

Université Joseph Fourier — C.N.R.S.

Cellule MathDoc - UMS 5638

BP 53 - 38041 Grenoble Cedex

www-mathdoc.ujf-grenoble.fr

COMITÉ de PILOTAGE

29 septembre 2003

Cellule de Coordination Documentaire Nationale pour les Mathématiques
UJF – CNRS

Cellule MathDoc – UMS 5638
Université Joseph Fourier
B.P. 53
38041 GRENOBLE Cedex 9
(France)

Fax : 04 76 63 56 11
accueil@mathdoc.ujf-grenoble.fr
www-mathdoc.ujf-grenoble.fr/

SOMMAIRE

INTRODUCTION.....	5
BILAN ET PERSPECTIVES.....	7
A. HISTORIQUE.....	7
B. PERSONNEL.....	7
C. BILAN.....	8
1. <i>Documentation</i>	8
2. <i>Coopération Zentralblatt Math</i>	8
3. <i>Développement de logiciels</i>	8
4. <i>Numérisation</i>	9
D. PERSPECTIVES.....	9
1. <i>Coopération Zentralblatt Math</i>	9
2. <i>Numérisation</i>	9
3. <i>Revue électronique</i>	10
PRINCIPALES ACTIONS DE LA CELLULE MATHDOC.....	11
A. DOCUMENTATION.....	11
1. <i>Catalogue Fusionné des Périodiques</i>	11
2. <i>Catalogue Fusionné des Ouvrages</i>	11
3. <i>Le service Simple de sommaires (sSs)</i>	12
4. <i>Index nationaux des Prépublications et Thèses de mathématiques</i>	12
5. <i>Serveurs d'information (web) de la Cellule MathDoc</i>	13
6. <i>La lettre d'information de la Cellule MathDoc</i>	14
7. <i>Édition de séminaires et conférences</i>	14
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>	17
C. DÉVELOPPEMENT DE LOGICIELS.....	17
1. <i>European Database Manager for Mathematics (EDBM)</i>	17
2. <i>Logiciel de Gestion Documentaire (LGD)</i>	18
D. NUMÉRISATION.....	21
1. <i>Introduction</i>	21
2. <i>Présentation générale du programme NUMDAM</i>	22
3. <i>Les réalisations</i>	23
4. <i>Les perspectives</i>	30
E. COLLABORATIONS NATIONALES ET INTERNATIONALES.....	31
MISSIONS.....	34
PUBLICATIONS.....	36

Introduction

Ce rapport couvre une période de deux ans depuis le dernier comité de pilotage et d'évaluation du 14 septembre 2001. Au cours de ces deux années, la Cellule MathDoc a vécu des changements profonds tant au niveau de ses missions que de sa composition.

Conçue à sa création pour assister les bibliothèques de mathématiques et les mathématiciens pour ce qui concerne la documentation électronique, elle s'est vu confier en 2001 la tâche de numériser les revues mathématiques françaises. Pour lui permettre de réaliser cette opération, son effectif a été porté de trois à cinq permanents mais il a été également nécessaire de recruter des vacataires et de s'adjoindre l'aide d'un enseignant-chercheur. Malgré ces renforts, la numérisation occupe une place très importante dans les activités de la Cellule et dans celles de tous ses membres.

Par ailleurs, Pierre Bérard, qui avait été à l'origine de la création de la Cellule et en avait assuré la direction depuis le début, a été appelé à d'autres tâches et remplacé par un nouveau directeur. Enfin, toujours au chapitre des changements, la Cellule doit quitter ses locaux du CICG pour intégrer de nouveaux bureaux dans la bibliothèque de l'Institut Fourier au printemps 2004.

Tout en développant le programme de numérisation, l'unité a poursuivi les actions entreprises précédemment, collaboration avec Zentralblatt Math, catalogues, services de sommaires (sSs), collaboration avec le Réseau National des Bibliothèques de Mathématiques (RNBM). Le service d'indexation des thèses qui était assuré pour les mathématiques par la Cellule a été intégré dans un service plus vaste assuré par le Centre pour la Communication Scientifique Directe – UPS 2275 (CCSD), la Cellule collaborant à la mise en place du serveur de thèses et de prépublications du CCSD. Le Logiciel de Gestion Documentaire (LGD), développé au départ par la Cellule sera continué dans une collaboration entre le RNBM, le réseau Mathrice des ingénieurs informatiques en mathématiques et la Cellule MathDoc.

Comme on le répète souvent, l'outil principal du mathématicien c'est la documentation et spécifiquement les livres et les revues. Dans ce domaine, de grands bouleversements ont lieu actuellement et les revues françaises doivent rapidement s'adapter à ces nouvelles données. Par les compétences qu'elle a acquises en publication électronique dans le cadre du projet de numérisation, la Cellule MathDoc est à même de leur apporter une aide efficace. C'est donc un chantier qui s'ouvre pour les prochaines années.

Une autre question qui se pose à la Cellule MathDoc, c'est l'avenir de la numérisation. La numérisation des revues françaises est en bonne voie et celle des grands séminaires va commencer. Du fait des compétences acquises, nous sommes sollicités pour plusieurs autres projets de numérisation et il faudra décider de ce qui est possible avec les moyens que nous avons et de ce qui ne l'est pas.

Pour terminer, je voudrais remercier toute l'équipe de la Cellule MathDoc pour le travail accompli et en particulier son directeur adjoint grâce à qui le changement de direction a pu se faire sans problèmes.

Yves Laurent, Directeur de la Cellule MathDoc
Septembre 2003

Bilan et perspectives

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'Unité 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 Math 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 Mathématiques (NUMDAM). 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. Phase active de NUMDAM I. Préparation 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 (TCE), CNRS, depuis juin 1996.
- Catherine Barbe-Zoppis, ingénieur (IR), CNRS, depuis novembre 2001.
- Hélène Bégnis, ingénieur (IE), CNRS, depuis août 2002.

Chercheurs/Enseignants-Chercheurs :

- Yves Laurent, directeur, directeur de recherches CNRS, depuis octobre 2002.
- Laurent Guillopé, directeur adjoint, professeur à l'Université de Nantes, depuis octobre 1995.
- Thierry Bouche, responsable du projet NUMDAM, Maître de Conférences à l'Université Joseph Fourier, depuis octobre 2000.

Vacataires :

- Un ou deux vacataires en permanence sur le projet NUMDAM.

Monique Marchand, dont le travail de secrétariat, de gestion et d'organisation est essentiel pour le fonctionnement de l'unité prendra sa retraite en janvier 2006. Il faudrait commencer la procédure de son remplacement dès l'an prochain.

C. BILAN

Nous diviserons l'activité de MathDoc en quatre grands thèmes : documentation, coopération avec Zentralblatt Math, développement de logiciels, numérisation. Les deux premiers sont à l'origine de la création de MathDoc et le travail accompli ces deux dernières années est dans le prolongement des résultats antérieurs. Le projet de numérisation, plus récent, a pris une importance grandissante et tous les membres de l'Unité y sont impliqués. On lira 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 RNBM, 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. Le principal changement est intervenu dans le domaine des thèses et prépublications avec la création en 2001 du CCSD chargé de les collecter dans tous les domaines scientifiques. La Cellule MathDoc a collaboré de manière significative avec le CCSD pour la mise au point de leur service spécialement en mathématiques.

2. Coopération Zentralblatt Math

La coopération avec le Zentralblatt Math a continué dans ses axes principaux, développement logiciel, fourniture de données bibliographiques de revues françaises, gestion des abonnés français et promotion. Ces actions se déroulent dans la perspective de la transformation du Zentralblatt Math en un grand instrument européen, géré dans ses aspects scientifiques, administratifs et techniques au niveau européen.

3. Développement de logiciels

Le développement du logiciel de recherche et d'indexation EDBM (très largement utilisé, voir ci-dessous) s'est poursuivi dans le cadre du contrat européen LIMES. Après un temps de flottement, le développement du Logiciel de Gestion Documentaire (LGD), destiné aux bibliothèques, a repris en coopération étroite avec le RNBM (plus spécifiquement la bibliothèque d'Orsay) et le réseau Mathrice.

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, se termine en septembre 2003 avec la fin de la numérisation des Annales de l'ENS. La seconde phase commencera immédiatement après avec un nouveau prestataire (la société Jouve).

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 S.M.F., Mémoires de la S.M.F., Annales Scientifiques de l'ÉNS). Les cinq premières sont dès à présent en place sur le serveur Web de NUMDAM. La deuxième phase débutera par la numérisation du Séminaire de Probabilités de Strasbourg, des Annales de l'IHP et de celles de l'Université de Toulouse. Ensuite, nous numériserons 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 (600dpi) doit permettre d'obtenir un résultat durable tout en permettant dans le futur de tirer profit des améliorations des matériels et logiciels de consultation. À ce jour, le projet NUMDAM est le seul à offrir à la fois une recherche 'plein-texte' et des liens hypertextes sur les bibliographies.

D. PERSPECTIVES

1. *Coopération Zentralblatt Math*

Si la coopération avec Zentralblatt Math se poursuit efficacement au niveau technique, elle n'a guère progressé en ce qui concerne la participation de MathDoc (comme des autorités européennes d'ailleurs) à la gestion qualitative de la revue. Une réunion a eu lieu au printemps à la Direction de la Recherche pour étudier ce problème qui devra continuer à être suivi de près.

Une échéance importante interviendra l'an prochain (mai 2004) avec la renégociation du contrat qui a permis aux bibliothèques françaises de conserver un tarif figé pendant quatre ans.

2. *Numérisation*

Après un temps de préparation assez long, le programme de numérisation avance maintenant très vite, ce qui pose deux problèmes, celui de la capacité de MathDoc à poursuivre ce rythme et celui du futur du projet.

Le travail de contrôle de qualité est maintenant bien au point et fonctionne sans à coups, même si le recours à des vacataires ne va pas sans poser des problèmes humains. La mise en place sur le serveur Web prend plus de temps que prévu et pour l'instant, chaque nouvelle revue pose des problèmes différents. Nous avons ainsi réussi à mettre en place rapidement les premières revues numérisées,

mais d'autres secteurs ont pris du retard, en particulier les demandes de cession de droits aux auteurs et la connexion avec les revues vivantes.

La qualité du travail accompli et les compétences développées au sein de l'Unité, dans un domaine encore peu exploré, font que nous sommes sollicités de plusieurs côtés pour des conseils ou pour réaliser un travail de numérisation. Si l'activité de conseil ne pose pas de problèmes, bien au contraire, puisque c'est une des missions de la Cellule MathDoc, la question de savoir si la Cellule MathDoc doit cesser la numérisation avec la fin du programme actuel ou au contraire devenir une agence de numérisation plus ou moins permanente, est cruciale. Pour travailler dans la durée, il lui faudra s'adapter en particulier au niveau du personnel.

3. *Revue électronique*

L'édition scientifique est en évolution très rapide actuellement. La concentration des maisons d'édition commerciales, l'augmentation exagérée du prix des revues, l'importance grandissante des serveurs de prépublications font évoluer les rapports chercheurs-éditeurs-bibliothèques et rechercher de nouveaux modes d'édition. Il nous semble qu'à l'heure actuelle, il est important de soutenir les revues académiques, à la fois garantes de qualité et de modération des prix.

L'autre bouleversement dans l'édition est, bien entendu, l'apparition d'un nouveau mode de diffusion par les réseaux informatiques. Si ce mode 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 trois 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. Le problème de la conservation revêt des aspects multiples selon qu'il s'agisse de conservation à court ou long terme. Dans tous les cas, c'est un problème entièrement nouveau par rapport à la conservation traditionnelle dans les bibliothèques. Enfin, devant le foisonnement d'Internet, la visibilité devient de plus en plus difficile pour un serveur Web isolé. Par ailleurs, 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, ... sont multipliés par la présence sur un même serveur du plus grand nombre de revues.

Par les compétences qu'elle a acquises dans le domaine de l'édition électronique, la Cellule MathDoc est prête à aider efficacement les revues françaises dans ces domaines.

Principales actions de la Cellule MathDoc

A. DOCUMENTATION

1. *Catalogue Fusionné des Périodiques*

Le catalogue fusionné des périodiques est une base de données relationnelle gérée par le logiciel libre MySQL. La base contient environ 8 000 journaux de mathématiques morts et vivants (y compris des journaux électroniques d'accès libre), et pour chaque titre précise :

- 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. De nouveaux journaux sont régulièrement ajoutés à la base.
- Les localisations de ce journal dans les différentes bibliothèques du Réseau National de Bibliothèques de Mathématiques, 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.

L'application informatique en elle-même n'a subi que des modifications mineures en deux ans, mais la base de données est régulièrement tenue à jour :

- Contacts avec les bibliothèques pour récupérer les nouvelles versions de leurs catalogues, réécriture parfois des programmes de retraitement des ces fichiers.
- 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 (Raymond Douet, bibliothèque J. Hadamard, Orsay) pour bâtir une véritable base de données des ouvrages des bibliothèques de mathématiques. La base "Catalogue Fusionné des Ouvrages" 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).

Sur le plan technique, l'application ressemble au catalogue des périodiques. Les différences essentielles sont :

- la méthode de catalogage (ici, les notices sont récupérées à partir des catalogues des bibliothèques, donc saisies par des bibliothécaires, et non à partir d'une base centralisée comme la base ISSN), d'où la nécessité absolue de pouvoir repérer les doublons (les bibliothèques de mathématiques possèdent pour une large part les mêmes ouvrages),
- le volume des données à traiter (environ 90 000 notices contre 8 000).

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

- l'analyse préalable,
- une contribution essentielle à 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.

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.

3. *Le service Simple de sommaires (sSs)*

L'application serveur de sommaires a été entièrement reprise et utilise maintenant le moteur EDBM (cf. ci-dessous). Autrefois basée sur le système WAIS, vieillissant et dont la maintenance n'est plus assurée, cette application était très lourde et posait de nombreux problèmes de maintenance et de mise à jour des données. Sa réécriture complète a permis d'obtenir un système beaucoup plus fiable, qui ne nécessite que peu d'interventions et dont l'évolutivité est grandement accrue.

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. Les données sont mises à jour de façon hebdomadaire par une procédure automatisée.

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.

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 des métadonnées (informations bibliographiques) sur différents serveurs, et leur concaténation au sein d'une même base de données.

La **récolte des 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 [OAIPMH](#), sur les sites du CCSD, [HAL](#) et [TEL](#).

Actuellement, pour être visibles dans l'index, les laboratoires de mathématiques ont deux possibilités :

- Maintenir un site de prépublications chez eux en utilisant l'ancienne méthode (mise en place de fichiers html contenant des "metatags"). Cette méthode a été décrite dans des rapports d'activité antérieurs, et s'est avérée relativement facile à mettre en place, mais pas très facile à entretenir pour un laboratoire. Les problèmes d'archivage, de maintenance et de changement de personnel "webmestre" dans les laboratoires en ont souvent montré les limites. La possibilité théorique existe d'installer dans un laboratoire un serveur de "eprint" plus moderne, plus dynamique, et conforme au protocole OAIPMH, tel que le logiciel "eprints.org", mais nous considérons que les problèmes d'installation et de maintenance risqueraient d'être lourds pour le laboratoire et leur conseillons l'utilisation des services du CCSD.
- Profiter des services centralisés (HAL et TEL) mis en place par le CCSD. Le laboratoire peut déposer ses documents au CCSD et disposer d'une page "personnalisée" permettant d'obtenir "sa" liste. La Cellule MathDoc a développé des programmes pour récupérer les métadonnées des publications de mathématiques déposées au CCSD, et permettre leur retraitement et insertion dans la base de données "index national".

La Cellule MathDoc est un partenaire privilégié du CCSD. Concernant les prépublications et thèses qui étaient autrefois déclarées par des individus sur le serveur de la Cellule MathDoc, cette possibilité a été supprimée et les utilisateurs sont redirigés vers les services du CCSD.

Une collaboration utile pour les deux parties s'est maintenue tout au long de ses deux premières années d'existence du CCSD. Elle a concerné dans un premier temps la mise en place du serveur TEL, puis HAL, et dans l'avenir probablement une récupération rétroactive des documents des laboratoires et l'aide et conseils aux laboratoires de mathématiques.

La base de données "index national" devrait être reconvertie prochainement pour passer de l'utilisation du logiciel actuel "harvest" à un système utilisant des technologies plus nouvelles et conformes au protocole OAIPMH.

5. *Serveurs d'information (web) de la Cellule MathDoc*

En 2002, la décision a été prise de séparer les serveurs d'information en deux :

- un serveur semblable à un serveur web de laboratoire, où la Cellule MathDoc donne des informations sur ses activités et projets, ses publications, son personnel, etc.

- un serveur pour le public, donnant accès à la documentation mathématique : applications de la Cellule MathDoc, comme d'autres sources d'informations connues. Ce serveur appelé "Portail Documentaire Mathématique" a été mis en place fin 2002, et semble apprécié par les utilisateurs, bibliothécaires comme mathématiciens. Des menus déroulants par rubrique permettent de prendre connaissance de toutes les rubriques du site sans quitter celle qu'on est en train de consulter, et sont présents dynamiquement sur toutes les pages du serveur. Un nouveau service, appelé "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 proposés sur ce serveur, la Cellule MathDoc maintient aussi une base de données comprenant les laboratoires et bibliothèques de mathématiques (coordonnées, responsables, site web...). Dans le même ordre d'idée, une base recensant les écoles doctorales et masters de mathématiques est en préparation.

Les deux serveurs sont conçus, et mis à jour, de manière beaucoup plus dynamique que le précédent.

6. *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 3 premiers numéros envoyés nous ont permis d'avoir un retour positif de la part des utilisateurs sur les services que nous proposons. De parution trimestrielle, elle est conjointement envoyée à une liste d'abonnés (environ 120 personnes actuellement, mathématiciens et bibliothécaires) et à des listes de diffusion telles que celle du réseau RNBM ou celle du réseau Mathrice.

7. *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.* le post 2001) seront rapidement insérées dans NUMDAM. Un cédérom, contenant un extrait statique de NUMDAM, sera réalisé à l'automne 2003 (action en coopération avec le Laboratoire J. Leray, avec soutien du BQR de l'Université de Nantes). Par ailleurs, l'édition dans le laboratoire nantais (qui n'est plus parmi les organisateurs des Journées, le GDR Analyse des équations aux dérivées partielles est basé depuis son dernier renouvellement à l'École polytechnique et les Journées ont quitté les bords de l'atlantique) va sans doute cesser. Il est souhaitable que l'insertion des futurs volumes JÉDP (cette collection de référence pour les ÉDPiste sera complétée significativement par la numérisation prévue des Séminaires Schwartz, notamment

celui consacré aux ÉDP qui continue) dans NUMDAM se poursuit : la mise en place d'un système pérenne d'alimentation de NUMDAM pour les publications courantes des séries archivées dans NUMDAM est un des axes de travail des prochains mois de la CMD.

La rédaction des notices de l'Index du Séminaire de probabilités se poursuit sous la direction de M. Émery. L'index, avec sa définition de champs particuliers, coexistera avec NUMDAM, qui accueillera les volumes du séminaire dans quelques mois. L'interface actuelle sera mise à jour, avec un renforcement des liens (NUMDAM, springerlink,...). Le maintien du mode d'affichage utilisant MathML (avec action de l'outil TeX4ht) est à l'étude. Il s'agit d'une bonne plate-forme de test et d'évaluation pour un affichage des formules mathématiques.

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.

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

La CMD est un des partenaires du contrat européen LIMES (2000-2004). Elle a participé à la rédaction du projet CITIZEMS soumis en juin 2003 dans le cadre du sixième plan cadre : en dépit d'une bonne évaluation, le projet n'a pas été retenu dans le premier lot du mois d'août 2003. Ceci est développé dans la quatrième section ci-dessous.

1. *Développement logiciel*

Développé par C. Goutorbe, l'ensemble logiciel EDBM l'a été à l'origine pour le serveur de la base Zentralblatt Math. 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 réseau (<http://mathdoc.ujf-grenoble.fr/ZMATH/zlink.html>), traite des lots de références bibliographiques légèrement balisées (auteur, titre, année de publication). Une extension pour traiter

des références 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.).

Une mise à niveau de la version EDBM 2.0 pour l'application ZM est en cours et sera livrée début 2004 : cet important travail tient compte des différents développements effectués à l'intérieur du projet LIMES et de certains projets internes à la CMD, en anticipant sur certains points (les structures de données n'ont pas encore été clairement définies par le partenaire berlinois par exemple).

2. Fourniture de données bibliographiques de revues françaises

Les revues concernées sont les Annales de l'Institut Fourier, les publications de la SMF, ainsi que celles de la SMAI (éditées par ÉDP-Sciences).

En 2001 (resp. 2002, début 2003), L. Guillopé a effectué 34 (30, 22 resp.) envois, correspondant à 196 (resp. 307, 145) notices. Courant 2002, les URL des articles disponibles sur le réseau ont été systématiquement rajoutées pour les revues ayant un système d'adressage au niveau de l'article, en plus des métadonnées standard (auteur, titre, volume, pages, résumé, mot-clés, classification).

L'ensemble des applications effectuant les migrations des formats des différents éditeurs sera renouvelé fin 2003 : les notices bibliographiques seront alors fournies au bureau berlinois en XML suivant la DTD définie au cours du projet LIMES et utilisée par les bureaux éditoriaux de Copenhague et Lecce. Ces développements sont en phase avec le projet d'alimentation en métadonnées pour les publications courantes des revues numérisées dans NUMDAM.

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 téléchargement de données du ZM dans Pascal. Le FIZ Karlsruhe et l'INIST ont demandé à la CMD d'étudier ces échanges qui posent des problèmes de compatibilité (forme et contenu des données). Cette étude est en cours.

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 public pour 2004 est de 5 300 euros, non compris les 15 % de réduction pour les abonnés de la SME). La renégociation de cet accord (similaire à celui obtenu sous l'égide de la DMV pour les abonnés allemands) devra être renégocié d'ici mai 2004, pour application sur les commandes de l'automne 2004 (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 : des actions de promotion auprès de certaines institutions (INRA, ENSAE,...) pourraient être envisagées.

L'amélioration (avec cohérence accrue) des accès pour les abonnés français est un axe de travail : l'authentification couplée avec les annuaires LDAP gérés par les laboratoires est en projet.

4. *Contrats européens*

Depuis avril 2000, et pour quatre ans, la CMD est 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 CMD, Claude Goutorbe travaille 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 CMD. En continuation de ce contrat s'achevant fin mars 2004, la CMD a participé à la définition du projet CITIZEMS (participation compatible avec les autres projets européens de la Cellule). Bien qu'il n'ait pas été retenu en août 2003, il est à espérer que ce projet sera finalement retenu dans le cadre du sixième plan de l'Union européenne. Dans le cas contraire, le processus d'eupéanisation, si difficile à réaliser concrètement, ne sera pas facilité.

C. DÉVELOPPEMENT DE LOGICIELS

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

a) Le logiciel

Développé à l'origine pour l'interface web de la base de données Zentralblatt Math, le logiciel EDBM (European Database Manager for Mathematics) est maintenant utilisé par de nombreuses autres applications développées par la Cellule MathDoc comme par d'autres institutions.

Les deux composants principaux en sont :

- un module d'indexation, permettant en particulier l'indexation de données de type textuel avec une bonne efficacité et un contrôle fin sur le contenu des index,
- un module de recherche efficace (en terme de rapidité) utilisant un langage de type "équations booléennes".

À ces deux modules de base, vient s'ajouter un ensemble d'utilitaires destinés à simplifier l'écriture d'applications, essentiellement les applications de type réseau (www).

b) Contrats d'utilisation

EDBM a fait l'objet de plusieurs contrats, sans contrepartie financière pour la CMD, avec des partenaires allemands : le tableau suivant en fait le récapitulatif.

	Accès	Organisme	Date	Durée
Zentralblatt Math	**	FIZ Karlsruhe	10/98	2
Jahrbuch	x	SUB Göttingen	05/99	--
MathDi	**	FIZ Karlsruhe	07/00	5
CompuScience	**	FIZ Karlsruhe	02/02	5
Euler	x	SUB Göttingen	05/02	1
Zentralblatt Math	**	FIZ Karlsruhe	09/03	4
Euler	x	Assoc. Euler	10/03	4

L'accès aux bases de données est soit réservé aux abonnés, avec en mode de démonstration (nombre limité de réponses seulement) pour les non abonnés (**), soit totalement libre (x).

c) Applications de la Cellule MathDoc utilisant EDBM

L'ensemble EDBM est utilisé pour de nombreuses applications de la Cellule MathDoc :

- Index du Séminaire de probabilités de Strasbourg
- Journal de mathématiques pures et appliquées (numérisation BnF)
- LiNuM (Livres numérisés mathématiques)
- Répertoire bibliographique des sciences mathématiques
- simple Serveur de sommaires (sSs)
- archives NUMDAM.

Les deux dernières applications utilisent la version 2 : elle a été développée lors de la rénovation du service sSs en juillet 2002, puis avec de notables renforcements lors de l'installation de la consultation à l'archive NUMDAM en décembre 2002. Les autres applications de la CMD utilisant la version 1 de EDBM seront mises à niveau de EDBM2 d'ici début 2004.

2. Logiciel de Gestion Documentaire (LGD)

a) Origine du projet

Le Logiciel de Gestion Documentaire ou LGD est un projet de la Cellule MathDoc en collaboration avec l'Institut Fourier de Grenoble et le Réseau National des Bibliothèques de Mathématiques (RNBM).

Il trouve sa motivation dans le contexte particulier des bibliothèques de mathématiques, confrontées à des problèmes de choix ou d'évolution de leur logiciel de gestion documentaire. Les systèmes intégrés de gestion de bibliothèque disponibles sur le marché sont très complets, mais s'avèrent souvent exhaustifs, complexes et surtout coûteux à l'achat comme en maintenance. Ce sont des outils génériques, non spécifiques au domaine des mathématiques et ils peuvent être ressentis comme contraignants par rapport à la souplesse et l'autonomie dont disposent les bibliothèques de recherche. Ce sont souvent des systèmes propriétaires (codage "privé" des données du fond documentaire, import / export du fond au bon vouloir des propriétaires du logiciel).

b) Développement initial d'un prototype par la Cellule MathDoc

Partant de cette analyse et pour répondre aux besoins précis de certaines bibliothèques de mathématiques, la Cellule MathDoc a entrepris le développement d'un logiciel de gestion documentaire.

Grâce à l'action du Réseau National des bibliothèques de Mathématiques, les bibliothèques concernées ont adopté des normes communes de catalogage (dérivées de la norme internationale MARC) et rendu possible l'échange de notices. Le travail conjoint des ingénieurs de la Cellule MathDoc et des bibliothécaires du RNBM a permis l'établissement d'un cahier des charges précis.

La première maquette du LGD a été développée entre juin 1999 et mars 2000 par Claude Goutorbe, à partir de logiciels libres (Python - MySql – GTK).

Cette version permet la gestion du catalogage, des lecteurs, les opérations de prêt. Il fournit aussi d'autres fonctionnalités utiles (import d'une base existante de type "texto", export des données, nettoyage des index, édition de codes barres, etc.).

Fonctionnel, il a été installé et rapidement mis en service à la bibliothèque de l'Institut Fourier de Grenoble en mars 2000. Depuis, diverses installations ont été réalisées comme à la bibliothèque du laboratoire de mathématiques de Montpellier, à celle d'Orsay ou celle du Centre de Mathématiques et Informatique (CMI) de Marseille.

c) Bilan actuel sur l'utilisation du LGD

Un certain nombre de bibliothèques de mathématiques ont fait la demande du LGD (*cf.* tableau récapitulatif). Suivant le cas, il a simplement été installé et mis en test, ou bien été amené à remplacer le système existant et a même donné lieu à des adaptations locales.

<i>Bibliothèques de mathématiques</i>	<i>Estimation du nbre d'ouvrages</i>	<i>Version du LGD</i>	<i>Usage</i>	<i>Plate-forme utilisée</i>
Institut Fourier	25 000	LGDO	en production	Linux
Montpellier	15 000	LGDO	en production	Linux
Besançon	-	LGDO	en test	Linux
Clermont-Ferrand	-	LGD1	en test	Linux
Orsay	39 000	LGD2	en production	Linux
IHÉS	-	LGD2	en test	Linux
CMI Marseille	15 000	LGD2	en test	Unix - Solaris
Rouen	10 000	LGD2	en test	Linux

Tableau récapitulatif des bibliothèques ayant fait la demande du LGD

Actuellement, on distingue trois versions du LGD.

LGD version 0

Il s'agit du LGD original actuellement en fonctionnement à la bibliothèque de l'Institut Fourier de Grenoble et à Montpellier. Depuis son installation, des fonctionnalités ont été ajoutées au sein de

l'Institut Fourier, comme l'interrogation publique du catalogue via Internet, un historique des prêts par lecteur, un outil de récolement pour les bibliothécaires.

En fonctionnement depuis 1999, le LGD a pu démontrer ses qualités et sa robustesse.

LGD version 1

Cette version se présente comme une réadaptation importante du logiciel initial. Elle a été réalisée à Clermont-Ferrand par Pascale Lefort. L'interface a été réécrite complètement (langage PHP) et la base de données notablement simplifiée. C'est une version assez éloignée de la version d'origine dans la mesure où les procédures fonctionnelles du logiciel ont été réécrites dans leur ensemble, l'interface initiale du LGD et ses fonctionnalités servant essentiellement de cahier des charges. D'un point de vue technique, la validation logicielle de cette version est indépendante de celle du LGD initial. Sa maintenance et les évolutions ne pourront être réinvesties sur la version initiale.

C'est un projet intéressant qui reprend donc les spécifications initiales du LGD, mais en l'adaptant aux récentes technologies logicielles d'Internet, moins robustes, mais plus faciles à appréhender et à interfacer avec les utilisateurs.

LGD version 2

La version 2 du LGD a été développée à la bibliothèque d'Orsay par Raymond Douet. Elle reprend la structure initiale du LGD et le même choix d'outils de développement. Le logiciel a été adapté aux besoins et aux usages des bibliothécaires d'Orsay.

Exception faite de l'utilisation d'une base intermédiaire journalière, toutes ces nouvelles fonctionnalités restent compatibles avec la version initiale du LGD.

d) Avenir du logiciel : projet d'uniformisation et de collaboration au sein du RNBM

En 2002, sous l'égide du Réseau National des Bibliothèques de Mathématiques et plus particulièrement de Raymond Douet d'Orsay, un groupe de travail concernant le LGD et son développement a pu voir le jour.

Ce nouveau projet réunit un ensemble d'ingénieurs informatique et de bibliothécaires (une dizaine de personnes, dont Catherine Zoppis pour la Cellule MathDoc) tous motivés pour faire vivre et évoluer ce logiciel de gestion documentaire pour les bibliothèques de mathématiques.

Plusieurs réunions de préparation se sont tenues entre les différents participants qui donnent lieu à des premiers travaux. Cette initiative a ainsi d'ores et déjà donné lieu au développement d'un portail commun consacré au LGD, à sa distribution, sa documentation et ses développements futurs (cf. <http://www.rnbnm.org/lgd/>). Avec l'aide de ce groupe de travail, toute nouvelle bibliothèque du RNBM qui le souhaite pourra ainsi acquérir les sources du LGD, l'installer et le tester.

Une première tâche importante pour les informaticiens concerne l'adoption d'une version commune du logiciel. Une fusion doit être réalisée entre les deux principales versions compatibles du LGD qui fonctionnent déjà, les versions 0 et 2 respectivement des bibliothèques de Fourier et d'Orsay. Il sera

alors possible de faire évoluer le logiciel tout en en faisant bénéficier chacune des bibliothèques l'ayant adopté. Une version stable du LGD est planifiée pour fin 2003. Elle sera accompagnée d'une documentation informatique ainsi que des procédures pour l'installation du logiciel et la validation des nouvelles fonctionnalités.

Dans le même temps, les bibliothécaires travaillent sur le cahier des charges d'un ensemble de modifications demandées, ainsi qu'à de nouveaux modules envisagés pour la gestion des commandes et la gestion des périodiques. Marie Chebance de l'Institut Fourier soumettra prochainement au RNBM une synthèse pour le cahier des charges du module de gestion des commandes.

Cette entreprise témoigne de l'intérêt réel des bibliothèques de mathématiques pour un logiciel de gestion documentaire, tel que le LGD. Elle confirme aussi la qualité de cet outil, sa fiabilité et les facilités d'évolution qu'il permet.

D. NUMÉRISATION

1. *Introduction*

Différents projets de numérisation de fonds anciens sont en cours dans le monde. De tels projets sont particulièrement importants pour les disciplines, comme les mathématiques, où la documentation se caractérise par sa pérennité. L'un des projets de numérisation les plus importants dans le monde académique est sans doute le projet américain JSTOR (une vingtaine de revues mathématiques américaines, dont certaines plus que centenaies, y sont en particulier disponibles). Il faut mentionner également le projet PROLA de l'American Physical Society, ainsi que d'autres projets à l'université Cornell, à l'université de Göttingen. Pour ce qui concerne la France, signalons l'important fonds disponible sur Gallica (BnF, Bibliothèque nationale de France).

La France produit une vingtaine de revues mathématiques de haut niveau, dont certaines se situent parmi les meilleures au plan international, et dont beaucoup sont plus que centenaies. Il était donc souhaitable de mettre en œuvre une opération importante de numérisation pour ces revues. Outre l'intérêt immédiat pour les mathématiciens comme pour les historiens des sciences, une telle opération augmentera la visibilité des revues mathématiques françaises dans le monde. Il est prévu en particulier que des liens soient établis entre les grandes bases de données mathématiques (*Jahrbuch*, *Zentralblatt Math* et *Mathematical Reviews*) et les fonds numérisés d'une part ; entre les fonds numérisés français et d'autres autres fonds mathématiques importants d'autre part, ceci par le biais d'accords de coopération.

La première phase touche à sa fin et a concerné six revues généralistes (numérisation de quelque 200 000 pages et de 8 000 articles).

2. *Présentation générale du programme NUMDAM*

a) *Les principes généraux*

- Respect de l'original : chaque page est scannée à suffisamment haute définition pour qu'aucun détail ne se perde. Ce sont ces images qui sont publiées, fac-similés fidèles, les enrichissements produits par MathDoc à l'occasion du changement de format étant placés à part.
- Respect de la propriété intellectuelle : un contrat est passé avec l'éditeur, qui fixe le créneau mobile et valide notre méthode de travail. Il s'assure également de l'acceptation des auteurs en leur demandant (avec notre aide) une cession de droits pour publication sur la Toile.
- Identification de l'origine des articles : chaque article comporte une page de garde rappelant le lieu de publication, une brève mention de copyright, et un bandeau identifiant la revue.

b) *Les moyens mis en œuvre*

Actuellement, l'ensemble du personnel de la Cellule MathDoc travaille plus ou moins régulièrement pour le projet NUMDAM :

- Thierry Bouche, conseiller scientifique chargé de la veille technologique et de la représentation dans diverses instances nationales ou internationales ;
- Hélène Bégnis, chef de projet, chargée des relations avec le prestataire de numérisation, du contrôle qualité et de l'encadrement du personnel vacataire ;
- Catherine Barbe-Zoppis, ingénieur développement, chargée de la mise en place des outils de contrôles des données (scripts de contrôle qualité exhaustif, base de données des auteurs, import/export des données, ...) ;
- Claude Goutorbe chargé du développement de l'interface de recherche et de l'établissement de liens externes (Zentralblatt Math et MathScinet) ;
- Elizabeth Cherhal qui a travaillé à la mise au point de notre structure XML des données.

Le programme NUMDAM, d'un point de vue technique, a demandé la production de nombreux développements logiciels et également l'installation et la gestion de deux nouveaux serveurs.

c) *Les choix techniques*

Nos choix techniques ont été conçus pour éviter au maximum deux écueils rencontrés par ce type de projet : si les standards sont trop bas, les données capturées sont rapidement dépréciées par rapport à des productions plus récentes et plus exigeantes ; si les standards sont trop complexes à maîtriser ou à gérer tout au long de la vie du projet, l'expérience prouve que l'on parvient rarement au-delà d'un prototype à moitié fonctionnel. Nous avons donc choisi :

- une qualité de capture des images de 600 dpi monochrome, désormais considérée comme standard au niveau international ;
- un format de catalogage ouvert avec un niveau de détail élevé et intègre.

À ce jour, nous n'exploitons pas encore, sur le site numdam.org, toutes les données recueillies.

d) *Les problèmes juridiques rencontrés*

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 la produit, pour une période allant jusqu'à 70 années après sa mort ; ainsi que du droit moral qui n'expire jamais. En règle générale, il ne peut céder ses droits qu'explicitement, ce qui implique qu'il ne peut pas céder des droits inexistantes lors de la cession. En conséquence, un auteur des années 1920 (ou ses héritiers) dispose aujourd'hui des droits pour toute exploitation de son œuvre relevant des technologies de l'information. L'éditeur dispose *a priori* des droits de l'édition originale sous forme papier, ainsi que des droits "artistiques" portant sur la représentation qu'il lui a conférée (mise en pages, typographie). De ce fait, il est nécessaire de retrouver les ayants droit de tous les auteurs pour les articles parus au XX^e au moins pour obtenir l'autorisation de présenter leurs travaux sur la Toile. Il est également souhaitable de travailler en concertation avec l'éditeur ou le détenteur des droits de l'édition scannée s'il existe encore.

Dans le cas des revues toujours vivantes, nous avons opté pour la solution suivante : demander aux auteurs de céder les droits liés à la publication électronique à leur éditeur, et conclure un contrat de diffusion avec l'éditeur. NUMDAM n'est pas un éditeur, mais juste un service de fourniture de documents numériques sous couvert des éditeurs, la formule est proche de celle d'un "imprimeur sur format numérique".

Pour faciliter la tâche des auteurs et des éditeurs, et obtenir un maximum d'autorisations, de façon à pouvoir démontrer sa volonté de respecter les droits des auteurs le cas échéant, MathDoc met au point une opération de *mailing* avec formulaire préimprimé de cession de droits à retourner à la revue signé. Les envois se feront par vagues successives en remontant dans le temps.

3. *Les réalisations*

a) *État d'avancement du programme*

Le programme NUMDAM a nécessité de longs mois de préparation avant la première mise en ligne des données. La chronologie ci-dessous permet de situer les grandes étapes :

1. de juin 2000 en décembre 2001 : phase d'écriture du CCTP, préparation des collections, choix du prestataire ;
2. de janvier 2002 en décembre 2002 : écriture des spécifications, banc d'essai, lancement de la production ;
3. de décembre 2002 en juillet 2003 : poursuite et finalisation de la production, affichage progressif des données de la production (*Annales de l'Institut Fourier, Journées Équations aux Dérivées Partielles, Publications Mathématiques de l'IHÉS, Bulletin de la SMF* et *Mémoires de la SMF*).

La première phase du programme NUMDAM concerne 6 revues généralistes françaises éditées soit par des éditeurs académiques, soit par des éditeurs commerciaux. La campagne de numérisation, débutée en janvier 2002 par les *Annales de l'Institut Fourier*, s'est terminée en août 2003 avec les *Annales Scientifiques de l'École Normale Supérieure*.

Titre	Volumes	Articles	Nb. pages	Nb. biblio	Nb. bibitem
<i>Annales de l'institut Fourier</i> (créée en 1949 - Association des A.I.F)	156	1 810	51 054	1 725 (~95 %)	25 623 (~14/article)
<i>Annales Scientifiques de l'É.N.S</i> (créée en 1864 - ÉNS contrat avec Elsevier)	295	1 864	68 442	817 (~43 %)	15 617 (~19/article)
<i>Bulletin de la S.M.F.</i> (créée en 1872 - SMF)	167	2 608	45 774	928 (~35 %)	13 157 (~14/article)
<i>Journées E.D.P.</i> (créée en 1974 - CNRS)	28	469	5 580	445 (~94 %)	5 335 (~11/article)
<i>Mémoires de la S.M.F.</i> (créée en 1964 - SMF)	134	396	18 118	374 (~94 %)	6 958 (~18/article)
<i>Publications Mathématiques de l'I.H.É.S.</i> (créée en 1959 - IHÉS contrat avec Springer)	92	344	17 424	327 (~95 %)	7 410 (~22/article)
Total	872	7 491	206 392	4 616	63 120

b) Processus de mise en ligne des revues

Le processus de la mise en ligne des revues s'appuie sur des métadonnées et des fichiers collectés à différents stades de la production pour chacun des volumes à numériser. Ces informations permettent à la fois l'archivage des données bibliographiques, l'organisation de l'affichage des volumes et articles, ainsi qu'une vérification de la concordance avec les données physiques issues de la numérisation et le contenu précis de chaque volume attendu. La mise en ligne des revues répond à une procédure précise dont le but est la gestion cohérente des données bibliographiques et de numérisation. D'un point de vue technique, cette procédure se décompose en quatre étapes :

- import des données initiales de catalogage ;
- réception et contrôle qualité des données issues de la numérisation ;
- import et analyse des métadonnées ;
- export complet des métadonnées.

1. Import des données initiales de catalogage

Le catalogage de chaque revue fournit un premier ensemble d'informations qui initialise la base de données globale de NUMDAM. Ce catalogage, réalisé par du personnel vacataire, se base sur l'utilisation de BibTeX. Un dépouillement est réalisé en parallèle à ce catalogage, dépouillement dont l'objectif est d'apporter une information globale sur la revue : qualité des volumes, particularité de la série, nombre estimé de pages, d'articles, de photos, de planches, etc.

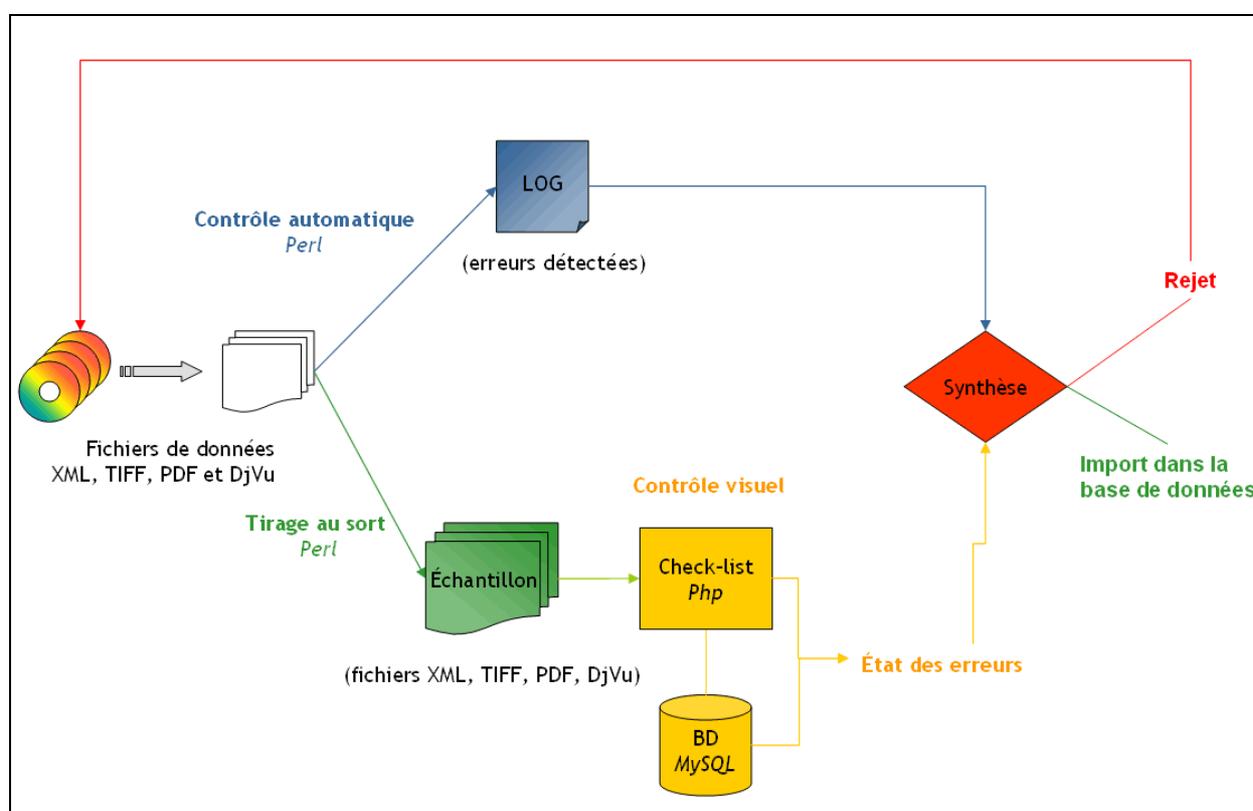
Les données initiales sont ensuite converties au format XML ("*Extensible Markup Language*" ou "Langage Extensible de Balisage"), métalangage présentant l'information encadrée entre des balises. Pour cela, une DTD (*document type declaration*) a été développée par nos soins afin de répondre au mieux à nos besoins. Structure globale et contenu de chaque volume sont ainsi détaillés précisément : numéro ISSN, année, série, tome, fascicule, éditeur, titres, auteurs, contributeurs et détails de chaque article (type d'article, numéro de première et dernière page, langue, relation avec d'autres articles

...). Ces informations sont ensuite importées dans une base de données relationnelle. Nous avons choisi un logiciel libre comme système de gestion, en l'occurrence MySQL.

2. Réception et contrôle qualité des données issues de la numérisation

Un travail important du programme NUMDAM, préalable à la mise en ligne des revues, concerne la mise en place d'une procédure, la plus efficace possible, de contrôle qualité des données issues de la numérisation. Après réception des métadonnées enrichies au format XML et des fichiers physiques de numérisation, nous procédons à deux phases de contrôle qualité. La première contrôle très finement un ensemble d'échantillons tirés au sort, alors que la deuxième procède à une série d'analyses automatiques de l'ensemble des métadonnées et fichiers.

Le schéma ci-dessous résume l'ensemble de la procédure mise en place.



Le contrôle automatique exhaustif des données

Différents traitements sont appliqués à l'ensemble des données de chaque volume numérisé. Il s'agit de s'assurer de l'intégrité et de la validité des données de numérisation et de contrôler leur cohérence avec les informations issues du pré-catalogage. Cette procédure travaille avec trois ensembles de données : celles issues du pré-catalogage, celles décrivant la numérisation des volumes et l'ensemble des fichiers physiques de numérisation. Les métadonnées de catalogage et de numérisation sont tout d'abord analysées : présence des renseignements attendus, cohérence interne des données (suite croissante des pages par exemple, déduction du nombre de pages d'après les numéros des premières et dernières pages, ...).

Vis à vis de l'étape de numérisation, nous nous assurons que tous les fichiers correspondant à un volume sont bien présents, correctement nommés et localisés dans l'arborescence physique prévue.

Leur format ('xml', 'tiff', 'pdf' ou 'djvu') et leur validité sont ensuite testés. Puis, nous extrayons le maximum d'information de ces fichiers pour les confronter aux données bibliographiques de catalogage :

- correspondance entre la nomenclature des volumes et des articles (identifiants) et les noms de répertoires ou fichiers physiques correspondants,
- nombre de pages des fichiers articles comparé aux données de catalogage fournissant les numéros des pages initiales et finales de chaque article,
- résolution et largeur des fichiers image,
- correspondance entre la langue de l'article et celle de l'outil de reconnaissance optique de caractère (OCR) utilisé pour cet article, ...

Ces deux étapes permettent ainsi de repérer différents problèmes et d'y remédier rapidement : corrections d'erreurs dans les données bibliographiques, réfection de données physiques endommagées ou manquantes.

Le contrôle manuel sur échantillon

Le contrôle manuel permet de vérifier la qualité visuelle des fichiers (monopages et articles), la qualité des bibliographies, la cohérence du nombre de pages,... . Ce contrôle s'effectue à l'aide d'une checklist (développée en php et couplée à une base de données mysql) dont l'objectif est de garder trace de tous les fichiers contrôlés et des erreurs détectées par ce contrôle. Actuellement, ce contrôle est effectué par une personne vacataire employée à temps partiel (50 % ou 80 % selon les besoins) en contrat à durée déterminée.

À chaque revue correspondent deux échantillons de données : un échantillon de fichiers monopages et un échantillon d'articles. La taille de ces échantillons est définie selon la norme NF X 06-021 (principes du contrôle statistique de lots), ainsi que le taux d'erreurs pour chaque échantillon. Généralement, le taux d'acceptation d'un lot est de 4 erreurs pour 315 pages contrôlées. Nous n'avons pas été amenés à rejeter un lot dans sa globalité. Nous avons seulement demandé la réfection des fichiers erronés.

À l'issue de ces deux contrôles, une synthèse est réalisée et est envoyée au prestataire sous forme d'une fiche de recette. La réfection des fichiers erronés est demandée sous huitaine.

Les principales erreurs constatées

À ce stade de la production (5 revues sur 6 sont actuellement en ligne, la 6^e est en cours de mise en ligne), nous pouvons déterminer une certaine typologie des erreurs relevées par les différentes phases du contrôle qualité. La plus fréquente concerne les fichiers djvu. Nous avons, en effet, constaté que bon nombre de fichiers djvu étaient 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.

Cette expérience va nous permettre d'aborder plus sereinement le deuxième marché qui démarre à la rentrée 2003.

3. Import et analyse des données

À l'issue de la phase de contrôle qualité, la base de données est enrichie avec les données bibliographiques de numérisation : paramètres de numérisation, lien entre chaque article et les fichiers physiques de numérisation qui lui correspondent, bibliographies balisées de l'article, etc. Nous pouvons encore travailler certains points qui permettent d'affiner la qualité globale des métadonnées.

– **Qualité des données** : une comparaison fine entre nos trois types de données (initiales et de numérisation tour à tour importées dans la base, puis exportées de la base, c'est-à-dire reconstruites) permet de révéler un certain nombre d'incohérences. Ce changement de format (XML versus MySQL) permet ainsi de repérer des incohérences dans la structure des volumes ou des articles, ou bien des modifications apportées dans les données bibliographiques post numérisation, qui peuvent être des erreurs ou des corrections. Cette conversion des données peut aussi signaler des caractères Unicode mal encodés. Nous avons fait le choix d'un codage des caractères au format international Unicode (stocké en UTF-8), jeu de caractère universel permettant notamment l'encodage des caractères cyrilliques, grecs mais aussi des diacritiques. Nous pouvons ainsi encoder, de façon exacte, noms d'auteurs et titres d'articles et vérifier l'écriture de cet encodage. Les articles sont d'autre part reclassés en fonction de leur type afin de fournir un affichage le plus logique de chaque volume (gestion des préliminaires, suppléments, pages volantes, errata, etc.).

– **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 nom d'auteur,... Nous avons, à cette fin, mis en place une base de données de référence de l'ensemble des auteurs. Cette base nous permet d'identifier précisément chaque auteur, de repérer et corriger des noms d'auteurs mal orthographiés suite à une erreur 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.

4. Export complet des données

En fin de traitement, nous obtenons un ensemble de données bibliographiques concernant chaque volume numérisé. Il décrit à la fois sa structure, son contenu et sa numérisation. Ces métadonnées peuvent ensuite être facilement maniées ou échangées via le format de balisage XML sous lequel elles sont exportées.

5. L'interface d'accès aux données et aux documents

Modes d'accès aux données bibliographiques et aux articles

Les données bibliographiques et les articles sont accessibles de deux manières différentes : d'une part par le feuilletage, d'autre part par l'interface de recherche. Ces deux accès sont basés sur l'outil EDBM (*European Database Manager for Mathematics*), développé par Claude Goutorbe, également utilisé par d'autres applications telles que Zentralblatt Math ou le Service de Sommaires. Des développements spécifiques ont été réalisés pour le programme NUMDAM, plus particulièrement pour la gestion des recherches croisées entre articles et bibliographies et pour l'établissement de liens entre la base NUMDAM et la base Zentralblatt Math (lien entre un article ou une référence bibliographique et son analyse dans Zentralblatt Math).

Le **feuilletage** volume par volume est 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 (format djvu et pdf) et aux analyses MathReviews et ZentralBlatt Math.

Exemple de feuilletage pour les Publications de l'IHÉS :

1959-1979	1980-2000
tome 1 (1959)	tome 26 (1965)
tome 2 (1959)	tome 27 (1965)
tome 3 (1959)	tome 28 (1966)
tome 4 (1960)	tome 29 (1966)
tome 5 (1960)	tome 30 (1966)
tome 6 (1960)	tome 31 (1966)
tome 7 (1960)	tome 32 (1967)
tome 8 (1961)	tome 33 (1967)
tome 9 (1961)	tome 34 (1968)
tome 10 (1961)	tome 35 (1968)
tome 11 (1961)	tome 36 (1969)
tome 12 (1962)	tome 37 (1969)
tome 13 (1962)	tome 38 (1970)
tome 14 (1962)	tome 39 (1970)
tome 15 (1963)	tome 40 (1971)
tome 16 (1963)	tome 41 (1972)
tome 17 (1963)	tome 42 (1973)
tome 18 (1963)	tome 43 (1974)
tome 19 (1964)	tome 44 (1974)
tome 20 (1964)	tome 45 (1975)
tome 21 (1964)	tome 46 (1976)
tome 22 (1964)	tome 47 (1977)
tome 23 (1964)	tome 48 (1978)
tome 24 (1965)	tome 49 (1979)
tome 25 (1965)	tome 50 (1979)

Figure 1 : feuilletage des Publications de l'IHÉS

Précédent: tome 33 (1967) Suivant: tome 35 (1968) Retour: liste des volumes
Grauert, Hans Affinoide Überdeckungen eindimensionaler affinoider Räume p. 5-35 Notice complète Texte intégral djvu pdf Analyses MR 39 #2342 Zbl 0197.17302
Serre, Jean-Pierre Groupe de Grothendieck des schémas en groupes réductifs déployés p. 37-52 Notice complète Texte intégral djvu pdf Analyses MR 39 #1152 Zbl 0195.90802
Mostow, G. D. Quasi-conformal mappings in SnS-space and the rigidity of hyperbolic space forms p. 53-104 Notice complète Texte intégral djvu pdf Analyses MR 38 #4679 Zbl 0189.09402
Segal, Graeme Classifying spaces and spectral sequences p. 105-112 Notice complète Texte intégral djvu pdf Analyses MR 38 #718 Zbl 0199.26404
Segal, Graeme The representation-ring of a compact Lie group p. 113-128 Notice complète Texte intégral djvu pdf Analyses MR 40 #1529 Zbl 0209.06203
Segal, Graeme Equivariant KS-theory p. 129-151 Notice complète Texte intégral djvu pdf Analyses MR 38 #2769 Zbl 0199.26202
Précédent: tome 33 (1967) Suivant: tome 35 (1968) Retour: liste des volumes

Figure 2 : table des matières du tome 34 (1968)

L'interface de recherche offre un accès multi-revues et peut se faire sur les champs suivants : "journal", "auteurs", "mots du titre", "années" et "plein-texte". Il est possible de restreindre la recherche à une ou plusieurs années et de faire un tri sur le nom d'auteur. En outre, l'interface de recherche du serveur numdam.org offre 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.

Les possibilités de recherche peuvent encore être améliorées et c'est un des objectifs pour l'année à venir. En effet, nous travaillons actuellement à la possibilité de localiser et surtout d'indiquer clairement à l'utilisateur dans quelle(s) page(s) de l'article se trouve le terme recherché dans le "plein-texte".

Le créneau mobile

Parallèlement à ces opérations purement techniques, la diffusion du texte intégral des articles est régie par le *créneau mobile*. Ce créneau mobile est une période, définie par et pour chacune des revues, durant laquelle les articles ne sont pas librement accessibles depuis le serveur www.numdam.org.

Revue	Créneau mobile	Disponibilité des données bibliographiques	Disponibilité du texte intégral
<i>Annales de l'IF</i>	5 ans	1949 – 2000	1949 – 1997
<i>Journées ÉDP</i>	Nul	1974 – 2000	1974 – 2000
<i>Publications de l'IHÉS</i>	5 ans	1959 – 2000	1959 – 1997
<i>Bulletin de la SMF</i>	10 ans	1872 – 2000	1872 – 1992
<i>Mémoires de la SMF</i>	10 ans	1964 – 2000	1964 – 1992

c) La gestion technique des droits d'auteurs

La base de données des auteurs décrite dans le point 3 nous sert également de support à la campagne de contact des auteurs, afin de nous conformer à la réglementation en vigueur. Notre objectif est de contacter l'ensemble des auteurs contemporains pour qu'ils puissent céder leurs droits. Nous avons donc fait des recherches sur les coordonnées de ces auteurs, afin de pouvoir les contacter. Cette opération de mailing consiste donc à envoyer par courrier électronique un contrat type par lequel l'auteur cède ses droits à la revue concernée pour les articles qu'il a publiés. Dans cette optique, cet outil nous sert également d'outil de gestion des réponses qui se feront par courrier.

Cette campagne se fera par vagues successives de manière à gérer plus efficacement le flux des réponses. Idéalement, nous souhaiterions contacter la quasi-totalité des auteurs contemporains et recueillir l'accord de la plupart d'entre eux. Dans le cas de refus de diffusion sur le Web, nous procéderons immédiatement au retrait du ou des articles concernés.

d) Actions de promotion de NUMDAM : les journées de Nantes (janvier 2003)

À l'occasion de l'ouverture du serveur NUMDAM, une conférence "JÉDP : un quart de siècle numérisé" a été organisée à Nantes les 7-8 janvier 2003, en collaboration avec le laboratoire J. Leray (UMR 6629 UN/CNRS) et le GDR Analyse des équations aux dérivées partielles (GDR 2434 CNRS).

Après deux demi-journées d'exposés mathématiques par J.-M. Bony, B. Helffer, N. Lerner, D. Robert, une après-midi a été consacrée à la présentation du programme NUMDAM et la discussion des problèmes juridiques du document numérique (avec la participation de A. Lucas, professeur de droit, expert international).

4. *Les perspectives*

a) *Le futur immédiat*

Nous souhaitons accroître le volume des données accessibles depuis le serveur NUMDAM, et des modes d'accès. Il nous reste aujourd'hui de nombreuses pistes de réflexion dans le cadre du programme NUMDAM, pour améliorer la visibilité, l'efficacité et l'utilité de ce service. Nous souhaitons principalement mener à bien deux chantiers : l'intégration de la production contemporaine des revues et la mise en place d'un point d'accès commun aux revues numérisées dans le monde.

Une de nos principales activités de l'année à venir est donc l'intégration dans le programme NUMDAM de la production contemporaine des revues de la phase I. Nous proposons de commencer ce travail avec les *Annales de l'Institut Fourier* qui nous serviront de laboratoire d'essai.

Nous menons actuellement une réflexion sur la mise en place d'un point d'accès commun international aux données bibliographiques et articles de revues, numérisés par d'autres programmes tels Gallica, Göttingen ou JSTOR, ainsi qu'aux articles récents produits de façon nativement numérique. Il s'agit de mettre en place une base de données contenant les notices bibliographiques de base ainsi que le lien vers l'hébergeur des articles, qui ne seront pas nécessairement d'accès libre. Cette démarche s'inscrit dans les objectifs à court terme de la bibliothèque numérique universelle de mathématiques (DML : *Digital Mathematics Library*), projet auquel les membres de la Cellule MathDoc ont été associés dès les origines (Pierre Bérard fut membre du *Steering Committee*, Thierry Bouche du *Technical working group* de la DML pendant sa phase initiale (NSF planning grant) ; ils sont actuellement membres du comité WDML de l'union mathématique internationale chargé de conduire la poursuite de ce projet).

b) *Le lancement d'un deuxième marché*

Dans la suite du premier marché NUMDAM, qui a concerné 6 revues françaises de premier ordre, nous avons lancé un deuxième marché dont l'objectif est la numérisation de grands séminaires et des séries non encore traitées.

Ce deuxième marché démarrera à la rentrée 2003 avec la numérisation du séminaire de Probabilité de Strasbourg (quelque 1 000 articles et 16 000 pages), puis des *Annales de l'IHP*. Suivront ensuite les *Annales de l'Université de Toulouse*, puis un certain nombre de grands séminaires.

c) *Les partenariats actuels et futurs du programme NUMDAM*

La participation à des projets régionaux (avec le programme BRAIN : Bibliothèque Rhône-Alpes d'information numérique), européens (DML-Europe) et internationaux (DML, EMANI) donne au programme NUMDAM une visibilité et un champ d'action plus vaste, ainsi que des perspectives de partenariats féconds.

Nous avons par exemple été contactés, suite à la conférence DML de Göttingen courant juin 2003, pour numériser une série majeure en mathématiques. Il est également question d'un projet de numérisation d'ouvrages dans le cadre de la restructuration de la bibliothèque de Jussieu. Cependant, ces projets demandent une réflexion tant sur le plan technique (une adaptation de notre modèle est

nécessaire, les contraintes n'étant pas les mêmes), juridique (les enjeux du droit d'auteur pour les ouvrages sont différents et plus ardues que ceux des articles de revues) qu'organisationnel.

E. COLLABORATIONS NATIONALES ET INTERNATIONALES

La Cellule MathDoc collabore avec de nombreux organismes et ces collaborations apparaissent tout au long du rapport. Nous rappellerons seulement ici les principales :

Réseau National des Bibliothèques de Mathématiques

C'est le partenaire naturel de la Cellule MathDoc pour tout ce qui concerne la documentation. Depuis quelques mois, nous collaborons également sur le logiciel de gestion documentaire. Sa transformation en GDS devrait faciliter les relations institutionnelles.

La Cellule MathDoc et le RNBM préparent un projet d'école thématique ayant pour thème, "La Documentation en mathématique". Le dossier administratif a été déposé en juillet 2003 auprès de la Formation Permanente de la Délégation Régionale du CNRS de Grenoble. Cette école se tiendra au CIRM (Centre International de Rencontres Mathématiques) à Marseille-Luminy, du 11 au 15 octobre 2004.

Une première école avait eu lieu en mars 1998.

Centre pour la Communication Scientifique Directe

Nous participons à la mise en place du serveur de thèses et de prépublications, en particulier pour ce qui concerne les mathématiques.

Réseau Mathrice

L'aide du réseau Mathrice nous a été précieuse pour résoudre divers problèmes techniques, comme la mise en place de l'adressage du serveur NUMDAM, nous collaborons également pour le logiciel de gestion documentaire.

Zentralblatt Math

Projets internationaux de numérisation

Plusieurs projets internationaux (Digital Mathematical Library, DML-Europe, EMANI) tentent de se mettre en place pour structurer au niveau international l'accès aux documents électroniques. La Cellule MathDoc y participe avec, entre autres, le RNBM, les bibliothèques de Cornell et Göttingen.

Enfin, une collaboration qui n'est pas encore apparue dans ce rapport :

Pôle associé "Numérisation concertée en mathématiques" avec la Bibliothèque nationale de France (BnF)

Le pôle (le seul à ce jour dans le domaine des mathématiques) a été créé en novembre 2002, par signature d'une convention entre la BnF, le CNRS et l'université Joseph Fourier. Sa nature scientifique est spécialement forte, puisque la quasi-totalité des pôles associés sont établis avec des organismes gérant en propre des fonds documentaires. Le cœur des actions concerne le développement des bibliothèques numériques : politique concertée de numérisation, amélioration des interfaces et enrichissement des métadonnées, archivage à long terme, analyse raisonnée des fonds en préparation de numérisation, etc.

Le projet LiNuM (Livres NUmérés Mathématiques), initié en janvier 2002, est une base de métadonnées décrivant des ouvrages mathématiques (monographies, oeuvres complètes, etc.) numérisés, avec accès direct à la version numérique de l'ouvrage librement accessible sur la toile. Actuellement, la base contient près de 2 000 références : outre le fonds mathématique de Gallica (y compris les ouvrages non visibles sur le réseau internet actuellement pour des questions de droits), il y a quatre collections importantes indexées : la *Digital Math Books Collection* de l'Université de Cornell, *Historical Math Collection* de l'Université de Michigan, *Mathematica* de la bibliothèque de Göttingen et la *Biblioteka Wirtualna Matematyki* du centre ICM à Varsovie.

Cet outil, signalé en bonne place dans l'article de A. Jackson (Notices AMS, 2003), a un public international (à peine 50 % d'utilisateurs français), avec 45 % des 800 (en moyenne) ouvrages sélectionnés mensuellement provenant du fonds Gallica. L'alimentation des nouveautés du site Gallica sera bientôt effectuée par fourniture de données intermarc par la BnF directement.

Le Journal de mathématiques pures et appliquées (JMPA) est, pour la période 1836-1880, numérisé dans Gallica : l'accès via l'interface Gallica se fait volume par volume. En mars 2001, un dépouillement article par article a été réalisé par la CMD, avec installation d'une interface d'interrogation et de consultation donnant accès directement à l'intégralité de tout article sélectionné. Des liens vers les recensions parues dans le Jahrbuch, si elles existent (le Jahrbuch a été créé en 1868), ont été introduits (ces données étant reprises peu après par le Jahrbuch sous forme de liens directs du Jahrbuch vers Gallica) ; des liens à partir du RBSM (*cf.* ci-dessous) ont été installés plus récemment.

La BnF a programmé la numérisation de la période 1881-1940 du JMPA : la base JMPA actuelle va être complétée, avec les liens croisés adéquats et une amélioration de la qualité du dépouillement (des vérifications importantes sont à faire, des vacations seraient utiles). Le prolongement de cette numérisation (où des négociations avec le propriétaire actuel, les éditions Elsevier France) est à étudier, ainsi que l'insertion à moyen terme dans NUMDAM (la différence des standards techniques adoptés par la BnF et NUMDAM pose problème) ou dans la WDML.

Créé sous la direction de Poincaré, le *Répertoire bibliographique des sciences mathématiques* [RBSM] a eu pour ambition à la fin du XIXe siècle de classer thématiquement les œuvres mathématiques pertinentes du siècle écoulé. Environ 20 000 documents ont été ainsi signalés dans le *Répertoire*. La saisie textuelle structurée du RBSM a été initiée par L. Rollet et P. Nabonnand (UMR 7117, Univ. Nancy 2 & CNRS) en 2001 et accompagnée rapidement d'un montage sur la toile par la CMD. En juillet 2003, où 25 % du RNSM est saisi, le laboratoire nancéen a obtenu un soutien financier de la Région Lorraine lui permettant d'achever la saisie rapidement. La BnF a par ailleurs décidé de la numérisation en mode image des fiches du Répertoire (un seul exemplaire complet des fiches est connu : il est dans une collection privée). En 2004, l'accès à la version numérique du RBSM combinera une base de données (interrogation sur divers champs, navigation hypertexte, divers critères de tris pour l'affichage) et l'ensemble des fiches originales numérisées en mode image. La CMD insérera cet affichage dans la WDML, notamment par des liens directs vers des documents de Gallica et de NUMDAM.

La CMD, en collaboration avec le RNBM, a, en juin 2003, effectué le signalement systématique et bien documenté de toutes les bibliothèques de mathématiques de France dans le Répertoire des bibliothèques et centres documentaires administré par la BnF.

Une présentation du Pôle a eu lieu lors de la Journée des pôles associés organisée à Marseille le 3 juillet 2003. L'exposé de Laurent Guillopé a souligné les coopérations entre la CMD et la BnF, existantes ou en devenir, dans trois projets : le JMPA, le programme NUMDAM et le RBSM. D'autres actions (archivage des serveurs web de la Cellule et des données numériques de NUMDAM, inventaires de séries de documents mathématiques, encadrement d'un chargé de recherche documentaire en poste à la BnF et dans une Unité de recherche,...) sont en cours.

Le Comité scientifique, composé à parité de trois représentants de la BnF et trois personnalités nommées par la Cellule MathDoc (Liliane Zweig représente le RNBM, Karine Chemla a été proposée par la SMF, Laurent Guillopé représente la CMD) s'est réuni pour la première fois le 1^{er} septembre 2003.

Missions

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

Catherine Barbe-Zoppis [CBZ], Hélène Bégnis [HB], Pierre Bérard [PB], Thierry Bouche [ThB], Elizabeth Cherhal [EC], Claude Goutorbe [CG], Laurent Guillopé [LG], Yves Laurent [YL], Monique Marchand [MM].

1. Réunions du projet européen LIMES

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

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

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

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

2. Réunions du comité de coordination du Zentralbaltt 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]

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]

4. Projet DML (Digital Mathematics Library)

24-31 juillet 2002 – Washington (États-Unis)

"Digital Mathematics Library" organisé par Cornell University et National Science Foundation, Washington. [PB]

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]

5. Congrès 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 Équations 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]

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. 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]

Publications

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

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

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

Rapport Comité de Pilotage
Cellule MathDoc

29 septembre 2003