blaise Pascal</i> (créée en 1994 - Université de Clermont)	19	165	2 554
<i>Annales de l'institut Henri Poincaré</i> (1930-1964 - IHP/Gauthier-Villars)	71	145	5 860
<i>Annales de l'institut Henri Poincaré, section B</i> (créée en 1964 -)	155	834	19 010
<i>Annales de l'institut Henri Poincaré (section C)</i> (créée en 1984 - IHP/Gauthier-Villars)	102	392	10 987
<i>Annales de l'université de Grenoble</i> (1945-1948 - Université de Grenoble)	3	47	1 006
<i>Séminaire de probabilités de Strasbourg</i> (créée en 1967 - Lecture notes, Springer)	37	1 255	17 352
Total de la seconde phase	687	3 872	92 820

(2) Les collections en cours de traitement représentent un total de 110 000 pages et 7 500 articles : *Annales de l'institut Henri Poincaré, section A, Annales de Gergonne, Revue de statistiques appliquées et les séminaires de l'IHP : Bourbaki, Chevalley, M.-P. Malliavin, Schwartz, Sophus Lie, Lelong, Louis de Broglie, Henri Cartan, Delange-Pisot-Poitou, Choquet, Dubreil, Paul Krée, Ehresmann, Grothendieck, Samuel, etc., Cahiers du Séminaire d'Histoire des Maths.* Leur mise en ligne est programmée pour l'année 2006 et devrait se faire de manière régulière sur toute l'année.

Titre	Volumes	Articles	Pages
<i>Annales de l'institut Henri Poincaré (section A)</i> (créée en 1964 – IHP/Gauthier-Villars)	287	1 371	31 482
<i>Séminaires de l'institut Henri Poincaré</i>		3 328	47 165
<i>Annales de Gergonne</i> (1810-1832)	22	1 083	8 701
<i>Revue de Statistiques Appliquées</i> (créée en 1953 - SFdS)	207	1 676	22 150
Total		7 458	109 498

(3) Le planning de l'année 2006 comporte quatre nouvelles revues dont deux sont des revues européennes de haut niveau (*Compositio Mathematica, Annales de l'école normale supérieure de Pise, Annales de Clermont et Mathématiques et sciences humaines*).

Une fois ce planning réalisé, le serveur Web www.numdam.org proposera quelque 500 000 pages et plus de 30 000 articles à ses utilisateurs.

4. Le processus de mise en ligne des revues

La mise en ligne d'une revue répond à une procédure finalement assez complexe.

4.1 Le contrôle des données

Un travail important dans ce processus de mise en ligne consiste à contrôler le plus efficacement possible les données issues de la numérisation. Nous avons adopté un double contrôle : l'un est exhaustif, l'autre sur échantillon.

- le contrôle exhaustif : une série d'analyses automatiques sont effectuées sur la totalité des fichiers livrés de manière à repérer rapidement toutes les erreurs éventuelles. Ce contrôle a été élaboré lors du premier marché, pour la production de la société AIS. Il a depuis subi différentes évolutions techniques et structurelles de manière à cadrer avec la production de la société Jouve.
- Le contrôle par échantillon : ce contrôle permet de vérifier finement la qualité visuelle des fichiers ainsi que la qualité des bibliographies. L'interface Web utilisée pour cette étape permet de garder trace de tous les fichiers contrôlés et des erreurs détectées.

La taille des échantillons de données ainsi que le nombre d'erreurs sont déterminés selon la norme NF X 06-021 (principes du contrôle statistique de lots). Généralement, le taux d'acceptation d'un lot est de 4 erreurs pour 315 pages contrôlées.

Lors du premier marché, nous n'avons pas été amenés à rejeter un lot de production (seule une réfection a été demandée). Cela n'a en revanche pas été le cas pour le deuxième marché, même si cette situation reste marginale.

À l'issue de ces deux contrôles, une synthèse est réalisée et est envoyée au prestataire sous forme de recette. Nous sommes dans l'obligation contractuelle de réaliser le contrôle sous

4 semaines. La réfection des fichiers est généralement demandée sous huit à quinze jours suivant la complexité du travail.

Nous avons pu établir une typologie des erreurs constatées dans ces deux marchés. Ces erreurs sont spécifiques à chacun des deux prestataires dans la mesure où ils ont développé leur propre méthode de production.

- Premier marché : l'erreur la plus fréquente concerne les fichiers DjVu. Ceux-ci étaient livrés incomplets, voire corrompus (c'est-à-dire ne possédant pas les caractéristiques du format DjVu). Les autres erreurs concernent essentiellement les fichiers monopages scannés à livre ouvert et présentant certaines déformations du texte.
- Second marché : les erreurs que nous avons détectées sont plus variées, plus systématiques et concernent tous les types de fichiers. Nous rencontrons régulièrement des articles incomplets ou avec des pages en trop, des fichiers PDF ou DjVu sans texte caché, des fichiers monopages avec des traces de doigts, des couvertures non détournées. Les traitements spécifiques sur les hors-textes ne sont pas toujours performants.

Précisons que, lors de ce second marché, la chaîne de production de la société Jouve a plusieurs fois été remaniée d'où l'apparition régulière de nouvelles erreurs et la réapparition d'erreurs précédemment corrigées. En effet, le remaniement n'intègre pas toujours l'ensemble des corrections apportées aux versions précédentes de la chaîne.

4.2. Le traitement complémentaire des données

Une fois une revue validée par nos soins, nous réalisons un ensemble de traitements complémentaires destinés à enrichir les données. Ces traitements sont de différentes sortes et portent (1) sur les métadonnées et (2) sur les fichiers utilisateurs (fichiers PDF et DjVu).

(1) les métadonnées

À travers l'import-export des données XML dans une base de données MySQL, les données sont enrichies avec les formules mathématiques au format Latex dans les titres, avec des relations de type erratum, addendum ou corrigendum entre les articles, avec la correction des noms d'auteurs et langues de catalogage.

Des liens internes et externes dans les articles et les bibliographies sont ajoutés par identification des articles dans d'autres bases de données (MathSciNet, Zentralblatt, etc.).

(2) les fichiers utilisateurs

L'export des métadonnées au format XML est complété par un export TeX des informations bibliographiques de chaque article. Cet export permet de générer les pages de garde qui sont systématiquement ajoutées aux articles et qui mentionnent, outre des informations juridiques et légales, la référence bibliographique de l'article et un lien vers la notice complète sur le serveur www.numdam.org.

Exemple de page de garde des Publications mathématiques de l'IHES :



4.3. La mise en ligne

Cette dernière étape consiste d'une part, à mettre à jour les pages dites statiques du site Web (page d'accueil et informations sur le créneau mobile en français et en anglais) et d'autre part à exploiter les métadonnées de la revue pour permettre le feuilletage et les recherches dans la nouvelle collection.

La partie critique du travail concerne l'indexation et l'exploitation des liens. Ce travail est effectué grâce au logiciel EDBM, initialement développé pour l'interface Web du Zentralblatt Math.

- **la génération du feuilletage** : le feuilletage volume par volume consiste en la reconstitution virtuelle de la table des matières de chaque volume avec des informations complémentaires telles que l'accès à la notice complète le cas échéant (résumé et bibliographie), aux articles et aux analyses MathReviews, Zentralblatt Math et Jahrbuch.

Exemple de feuilletage pour les Annales scientifiques de l'ÉNS :

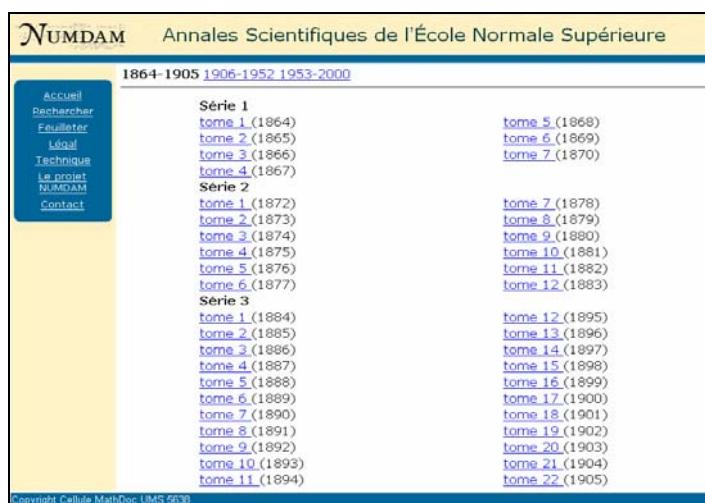


Figure 1 : feuilletage des Annales scientifiques de l'ÉNS

NUMDAM Annales Scientifiques de l'École Normale Supérieure (3) tome 11 (1894)	
Précédent: tome 10 (3) (1893) Suivant: tome 12 (3) (1895) Retour: liste des volumes	
Accueil Rechercher Feuilleter Légal Technique Le projet NUMDAM Contact	Elliot Sur les cas d'intégrabilité du mouvement d'un point dans un plan p. 9-22 Texte intégral dhu pdf
	Weill, G. Sur les substitutions orthogonales à déterminant $\epsilon(-1)$ p. 23-35 Texte intégral dhu pdf Analyses JFM 25.1685.01
	Kluwyer, J. C. L'évaluation des intégrales et des fonctions elliptiques au moyen de la transformation du second degré p. 37-73 Texte intégral dhu pdf Analyses JFM 25.0776.01
Bibliothèque PDF et Doc	Cohen, E. Sur la fonction $\zeta(s)$ de Riemann et sur des fonctions analogues p. 75-164 Texte intégral dhu pdf Analyses JFM 25.0702.01
	Greenhill, A. G. Les modules dans la multiplication complexe des fonctions elliptiques p. 165-248 Texte intégral dhu pdf Analyses JFM 25.0785.02
	Grévy, A. Étude sur les équations fonctionnelles p. 249-323 Texte intégral dhu pdf Analyses JFM 25.0671.02
	Maltezos, C. Les enveloppes solides minces, les doches p. 325-375 Texte intégral dhu pdf Analyses JFM 25.1295.02
	Gomez, Désiré Notice sur Henri Sainte-Claire Deville p. 3-70 (supplément) Texte intégral dhu pdf
Précédent: tome 10 (3) (1893) Suivant: tome 12 (3) (1895) Retour: liste des volumes	
Copyright Cellule MathDoc UMS 5638	

Figure 2 : table des matières du tome 11, fascicule 3 de l'année 1894

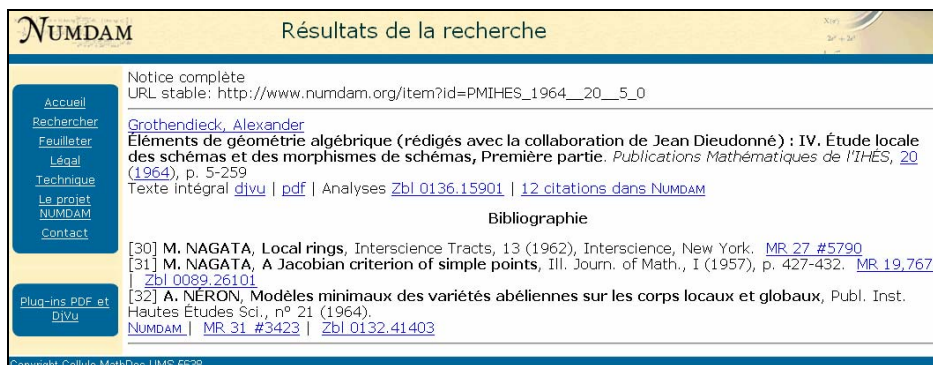
Une interface de recherche avancée offre également un accès multi-revues et multi-critères. Les champs suivants sont interrogeables : journal, auteur, mots du titre, années et plein-texte. L'interface propose également la possibilité d'affiner la recherche en utilisant les bibliographies. Il est alors possible de combiner une recherche sur un auteur qui en cite un autre, ce qui ouvre des perspectives intéressantes lors de recherches bibliographiques.

NUMDAM Interface d'interrogation et d'accès aux documents	
Guide pour la recherche	
Accueil Rechercher Feuilleter Légal Technique Le projet NUMDAM Contact	Rechercher un article <i>Au moins l'un de journal, auteur, titre, plein texte doit être rempli</i>
	Journal: <input type="text" value="Tous"/>
	Auteur(s): <input type="text" value="Ex: Brelot, M; Choquet"/> Index
	Mots du titre: <input type="text"/>
	Années: <input type="text" value="entre"/> et <input type="text"/>
	Plein-texte: <input type="text"/>
	Les expressions ci-dessus apparaissent sur la même page <input type="checkbox"/>
	Tri par nom d'auteur <input type="checkbox"/>
	Affiner la recherche en utilisant les bibliographies
	Auteur(s): <input type="text"/> Index
Mots du titre: <input type="text"/>	
N'importe où: <input type="text"/>	
Années: <input type="text" value="entre"/> et <input type="text"/>	
<input type="button" value="Rechercher"/> <input type="button" value="Annuler"/>	
Copyright Cellule MathDoc UMS 5638	

Figure 3 : interface de recherche avancée

- **l'ajout de liens internes et externes** : au niveau des articles et de chaque référence bibliographique, un travail d'identification a été accompli vers les bases de données Zentralblatt Math, MathReviews et Jahrbuch rendant ainsi accessible via un lien les analyses correspondantes dans ces trois bases de données. Dans les bibliographies, des liens vers les articles présents dans NUMDAM sont aussi ajoutés.

Exemple de notice complète avec les analyses MR et ZM et les liens vers NUMDAM :



The screenshot shows a web page titled "NUMDAM Résultats de la recherche". On the left is a navigation menu with links: Accueil, Rechercher, Feuilletter, Légal, Technique, Le projet NUMDAM, and Contact. Below the menu are links for "Plug-ins PDF et DjVu". The main content area displays a search result for "Grothendieck, Alexander" with the title "Éléments de géométrie algébrique (rédigés avec la collaboration de Jean Dieudonné) : IV. Étude locale des schémas et des morphismes de schémas, Première partie. Publications Mathématiques de l'IHÉS, 20 (1964), p. 5-259". It includes a stable URL, a full-text link, and a bibliography section with three entries: [30] M. NAGATA, Local rings; [31] M. NAGATA, A Jacobian criterion of simple points; and [32] A. NÉRON, Modèles minimaux des variétés abéliennes sur les corps locaux et globaux.

5. Développements spécifiques

L'évolution de la fréquentation du serveur NUMDAM et sa visibilité au niveau national et international nous ont amené à réaliser des développements spécifiques. Les objectifs de ces développements sont :

- la diffusion plus large des données ;
- une amélioration des fonctionnalités de recherche ;
- un partenariat privilégié avec les revues.

5.1 Diffusion large des données

Les développements correspondants ont été de deux ordres : la mise en place d'un serveur OAI et la mise en place d'une interface spécifique permettant l'indexation fine des articles par le moteur de recherche Google.

5.1.1 Mise en place d'un serveur OAI

Dans le but d'assurer une meilleure visibilité du service NUMDAM, nous avons donc développé et installé un serveur d'archives ouvertes. Ce serveur expose les métadonnées des articles numérisés, à travers le protocole OAI-PMH.

Ces données sont périodiquement « ramassées » par plusieurs sites intérateurs de données, et réexposées par ces services à travers leurs propres interfaces de recherche, qui fonctionnent ainsi comme autant de points d'accès aux articles numérisés. En particulier, la base de données MathSciNet a pu s'enrichir de références anciennes en provenance de NUMDAM grâce à ce service.

5.1.2 Indexation fine des articles par Google

Lors d'échanges avec le moteur de recherche Google, nous avons été sollicités pour réaliser un système d'indexation fine du texte intégral des articles. Jusqu'alors, le texte intégral des articles n'était pas indexé par les moteurs de recherche du fait de la protection appliquée sur les fichiers PDF.

L'idée est d'inclure le texte complet des articles dans leur notice complète lorsque cette notice est « consultée » par un moteur de recherche, texte qui est donc indexé. Ce travail, réalisé à la demande de Google, est aussi utilisé par d'autres moteurs tels que Yahoo.

Il est à noter que l'utilisation de NUMDAM s'est accrue considérablement dès lors que cette indexation a été mise en place. En fait, l'accès à NUMDAM à partir d'une recherche sur Google a environ doublé.

5.2 Amélioration des fonctionnalités de recherche

Nous avons le souci constant d'améliorer les fonctionnalités de recherche proposées par le serveur NUMDAM. Dans cette optique, nous avons mis en place un système de recherche avancé du plein-texte qui propose une localisation des résultats page par page. En outre, nous avons mis à jour l'application SemProba.

5.2.1 Recherche plein-texte et pages

Il est désormais possible, lors d'une recherche dans le texte intégral de l'article, de spécifier que les mots ou les expressions recherchées doivent apparaître sur la même page.

La liste des articles qui satisfont aux critères de recherche inclut dans ce cas, pour chaque article, un menu présentant la liste des pages pertinentes, l'utilisateur ayant la possibilité d'afficher directement la page correspondante du fichier numérique PDF ou DjVu.

5.2.2 Application SemProba

À l'occasion de la mise en ligne du Séminaire de probabilités de Strasbourg, il a été mis en place un couplage bi-directionnel entre la collection du séminaire numérisée par Numdam, et l'index du séminaire de probabilités, une base de données produite à Strasbourg qui contient des commentaires, des critiques et des descriptions parfois informelles des exposés du séminaire.

5.3 Développement d'un partenariat privilégié avec les revues

Les accords que nous concluons avec les revues stipulent que nous devons leur fournir l'ensemble des métadonnées et les articles numérisés. Dans cette optique, un cdrom contenant l'intégralité des actes des Journées équations aux dérivées partielles a été produit à partir de la base de données NUMDAM et distribué aux participants des journées ÉDP en 2004.

6. Perspectives immédiates et futures

6.1 Conclusion de ces quatre années :

À l'issue du premier marché, notre ambition était d'accroître le volume des données accessibles depuis le serveur NUMDAM et de multiplier les modes d'accès.

Le premier objectif a été atteint grâce au lancement du deuxième marché qui nous permettra à terme de doubler le nombre de pages numérisées et tripler le nombre d'articles en ligne. Avec la plus grande part des revues académiques de mathématiques, un ensemble significatif de séminaires, et deux revues européennes, on peut considérer que nous sommes sur le point d'atteindre le but initial du programme. Parmi les revues françaises vivantes qui ne publient que des mathématiques, il ne reste guère que celles qui sont la propriété d'Elsevier (*Journal de mathématiques pures et appliquées*, *Bulletin des sciences mathématiques*) ou sous contrat avec E.D.P.S. (revues de la S.M.A.I.)

Concernant les modes d'accès, le développement de la mini-DML a permis la mise en place expérimentale d'un point d'accès unique aux données bibliographiques et articles des revues numérisées par d'autres programmes tels que Gallica (grâce à notre programme Callica-math), JSTOR, projet Euclid, ainsi qu'aux prépublications (Arxiv).

6.2 Prévisions de numérisation

Le présent marché se terminera début 2007 avec la numérisation de deux revues européennes de haut niveau (*Compositio Mathematica* et *Annales de l'école normale supérieure de Pise*). A ces deux revues s'ajoutent deux revues françaises qui sont les *Annales de l'université de Clermont* et *Mathématiques et sciences humaines* (EHESS).

Nous recevons de temps à autre des suggestions de la part d'utilisateurs du site. Pour l'instant, nous avons répondu favorablement aux demandes émanant des revues elles-mêmes, lorsqu'elles répondent à des critères raisonnables (nous avons refusé la demande d'une revue hollandaise mineure) et adoptent le modèle du créneau mobile avec une valeur acceptable. Si la mise en ligne de *Compositio* et des *Annales de Pise* suscite des demandes venant de l'étranger, il sera nécessaire de définir une politique de partenariat pour envisager des cofinancements.

Pour le domaine français, Pierre Bérard a réalisé en 2002 un inventaire très détaillé¹⁶ qui donne une idée du chemin qu'il reste à parcourir si l'on vise l'exhaustivité. Pour ce qui concerne les revues toujours vivantes, les plus importantes sont les deux revues propriété d'Elsevier; un premier accord informel d'échange de données n'a pas abouti car Elsevier semble avoir renoncé provisoirement à effectuer la numérisation et peut-être allons-nous proposer de le faire.

Par ailleurs, certaines revues défuntées plus généralistes sont encore très consultées, par exemple le *Bulletin astronomique* édité à l'Observatoire de Paris qui nous a été proposé et vivement recommandé par l'équipe de mathématiciens de cet institut. Cette revue était d'ailleurs la suite du *Bulletin général et universel des annonces et des nouvelles scientifiques* ou *Bulletin de Férussac* (1823-1831) dont la série 1 consacrée aux mathématiques et sciences connexes comporte de nombreux textes de première importance (Fourier, Sturm, Abel, Galois, etc.).

Nous avons également deux projets d'éditions en fac-similé d'œuvres complètes (Schwarz et Lichnerowicz). Il s'agit d'effectuer la numérisation de l'ensemble de l'oeuvre avec un affichage sur le site de NUMDAM et la création d'un CD-ROM, tandis qu'une partie des fichiers numérisés seraient utilisés pour faire une édition papier.

E. ÉDITION : LE PROJET CEDRAM

Centre d'Édition et Diffusion des Revues Académiques de Mathématiques

1. Enjeux : évolution de l'édition scientifique

On a observé depuis quelques années une concentration rapide des maisons d'édition commerciales qui éditent les revues scientifiques au point qu'il n'en reste plus actuellement en France que trois pour publier des revues de mathématiques. Les petites maisons d'édition françaises comme Gauthier-Villars ont été rachetées par Elsevier et Springer-Verlag a rejoint le groupe de Kluwer.

Le résultat a été une augmentation rapide du prix des revues, tout particulièrement des plus renommées, entraînant une augmentation des coûts insupportable pour les bibliothèques qui ont dû réduire leur catalogue de revues. Dans le même temps, les revues dites académiques, c'est-à-dire éditées par des laboratoires, des universités ou des sociétés savantes, ont pu maintenir des tarifs raisonnables tout en dégagant des revenus indirects par un système d'échanges. Les

¹⁶ <http://www.numdam.org/historique/Inventaire/numdam-inventaire-revues-fr-2002oct13.pdf>

Annales de l'institut Fourier par exemple, reçoivent un soutien de la part du CNRS qui met à leur disposition une secrétaire de rédaction mais en contrepartie permettent à la bibliothèque de l'institut Fourier de faire des économies importantes grâce aux revues reçues en échange des Annales. Ce système d'échanges permet également d'aider certains laboratoires étrangers.

Il paraît donc souhaitable de soutenir les revues académiques et de favoriser la prise en charge publique des revues d'origine académique éditées aujourd'hui par des éditeurs commerciaux. Pas toujours très satisfaites de la prestation de ces éditeurs commerciaux sur le plan technique, plusieurs revues envisagent cette solution. En particulier, l'une des plus prestigieuses revues françaises de mathématiques, les Annales scientifiques de l'École normale supérieure, dont le contrat avec Elsevier vient à échéance dans deux ans, envisage de changer d'éditeur et souhaite des propositions de type académique.

L'autre événement dans l'édition est l'apparition de l'édition électronique qui est un nouveau mode de diffusion mais bouleverse également l'édition elle-même et les habitudes des lecteurs. Si ce mode de diffusion n'a pas encore supplanté l'édition traditionnelle, il devient de plus en plus important et toutes les revues se doivent maintenant d'avoir leur version électronique. Les problèmes qui se posent sont de plusieurs ordres : édition, conservation, visibilité.

Le problème de l'édition, c'est-à-dire de la création du support électronique, se pose avant tout pour les petites revues qui n'ont pas toujours les moyens humains nécessaires pour mettre en place un tel service et surtout pour suivre les évolutions techniques très rapides. En outre, on peut raisonnablement considérer que la commercialisation des revues au format papier est menacée à moyen terme, et que la production d'une édition électronique de qualité est une condition impérative de développement, peut-être de survie.

Le problème de la conservation de l'édition électronique revêt des aspects multiples selon qu'il s'agisse de conservation à court ou à long terme. Dans tous les cas, c'est un problème entièrement nouveau par rapport à la conservation traditionnelle dans les bibliothèques. Cette conservation doit être organisée rigoureusement et surtout doit être évolutive donc gérée en permanence. Un fois publiée sur un support papier, une revue ou un livre n'évolue plus et il suffit de le conserver à l'abri alors qu'une publication électronique doit pouvoir être reprise pour s'adapter à de nouveaux formats ou tirer parti d'un nouveau mode de diffusion.

Devant le foisonnement d'Internet, la visibilité devient de plus en plus difficile pour un serveur Web isolé. Comme l'ont bien compris les éditeurs commerciaux, les avantages que présente l'édition électronique en ce qui concerne la recherche d'articles, les liens entre articles, etc., sont multipliés par la présence sur un même serveur du plus grand nombre de revues. Cependant, ces grands éditeurs commerciaux pratiquent une politique de vente groupée de toutes leurs revues à des consortiums très larges, ce qui risque de ne laisser ni place ni financement pour les revues indépendantes.

Actuellement, les véritables portails d'accès à la littérature mathématique vivante sont les grandes bases de données MathSciNet et Zentralblatt (à quoi il faut peut-être ajouter l'incontournable Google ...). Pour qu'une revue se développe, il faut qu'elle y soit référencée, et que ses articles y soient atteignables d'un clic. Ce qui différencie un portail spécialisé d'un site isolé, c'est l'étendue des outils de navigation fournis. Un portail d'accès aux revues favorise la diversité des modes d'accès, en indexant la totalité des articles publiés et leurs bibliographies, transformant une masse de textes isolés en un réseau vivant. De la sorte, l'accès à un article devient une publicité pour tous les articles liés, que ce soit parce qu'ils y sont cités, parce qu'ils le

citent, ou parce qu'ils partagent un auteur. Un exemple à suivre du point de vue de la qualité du référencement et de l'hypertexte est le projet Euclid de la bibliothèque universitaire de Cornell (USA). En dehors de l'efficacité de ces modes de promotion et de diffusion, un tel portail implique plus largement la création d'un pôle fédérateur où les modes de production et d'archivage seront aussi améliorés de par la mise à dispositions de services et d'outils particulièrement adaptés aux exigences des revues de mathématiques.

2. Le projet : créer un pôle pour les revues de mathématiques

Après plusieurs réunions préparatoires réunissant l'ensemble des responsables de revues françaises et les représentants du ministère de la Recherche et du CNRS pour les mathématiques, il a été décidé de créer à Grenoble un pôle d'édition et de diffusion pour les revues de mathématiques. Ce pôle est organisé au sein de la Cellule MathDoc en collaboration avec les Annales de l'institut Fourier, revue de mathématiques de haut niveau éditée à Grenoble. Il doit bénéficier de l'expérience en édition électronique acquise par la Cellule MathDoc au cours de la réalisation du programme de numérisation NUMDAM et de l'expérience en édition traditionnelle et diffusion des Annales de l'institut Fourier.

Le pôle s'adresse en premier lieu aux revues françaises de mathématiques soutenues par le CNRS. Au-delà, le projet pourrait être amené à se développer dans diverses directions : en mathématiques vers l'échelon européen ; en France vers d'autres revues composées en LaTeX.

Par ailleurs, le serveur du pôle inclura les données de NUMDAM pour offrir aux utilisateurs sous une même interface les documents anciens et récents. Un moteur de recherche commun permettra d'interroger tous les articles diffusés par la Cellule MathDoc.

L'ambition de ce projet est d'augmenter la visibilité et l'impact (édition électronique sur un serveur commun ; gestion, promotion et échanges des abonnements papier) des revues. Pour y répondre, ce pôle doit être complet, c'est-à-dire aborder tous les aspects de l'édition, et flexible c'est-à-dire qu'il doit offrir des services modulaires que chaque revue doit pouvoir choisir à la carte. Il est donc prévu de distinguer les modules suivants :

- aide à l'édition (outils de gestion du secrétariat et des flux éditoriaux) ;
- formats (formats LaTeX, standards de métadonnées) ;
- édition (mise aux normes, correction) ;
- diffusion électronique (serveur web avec contrôle des accès pour les articles récents) ;
- archivage (de tous les éléments de production et de gestion des revues, et de la production sur le long terme) ;
- diffusion papier (impression, routage) ;
- promotion et gestion des abonnements.

Le niveau d'intégration des revues dans le pôle pourra varier en fonction de nombreux paramètres. Il est clair que les revues doivent garder la maîtrise scientifique de leur contenu, c'est-à-dire qu'elles conserveront leur indépendance éditoriale (comité de rédaction, secrétariat, relations avec les auteurs et les rapporteurs), et collaboreront au sein du pôle pour définir la mise aux normes des articles et la chaîne de production qui aboutira à la publication électronique et papier. Les modèles économiques seront divers, ils dépendront de l'ampleur des tâches prises en charge par le pôle et nécessiteront la signature de conventions entre la Cellule MathDoc et les revues.

3. Avancement du projet

Le projet a été véritablement lancé en mars 2004 par la décision de créer un poste d'ingénieur d'études affecté à la Cellule MathDoc. Le poste a été pourvu en février 2005, date de la prise de fonctions de Laure Heigéas. Quelques mois ont été nécessaires pour qu'elle se forme aux technologies utilisées à la Cellule MathDoc, et pour prendre la mesure du projet. C'est finalement au cours de l'été 2005, après la visite de Claude Sabbah à Grenoble, que les premiers travaux concrets ont été entrepris en vue de l'affichage des trois premières revues début 2006.

3.1 Organisation

Le personnel du projet est donc constitué d'un poste d'ingénieur d'études dédié pour les développements spécifiques et la maîtrise d'oeuvre, avec l'appui des autres personnels de la Cellule MathDoc, notamment l'équipe de NUMDAM, pour l'expertise acquise dans la gestion des métadonnées et la mise en ligne des articles.

Un directeur du projet est nommé. Son rôle est de diriger la conduite du projet, c'est-à-dire de faire la synthèse des souhaits des parties prenantes, et d'arrêter un calendrier et des objectifs réalistes.

Par ailleurs, un comité de pilotage a été mis en place. Il est présidé par Claude Sabbah, ancien responsable des publications de la Société mathématique de France et promoteur du projet depuis son origine, et s'est réuni pour la première fois en octobre 2005. Le comité de pilotage est constitué de représentants des tutelles (CNRS, MEN, UJF), des revues participant effectivement au pôle, du directeur de la Cellule MathDoc, assisté par le directeur du pôle. Le comité de pilotage est un organe consultatif qui est tenu informé par le directeur du pôle. Il permet la concertation des différents partenaires et propose les grandes orientations. Il évalue les priorités et le calendrier d'exécution des tâches effectuées pour le pôle. Il se réunit si nécessaire tous les six mois et au minimum une fois par an, il peut également débattre par le truchement d'une liste de diffusion d'accès fermé mais archivée. Le comité de pilotage pourra également exercer une fonction prospective, par des relations avec des revues susceptibles d'adhérer au pôle, mais aussi en suggérant des développements pour que l'offre du pôle reste concurrentielle.

Lors du premier comité de pilotage, Thierry Bouche a été nommé directeur du pôle d'édition, un calendrier a été établi pour les premiers mois de l'année 2006 et le pôle a été baptisé : CEDRAM. Ainsi, l'URL du portail est www.cedram.org et le site de chaque revue associée au pôle peut avoir comme URL revue.cedram.org.

Les premières revues qui vont intégrer CEDRAM sont les suivantes :

- *Annales de l'institut Fourier* (AIF dans la suite) ;
- *Annales mathématiques Blaise-Pascal* (AMBP) ;
- *Annales de la faculté des sciences de Toulouse* (AFST).

La SMF et la SMAI participent également au comité de pilotage en tant qu'émanations de la communauté mathématique française et comme éditeurs de revues.

3.2 Problématique

Parmi les différents modules envisagés plus haut, l'effort initial s'est porté sur l'édition électronique des trois premières revues intégrées au pôle. Pour deux de ces revues, l'édition électronique est conçue comme un bonus réservé aux abonnés papier tandis que pour la troisième

elle est en accès libre. Pour ces revues, l'intervention du CEDRAM peut donc être vue comme une subvention CNRS-UJF en vue de les doter d'une édition électronique concurrentielle du meilleur niveau.

Prendre en charge l'édition électronique d'une revue de mathématiques demande de nombreuses tâches qui font appel à des compétences très diversifiées. Ces tâches ont été menées en parallèle dans la mesure du possible, de façon à accélérer le processus.

En premier lieu, il faut que chaque revue dispose d'un serveur web qui reproduise fidèlement son image et assure son identité : cela demande le développement d'une architecture propre et d'une charte graphique, mais aussi une bonne entente avec l'administration de la revue et la définition des modes d'interaction (le contenu textuel du site reste sous la seule responsabilité du comité de rédaction de la revue, qui doit donc pouvoir le modifier, par exemple). Ces pages statiques doivent néanmoins pouvoir abriter les pages dynamiques qui permettent aux utilisateurs d'accéder aux articles eux-mêmes (sommaires, résultats de recherches, métadonnées des articles).

Pour alimenter ces pages dynamiques, le point le plus critique est certainement la définition des formats de métadonnées, puisque c'est ce qui détermine les fonctions possibles du serveur : il faut à la fois une structure très riche et détaillée de façon à offrir une interface puissante, mais il faut éviter de créer une usine à gaz trop coûteuse à produire et impossible à exploiter. L'idée qu'on ne pouvait pas faire moins que NUMDAM a semble-t-il été l'objet d'un consensus : c'est déjà beaucoup demander !

Il faut alors établir des normes de fabrication qui assurent la production de ces métadonnées, synchronisée avec celle des articles composés. Comme il ne s'agit pas de créer ex nihilo des revues nouvelles, il faut prendre en compte l'existant et parfois s'y adapter entièrement.

En outre, pour offrir le meilleur service aux utilisateurs, il faut fondre dans une seule interface des données hétérogènes : archives numérisées dans le cadre de NUMDAM, production « nativement numérique » orientée papier des années qui ont suivi, et enfin les produits de la nouvelle chaîne définie par le pôle.

3.3 État des lieux

3.3.1 Production des articles

L'objectif poursuivi a été la définition d'une chaîne de production évoluée automatisant tout ce qui peut l'être, de façon à éviter les erreurs manuelles et les duplications d'informations. Il est apparu très rapidement que, si l'on pouvait alimenter assez facilement un site de publication électronique dès lors que l'on dispose du texte des articles (PDF) et des métadonnées correspondantes, il n'est pas simple de synchroniser ces éléments si l'on ne dispose pas d'une source unique faisant autorité. Par pragmatisme, du fait que l'objet qui fait foi dans le domaine scientifique est l'article imprimé, il a été décidé de privilégier le fichier latex et sa bibliographie bibtex pour produire toutes les données d'exploitation.

Un système de production automatisée de revue en latex a été écrit pendant l'été 2005 par Thierry Bouche. Ce système permet de reproduire fidèlement la maquette de plusieurs revues à partir d'une source normalisée, et de produire les métadonnées exactes sans aucune duplication des informations. Les numéros de page et les sommaires sont produits à la volée. On dispose en sortie de tous les éléments nécessaires pour l'impression et la mise en ligne.

Ce système est entièrement opérationnel pour les *Annales de l'institut Fourier* : il servira à produire le premier fascicule du tome 56 (2006). Il fonctionne pour les deux autres revues, mais n'a pas encore été testé en grandeur nature. Une déclinaison simplifiée du système a permis de récupérer la production post-NUMDAM des AMBP (2003-2005). Les AFST ont un contrat de sous-traitance pour l'année 2006 et ne pourront donc pas basculer avant 2007.

De façon à ne pas avoir de défaut de couverture dans l'histoire de la revue, un effort particulier a été consacré à la récupération des articles nativement numériques produits dans l'intervalle entre la numérisation par NUMDAM et la production future selon les nouvelles normes CEDRAM. Yves Laurent et Laure Heigéas ont testé pour les AIF (2001-2005) des solutions qui s'avèrent assez coûteuses tant en développement qu'en contrôle humain. En effet, le peu de normalisation des documents reçus en entrée ne permet pas d'automatiser efficacement la récupération des données et exige un contrôle systématique de toutes les données, celles-ci étant pour la plupart à reprendre à la main. Dans certains cas, il sera peut-être nécessaire de numériser des volumes récents dans NUMDAM s'ils ne peuvent pas être facilement intégrés dans le serveur, cela même si le PDF de l'article est exploitable. Cette dernière solution est actuellement testée pour les AFST (2003-2005).

Les AIF ont organisé une journée de formation à ces nouveaux outils à laquelle ont assisté les sous-traitants actuels de la revue, ainsi que deux autres sociétés candidates qui pourront offrir leurs services aux autres revues.

3.3.2 Exploitation des données articles

Laure Heigéas a fixé les formats pivot XML (DTD : définition de la structure d'un fichier XML) afin de normaliser les métadonnées issues des fichiers latex et bibtex, et a écrit des scripts Perl qui permettent de préparer l'exploitation de ces données par EDBM, qui est le coeur de notre interface d'accès aux articles.

Claude Goutorbe a dû paramétrer EDBM pour prendre en compte les nouveaux formats de données, et gérer les articles qui ne sont pas en accès libre. Les expériences du Zentralblatt et de NUMDAM ont été précieuses pour mener à bien ce travail délicat en quelques mois. La base de données d'EDBM est constitué des articles anciens et courants : les archives NUMDAM sont intégrées de façon transparente à la production actuelle, les articles postérieurs à NUMDAM mais non produits par CEDRAM étant mis aux nouvelles normes de façon parfois acrobatique.

En conclusion, à l'heure d'écrire ce rapport, nous disposons des articles et de leurs métadonnées pour la totalité des AMBP et des AIF, la période 2003-2005 des AFST étant en cours de finalisation. L'objectif d'afficher ces trois revues début 2006 est réaliste.

3.3.3 Sites Web

Laure Heigéas a travaillé à la conception de l'architecture d'un site de revue. Le contenu du site Web d'une revue peut être représenté en deux parties : le contenu statique et le contenu dynamique. Le contenu statique concerne principalement les informations stables relatives à la revue elle-même : présentation des sujets traités, du comité de rédaction, indications aux auteurs pour la soumission d'articles... Ces pages statiques sont directement modifiables par la revue elle-même via une interface de mise à jour. L'interface actuellement utilisée est le logiciel libre SPIP, (alias système de publication pour l'Internet), un système de gestion de contenu ou CMS (de l'anglais Content manager System).

Deux sites web sont en cours de finalisation : AIF et AMBP. Là encore, les AFST sont moins avancées. Le site portail du projet (www.cedram.org), qui donnera les informations institutionnelles et un accès commun à toutes les revues, est en voie d'achèvement.

3.4 Chaîne de production nouvelle

Voici une description de la nouvelle chaîne de production des revues. La répartition des tâches entre la Cellule MathDoc et les revues décrite ci-dessous correspond au cas des trois premières revues. Dans le cas où l'édition elle-même serait plus intégrée au pôle, la ligne de partage serait différente, mais les opérations seraient les mêmes.

L'auteur soumet son article à la revue de son choix. Lorsque le texte est accepté par la revue, l'objet des opérations suivantes est d'associer à chaque article

- un identifiant unique qui permettra sa gestion jusqu'au moment de sa publication ;
- un PDF servant à la fois à la publication en ligne et à l'impression ;
- un ensemble de métadonnées permettant d'alimenter toutes les fonctions attendues de l'édition électronique.

L'identifiant est un numéro quelconque, la seule contrainte est qu'il soit unique : c'est par exemple un numéro d'ordre parmi les articles acceptés de l'année : RevueX200600001. Lorsque l'article sera publié, il se verra affecter un identifiant public plus commode, reflétant ses données bibliographiques et en général calculable, similaire à l'identifiant NUMDAM qui a largement fait ses preuves.

Les métadonnées essentielles sont insérées dans l'article lui-même (source latex) à l'aide d'un balisage simple ; la métadonnée complexe qu'est la bibliographie est structurée dans une base bibtex. Le format d'entrée est un fichier latex structuré pour l'essentiel selon les usages bien établis de l'AMS (également employés par la SMF). L'auteur fournit dans la mesure du possible son source latex et un fichier bibtex, lesquels sont mis aux normes de la revue (relecture, corrections typographiques, maquette, marche pour la présentation des énoncés et de la bibliographie, etc.) et validés sur épreuves par l'auteur et la rédaction.

La responsabilité de la revue est donc de fournir des données conformes aux normes établies par le pôle. La Cellule MathDoc n'intervient que comme soutien logistique : c'est la revue qui a un contrôle total sur le contenu qu'elle publie et la forme qu'il revêt.

Lorsque la revue est prête à publier un nouveau fascicule, elle fournit à la Cellule MathDoc les éléments nécessaires :

- pour chaque article tous les fichiers nécessaires au traitement de cet article (fichier source latex, évidemment, mais aussi éventuels inputs, images et .bib), selon des normes préalablement établies ;
- les données qui définissent le fascicule :
 - la tomaison,
 - la liste ordonnée des articles,
 - les éventuels ajouts éditoriaux ou non (éditoriaux, éléments pour un numéro spécial, comme photographies, avis au lecteur, publicités...).

Ces données sont exportées dans un fichier latex dont la compilation produit à la fois la totalité des pages intérieures du fascicule papier, les éléments pour la couverture et les tirés à part, les PDF écran de chaque article, et la forme brute des métadonnées. MathDoc retourne alors à la revue le PDF destiné à l'imprimeur, qu'elle valide et fait imprimer, et réalise un affichage web

du fascicule dont l'accès n'est autorisé qu'à la rédaction de la revue (revuetest.cedram.org), qui valide cette « impression sur format numérique » et signe une sorte de « bon à mettre en ligne » avec date d'effet, date à laquelle MathDoc bascule le fascicule sur le site Web de la revue (revue.cedram.org).

Les rapports de la revue avec MathDoc sont donc essentiellement ceux qu'elle a déjà avec un imprimeur classique. La contrainte imposée par des formats structurés proprement est partiellement contrebalancée par une meilleure efficacité de la chaîne. Par exemple, les textes imprimés et mis en ligne étant rigoureusement les mêmes, une seule vérification suffit. De plus, il n'y a pas de risque d'erreurs de dernière minute car le fascicule est entièrement reconstruit, et tous les numéros de page recalculés, de même que le sommaire est constitué de façon automatique par le programme.

Comme cela est fait pour NUMDAM, MathDoc enrichit la notice complète de chaque article par tous les liens possibles : rebonds sur les mots clés, la classification mathématique, les auteurs ; liens de la bibliographie vers MathSciNet et ZM, vers CEDRAM ou NUMDAM quand cela est possible. D'autres liens directs grâce à la mini-DML quand celle-ci sera en production. Ces liens sont appelés à être régénérés périodiquement (ne serait-ce que parce que les articles publiés finiront par avoir des analyses, et les prépublications citées par être publiées). Un serveur OAI permettra de diffuser les données bibliographiques des articles disponibles, assurant une mise à jour et un référencement optimaux.

A contrario, le PDF est figé pour l'éternité, archivé, et ne pourra être amendé que par la publication d'un nouvel article, éventuellement par d'autres ressources à travers des liens ajoutés à sa notice complète. Comme dans le cas de NUMDAM, les PDF auront une page de garde comportant notamment un lien vers cette notice complète, qui reste le point névralgique comportant toutes les données liées à l'article, et qui est donc appelée à évoluer au fil de la vie de celui-ci.

La Cellule MathDoc est donc gestionnaire de tous les moyens informatiques, du serveur « portail », et développe un système fiable d'archivage sur le long terme et des outils simplifiant la tâche des revues. Par exemple, un logiciel de gestion du secrétariat de rédaction, le logiciel RUCHE, logiciel libre développé pour le secrétariat de la revue *Traitement du signal*, est en cours d'adaptation pour *les Annales de l'institut Fourier* et sera utilisé en janvier 2006. Ce logiciel RUCHE permet à la fois une gestion souple de la revue, et un meilleur interfaçage avec la chaîne de production du CEDRAM en aval.

3.5 Commercialisation et droits d'accès

Les premières revues participant au pôle disposent de leur propre structure de diffusion papier et de commercialisation des abonnements. Ainsi le CEDRAM ne se voit actuellement pas en charge de ce module, mais la Cellule MathDoc étudie actuellement en concertation avec l'UJF et sa filiale de droit privé Floralis¹⁷ l'intégration des services commerciaux dans le pôle.

¹⁷ <http://www.floralis.fr/>

Le modèle choisi par toutes les revues académiques françaises est de proposer l'abonnement électronique comme bonus gratuit à l'abonnement papier. De plus, les revues se sont généralement prononcées pour un créneau mobile de 5 ans à l'occasion de leur numérisation. La participation au pôle ne modifie rien à cet état de fait : les revues se chargent de collecter les adresses IP autorisées à consulter la revue sur le site dont les accès sont contrôlés par la Cellule MathDoc, elles les fournissent au personnel du CEDRAM. Les revues restant maîtres de leur fabrication, elles en assurent les frais par le recouvrement des abonnements et toutes autres formes de financement.

Par ailleurs, un utilisateur non abonné pourra naviguer dans l'intégralité du site (les métadonnées) et utiliser les liens fournis, ce qui assure la meilleure visibilité à la revue et une indexation correcte (grâce aux résumés et aux mots clés) dans les moteurs de recherche. Seuls les abonnés pourront accéder au PDF des articles protégés par le créneau mobile, les autres seront cordialement invités à s'abonner.

D'autre part, sur les sites du CEDRAM comme dans le cadre de NUMDAM, il est possible sur requête de fournir le texte intégral des articles - mais pas le PDF - à quelques moteurs de recherche qui pourront ainsi indexer beaucoup plus finement le contenu de la revue.

3.6 Démarrage

L'inauguration des serveurs des trois premières revues et du portail www.cedram.org est programmée pour le début de l'année 2006. Au vu des avancées significatives des derniers mois, cet objectif sera très certainement atteint.

La mise en ligne de ces trois revues peut être considérée comme le point de départ d'un véritable pôle diffusant un contenu mathématique solide et suffisamment diversifié. Elle fournira des éléments tangibles pour mesurer la charge de travail correspondante, le modèle de travail adopté, et l'intérêt des utilisateurs. Si, comme nous avons la faiblesse de le croire, les mathématiciens et les éditeurs des revues apprécient le service rendu, elle permettra d'entamer des négociations fructueuses avec les revues qui hésitent encore.

4. Futur du projet CEDRAM

4.1 Développements envisagés

Au plan technique, les prochaines étapes seront la mise en place d'un système de conservation et d'archivage à long terme à la fois pour les sources TeX des articles - avec l'installation TeX correspondante - et pour les versions imprimables.

Par ailleurs, le pôle pourra mettre en place des outils communs pour la gestion et la préparation des articles en amont. L'objectif étant que la chaîne de production soit aussi fluide que possible, et qu'un standard soit utilisé par les revues le plus tôt possible de façon à disposer ensuite d'un maximum d'automatismes permettant d'éviter les erreurs ou les contretemps. Par exemple dans le cas de la mise aux normes des articles, la validation des sources latex+bibtex produits doit être possible par la revue elle-même. Du fait que, si les normes sont respectées, tous les traitements ultérieurs sont automatiques, il est assez facile de concevoir un outil de validation qui, à partir des éléments normalisés pour un article, produirait le PDF destiné à l'impression ainsi que la partie du sommaire correspondant à l'article, mais aussi la page HTML de métadonnées qui formera (à quelques liens près) la « notice complète » de l'article sur le serveur. Une interface

Web très simple pourra être développée rapidement : elle permettra par exemple le téléchargement d'une archive contenant tous les éléments techniques pour la fabrication d'un article, et retournera la notice HTML comportant le lien sur le PDF. Pour les revues disposant d'un serveur Linux, un script un peu plus lourd à installer pourrait être proposé.

Pour les revues souhaitant une indépendance maximale, d'autres outils seront développés à la demande (accès sécurisé aux serveurs de production, interface de mise à jour des pages Web ou des listes de numéros IP autorisés, etc.).

4.2 Perspectives d'évolution

Pour convaincre certaines revues, il est nécessaire que le CEDRAM puisse assurer l'ensemble des tâches d'édition, diffusion et promotion, y compris la vente et la gestion des abonnements. De plus, il est fort probable que la demande d'abonnements uniquement électroniques progresse, possiblement de rabais pour des abonnements groupés. Les économies de papier sont minces par rapport au coût d'une mise aux normes évoluée autorisant une édition électronique efficace.

Il nous faut donc mettre en place maintenant une structure capable de gérer les finances d'une revue, et d'équilibrer les comptes du pôle pour le pérenniser à long terme, ce qui pourra se faire en partenariat avec la filiale Floralis de l'université Joseph Fourier. Un poste d'administrateur semble nécessaire pour intégrer l'ensemble, et assurer une réactivité satisfaisante.

ANNEXES

A. MISSIONS

Principales missions et interventions, à compter du dernier comité de pilotage (14 septembre 2001).

Catherine Barbe-Zoppis [CBZ], Hélène Bégnis (Falavard) [HB, HF], Pierre Bérard [PB], Thierry Bouche [ThB], Elizabeth Cherhal [EC], Isabelle Costerg [IC], Claude Goutorbe [CG], Laurent Guillopé [LG], Laure Heïgéas [LH], Yves Laurent [YL], Monique Marchand [MM].

Missions invités : Claude Sabbah [CS], Christian Gérini [ChG], Bernard Guidoux [BG]

1. Réunions des projets européens LIMES et DML-EU (Digital Mathematics Library Europe)

6-7 octobre 2001 – Santiago (Espagne) [CG, LG]

03-05 mai 2002 – Lecce (Italie) [CG, LG]

24-31 juillet 2002 – Washington (États-Unis), "*Digital Mathematics Library*" organisé par Cornell University et National Science Foundation, Washington. [PB]

25-26 octobre 2002 – Nantes [PB, CG, LG]

31 janvier 2003–02 février – Berlingen (Suisse), Digital Mathematical Library Project in 6FP: Brainstorming weekend Berlingen. [LG]

19-20 mars 2003 – Grenoble, Journées DML. [CBZ, HB, PB, ThB, EC, CG, LG, YL]

21-22 mai 2003 – Göttingen (Allemagne) [PB, ThB]

02-04 mai 2003– La Haye (Pays-Bas) [CG, LG]

23-28 octobre 2003 – Berlin (Allemagne) [CG, LG]

19-21 mars 2004 – Zurich (Suisse) [CG, LG]

14 novembre 2005 – Berlin (Allemagne) [ThB, YL]

2. Réunions du comité de coordination du Zentralblatt Math

6 novembre 2001 – Heidelberg (Allemagne) [LG]

13-14 mai 2002 – Heidelberg (Allemagne) [LG]

06 novembre 2002 – Karlsruhe (Allemagne) [LG]

05 mai 2003 – Heidelberg (Allemagne) [LG]

13-16 novembre 2003 – Heidelberg (Allemagne) [LG]

27 mai 2004 – Karlsruhe (Allemagne) [LG, YL]

12 novembre 2004 – Heidelberg (Allemagne) [YL]

27-28 janvier 2005 – Heidelberg (Allemagne) [YL]

20-25 février 2005 – Berlin (Allemagne) [CG]

9 mai 2005 – Heidelberg (Allemagne) [YL]

25 novembre 2005 – Berlin (Allemagne) [YL]

De plus, une réunion s'est tenue au Ministère sur l'avenir de cette coopération le 11 juin 2003 à Paris. [LG, YL]

3. Projet EMANI (Electronic Mathematics Archive Network Initiative)

24-31 juillet 2002 – Washington (États-Unis) [PB]

22-23 novembre 2002 – Göttingen (Allemagne) [YL]

04-05 avril 2003 – Paris [HB, LG, YL]

1er août 2003 – Berlin (Allemagne) [EC, YL]

30 mars au 4 avril 2004 – Pékin (Chine) [EC]

27-28 janvier 2005 – Heidelberg (Allemagne) [YL]

29 octobre – 2 novembre 2005 – Cornell University (USA) [YL]

4. Pôle des revues

- 26 janvier 2004 – Paris [ThB, EC, LG, YL]
- 2 mars 2004 – Paris [LG, YL, ThB]
- 14 avril 2004 – Paris [EC, LG]
- 1^{er} juin 2005 – Paris [YL]
- 7 juillet 2005 – Grenoble [CS]
- 10 octobre 2005 – Grenoble [CS]

5. Projet NUMDAM (NUMérisation de Documents Anciens Mathématiques)

- 1-2 octobre 2003 – Lens [CB-Z, HB, ThB, YL]
- 26 novembre 2003 – Isle d'Abeau [YL]
- 16 avril 2004 – Lens [HB, ThB]
- 1^{er} février 2005 – Nîmes [ThB, HF, ChG]
- 5 juillet 2005 – Lens [ChG, BG]
- 8 novembre 2005 – Lens [IC, HF]

6. Collaboration avec la Bibliothèque Nationale de France

- 3 juillet 2003 – Marseille
Journée des Pôles associés organisée par la Bibliothèque nationale de France. *"La coopération entre bibliothèques face aux enjeux du web et de la numérisation partagée"*. Exposé de L. Guillopé : *"Le projet MathDoc/BnF : Enrichissement des accès, mise en réseau"*. [CBZ, HB, LG]
- 5-6 février 2004 – Paris
Journées de la Bibliothèque Nationale de France [LG]
- 17 mars 2004 – Paris
Journée à la Bibliothèque Nationale de France [LG]
- 1-2 juillet 2004 – Paris
Journées de la Bibliothèque Nationale de France [EC]
- 20-21 janvier 2005 – Paris
Collaboration entre la Cellule MathDoc et la BNF dans le cadre du pôle associé « Numérisation concertée en Mathématiques » [EC, YL, ThB]
- 7 juin 2005 – Paris
Journée « Formats et métadonnées » organisée par la BNF [EC, LH]
- 29 juin – 1^{er} juillet 2005 – Rennes
« 9e Journées des Pôles associés » organisées par la Bibliothèque Nationale de France [EC]

7. Congrès et journées en France

- 13 octobre 2001 – Paris
Réunion Société Mathématique de France. Exposé : NUMDAM. [PB]
- 5-6 novembre 2001 – Lyon
Conférence *"Long Term Archiving of Digital Documents in Physics"* organisée par le Centre pour la Communication Scientifique Directe, CNRS. [CBZ, PB, EC, LG]
- 14 décembre 2001 – Paris
Journée *"Le document électronique"* organisée par le Réseau National des Bibliothèques de Mathématiques et la Cellule MathDoc, en l'honneur du départ à la retraite de Geneviève Sureau, directrice de la bibliothèque de mathématiques Jacques Hadamard d'Orsay et co-responsable du RNBM. [PB, EC, LG, MM]
- 7 mars 2002 – Lyon
"Valorisation des Documents et Numérisation des Collections (ISDN)", Exposé de Th. Bouche : *"NUMDAM"*, ENS, LSH. [ThB]
- 18-20 juin 2002 – Nancy
Journées des Documentalistes CNRS, INIST Nancy. Exposé d'E. Cherhal : *"Open Archives, OpenUrl, des initiatives pour l'interopérabilité des applications documentaires"*. [HB, EC]

- 23-24 octobre 2002 – Paris
Journées "Mathrice", IHP. [CBZ, HB, EC, LG]
- 07-08 janvier 2003 – Nantes
"Journées Equations aux Dérivées Partielles : un quart de siècle numérisé", organisées par la Cellule MathDoc et le laboratoire J. Leray. Exposé de Y. Laurent : "Le programme NUMDAM". [HB, PB, ThB, LG, YL]
- 23-24 janvier 2003 – Paris
"Libre accès à l'information scientifique et technique : état de l'art et perspectives", INIST/INSERM/ICSTI, Ministère de la Recherche. Exposé de L. Guillopé : "Mathématiques et accès (libre) aux bases de données". [EC, LG]
- 11-12 février 2003 – Nice
Congrès "Applied Mathematics and Applications of Mathematics". Table ronde *Electronic databases "NUMDAM: year 0"*. [LG]
- 26 février 2003 – Paris
"Autour des archives anciennes de Bourbaki", Séminaire d'Histoire des Mathématiques, "Les connexions de NUMDAM". [LG]
- 24 mars 2003 – Paris
Assemblée générale du RNBM (Réseau National des Bibliothèques de Mathématiques). [EC, YL]
- 14 juin 2003 – Paris
Assemblée Générale de la Société Mathématique de France. Exposés : "NUMDAM : le programme" et "NUMDAM : portrait de groupe". [LG]
- 1^{er} octobre 2003 – Paris
Journée du CCSD au Collège de France. Exposé de T. Bouche « le programme NUMDAM » [ThB, EC, LG, YL]
- 13-14 octobre 2003 – Paris
Journées à l'IHP et à l'EHESS [LG]
- 22 octobre 2003 – Paris
Journée « Mathrice », organisée à l'institut Henri Poincaré [CB-Z]
- 7 novembre 2003 – Lyon
Journée du CCSD. [EC]
- 18 novembre 2003 – Paris
Journée du Réseau National des Bibliothèques de Mathématiques [EC]
- 27 février 2004 – Montpellier
Journée ABES [EC]
- 20-22 mai 2004 – Poitiers
Congrès d'histoire des sciences et des techniques organisé par la Société française d'histoire des sciences et des techniques et par l'université de Poitiers [ThB, LG]
- 6-8 juin 2004 – Forges les Eaux
Journées JEDP [LG]
- 21-23 juin 2004 – La Rochelle
Premières rencontres Numérisation et Patrimoine. Semaine du Document Numérique organisée par le Réseau Thématique Pluridisciplinaire "Document" du CNRS et par l'université de La Rochelle [CB-Z, HB, LG]
- 16 novembre 2004 – Lyon
Journée d'études « Indexation des ressources pédagogiques numériques » [EC]
- 7-11 mars 2005 – Pont-à-Mousson
Congrès « EuroTex 2005 » organisé par l'association Gutenberg [ThB, LH]
- 19 mars 2005 – Paris
Archives Bourbaki, Séminaire Bourbaki [YL]
- 21 mars 2005 – Marseille
Présentation du logiciel KOHA au CMI [EC]

- 1^{er} juin 2005 – Paris
Journée « Documentation mathématique » à l'IST/CNRS avec M. Romary [YL]
- 9 juin 2005 – Lyon
Journée au CCSD [EC]
- 13-15 juin 2005 – Paris
CA Gutenberg et journée NUMDAM avec la société des Statistiques à l'IHP [ThB, HF]
- 20-23 juin 2005 – Nancy
« Les rencontres 2005 des Professionnels de l'IST » [EC, IC, HF]
- 12 juillet 2005 – Orsay
Journée de travail sur le « Catalogue fusionné des ouvrages »
- 13 juillet 2005 – Lyon
Journée « TEL » au CCSD [EC]
- 3 octobre 2005 – Paris
Journée « circulation des savoirs », sommet mondial de la société de l'information. Exposé sur NUMDAM à la Bibliothèque Nationale de France [ThB]
- 20 octobre 2005 – Paris
Journée « Mathrice » à l'IHP [EC]
8. Congrès à l'étranger
- 06-09 janvier 2002 – San Diego (États-Unis)
Congrès "*Joint Mathematics Meetings*". (AMS, MAA, SIAM). Informal meeting on mathematics digitization. [PB, LG]
- 17-21 mars 2002 - Ann Arbor (États-Unis), "*Linking and searching in distributed digital libraries*". [ThB, EC]
- 20-21 juin 2002 - Essen (Allemagne), "Workshop on digitisation of mathematics journals". Exposé : "The NUMDAM program". [ThB]
- 01-04 juillet 2002 – Graz (Autriche)
Conférence annuelle LIBER (Ligue des Bibliothèques Européennes de Recherche) : "*European libraries as portals to information*". Participation à la table ronde "*Research organisations and learned academies : views on the future of research libraries in the electronic age*". [LG]
- 18-23 août 2002 – Glasgow (Écosse)
68e Congrès IFLA (International Federation of Library Associations and Institutions) et assemblée générale, "*Bibliothèques pour la vie : démocratie, diversité, fourniture*". Exposé : "*Documentation mathématique : enjeux pour l'ère numérique*". [PB]
- 17-19 novembre 2002
OpenArchives Initiative (OAI) Workshop, CERN, Genève (Suisse), "Cellule MathDoc, CCSD, two teams collaborating on promoting scientific documentation". [EC]
- 20-23 mai 2003 - Göttingen rencontres DML/DML-EU "www.numdam.org : *questions & answers*". [ThB]
- 1-5 septembre 2003 – Madrid (Espagne)
"Ciencia en el Siglo XXI, El futuro de las Publicaciones Electronicas", Exposé : "La documentation électronique en France". [YL]
- 23-28 avril 2004 – Minneapolis (USA)
Conférence intitulée « Enhancing Searching of Mathematics » à l'institut de Mathématiques et ses applications de Minneapolis [LG]
- 24-28 juin 2004 – Stockholm (Suède)
Conférence WDM-EMANI « New Developments in Electronic Publishing of Mathematics » Exposé de Y. Laurent « NUMDAM, digitising French mathematical serials », exposé de T. Bouche « The mini-DML project » [ThB, LG, YL]
- 12-18 avril 2005 – Berkeley (USA)
Workshop on "Conference on the Digital Mathematics Library Project" organisé par le Mathematical Sciences Research Institute. 2 exposés de T. Bouche : "Digitization and Metadata Overview", "The NUMDAM program" [ThB]
- 7-9 juin 2005 – Louvain (Belgique)
9th ICC International Conference on Electronic Publishing « Elpub2005 » [EC, LH]

B. PUBLICATIONS

P. Bérard, Documentation issues for mathematics in the digital age, *International Journal of Special Libraries*, 36, 3 (2002) 147-161.

P. Bérard, G. Sureau, B. Teissier, La documentation mathématique à l'aube du 21e siècle, *Lettre SPM CNRS*, 38 (2002) 7-9.

L. Guillopé, Mathematics and databases: open access, *Information Services and Use*, 23, 2-3 (2003) 127-131.

T. Bouche, Introducing the mini-DML project, *actes de ECM4 Satellite Conference*, Stockholm, (2004) 19-29

E. Cherhal, Numérisation/conservation partagée : la collaboration BnF/Cellule MathDoc ; pistes de travail pour de nouveaux services, *actes des journées des pôles associés*, Paris, Juillet 2004.

E. Cherhal, Présentation des standards : le Dublin Core, *actes de la journée « Indexation des ressources pédagogiques »*, Lyon, novembre 2004.

E. Cherhal, État de l'art en matière de normes et approche critique : de Marc à Dublin Core, *actes de la journée « Indexation des ressources pédagogiques »*, Lyon, novembre 2004.

E. Cherhal, La Cellule MathDoc, *Arabesques*, 37 (2005) 9-10.

Y. Laurent, MathDoc, Spécificités et évolutions, *Arabesques*, 37 (2005) 8-9.

E. Cherhal, L. Heigéas, MathDoc and the Electronic Publishing of Mathematics, *actes de la conférence EIPub*, Leuven, Belgique, juin 2005.

C. BUDGET

Récapitulatif des budgets 2002-2005						
Recettes						
	Ministère (UJF)		U. Nantes	CNRS		TOTAL
	UJF	Europe	AS	CNRS	NUMDAM	
2002	19 272	29 360	1 377	14 940	66 102	131 051
2003	32 943	13 502	1 967	14 872	123 211	186 495
2004	44 147	1 704		15 429	26 373	87 653
2005	44 041	35 256		15 635	40 827	135 759
Dépenses						
	Equipement	Personnel	Fonctionnement	Missions	Numérisation	TOTAL
2002	19 207	33 965	13 451	29 137	35 291	131 051
2003	25 735	19 214	16 771	25 356	99 419	186 495
2004	14 245	24 190	18 786	19 370	11 062	87 653
2005	10 883	44 679	12 898	10 495	56 805	135 759
Moyenne	17 518	30 512	15 477	21 089	50 644	135 240

1. Recettes

a) UJF et CNRS : crédits récurrents

b) Université de Nantes : action spécifique.

c) Europe : contrats européens Euler (European Libraries and Electronic Resources in mathematical sciences) de 1998 à 2000 et LIMES (Large Infrastructure in Mathematics Enhanced Services) de 2000 à 2004.

d) NUMDAM : Action spécifique destinée à financer le projet de numérisation. Le crédit initial de 379 226 € sera entièrement dépensé à la fin de la deuxième phase de NUMDAM début 2007.

2. Dépenses

a) La ligne "Numérisation" correspond aux versements effectués aux sociétés AIS et Jouve qui effectuent la numérisation. Elle ne comprend pas d'autres frais comme ceux de personnel. Les sommes dépensées varient beaucoup d'une année à l'autre correspondant à des phases de mises en œuvre ou de production continue. Ces irrégularités expliquent celles du total par année, les autres activités étant relativement stables globalement.

b) L'équipement est essentiellement informatique, sauf pour l'année 2003 (et dans une moindre mesure 2002) où une part importante des dépenses correspond à l'installation de l'unité dans ses nouveaux locaux et donc à des achats de mobilier.

c) Les dépenses de personnel correspondent d'une part à des décharges de services (Laurent Guillopé à Nantes et Thierry Bouche à Grenoble) et d'autre part à des contrats temporaires pour le contrôle de qualité de NUMDAM. Elles ont été financées en partie grâce aux contrats européens.

d) Le pic de missions en 2002-2003 est lié au contrat européen LIMES et financé par lui.

e) Dans le fonctionnement, une somme importante (5000 €) correspond à des abonnements à des bases de données qui permette de fournir des données utiles aux mathématiciens sur le serveur web (par exemple les sommaires des revues).

f) Dans les budgets détaillés 2004 et 2005 en annexe apparaît une ligne ANGD qui n'a pas été reprise dans ce récapitulatif. Il s'agit d'une subvention de la formation permanente du CNRS et de la région Provence-Alpes-Côte d'Azur qui a servi à organiser une école thématique en collaboration avec le RNBM et dont le reliquat a été utilisé en 2005 pour l'organisation d'une réunion du RNBM.

BUDGETS 2002-2005 – Recettes / Dépenses**BUDGET 2002****RECETTES**

Origine	Objet	TTC	HT	HT hors BQR (7,5%)	Totaux
Ministère (UJF)	Reliquats 2001			-744	
	Subvention 2002			18 865	
	Reversement frais de gestion EULER			1 783	
	Fonctionnement	15 245	12 747	11 791	
	Vacations	9 147	7 648	7 074	
Total					19 904
Contrats UE (UJF)					
EULER	Solde 2001			3 626	
Total					3 626
LIMES	Solde 2001 ⁽¹⁾			8 215	
	2002 ⁽²⁾			34 421	
Total					42 636
Université de Nantes					
	Action spécifique ⁽³⁾		4 000	3 344	
					3 344
CNRS	Reliquats 2001			1 694	
	2002 - Soutien de base			15 500	
Total					17 194
NUMDAM (CNRS)	Action spécifique				
	Solde 2001			364 665	
Total					364 665
Total Recettes 2002					451 369

BUDGET 2002

Réalisations Dépenses

	Recettes	Ministère (UJF)			Univ. Nantes	CNRS		Total
		UJF	Euler ⁽¹⁾	Limes	AS	CNRS	NUMDAM	
		19 904	3 626	42 636	3 344	17 194	364 665	
Code	Nature crédits	Dépenses						
	Équipement							
218	Équipement ^(*)	0	2 000	6 372			7 545	15 917
615	Équipement non inventorié et Maintenance	0		1 009				1 009
618	Abonnements et Adhésions – Livres	602		171		1 227	191	2 191
623	Publicité - Publications					0	90	90
	Marché Sté A.I.S.						35 291	35 291
	Sous-total	602	2 000	7 552		1 227	43 117	54 498
	Personnel							
621– 641	Frais de personnel			11 403			9 015	20 418
	Décharges service : P. Bérard, Th. Bouche, L. Guillopé	11 347						11 347
622	Honoraires (Tosca Consultants et Me Marter)			2 200				2 200
	Sous-total	11 347	0	13 603		0	9 015	33 965
606	Fonctionnement							
	Petit matériel - Consommables - Divers	1 847		1 315		2 478	982	6 622
	Téléphone - Location Bureaux	712		826		5 291		6 829
	Sous-total	2 559	0	2 141		7 769	982	13 451
625	Missions							
	Frais de réception	1 595	350			26		1 971
	Administration	676		551	1 377	2 355		4 959
	Bibliothèques					997	566	1 563
	Coopération franco- allemande	1 456				370		1 826
	Colloques - Congrès	1 037	198	61		1 753	4 359	7 408
	NUMDAM					443	8 063	8 506
	LIMES			2 904				2 904
	Sous-total	4 764	548	3 516	1 377	5 944	12 988	29 137
TOTAL		19 272	2 548	26 812	1 377	14 940	66 102	131 051
Solde disponible au 31/12/2002		632	1 078	15 824	1 967	2 254	298 563	320 318

Frais de Personnel

3 enseignants-chercheurs à 50 %

1 vacataire à temps complet / 12 mois

^(*) dont 4 856 € / LIMES pour participation achat serveur Nantes.

⁽¹⁾ EULER : période hors contrat.

BUDGET 2003
RECETTES

Origine	Objet	TTC	HT	Totaux
Ministère (UJF)	Reliquats 2002	1 225		
	Subvention 2003 ⁽¹⁾	40 700		
Total				41 925
Contrats UE (UJF)				
EULER	Solde 2002		1 078	
Total				1 078
LIMES	Solde 2002 ⁽²⁾		36 077	
	2003 ⁽³⁾		58 143	
Total				94 220
Université de Nantes				
	Action Spécifique ⁽⁴⁾ / Solde 2002		1 967	
Total				1 967
CNRS				
	Reliquats 2002		2 252	
	2003 - Soutien de base		15 457	
Total				17 709
NUMDAM (CNRS)				
	Action spécifique ⁽⁵⁾			
	Solde 2002		304 099	
Total				304 099
Total Recettes 2003				
				460 998

^[1] Quadriennal 2003 – 2006 (Totalité de la subvention versée en crédits de Fonctionnement).

^[2] LIMES Solde 2002 = (Solde 2001 + Encaissement 2002) – Dépenses 2002.

^[3] Encaissements correspondant aux *Cost Statements* des périodes : avril 2001-mars 2002 (42 178 €) et avril 2002-mars 2003 (15 965 €).

^[4] Subvention ministérielle attribuée à l'Université de Nantes au titre des actions spécifiques pour l'exercice 2002

(participation aux frais de déplacement de L. Guillopé, directeur-adjoint de la Cellule MathDoc).

^[5] TOTAL Action Spécifique NUMDAM : 379 226 € (1^{ère} notification 13/07/2000 : 228 673 € ; 2e notification : 05/03/2001 : 150 553 €).

BUDGET 2003

DEPENSES

	Recettes	Ministère (UJF)			U.Nantes	CNRS		Total
		UJF	Euler ⁽¹⁾	Limes	AS	CNRS	NUMDAM	
		41 925	1 078	94 220	1 967	17 709	304 099	
Code	Nature crédits	Dépenses						
	Équipement							
218	Équipement informatique	15 231				827	1 145	
	Équipement Bureaux IF	8 532						
615	Équipement non inventorié et Maintenance							
	Sous-total	23 763	0	0	0	827	1 145	25 735
	Personnel							
64	Frais de personnel	2 405		9 002			6 807	
	Décharges service							
	Honoraires (Me Marter)			1 000				
	Sous-total	2 405	0	10 002	0	0	6 807	19 214
606	Fonctionnement							
	Petit matériel - Consommables	2 226				420	1 851	
	Téléphone - Location Bureaux	646				4 981		
618	Abonnements, Adhésions, Livres					1 631	4 250	
623	Publicité - Publications						766	
628	Numérisation / Sté AIS						99 419	
	Sous-total	2 872	0	0	0	7 032	106 286	116 190
625	Missions							
	Frais de réception	656	18			510	393	
	Administration	846	423		1 967	2 696		
	Bibliothèques	613	236			1 345	1 204	
	Coopération franco-allemande	0	398	2 425		49		
	Colloques - Congrès	772				1 441	1 418	
	NUMDAM	1 016				66	5 958	
	LIMES					906		
	Sous-total	3 903	1 075	2 425	1 967	7 013	8 973	25 356
TOTAL		32 943	1 075	12 427	1 967	14 872	123 211	186 495
Solde disponible au 31/12/2003		8 982	3	81 793	0	2 837	180 888	274 503

⁽¹⁾ EULER : période hors contrat.

Frais de personnel :

Personnel enseignant : 1 enseignant à 50 % + 1 enseignant à 25 %

Vacataires/NUMDAM : Luc Courbon : janv.-févr., 50 % (1 200 €) ;

Virginie Imbrogno, janv.-sept., 80 % (11 396 €) ; Estelle Nivault : oct.-déc., 50 % (2 635 €).

Régularisation complément APE : 578 €.

Allocations Perte d'Emploi : Patricia Aiello : juil.-août (1 298 €) ; V. Imbrogno, oct.-nov. (1 107 €).

BUDGET 2004
RECETTES

Origine	Objet	TTC	HT	Totaux
Ministère (UJF)	Reliquats 2003	8 982		
	Subvention 2004 ⁽¹⁾	40 700		
Total				49 682
Contrat UE (UJF)				
LIMES	Solde 2003		81 793	
	2004 ⁽²⁾		0	
Total				81 793
CNRS				
	Reliquats 2003		2 837	
	2004 - Soutien de base		15 500	
Formation Permanente	ANGD ⁽³⁾ "Documentation en mathématiques"		19 000	
Conseil Général - Départ. des Bouches- du-Rhône	Subvention École "Documentation en mathématiques"	2 287	1 912	
Total				39 249
NUMDAM (CNRS)				
	Action spécifique			
	Solde 2003		180 888	
Total				180 888
Total Recettes 2004				351 612

^[1] Quadriennal 2003-2006 (Totalité de la subvention versée en crédits de Fonctionnement).

^[2] Le virement correspondant au *Cost Statement* de la période avril 2003 – mars 2004 a été versée en juillet 2005.

^[3] ANGD (Action Nationale à Gestion Déconcentrée), CIRM Luminy, 11-15 octobre 2004.

BUDGET 2004

DEPENSES

	Recettes	Ministère (UJF)		CNRS			Total
		UJF	Limes	CNRS	NUMDAM	ANGD *	
		49 682	81 793	18 337	180 888	20 912	
Code	Nature crédits						
	Équipement						
218	Équipement informatique	4 790					
	Équipement Bureaux IF	8 655					
615	Équipement non inventorié et Maintenance	800					
	Sous-total	14 245	0	0	0	0	14 245
	Personnel						
64	Frais de personnel	10 603	1 272		8 586	114	
	Décharges service	3 729					
	Sous-total	14 332	1 272	0	8 586	114	24 190
606	Fonctionnement						
	Petit matériel - Consommables Divers	4 382		2 000	1 118	1 826	
626	Téléphone - Location Bureaux	550		2 452		1 500	
618	Abonnts, Adhésions, Livres	424		2 319	4 505	500	
623	Publicité - Publications				191	500	
624	Transports de biens	845				500	
628	Numérisation / Sté JOUVE				11 062		
	Sous-total	6 201	0	6 771	16 876	4 826	29 848
625	Missions						
	Frais de réception	1 379				11 802	
	Administration	316		2 503			
	Bibliothèques	351		789	64	4 170	
	Coopération franco-allemande	424		1 470			
	Colloques - Congrès	5 893		1 243			
	NUMDAM			2 653	847		
	LIMES	1 006	432				
	Sous-total	9 369	432	8 658	911	15 972	19 370
TOTAL		44 147	1 704	15 429	26 373	20 912	87 653
Solde disponible au 31/12/2004		5 535	80 089	2 908	154 515	0	263 959

Vacataires/NUMDAM : Estelle Nivault : janv.-mars, 50 % (2 647 €)

Isabelle Costerg : mi-avril-mi-oct., 50 % (5 302 €) et mi-oct.-déc., 100 % (4 348 €)

Régularisation complément APE : 637 €

Allocations Perte d'Emploi : V. Imbrogno, déc. 2003 à nov. 2004 (7 527 €).

* Subvention Formation Permanente (19 000 €) + Subvention Conseil Général Dépt. Bouches-du-Rhône (1 912 €)

BUDGET 2005
RECETTES

Origine	Objet	TTC	HT	Totaux
Ministère (UJF)	Reliquats 2004	4 951		
	Subvention 2005 ⁽¹⁾	40 700		
Total				45 651
Contrat UE (UJF)				
LIMES	Solde 2004		80 089	
	2005 ⁽²⁾		23 475	
Total				103 564
CNRS				
	Reliquats 2004		2 908	
	2005 - Soutien de base		18 000	
Total				20 908
NUMDAM (CNRS)				
	Action spécifique			
	Solde 2004		154 515	
Total				154 515
Total Recettes 2005				324 638

⁽¹⁾ Quadriennal 2003–2006 (Totalité de la subvention versée en crédits de Fonctionnement).

⁽²⁾ Solde du contrat : encaissement correspondant au dernier *Cost Statements*, période avril 2003 – mars 2004.

BUDGET 2005

DEPENSES

	Recettes	Ministère (UJF)		CNRS			Total
		UJF	Limes	CNRS	NUMDAM	ANGD	
		45 651	103 564	20 908	154 515		
Code	Nature crédits						
	Équipement						
218	Équipement informatique	4 797		820			
	Équipement Bureaux IF ⁽¹⁾	4 647				114	
615	Équipement non inventorié et Maintenance	619					
	Sous-total	10 063	0	820	0	114	10 883
	Personnel						
64	Frais de personnel ⁽²⁾	23 466	17 466				
	Décharges service	3 747					
	Sous-total	27 213	17 466	0	0	0	44 679
606	Fonctionnement						
	Petit matériel - Consommables Divers	1 281		2 183	1 509		
626	Téléphone	750					
618	Abonnts, Adhésions, Livres	267		5 255	303		
623	Publicité - Publications						
624	Transports de biens						
628	Numérisation / Sté JOUVE		17 790		39 015		
	Frais d'infrastructure	1 350					
	Sous-total	3 648	17 790	7 438	40 827	0	69 703
625	Missions						
	Frais de réception	497					
	Administration						
	Bibliothèques	647		208		1 142	
	Coopération franco-allemande	440		1 647			
	Colloques - Congrès	1 233		2 606			
	NUMDAM	300		2 560			
	POLE REVUES			356			
	Sous-total	3 117	0	7 377	0	1 142	10 495
TOTAL		44 041	35 256	15 635	40 827	1 256	135 759
Solde disponible au 01/09/05		1 610	68 308	5 273	113 688	-1 256	188 879

Frais d'infrastructure à prévoir : 2004 (485 €) et 2005 (970 €)

⁽¹⁾ Meuble sous lavabo + Signalétique extérieure + Alimentation stores électriques (remboursé par Sce Intérieur-UJF/2004)

⁽²⁾ UJF: APE Virginie Imbrogno + CDD Isabelle Costerg (de janv. à oct. 2005)

⁽²⁾ LIMES: CDD Philippe Jacquier-Roux (de juin à déc. 2005) + Isabelle Costerg (oct. à déc. 2005)

Glossaire par ordre alphabétique

AMS	American Mathematical Society
BNF	Bibliothèque Nationale de France
BRAIN	Bibliothèque Rhône-Alpes d'Informations Numériques
CCSD	Centre pour la Communication Scientifique Directe
CEA	Commissariat à l'Énergie Atomique
CEDRAM	Centre d'Édition et de Diffusion de Revues Académiques de Mathématiques
CITIZEMS	Comprehensive Information system Through Integration of the Zentralblatt-MATH Europe-based database in the Mathematical Sciences
CNRS	Centre National de la Recherche Scientifique
DML	Digital Mathematical Library
DML-Eu	Digital Mathematical Library Europe
EDBM	European DataBase manager for Mathematics
ÉDP	Équations aux Dérivées Partielles
ÉDPS	Édition Diffusion Presse Sciences
ÉHESS	École des Hautes Études en Sciences Sociales
ÉNS	École Normale Supérieure
EULER	EUropean Libraries and Electronic Resources in Mathematical Sciences
FIZ Karlsruhe	FachInformationsZentrum Karlsruhe
GDZ	Göttinger DigitalisierungsZentrum
GRETSI	GRoupe d'Étude du Traitement du Signal et des Images
HAL	Hyper Article en Ligne
IHÉS	Institut des Hautes Études Scientifiques
IHP	Institut Henri Poincaré
INIST	INstitut de l'Information Scientifique et Technique
INRIA	Institut National de Recherche en Informatique et en Automatique
IRMA	Institut de Recherche Mathématique Avancée
JFM	Jahrbuch über die Fortschritte der Mathematik
JSTOR	Journal STORage
LGD	Logiciel de Gestion Documentaire
LIMES	Large Infrastructure in Mathematics Enhanced Services
LINUM	Livres NUMérisés Mathématiques
M2AN	Modélisation Mathématique et Analyse Numérique
MEN	Ministère de l'Éducation Nationale
MR	Mathematical Reviews
MSRI	Mathematical Sciences Research Institute
NUMDAM	NUMérisation de Documents Anciens Mathématiques
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
PMB	Pour Ma Bibliothèque
PROLA	Physical Review OnLine Archives
RNBM	Réseau National des Bibliothèques de Mathématiques
SMAI	Société de Mathématiques Appliquées et Industrielles
SME	Société Mathématique Européenne
SMF	Société Mathématique de France
SUDOC	Système Universitaire de DOCumentation
TEL	Thèses En Ligne
UJF	Université Joseph Fourier
WAIS	Wide Area Information Servers
WDML	World Digital Mathematics Library
ZM	Zentralblatt Math

