



Université Joseph Fourier —
C.N.R.S.
**Cellule MathDoc - UMS
5638**

RAPPORT D'ACTIVITÉ

Cellule MathDoc – UMS 5638

2005 – 2009

Cellule de Coordination Documentaire Nationale pour les Mathématiques
UJF – CNRS

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

Tél. : 04 76 63 56 36
Fax : 04 76 63 56 11
ums5638@math-atlas.ujf-grenoble.fr
www-mathdoc.ujf-grenoble.fr

Table des matières

Introduction.....	3
A.Rapport d'activité 2005-2009.....	5
I.Présentation de l'unité.....	5
II.Bibliothèque numérique.....	5
1.Numérisation.....	6
2.Acquisition.....	10
3.Bibliothèques virtuelles.....	11
4.Infrastructures.....	12
III.Édition électronique.....	13
IV.Services vers les bibliothèques.....	17
1.Catalogues fusionnés.....	17
2.Actions de collaboration avec le RNBm.....	17
V.Visibilité.....	18
1.Accès public par le Web.....	18
2.Coopération européenne et internationale.....	19
3.Coopération avec le Zentralblatt.....	20
4.Activités de conseil et d'expertise.....	20
B.Projets.....	22
I.Perspectives et difficultés.....	22
II.Bibliothèque numérique.....	23
1.Projets européen EuDML.....	23
2.Numérisation.....	24
3.Acquisition.....	25
4.Bibliothèques virtuelles.....	26
III.Édition.....	26
IV.Archivage.....	27
V.Réorganisation de la documentation mathématique.....	28
VI.Relations nationales et internationales.....	28
C.Annexe.....	30
II.Publications.....	41

Introduction

La Cellule MathDoc est une Unité Mixte de Service spécialisée dans la documentation mathématique. Sa mission concerne l'ensemble des laboratoires et des bibliothèques de mathématiques et, par sa vocation nationale, elle est dans une position originale et ne peut guère se classer dans les structures fédératives comme d'autres UMS. En fait son fonctionnement s'apparente plutôt à celui d'une UMR.

L'unité a été créée en 1995 avec pour première mission l'aide à l'informatisation des bibliothèques. Cette mission se poursuit par la gestion de ressources communes et l'organisation régulière de journées de formation pour les bibliothécaires.

La deuxième mission était la coopération avec le Zentralblatt Math. Elle s'est traduite par la fourniture de logiciels à Zentralblatt et se poursuit à plusieurs niveaux, échanges de données, partages de compétences, organisation.

L'étape suivante fut le programme de numérisation. Dans le projet initial, il s'agissait d'aider les revues de mathématiques françaises en numérisant leurs numéros anciens. En 2000, MathDoc lança le projet Numdam avec dès le départ des exigences élevées en termes de qualité de la numérisation et de qualité des métadonnées associées. Neuf ans plus tard, on peut considérer que le but est atteint (ce qui ne signifie pas qu'il ne reste rien à numériser!) et le projet Numdam est une référence dans le monde entier.

Si dans le rapport quadriennal précédent, un grand chapitre était consacré à la numérisation, dans celui-ci il est remplacé par un chapitre intitulé "Bibliothèque Numérique". En effet, autour du projet Numdam se sont organisés plusieurs autres services ou projets qui marquent l'évolution de notre travail vers une offre globale. C'est ainsi que nous avons obtenu de la plupart des éditeurs la fourniture des articles récents des revues que nous offrons en libre accès avec un certain délai après parution. Parmi les autres services offerts, citons des catalogues et des moteurs de recherches (Linum, mini-DML) et surtout le projet Cedram d'édition de journaux électroniques lancé en 2005.

La situation de la documentation mathématique continue à évoluer rapidement. En ce qui concerne les revues, la version électronique tend à supplanter la version papier qui bientôt ne servira plus que d'archivage. En réalité, dans ces conditions la version papier ne sera plus rentable et risque de disparaître bientôt. Le danger de cette situation est qu'elle transfère aux éditeurs commerciaux, une des fonctions majeures des bibliothèques, la conservation des documents.

L'INSMI, le nouvel institut des mathématiques au CNRS a entrepris une réflexion sur l'avenir des bibliothèques de mathématiques et plus particulièrement sur le devenir des revues au sein de ces bibliothèques. Avec ses partenaires du RNBM (Réseau National des Bibliothèques de Mathématiques) et de MATHRICE (Réseau des ingénieurs informatiques), MathDoc participera à cette évolution qui concernera à la fois les abonnements, les accès électronique et la conservation. Après l'édition électronique, notre prochain projet sera donc la mise en place d'un système d'archivage à moyen et à long terme suffisamment sécurisé. Il devra conserver aussi bien nos propres productions que celles des autres revues académiques françaises mais également les productions des éditeurs commerciaux avec qui des accords devront être passés dans le cadre des consortiums.

La coordination au niveau européen et au niveau international des différents acteurs de la numérisation, l'établissement de normes pour les échanges de données sont évidemment très souhaitable. C'est un sujet qui avance lentement même si un grand nombre de relations bilatérales ont pu être établies. Nous devons continuer nos efforts dans ce sens et le projet européen EuDML qui vient de débiter pour trois ans nous aidera à avancer dans cette direction.

Pour terminer, je voudrais parler des effectifs de l'unité qui malgré des renforts venus régulièrement du CNRS sont encore bien faibles. Cette année 2009 est particulièrement difficile avec un congé de longue durée et un départ à la retraite (poste universitaire) non remplacés.

Yves Laurent
Directeur de l'UMS 5638

A. Rapport d'activité 2005-2009

I. Présentation de l'unité

La Cellule MathDoc est une Unité Mixte de Service (UMS 5638) sous la co-tutelle du Centre National de la Recherche Scientifique et de l'Université Joseph Fourier.

Personnel permanent :

- Claude Goutorbe, ingénieur (IR) Université Joseph Fourier, depuis octobre 1995.
- Catherine Barbe-Zoppis, ingénieur (IR) CNRS, depuis novembre 2001.
- Hélène Falavard, ingénieur (IE) CNRS, depuis août 2002. [en congé post-natal]
- Myriam Charles, assistante (TCE) CNRS, depuis janvier 2005.
- Frank Lontin, ingénieur (IE) CNRS depuis janvier 2008
- Isabelle Costerg, technicienne CNRS depuis décembre 2009

Contractuels :

- Nicolas Houillon ingénieur (IR) sur le projet européen

Elizabeth Cherhal, ingénieur (IE) Université Joseph Fourier, depuis novembre 1995 a pris sa retraite en avril 2009. Elle n'a pas été remplacée pour l'instant.

Direction :

- Yves Laurent, directeur de recherches CNRS, est directeur de l'unité depuis octobre 2002. Il ne souhaite pas son renouvellement pour le prochain quadriennal.
- Thierry Bouche, Maître de Conférences à l'université Joseph Fourier, responsable du projet NUMDAM depuis octobre 2000, est directeur adjoint. Il est proposé comme prochain directeur.

II. Bibliothèque numérique

Lancé comme un projet pilote de numérisation des principales revues académiques françaises de mathématiques, le projet Numdam a débordé largement de son cadre initial pour devenir une véritable bibliothèque numérique. Les efforts récents se sont portés naturellement sur la poursuite du programme de numérisation, qui est entré dans sa troisième phase début 2008, pour une couverture toujours plus étendue des sciences mathématiques (revues de mathématiques appliquées ou de statistiques, revues françaises anciennes, thèses de l'entre-deux guerres,...). Mais il aura fallu beaucoup d'efforts pour que Numdam ne soit pas seulement une archive morte, mais une bibliothèque vivante qui acquiert la production des éditeurs partenaires au fil du temps.

La première étape fut l'accord obtenu avec Elsevier, qui nous a permis de mettre en ligne les références de tous les articles parus jusqu'en 2006 dans les revues déjà présentes sur Numdam. Suivirent Springer, EDP Sciences et les revues que nous publions sur la plateforme CEDRAM. Aujourd'hui, plus de 10 % du contenu navigable sur le serveur www.numdam.org a été fourni directement par les éditeurs. De nouvelles chaînes d'acquisition ont été développées, qui permettent d'ingérer les formats propriétaires des éditeurs dans notre base de

données, assurant à leur production une meilleure visibilité, et la garantie d'un archivage pérenne.

Par rapport à d'autres centres de documentation électronique, l'originalité de Numdam peut se résumer en quelques points:

- Validation scientifique: ne sont présentés que des articles ou des thèses qui ont été validés et publiés, à l'exclusion de toute prépublication
- Accès libre: l'accès aux archives est entièrement libre et gratuit (pour les articles les plus récents l'accès est libre après un délai de cinq ans ou moins après la publication).
- Collections complètes: les revues sont numérisées depuis leur origine (parfois le milieu du 19e siècle) et les articles récents sont intégrés au fur et à mesure de leur publication.
- Hyper-liens: un très gros travail est fait pour créer des liens multiples entre les différents articles de nos collections et des autres collections aussi bien à partir des titres que des bibliographies.

Le programme de numérisation de revues de recherche en mathématiques Numdam occupe une place originale et importante parmi les nombreux programmes de numérisation en cours de par le monde, et son modèle de production fait désormais référence au niveau national et au niveau international.

Comme nous le verrons dans le chapitre consacré à la coopération, l'organisation d'un réseau supranational des bibliothèques numériques de mathématiques se heurte à beaucoup de difficultés. Cependant une nouvelle impulsion va être donnée avec le démarrage en février 2010 du projet européen EuDML.

Sans attendre, la Cellule MathDoc avait développé plusieurs projets d'intégration, portant à quelque 50 000 références l'offre française visible dans ses moteurs de recherche mini-DML (articles) ou LiNum (livres) fédérant les contenus de Gallica (BNF: Bibliothèque Nationale de France), TEL, HAL (CCSD: Centre pour la Communication Scientifique directe), ArXiv pour une couverture internationale de l'ordre de 300 000 articles et 3 000 livres.

1. Numérisation

a. Présentation

Le programme Numdam a été lancé en 2000 grâce à des fonds provenant de la direction de la Recherche. Sa première mission fut de définir les fonctions attendues par les chercheurs d'une publication électronique des archives de revues de recherche en mathématiques, puis de concevoir une interface puissante et ergonomique pour l'exploitation de ces données.

Les enjeux spécifiques d'une opération de numérisation de revues mathématiques se fondent sur les particularités de la documentation mathématique:

- La documentation mathématique ne se périmé 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 chimistes, les physiciens ou les économistes 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 les rendre accessibles à tous les utilisateurs potentiels, mathématiciens ou non.

- Elle est valide comme un tout formant un réseau (international) de références. Il est donc essentiel de permettre aux lecteurs de remonter dans l'histoire d'un résultat et de proposer le maximum de liens significatifs.

La plateforme Numdam donne accès à un fonds numérisé d'une très grande richesse. Constitué des revues mathématiques françaises de haut niveau dont certaines se situent parmi les meilleures au plan international, de grands séminaires, et de quelques grandes revues européennes, il contribue à la valorisation de la recherche française au niveau international.

À ce jour ce sont plus de 650 000 pages et 30 000 articles enrichis de liens établis vers les grandes bases de données mathématiques (Jahrbuch, Zentralblatt et Mathematical Reviews) qui sont accessibles en texte intégral. Nous avons en outre de l'ordre de 100 000 pages et 10 000 articles déjà numérisés, en cours de traitement en vue de leur mise en ligne.

b. Bilan des réalisations (années 2005-2009)

Le premier marché Numdam, contracté avec la société AIS, a nécessité de longs mois de tests et de préparation avant d'aboutir à la première mise en ligne de données sur le serveur www.numdam.org. Les grandes étapes de vie de ce programme depuis son lancement se déclinent de la façon suivante :

- de juin 2000 à septembre 2003 : démarrage et première phase de production de Numdam
- d'octobre 2003 à juin 2007: le marché Numdam II est lancé avec la société Jouve. En parallèle, une opération « BRAIN » (Bibliothèque Rhône-Alpes d'Informations Numériques) est menée de mai 2004 à décembre 2004,
- de janvier 2008 à janvier 2011 : un nouveau marché Numdam III est passé avec la société DIADEIS. Une opération ponctuelle sur fonds CNRS permet de numériser la revue ancienne *Nouvelles annales de mathématiques* en concertation avec un projet d'historiens des mathématiques.

Le premier marché avait permis de rédiger un cahier des charges spécifiques et de valider ces choix lors de l'exploitation des données recueillies sur le site numdam.org. Les marchés suivants n'ont pas remis en cause ces choix initiaux. Cependant, chacun d'entre eux a nécessité une adaptation du cahier des charges, la réalisation d'un banc d'essai et un suivi de la finalisation de la chaîne de production avec les mises au point nécessaires.

À la fin de la première phase, nous avons numérisé 6 grandes revues généralistes françaises (*Annales de l'institut Fourier*, *Journées ÉDP*, *Publications mathématiques de l'IHÉS*, *Bulletin et Mémoires de la SMF*, *Annales Scientifiques de l'ÉNS*) pour un total d'environ 200 000 pages et 8 000 articles. Au cours des années suivantes, nous avons étendu le spectre, notamment scientifique, des collections, à travers la numérisation de plus de 380 000 pages et 20 000 articles supplémentaires lors de la phase II dont tous les résultats sont actuellement en ligne.

Numdam II

Titre	Volumes	Articles	Pages
Annales de la faculté des sciences de Toulouse (créée en 1887 – Université de Toulouse)	207	1 034	36 051
Annales mathématiques Blaise Pascal (créée en 1994 – Université de Clermont)	19	165	2 554
Annales de l'institut Henri Poincaré (1930-1964 – IHP/Gauthier-Villars)	71	145	5 860
Annales de l'institut Henri Poincaré, section B (créée en 1964 -)	155	834	19 010
Annales de l'institut Henri Poincaré (section C) (créée en 1984 - IHP/Gauthier-Villars)	102	392	10 987
Annales de l'université de Grenoble (1945-1948 – Université de Grenoble)	3	47	1 006
Séminaire de probabilités de Strasbourg (créée en 1967 – Lecture notes, Springer)	37	1 255	17 352
Total phase II avant 2005	687	3 872	92 820
Annales de l'institut Henri Poincaré (section A) (créée en 1964 – IHP/Gauthier-Villars)	287	1 371	31 482
Annales de Gergonne (1810-1832)	22	1 083	8 701
Revue de Statistiques Appliquées (créée en 1953 - SFdS)	207	1 676	22 150
Séminaires de l'IHP	218	2 383	30 000
Séminaires de l'X, Collège de F., Bourbaki	119	1 959	34 028
Compositio Mathematica	275	1 934	39 442
Annali della Scuola Normale Superiore di Pisa - Classe di Scienze	253	1 731	46 000
Cahiers de topologie et géométrie différentielle catégoriques	157	691	15 700
Rendiconti del seminario matematico di Padova	108	1 981	33 300
Journal de théorie des nombres de Bordeaux	36	529	10 000
Annales de la Faculté des sciences de l'Université de Clermont, Séries Mathématiques (1962-1992)	28	257	4 000
Annales de la Faculté des sciences de l'Université de Clermont, Série Probabilités et statistiques (1983-1991)	9	71	1 100
Mathématiques et Sciences humaines (EHES) (1962-1999)	148	763	12 000
Total phase II 2005-2007	1867	16 429	287 903

Numdam réalise sa troisième phase qui a commencé fin 2007 et courra sur un peu plus de trois ans. Son développement se poursuit sur les bases des spécificités techniques et orientations scientifiques originelles. L'obtention d'une production de qualité conforme aux spécifications du cahier des charges a été un peu laborieuse. Par la suite, nous sommes parvenus à atteindre un rythme de production stable et comparable à celui des marchés précédents. Au printemps 2010, la production est déjà bien avancée, mais seules les archives

de la revue M2AN et du séminaire TSG sont déjà publiées sur le site web de Numdam. Cela est dû pour part à la spécificité des premières revues anciennes par lesquelles le programme a débuté (NAM 7 000 articles pour 50 000 pages, chaque titre doit être relu), mais surtout au déficit actuel en personnel. Pour les articles récents fournis par leurs éditeurs, mais ne disposant pas de métadonnées suffisantes pour intégrer le système, nous avons utilisé dans certains cas notre opérateur pour la génération des métadonnées au format Numdam.

Bien que la numérisation proprement dite soit sous-traitée à travers une série de marchés publics, il faut souligner que les ressources internes de la Cellule MathDoc sont mobilisées de façon assez intense par la conduite du programme : collation, dépouillement et préparation des collections avant numérisation ; réception et contrôle de la qualité de la production avant facturation ; adaptation des contrôles automatiques et de la structure de la base de données aux spécificités des séries traitées ; maintenance et adaptations de l'indexation et de l'interface publique d'accès aux collections. On peut estimer que pour chaque euro dépensé dans le marché, MathDoc investit un autre euro en ressources propres, notamment humaines. Ceci porte le coût consolidé moyen de la page numérisée et diffusée à un euro.

Numdam III

Titre revue	Volumes	Articles	Pages
Nouvelles annales de mathématiques : journal des candidats aux écoles polytechnique et normale (1842-1927)	84	8703	50 700
Bulletin des sciences mathématiques et astronomiques (1870-1883)	19	608	9 600
Œuvres complètes de Laurent Schwartz	59	61	1 400
Thèses de l'entre deux guerres (1914-1945)	264	264	27 680
Modélisation Mathématique et Analyse Numérique (nom actuel) (1967-2000)	153	963	21 080
Revue française d'automatique, d'informatique et de recherche opérationnelle (1967-2000)	123	795	13 700
Revue française d'automatique informatique - Recherche opérationnelle Informatique théorique (1974-2000)	120	651	12 440
Séminaire de Théorie spectrale et géométrie (1982-2004)	23	265	3 790
Cahier du séminaire d'histoire des mathématiques, 2 ^e série (1991-1993)	3	19	370
Diagrames (1979-2003)	49	117	4630
Géologie alpine (1890-2001)	102	856	20772
Statistiques et analyse de données (1976-1991)	46	231	4920
Total numérisé début 2010	1045	13533	147982
ESAIM: Control, Optimisation and Calculus of Variations 1996-2000	5	95	2 500
ESAIM: Probability and Statistics 1997-2000	4	39	1 010
Revue de statistique appliquée 2002-2003	8	42	790
Total renumérisé mi-2009	17	176	4 300
Total 2005-2009	2929	30138	463184

2. Acquisition

Dès la définition du projet Numdam, il nous était apparu qu'une composante importante de l'opération devait être l'acquisition de la production des éditeurs partenaires de façon à constituer une archive exhaustive et uniforme.

Cet aspect était formulé de façon assez vague dans les contrats qui ont été signés, et il a donc fallu user de beaucoup de persuasion et de l'appui des rédactions pour parvenir à débloquent la situation courant 2006. Les données nous étant fournies aux formats propriétaires des éditeurs, nous avons mis en place plusieurs chaînes d'acquisition basées généralement sur XML. Une synergie intéressante avec le projet Cedram est apparue lorsque nous avons voulu récupérer certaines revues qui avaient un XML insuffisant (par exemple sans bibliographies) mais pour lesquelles l'éditeur a bien voulu nous transmettre les sources LaTeX : les outils du Cedram (et la chaîne d'acquisition Cedram vers Numdam) ont alors été utilisés avec succès ! Pour les articles récents (protégés par le créneau mobile), nous contribuons à leur visibilité en fournissant l'ensemble des métadonnées à nos utilisateurs, et en les enrichissant de liens. Nous fournissons aussi le lien direct vers l'article sur le site de l'éditeur (DOI ou équivalent), même lorsque le texte intégral est téléchargeable depuis Numdam, après l'expiration du créneau mobile (délai de cinq ans ou moins après la parution).

Acquisition des nouveautés dans Numdam (2006-2009)

Titre revue	Années	Articles	Pages
Elsevier/ScienceDirect (XML)	23	991	23590
Annales scientifiques de l'ÉNS	2001-2007	264	6860
Annales de l'institut Henri Poincaré, section B	2001-2007	276	6200
Annales de l'institut Henri Poincaré, section C	2001-2009	579	10530
Cedram (XML)	47	1386	36937
Annales de la faculté des sciences de Toulouse	2003-2009	189	5063
Annales de l'institut Fourier	2001-2009	714	22641
Ann. math. Blaise-Pascal	2003-2009	110	2420
J. théorie des nombres Bordeaux	2004-2009	270	5540
Journées ÉDP	2001-2006	84	1100
Séminaire X-EDP	1997-2008	265	4000
Springer/Springerlink (XML)	5	40	2572
Publ. math. I.H.É.S.	2003-2007	40	2572
EDP Sciences (LaTeX)	45	1584	29873
ESAIM : M2AN	2001-2009	447	10070
RAIRO : RO	2001-2009	315	3556
RAIRO : ITA	2001-2009	285	5023
ESAIM: COCV	2001-2009	360	7860
ESAIM: Probability & Statistics	2001-2009	169	3364
IMS/Euclid	1	53	1186
Annales de l'institut Henri Poincaré, section B	2008	53	1186
Autres (LaTeX)	14	334	8361
Publ. math. I.H.É.S.	2001-2002	16	936
Journal de la SFdS	2007-2008	46	1626
Annali della Scuola Normale Superiore di Pisa	2002-2003	64	1790
Rendiconti di Padova	2002-2009	208	4260
Total acquisition nativement numérique	135	4388	102770

3. Bibliothèques virtuelles

La collection 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 œuvres complètes).

Dans le cadre du pôle associé BNF-Cellule MathDoc (2003-2007), un nouveau service «Gallica-Math» a été mis en place pour offrir un accès au niveau article à un choix de collections numérisées par la BNF. L'application spécifique permet un feuilletage de la table des matières de chaque volume et un système de recherche complet est mis en place au travers de la mini-DML (voir paragraphe suivant). Les collections sont complétées au fur et à mesure en fonction des nouvelles numérisations à la BNF.

Sont actuellement consultables sur ce serveur :

- les œuvres complètes de nombreux mathématiciens (1989 textes répertoriés),
- le *Journal de mathématiques pures et appliquées* (JMPA) : 2081 articles catalogués de 1836 à 1934,
- le RBSM, *Répertoire bibliographique des sciences mathématiques* (1894-1912), une des premières bibliographies dédiées aux mathématiques en 2000 fiches et 18 792 références.

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).

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 Jahrbuch (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 LiNuM qui permet un accès unifié à 2849 volumes provenant 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 soit par l'envoi de catalogues par les institutions concernées, soit par moissonnage. Il demande un travail de maintenance relativement léger, mais essentiellement 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 établi un système léger, le projet mini-DML. 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 avec la bibliothèque de Göttingen et le projet Euclid pour mettre en place les serveurs OAI avec des standards de métadonnées simplifiés ont permis de proposer un

standard de fait : http://projecteuclid.org/collection/euclid/documents/metadata/dml_dc.html.

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, GDZ, HAL, TEL, Numdam, Cedram...) et d'autres collections pour lesquelles la Cellule MathDoc a, elle-même, produit un catalogue exploitable (Gallica-math, Fundamenta Mathematicae sur ICM Varsovie...). Un peu plus de 300 000 articles sont ainsi référencés, dont une part importante échappe aux grandes bases de données (ressources trop anciennes ou trop fragmentées). Une conséquence intéressante est que les articles référencés dans la mini-DML peuvent être directement liés dans les notices du Cedram ou de Numdam depuis les bibliographies où ils sont cités.

4. Infrastructures

L'outil informatique est au cœur du développement de la Cellule Mathdoc, aussi bien pour la création que la diffusion des documents. L'infrastructure informatique constitue une part importante des investissements de la Cellule MathDoc.

Pour ce qui concerne la diffusion, nous avons choisi d'utiliser l'infrastructure de l'Université Joseph Fourier plutôt que de faire appel à un opérateur extérieur. Cela nous facilite énormément le travail de mise en place des documents mais nous demande une vigilance constante pour parer à toutes les attaques qui fourmillent sur internet. La continuité du service en toutes occasions est assurée grâce à la compétence et la disponibilité des ingénieurs de l'Institut Fourier qui gèrent notre réseau.

Actuellement, l'unité utilise trois serveurs :

- Un serveur de développement où la création et la mise à jour des sites sont réalisées puis testées. Ce serveur basé sur un système d'exploitation Linux mandriva 2006, comprend deux processeurs Xeon à 2,8 Ghz, 2 Go de mémoire vive, 500 Go d'espace disques et un lecteur de bandes.
- Une fois les procédures de test validées sur le site de développement, les sites sont transférés sur le serveur web qui est basé sur le système d'exploitation Ubuntu 6.06 LTS avec deux processeurs Xeon à 2, 8 Ghz, 2 Go de mémoire vive et de 400 Go de disques.
- Un troisième serveur plus récent est basé sur deux Xeon quadcore, 20 Go de Ram, 600 Go d'espace disques. Il lui est attachée une baie de disques disposant de 4 To d'espace de stockage.

Ce serveur est le premier de l'unité à s'appuyer sur la virtualisation qui permet d'isoler les services et d'assurer le mirroring des autres serveurs.

Le dernier serveur est la première étape pour la mise en place d'une plateforme de virtualisation complète dont les objectifs sont :

- faciliter la mise en place de mirroring matériel,
- faciliter l'évolution et les mises à jour des systèmes d'exploitation et des logiciels en minimisant l'interruption de services,
- faciliter les migrations matérielles en désolidarisant le logiciel du hardware,

- associé au système de sauvegarde et d'archivage, améliorer la reprise sur incident avec un minimum d'interruption de service.

III. Édition électronique

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 maisons d'édition françaises de taille moyenne comme Gauthier-Villars ont été rachetées par Elsevier, et Springer-Verlag a fusionné avec 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 *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.

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, la commercialisation des revues au format papier est menacée à moyen voire à court terme et la production d'une édition électronique de qualité est une condition impérative de développement et de survie pour ces revues.

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. Une 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. 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 ajouter de plus en plus 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 sont aussi améliorés de par la mise à disposition de services et d'outils particulièrement adaptés aux exigences des revues de mathématiques.

Après plusieurs réunions préparatoires réunissant l'ensemble des responsables de revues françaises, les sociétés savantes et les représentants des mathématiques au ministère de la Recherche et au CNRS, il a été décidé en 2005 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 bénéficie de l'expérience en édition électronique acquise par la Cellule MathDoc, notamment au cours de la réalisation du programme 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 pourra être amené à se développer dans diverses directions : en mathématiques vers l'échelon européen ; en France vers d'autres revues composées en LaTeX.

L'ambition de ce projet est d'augmenter la visibilité et l'impact des revues (édition électronique sur un serveur commun ; gestion, promotion et échanges des abonnements papier). Pour y répondre, ce pôle, baptisé Cedram (Centre de diffusion de revues académiques mathématiques), a souhaité aborder tous les aspects de l'édition et rester flexible, c'est-à-dire qu'il offre des services modulaires que chaque revue peut choisir à la carte. Les modules suivants sont opérationnels :

- aide à l'édition (RUCHE, outil de gestion du secrétariat et des flux éditoriaux) ;
- production (CEDRICS, système de production des volumes – électronique et papier – en LaTeX, produisant des PDF hypertextes et des métadonnées structurées en XML, les formules mathématiques dans un format LaTeX ou MathML) ;
- diffusion électronique (une instance dédiée d'EDBM gérant les serveurs 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).

Un comité de pilotage a été mis en place. Il 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, directeur adjoint de la Cellule. 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. 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.

Les premières revues qui ont intégré le Cedram lors de son démarrage en 2006 sont les suivantes :

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

Depuis, l'offre s'est un peu étoffée :

- *Journal de théorie des nombres de Bordeaux*
- MathS In Action (nouvelle revue de la SMAI)
- Actes des journées Équations aux dérivées partielles (GDR 2434)
- Actes du séminaire Équations aux dérivées partielles (École polytechnique, UMR 7640)
- Actes du séminaire de Théorie spectrale et géométrie (Grenoble, UMR 5582)
- Actes et cours des rencontres du CIRM (Luminy, UMS822 CNRS/SMF).

Le premier 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 premier système de production automatisée de revue en LaTeX a été écrit pendant l'été 2005 par Thierry Bouche. Ce système permettait de reproduire fidèlement la maquette de plusieurs revues à partir d'une source normalisée, et de générer des métadonnées à jour sans duplication des informations. Les numéros de page et les sommaires sont produits à la volée, bien que chaque article soit compilé séparément. On dispose en sortie des éléments nécessaires pour l'impression et la mise en ligne.

Produire du XML directement à partir de LaTeX n'étant pas d'une fiabilité exemplaire, un grand nombre de contrôles et de retraitements étaient nécessaires avant de pouvoir exploiter les métadonnées, ce qui était au final assez coûteux. Une stratégie différente reposant sur le convertisseur Tralics développé à l'INRIA a été testée au cours de l'année 2006. Cette deuxième version du système ne remet pas en cause la conception du premier, puisque l'interface utilisateur n'a pas été modifiée. Ce système a permis d'accueillir de

nouvelles séries sans alourdir la charge de travail.

Il n'a pas été difficile de l'adapter à toutes les situations rencontrées à ce jour, et bien au-delà, puisque il s'est révélé un moyen extrêmement efficace de générer des métadonnées à partir de sources TeX de toutes provenances, y compris à seule fin d'intégrer de nouveaux articles dans Numdam sans passer par la chaîne de numérisation. Ainsi, aux 1900 articles produits pour le Cedram, il faut en ajouter 2400 qui ont été intégrés dans Numdam grâce à cet outil. Ce système CEDRICS est diffusé en logiciel libre (licences LPPL et Cecill) et est utilisé pour la production des *Cahiers GUTenberg* en France, et par le projet DML-CZ en République Tchèque.

Claude Goutorbe a paramétré 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 constituée 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 ont été mis aux nouvelles normes de façon parfois acrobatique. Ces normes ont été profondément modifiées en 2007 avec l'utilisation d'un XML plus riche produit par Tralics, comportant les formules en LaTeX et en MathML, des bibliographies provenant de bibtex ou non.

Pour préserver l'image de chaque revue, une architecture peu contraignante a été développée, qui repose sur un site « statique » géré par SPIP, dans lequel s'intègre l'affichage dynamique des articles gérés par EDBM, avec une même charte graphique. 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. Les actes de séminaires sont intégrés plus simplement au site Cedram, ce qui permet d'être plus réactif à moindre coût.

Bon exemple de la synergie attendue entre MathDoc et les *Annales de l'institut Fourier*, le logiciel RUCHE, logiciel développé par la société « n.b. nota bene » pour le secrétariat de la revue Traitement du signal, a été adapté pour le Cedram. Les *Annales* ont défrayé la société éditrice du logiciel, ce qui permet désormais de le diffuser comme logiciel libre (licence Cecill) ; la Cellule MathDoc a embauché en CDD son développeur principal pour qu'il l'adapte aux besoins d'une revue de maths, et prépare une version bilingue français-anglais. La traduction a été effectuée par le secrétariat des *Annales*.

	Scan/pdfs/ps/ai		Bim-digital recovery			Future-ready bim-digital	
	1997	2000	2001	2002	2003	2004	2005
AF	NUMDAM 1,811 art. 51,064 p.		Mixed strategies 359 art. 11,054 p.			CEDRICS (r, -B) 350 art. 11,323 p.	
AFSI	NUMDAM		1,109 art. 38,400 p.		NUMDAM again		CEDRICS (r, -B) CEDRICS (r, -B) 74 art. 834 p. 47 art. 1840 p.
Aues	NUMDAM 177 art. 1,632 p.		CEDRICS (r, -B) 45 art. 922 p.			CEDRICS (r, -B) 59 art. 1341 p.	
IME	NUMDAM 431 art. 7,700 p.		CEDRICS (r, -B) 98 art. 1,818 p.			CEDRICS (r, -B) 123 art. 2,424 p.	
IEDP	NUMDAM 469 art. 3,580 p.		Mixed strategies 43 art. 400 p.		CEDRICS (r, -B) 13 art. 204 p.		CEDRICS (r, -B) 24 art. 510 p.
ISG	NUMDAM 244 art. 3,800 p.		CEDRICS (r, -B) 13 art. 240 p.			CEDRICS (r, -B) 13 art. 240 p.	
SEDP	NUMDAM 640 art. 9,400 p.		CEDRICS (r, -B) 277 art. 4,114 p.				
Total	NUMDAM metadata 4,903 art. 117,544 p.		CEDRICS metadata 1,107 art. 25,848 p.				

IV. Services vers les bibliothèques

1. Catalogues fusionnés

Depuis 1998, MathDoc a mis en place un catalogue fusionné des périodiques. Cet outil, apprécié de tous, a été décrit dans un précédent rapport et a été amélioré d'année en année. Ainsi en 2007 et 2008, les améliorations suivantes y ont été apportées :

- Les URL (adresses des sites de journaux) bénéficient d'un véritable catalogage. On indique si le journal est gratuit ou sous abonnement, s'il y a une archive, si l'on peut y accéder par le portail Bibliosciences, etc. Des listes téléchargeables de journaux en ligne sont extraites du catalogue et affichées sur le site « portail documentaire » de MathDoc.
- Des listes dérivées du catalogue peuvent être téléchargées par les bibliothèques, donnant l'état des collections papier pour les abonnements en cours. Cette amélioration est utile pour le RNBM pour savoir si des abonnements peuvent être supprimées, et pour préparer un plan de conservation partagée.
- Une collaboration s'est instaurée avec la bibliothèque de l'institut Fourier, qui dispose de son point d'entrée particulier dans le catalogue des périodiques. Des programmes d'affichage spécifiques pour l'institut Fourier ont été développés.

Une nouvelle interface est en cours d'élaboration qui permettra aux bibliothèques de rentrer directement leurs données.

Le Catalogue fusionné des ouvrages (CFO) a été refait, en collaboration avec MathDoc, par un ingénieur de la bibliothèque Jacques Hadamard à Orsay. L'intervention de MathDoc a porté en particulier sur la conception des filtres permettant d'interpréter les fiches au format UNIMARC.

2. Actions de collaboration avec le RNBM

Outre les catalogues fusionnés, MathDoc a fait un important travail de soutien aux bibliothèques : actions de formation, conseils, participation à la vie du réseau. C'est ainsi que nous avons participé à l'organisation et à l'enseignement dans 3 actions de formation, journées RNBM (2006 et 2008) et école thématique (une semaine au CIRM à Luminy en 2007). Nous avons également participé, avec nos collègues du RNBM, à une action de formation dans le cadre d'une école de documentation Franco-Maghrébine en novembre 2008.

Mise en ligne de collections numérisées par les bibliothèques de mathématiques.

Les bibliothèques font parfois numériser leurs propres collections : par exemple la bibliothèque de Polytechnique avec les archives Laurent Schwartz, ou celle d'Orsay avec les « publications mathématiques d'Orsay ». Une fois la numérisation terminée, elles souhaitent profiter de l'expertise de MathDoc, et de la visibilité de ses serveurs. La collection des publications mathématiques d'Orsay a pu être traitée et affichée sur le portail documentaire de MathDoc où elle bénéficie d'une visibilité accrue et d'une indexation dans le moteur « miniDML ». Il est envisagé de l'intégrer à terme dans les collections Numdam, mais cela demandera un surcroît important de travail, car les métadonnées beaucoup plus détaillées devront être produites pour cette collection.

Les archives de Nicolas Bourbaki ont été numérisées par la défunte unité du CNRS « Archives de la création mathématique » (UPS 2065). Ce fonds unique a ensuite été transféré

à la Cellule MathDoc, qui a restructuré profondément son catalogue et en a réalisé le premier affichage public sous le contrôle de l'association. Au moment de la rédaction, une convention est en cours d'écriture entre MathDoc et la bibliothèque de l'école Polytechnique pour une utilisation similaire des archives personnelles de Laurent Schwartz.

V. Visibilité

1. Accès public par le Web

La visibilité de la Cellule MathDoc est en premier lieu celle de ses serveurs Web (au point qu'à l'étranger on a un peu de mal à faire la différence entre le programme Numdam et l'unité MathDoc).

Nous avons essentiellement trois serveurs, le portail documentaire, le serveur Numdam et le serveur Cedram.

Le portail documentaire (<http://portail.mathdoc.fr>) s'adresse en priorité aux bibliothécaires et mathématiciens français. Il contient le catalogue fusionné des périodiques (très utilisé par les bibliothèques de mathématiques), Gallica-Math (portail vers les archives de la BNF), LiNum (répertoire de livres numérisés très utilisé par les historiens) et plusieurs annuaires.

Le site de Numdam (<http://www.numdam.org>) est de loin le plus connu. Il présente l'ensemble des archives numérisées par MathDoc et reçoit de l'ordre de 100 000 visites par mois, avec plus de 150 pays différents qui visitent le site principalement depuis France, Italie, Allemagne, USA, Japon, reste de l'Europe et francophonie. Il se télécharge environ 100 000 articles par mois (dont 85% au format PDF et 15 % au format DjVu).

Le site de Cedram (<http://www.cedram.org>) présente les éditions électroniques des revues et présente un site distinct par revue. Celui des Annales de l'Institut Fourier (<http://aif.cedram.org>) a de l'ordre de 1000 visites par mois et 400 articles téléchargés.

La visibilité du Cedram est plus faible que celle de Numdam. Cela est dû avant tout au fait que les articles de Cedram sont pour la plupart réservés aux abonnés alors que Numdam est entièrement libre mais aussi au fait que les revues préfèrent avoir chacune un site bien individualisé. L'équilibre entre individualisation des sites et mise en commun des ressources et des accès est délicat à trouver.

Pour en revenir à Numdam, on peut également tenter de mesurer son impact en effectuant des recherches sur Google. On s'aperçoit que Numdam arrive largement en tête dans toutes les recherches d'articles qui sont présentés simultanément sur Numdam et d'autres serveurs (Elsevier, SMF, ArXiv, etc...)

La notoriété de Numdam et la visibilité que ce programme apporte à l'Université de Grenoble dans le monde entier peuvent être constatées par les références que l'on trouve un peu partout, par exemple:

<http://www.science.gouv.fr/fr/bibliotheques-numeriques/bdd/t/17/web/mathematiques>
(deux citations : Gallica-Math et Numdam)

<http://bibnum.over-blog.com/article-6995465.html>

<http://www.nul.nagoya-u.ac.jp/db/dbdb/scimath.html>

Une citation dans le livre /Google and the myth of universal knowledge/, de Jean Noël Jeanneney et Teresa Lavender Fagan :
“MathDoc, located in Grenoble, has become a valuable tool for mathematicians.”

On peut aussi lire sur le site du CEIC de l'IMU parmi les sujets qui devraient être traités plus en profondeur :

"The NUMDAM experience: The high standards adopted, the quality, and the volume of output of Cellule Mathdoc at Grenoble make NUMDAM a model numérisation=retrodigitisation effort. See www.numdam.org/."

La visibilité de Numdam c'est aussi la reconnaissance par l'ensemble des acteurs de la qualité du modèle et de l'ensemble du travail. Numdam ou Cedram sont cités dans 15 des 30 articles publiés lors des conférences DML (Digital Mathematical Library) en 2008 et 2009.

Pour terminer ce paragraphe, faisons-nous plaisir avec quelques citations:

« Many of these journals have been amalgamated into digital libraries, some of which are quite spectacular. The leading example is the Numdam collection in Grenoble, France. »
Michael Doob (2008)

« We take inspiration mainly from the French initiatives Numdam and Cedram. »
Petr Sojka & Jiri Rakosnik (2008)

« I am writing you because I know Numdam has done such a great job with digitizing old French math. »
David Mumford (2007)

« Numdam would be an excellent place for this [ICM proceedings], since everything there is done exactly the right way! »
Ulf Rehmann (2006)

2. Coopération européenne et internationale

Le vaste projet de la “Digital Mathematical Library” et ses divers avatars comme la DML-EU ont suscité de très nombreuses réunions. S'ils n'ont pas donné beaucoup de résultats opérationnels, ils ont permis d'établir des relations suivies entre les différents acteurs, des échanges très fructueux d'ordre aussi bien “politique” que technique.

Nous citerons en particulier le projet EMANI qui n'a pas donné tous les résultats escomptés du fait du changement de stratégie de Springer après son rachat mais a permis des échanges de vues extrêmement enrichissant entre les différents partenaires:

- Cornell University Library (USA)
- SUB Göttingen (Allemagne)
- Tsinghua University Library (Chine)

- MathDoc (France)
- Zentralblatt Math (Allemagne)

Au niveau européen, tout un groupe s'est formé regroupant des acteurs d'Allemagne, de Grande-Bretagne, d'Espagne, de Pologne, du Portugal, de Tchécoslovaquie et de France pour tenter de lancer des projets européens. Comme nous le verrons ci-dessous, un projet européen a débuté en février 2010 mais déjà, les relations établies ont permis de nombreux échanges (voir le paragraphe sur les activités de conseil et d'expertise).

3. Coopération avec le Zentralblatt

À la suite du projet LIMES (2001-2004), la coopération avec le Zentralblatt MATH (ZM) s'est poursuivie par une assistance à la mise en place de la version la plus récente du moteur de recherche développé par la Cellule MathDoc. L'intégration de ce logiciel dans le nouveau système de production et le nouveau portail d'interrogation (accessible publiquement depuis septembre 2006) ont été réalisés par l'équipe berlinoise du Zentralblatt avec un support technique initial de la Cellule MathDoc au cours de l'année 2005.

Nous avons également développé un nouvel outil permettant d'associer à une référence bibliographique l'identifiant ZM lorsqu'il existe. Largement utilisé pour leurs besoins internes par les projets Numdam et Cedram, ce service devra être intégré à l'offre standard du portail Zentralblatt MATH dans le futur.

Au cours de la dernière réunion technique (février 2009) nous avons examiné divers points d'intérêt commun: les développements futurs du moteur de recherche, l'échange de données (alimentation de la base Zentralblatt grâce aux serveurs OAI de la Cellule MathDoc, mise en place systématique dans cette même base de liens vers les articles de la bibliothèque numérique), outils permettant la conversion de TeX vers MathML. Ces deux derniers points sont effectifs : 25 000 liens sur Numdam ont été ajoutés à la base et toutes les notices disposent d'une version MathML produites avec Tralics grâce à nos indications.

4. Activités de conseil et d'expertise

Dans le bassin grenoblois

Au niveau local ou régional, des ingénieurs de la Cellule MathDoc ont été associés à la mise en place du CCSD, en particulier du serveur de thèses en ligne TEL, qui s'est appuyé sur l'expérience de MathDoc en la matière, car la Cellule maintenait depuis quelques années un service recensant les thèses de mathématiques soutenues en France. En 2008, MathDoc a décidé de cesser son propre service pour s'en remettre à TEL. Dans la poursuite de cette activité, un ingénieur de MathDoc a contribué significativement à la mise en place de l'archivage des thèses dans TEL à l'UJF.

Dans le domaine de la numérisation qui est un de nos champs d'expertise reconnus, nous avons été sollicités par la bibliothèque de l'OSUG, qui dépend de l'Université Joseph Fourier pour intégrer la *Revue de géologie alpine* à la chaîne de production Numdam III. La numérisation est terminée et nous travaillons actuellement aux corrections et à la mise en ligne.

La Cellule MathDoc a organisé à Grenoble en 2007 une formation LaTeX à destination des chercheurs et ITA des laboratoires du CNRS, suivie par une vingtaine de personnes.

Numérisation du *Journal de Physique*

La société EDP Sciences a bénéficié de notre expertise pour la numérisation des archives du *Journal de Physique* (48 500 articles sur 300 000 pages, en utilisant notre chaîne de production et une légère modification de notre DTD chez Jouve).

Programme de numérisation de la bibliothèque Mathématiques Recherche Paris Centre

Nous avons également été sollicités par la bibliothèque mathématiques Recherche de Paris Centre (ex-Jussieu), qui doit se réorganiser sur deux sites géographiques différents pour Paris 6 et Paris 7. Cette bibliothèque a passé un accord avec l'éditeur Springer lui permettant de numériser et de mettre en ligne une partie de son fonds. Des ouvrages mathématiques essentiels pourront ainsi demeurer accessibles aux deux bibliothèques, ainsi qu'à l'ensemble des bibliothèques mathématiques du réseau RNBM.

Nous encadrons la mise en place du projet de numérisation ainsi que sa mise en œuvre. Nous avons proposé un ensemble de choix techniques adaptés, pour le dépouillement des données, la structure des métadonnées, ainsi que les spécifications techniques de numérisation. Ceci nous a amené à réécrire la DTD (Document Type Definition) pour l'adapter au cas particulier de la numérisation de monographies et à la rédaction du cahier des charges correspondant pour l'appel d'offre.

Nous avons participé à l'analyse comparative des offres, puis au choix du prestataire. Catherine Barbe-Zoppis, ingénieur développement, a supervisé, en tant qu'interlocuteur de référence du projet à la Cellule MathDoc, la rédaction du cahier des spécifications techniques qui permettra le réglage initial de la chaîne de numérisation.

Cette collaboration comprend aussi l'adaptation de notre procédure de contrôle qualité utilisée pour le programme Numdam et la formation du personnel de la bibliothèque Recherche à son utilisation. Un technicien a été recruté pour 6 mois en CDD par la bibliothèque de Paris Centre et est hébergé par MathDoc où il a commencé à travailler à la mise en ligne des documents.

Europe et international

Participant à tous les projets de bibliothèque numérique de mathématiques supranationaux (DML) depuis l'an 2000, la Cellule MathDoc a été appelée à conseiller, évaluer, et collaborer avec un grand nombre d'entre eux au cours des dernières années. Par exemple, Thierry Bouche a participé en observateur à la réunion du Committee on Electronic Information and Communication (CEIC) de l'Union mathématique internationale (IMU) qui s'est tenue à Grenoble en mars 2005, puis il a participé à la rédaction des deux textes qui allaient être promulgués par l'assemblée générale de l'IMU l'année suivante (Some Best Practices for Retrodigitization ; Digital Mathematics Library: A Vision for the Future). Un mois plus tard, il était invité à Berkeley pour faire partie du panel d'experts associés à la définition d'un vaste projet de numérisation (finalement infructueux) piloté par l'AMS et le MSRI pour le compte de la fondation Moore. Cette situation se reproduit d'une certaine façon en 2009, à la demande du groupe PESC de l'ESF qui souhaite soutenir avec la SME un projet de bibliothèque virtuelle européenne de mathématiques dans les infrastructures de recherche du 7^e PCRD, et qui l'ont mandaté pour concevoir le plan de travail. La direction scientifique des deux dernières propositions EuDML soumises à la commission européenne (programme eContentplus, puis CIP ICT PSP) sous les auspices de la SME a été assurée par la Cellule MathDoc, qui n'a cependant pas pu rassembler localement les moyens humains pour assumer

la coordination de la gestion du projet, et a donc dû la déléguer à un collègue portugais (tout en conservant la coordination scientifique).

Nous avons été invités à Madrid, à Prague pour orienter des projets émergents et évaluer leurs choix. Nous avons en outre reçu la visite à Grenoble de projets suisse, canadien, tchèque, espagnol, italien, et eu l'occasion de nous entretenir avec de nombreux autres lors de nombreuses conférences internationales. Nous avons représenté le nœud français (bibliothèque Hadamard, Orsay/Cellule MathDoc) du projet EMANI (2002-2006, coord. Bernd Wegner) et organisé une rencontre d'une demi-journée à la MSH de Saint Martin d'Hères en mars 2005 en présence des membres du CEIC (3 exposés : Michael Doob, Petr Sojka, Th. B.).

B. Projets

I. Perspectives et difficultés

La Cellule MathDoc est une petite structure, elle est unique dans son genre et de plus sa mission concerne la documentation électronique, un domaine en évolution très rapide. Elle doit se remettre en cause régulièrement et s'adapter aux modifications de son environnement. Enfin, elle a été conçue comme une unité à vocation nationale alors que la LRU favorise plutôt les missions à l'échelle de l'université.

La première question qui se pose est celle de son champ d'action disciplinaire. Doit-elle se cantonner aux mathématiques ou élargir son domaine d'intervention ?

Il faut considérer les moyens qui lui sont alloués. D'une part, ils sont assez modestes, surtout en personnel, donc ne lui permettent pas de tout embrasser et d'autre part ils ont été fournis essentiellement sur les crédits et le personnel des mathématiques aussi bien du côté du ministère que du CNRS. La LRU change quelque peu ce point de vue puisque une part importante des crédits viendra de l'université donc ne sera plus liée spécifiquement aux mathématiques même si actuellement l'essentiel du personnel vient du CNRS.

Le deuxième point de vue, et le plus important, est celui de l'efficacité. Le bilan des 13 ans d'existence de l'unité démontre l'intérêt d'une petite structure ayant un domaine d'intervention limité, d'autant que la documentation mathématique a ses spécificités. Cependant, les membres de l'unité ont acquis au cours de ces années un savoir-faire unique en particulier dans le domaine de la numérisation. Il est donc tout naturel d'en faire bénéficier les autres domaines qui rencontrent les mêmes problèmes. Nous avons déjà commencé un peu cette diversification et nous envisageons de l'accentuer dans les prochaines années (en coopération avec le SICD notamment).

L'autre question est celle de la mission nationale de la Cellule MathDoc. L'unité a été créée très clairement et très explicitement dans ce but. Cependant, vu de Paris il était parfois difficile d'imaginer qu'une mission nationale puisse s'exercer à partir d'une ville de province et, surtout à ses débuts, elle a dû défendre cette idée de mission nationale et démontrer qu'elle ne travaillait pas que pour Grenoble. Le point de vue des étrangers est différent. Ils ne sont guère habitués aux structures nationales comme le CNRS et par contre il est tout naturel pour eux qu'une université monte une structure à vocation nationale. Pour eux, nous avons toujours été "la Cellule MathDoc de l'Université de Grenoble". Un certain nombre de demandes locales nous incite à consacrer une part de nos efforts à des objectifs plus régionaux.

Cet élargissement de nos missions pose le problème des moyens et singulièrement celui des moyens humains. Or l'unité traverse actuellement une passe difficile. Un ingénieur de l'université a pris sa retraite et n'a pas été remplacé, une autre membre du CNRS est en congé de maternité et va prendre un congé post-natal mais le CNRS ne donne plus de crédits aux unités pour engager un personnel temporaire en remplacement, enfin plusieurs membres de l'unité se sont mis en temps partiel. Le CNRS a fait de gros efforts pour donner du personnel à MathDoc et il serait absolument nécessaire que l'université participe à cet effort à travers l'affectation de personnel ITA, mais aussi en incitant les mathématiciens du site à participer aux orientations scientifiques de ses activités en offrant des décharges de service ciblées.

Les projets qui suivent s'inscrivent dans la continuité de ce qui a été fait jusqu'à présent (numérisation, édition) mais engagent également de nouvelles actions (archivage pérenne) et de nouvelles collaborations.

II. Bibliothèque numérique

1. Projets européen EuDML

Le projet DML a été formulé aux alentours de l'an 2000, mais il n'a guère avancé depuis. Par contre, beaucoup de projets de numérisation ont vu le jour à travers le monde, comprenant des projets généralistes avec une forte dose de mathématiques (JSTOR, bibliothèque universitaire de Göttingen, Gallica, à quoi il faut ajouter les projets purement commerciaux comme ceux d'Elsevier, Springer, Cambridge university press, etc.) aussi bien que des projets spécifiques pour les mathématiques (project Euclid, NUMDAM, DML bulgare, italienne, espagnole, serbe, russe, tchèque, suisse, polonaise, portugaise, japonaise ou canadienne...). La volonté de structuration de ces efforts indépendants s'est heurtée très tôt à la difficulté de définir une gouvernance équilibrée au niveau mondial, et à l'absence de financement suffisant pour réaliser une partie substantielle du projet de façon globale. La ligne directrice de nos actions à venir sera de faire avancer ce projet « bottom-up » plutôt que de façon centralisée et pyramidale qui a montré son inefficacité.

Au cours des dix dernières années, la proportion du corpus mathématique disponible au format électronique a crû de façon continue, atteignant actuellement une masse assez considérable couvrant de l'ordre du quart des références connues. Les collections numériques sont produites (et détenues) pour moitié par des bibliothèques de recherche et pour moitié par des sociétés commerciales. Cet état de fait pose de nombreuses questions non résolues pour lesquelles la Cellule MathDoc se doit d'apporter des réponses pertinentes au niveau national, et compte contribuer à l'émergence de solutions concertées au niveau européen, puis mondial.

Le projet EuDML a formellement démarré le 1er février 2010. Il a pour ambition de construire une implémentation pilote du chapitre européen de la DML. Étant financé dans le cadre du programme CIP ICT PSP qui met l'accent sur l'innovation plutôt que la recherche, il va principalement se concentrer sur les défis technologiques, qu'il devra résoudre en intégrant des solutions éprouvées : mise à niveau des métadonnées de formats très hétérogènes, intégration dans une base de données unique, développement d'un point accès unique aux collections distribuées dans toute l'Europe, enrichi d'une interface utilisateur personnalisée,

travaux pour une plus grande accessibilité des textes et une meilleure interopérabilité. La stratégie est d'espérer un effet boule de neige bâti sur la réussite de ce pilote. Assemblant les collections de toutes les bibliothèques numériques publiques européennes de mathématiques, et soutenu par au moins un éditeur commercial (EDP Sciences), il aura un poids suffisant pour infléchir durablement la construction de la DML. Il est également probable qu'il redonne une impulsion au projet mondial puisqu'il s'agit de la première avancée significative vers l'objectif d'intégrer les collections existantes au-delà des frontières nationales.

Géré et coordonné techniquement par l'IST à Lisbonne, scientifiquement par la Cellule MathDoc, le projet EuDML bénéficie d'un financement communautaire à hauteur de 1,6 M€ (dont 1/8 pour MathDoc via ses deux tutelles UJF et CNRS). Il rassemblera les efforts de 13 partenaires européens, technologiques, scientifiques, et détenteurs de contenus mathématiques numériques (natifs ou rétrospectifs). Les travaux se poursuivront jusqu'en janvier 2013.

Du fait de la taille réduite de l'équipe, de nombreux développements technologiques très importants pour l'avenir n'ont pas pu être pris en charge en interne : nous attendons des collaborations internationales qu'elles soient l'occasion de transferts technologiques fructueux, voire de création de nouvelles technologies modifiant profondément la façon dont les mathématiciens et les autres scientifiques s'approprient la littérature mathématique.

Citons trois exemples.

- Des logiciels de reconnaissance optique de caractères pour les formules mathématiques sont actuellement en cours de développement, notamment en Angleterre et au Japon. Si l'on dispose de l'intégralité du texte des articles (y compris les formules), de nouvelles voies d'exploration du corpus seront ouvertes, que les chercheurs ne manqueront pas d'explorer.
- Les métadonnées des articles de NUMDAM, essentiellement déduites du contenu des articles numérisés, sont en général dans la langue principale de l'article. Ceci rend une partie considérable du corpus invisible à une recherche en anglais, par exemple. Des outils utilisant le contexte donné par les liens (notamment vers le Zentralblatt), le texte intégral et les formules devraient permettre de fournir suffisamment de mots clés pour franchir ces barrières linguistiques à moindre coût.
- Il reste beaucoup à faire pour rendre accessibles aux déficients visuels les textes scientifiques en général, et mathématiques en particulier. Rendre accessible le contenu produit par les éditeurs commerciaux (PDF numérique uniquement graphique), par la numérisation (PDF image, texte caché sans formules), ou à partir des sources LaTeX est un objectif naturel en soi. Il pourrait en outre convaincre les éditeurs de collaborer plus étroitement pour bénéficier de ce genre de service qu'ils n'ont pas les moyens de développer en interne. Il faut souligner que ces avancées, plus faciles à mener sur un corpus réduit et relativement homogène, auraient des retombées considérables non seulement pour l'accessibilité de l'offre didactique ou de la documentation technique, mais aussi parce qu'elles rendraient le corpus lui-même plus apte à des traitements automatisés.

Le projet EuDML est un laboratoire où certaines de ces idées seront testées (il y a une tâche dédiée à la génération de texte MathML à partir d'images scannées ou de PDF éditeurs, une autre à l'accessibilité pour les handicapés visuels des textes mathématiques ainsi obtenus). Pour aller plus loin il faudra participer à des projets de recherche sur ces questions.

2. Numérisation

Nous prévoyons de poursuivre l'activité de numérisation au-delà du marché Numdam III en cours. Un domaine dans lequel Numdam n'est pas encore présent est celui du livre (rappelons toutefois les mémoires de la SMF, et les thèses de l'entre-deux guerres ainsi que le projet de numériser tout ou partie de la collection « Mémorial du XXe siècle » parue chez Gauthier-Villars).

Du fait que Gallica ne traite que des ouvrages du domaine public, il existe une forte demande pour des ouvrages du vingtième siècle, notamment des notes de cours de troisième cycle, ou des monographies publiées de façon relativement informelle par des laboratoires. Ces ouvrages, qui sont parfois des références essentielles pour certaines recherches en cours, sont rares et n'ont pas l'impact qu'ils méritent. La numérisation de tels documents a jusqu'à présent toujours été repoussée au profit des revues qu'il restait à traiter sur un plan d'égalité avec celles qui étaient déjà intégrées. Elle pose par ailleurs de nombreux problèmes nouveaux par rapport à la numérisation en série que nous avons pratiqué jusqu'à présent : chaque ouvrage demande un travail préparatoire important pour recueillir les autorisations nécessaires des ayants-droit, et il est évident qu'il faut opérer une sélection sur critères scientifiques

De plus, devant la difficulté matérielle à constituer des lots homogènes et de taille suffisante pour la sous-traitance de tels documents, nous envisageons de mettre en place une chaîne plus légère et réactive de « numérisation à la demande », ce qui pourrait se faire de façon concertée avec le SICD de l'UJF, qui est par ailleurs demandeur de notre expertise en numérisation, notamment pour valoriser son fonds important en physique.

Au-delà de ce programme, nous recevons régulièrement des sollicitations, par exemple de la part de revues ou de séminaires confidentiels, mais aussi de l'étranger. Pour les documents publiés en France, nous voulons les traiter sur un pied d'égalité avec ceux qui ont été accueillis en priorité, après une validation scientifique légère. Nous ne traitons pas les demandes en temps réel, ce qui nous permet de gérer les priorités, mais nous avons d'ores et déjà un stock important de suggestions qu'il n'y a pas de raison de rejeter.

La question des documents étrangers pose des questions un peu différentes, du fait qu'ils doivent apporter un bénéfice à la base toute entière en terme de reconnaissance ou de visibilité. Nous accueillerons en général les revues étrangères de qualité à leurs frais. Certaines opérations sont particulièrement stratégiques car elles contribuent très positivement à notre visibilité internationale, à notre image d'ouverture et de soutien des publications académiques. Par exemple, nous avons été sollicités par l'union mathématique internationale pour être l'un des sites qui diffuseront la collection numérisée complète des actes des congrès internationaux. De même que nous avons hébergé la collection des publications d'Orsay et que nous prévoyons de l'élever *in fine* aux normes Numdam, d'autres collections pourraient être contribuées, notamment en provenance de la bibliothèque de Cornell.

3. Acquisition

Nous continuerons à soutenir notre politique d'acquisition de la production des éditeurs dont les revues sont archivées sur Numdam. Ceci est un travail d'intensité modérée, mais qui demande de l'énergie pour faire vivre les accords et pour nous adapter aux modifications de formats et de politique de la part des éditeurs. Par exemple, les *Annales scientifiques de l'ENS* revue publiée par Elsevier jusqu'en 2007, sont désormais publiées par la SMF, dont nous souhaitons également archiver les revues présentes sur Numdam, ainsi

qu'*Astérisque*, actuellement archivée nulle part. La revue *Compositio mathematica* nous a demandé de mettre en place une chaîne mais pour l'instant la collaboration avec Cambridge university press s'est révélée impossible.

Pour ce qui est des revues françaises vivantes, notre cœur de métier, la couverture par Numdam est presque exhaustive, mais l'acquisition de routine est loin d'être assurée. Le cas de revues récentes comme le *Journal of the Institute of Mathematics of Jussieu* n'a pas été véritablement abordé à ce jour. La question d'une archive pérenne de la production académique française, indépendante des aléas politiques et économiques des éditeurs, devra être reconsidérée en tenant compte de l'expérience accumulée.

4. Bibliothèques virtuelles

La question de bibliothèques virtuelles performantes et exhaustives est actuellement de nouveau très en vogue, après un temps d'essoufflement suite aux échecs des différents projets transnationaux. Le projet EuDML devrait intégrer les résultats de nos projets pilotes mini-DML et LiNum, et porter les acquis de NUMDAM en tant que bibliothèque numérique de référence au niveau européen.

III. Édition

Le pôle d'édition avait été conçu comme un soutien structurant pour les revues académiques de mathématiques au niveau national. Une des motivations était la difficulté de suivre les évolutions du secteur pour des revues reposant pour l'essentiel de leur fabrication sur une large part de bénévolat et de sous-traitance professionnelle. Le projet Cedram ne réalise au final qu'une partie de ces objectifs : la normalisation des formats, conçue entre autres pour permettre un archivage pérenne et fiable des sources, n'a été adoptée que par le noyau dur des revues présentes sur le serveur.

On peut considérer que les réunions préparant le pôle, et la numérisation des archives par Numdam, ont facilité grandement le départ d'Elsevier des deux revues qui l'on fait en réduisant les risques d'une telle opération au minimum. Mais le pôle à peine formé n'a pas été jugé capable de remplacer au pied levé un éditeur professionnel. En revanche, une revue de la SMAI a été créée sans difficulté majeure, et l'offre s'est étendue aux séminaires de grands laboratoires français. Cette offre s'est encore élargie par la création de deux séries du CIRM de Luminy (UMS CNRS/SMF) : les Actes des rencontres du CIRM, et les notes de cours du CIRM, Il y a également un projet de lancement d'une nouvelle revue éditée par la Fédération de recherche mathématique du Nord-Pas-de-Calais (FR 2956)

Le Cedram a produit une suite logicielle qui est fonctionnelle et a pu être appropriée par les revues sans grand obstacle. Pour un coût de production généralement moindre, les revues participantes bénéficient d'une édition en pointe, notamment un site web dynamique avec affichage bilingue et dual TeX/MathML. Pour être performante, la mise en œuvre du Cedram repose néanmoins sur des formats un peu contraignants : bien que nos outils aient été conçus comme modulaires, on constate que les revues sont soit intégrées au site, soit gérées de façon totalement indépendante. La solution pour l'intégration continue des archives des revues françaises dans Numdam reste à mettre au point pour les revues qui ne sont pas intégrées au projet. L'expérience avec les revues fournies par EDP Sciences montre cependant des voies prometteuses, si le source LaTeX n'est pas trop mal structuré. Ces voies seront explorées plus systématiquement, mais il faut souligner que les principaux obstacles vers un

support homogène pour la production des revues françaises en vue de leur archivage pérenne sont avant tout de nature politique.

Sur le plan technique, le Cedram est une réussite. Les principaux développements envisagés à court terme, sont la production de texte intégral XML/MathML traduits directement à partir des sources LaTeX des articles et la génération de PDF plus accessibles. Ces deux développements permettraient de mieux intégrer la production du Cedram dans un futur “web sémantique mathématique” dont EuDML pourrait préfigurer certains aspects. Des contacts sont déjà en cours à Sophia-Antipolis, Melbourne et Brno sur ces sujets qui restent spéculatifs.

IV. Archivage

L'archivage des documents numériques et spécifiquement l'archivage à long terme c'est-à-dire à l'horizon de plusieurs dizaines d'années est une nécessité qui va rapidement s'imposer. Pour l'instant, bibliothécaires et chercheurs accordent plus de confiance dans les archives papiers mais outre le fait qu'il serait peu raisonnable de numériser indéfiniment les mêmes oeuvres, de plus en plus de documents électroniques n'ont pas leur équivalent papier.

Pour nous, cet archivage concerne trois secteurs. Tout d'abord notre propre production bien entendu, ensuite la production des revues françaises académiques et enfin la production des éditeurs commerciaux.

Concernant ceux-ci, on voit bien qu'ils ont transformé la vente de revues papier en la “location” de revues électroniques. C'est-à-dire qu'ils font payer l'accès aux documents électroniques sans fournir l'archive. Cette situation n'est pas saine car on a vu dans le passé que les archives des éditeurs ne survivaient pas à la fermeture ou au rachat de la maison d'édition. Il faut donc absolument, comme cela commence à se faire que tous les contrats passés avec les éditeurs comprennent la fourniture des archives.

Le problème de l'archivage à long terme n'est pas tellement comme on l'entend souvent celui de la pérennité des supports. Bien sûr, il est nécessaire de recopier à intervalle régulier les archives d'un support à un autre en fonction des progrès technologiques. Mais comme la capacité des supports augmente plus vite que les volumes de données, l'opération est en fait de plus en plus rapide.

Les vrais problèmes se situent au niveau du suivi des formats et des logiciels qui les comprennent et sans doute plus encore au niveau de l'organisation initiale de l'archive. Des expériences récentes, aussi bien à la petite échelle des Annales de l'Institut Fourier qu'à plus grande échelle avec l'éditeur Springer, ont montré que des archives insuffisamment organisée se révèlent inutilisables quelques années plus tard.

Pour ce qui est de l'organisation initiale des archives, c'est assez naturellement au niveau de la Cellule MathDoc que cela doit se faire. Il faut tout d'abord choisir le format d'archivage. Si le pdf s'impose pour un archivage à moyen terme, un archivage à long terme doit être capable de répondre aux évolutions technologiques et se doit donc de concerner les sources (images scannées ou sources TeX). Il est impératif de conserver l'ensemble des données permettant de reconstruire le texte ce qui est beaucoup moins facile qu'il n'y paraît. Enfin cette organisation doit être suffisamment simple et explicite pour que l'archive puisse

être utilisée sans le concours de ceux qui l'ont fabriquée.

La conservation définitive des archives se fera de manière centralisée. A ce niveau il faut fixer des formats de référence et en assurer la continuité. L'Allemagne est en avance dans ce domaine avec de grands projets nationaux. En France, l'expérience la plus intéressante est celle du CINES à Montpellier.

Cependant si on ne veut pas répéter la triste expérience de la bibliothèque d'Alexandrie, il faut créer plusieurs centres d'archivages indépendants.

V. Réorganisation de la documentation mathématique

Comme nous l'avons signalé en introduction, une réflexion est en cours au niveau de l'INSMI (INstitut des Sciences Mathématiques et de leurs Interactions) sur l'avenir de la documentation mathématique. Cela concerne les abonnements aux revues, le type d'accès aux revues électroniques, l'archivage papier et l'archivage électronique.

MathDoc participe à ces réflexions avec ses partenaires du RNBM et de Mathrice. Du point de vue opérationnel, nous sommes plus particulièrement concernés par l'accès (création d'un portail spécifique) et l'archivage à moyen et à long terme.

VI. Relations nationales et internationales

Dans le bassin grenoblois

Une collaboration pour la numérisation concertée en physiques et mathématiques est prévue avec le SICD de l'Université Joseph Fourier.

La mise en ligne de la Revue de géologie alpine numérisée par nos soins se fera en concertation entre MathDoc et l'OSUG.

L'Institut Fourier nous a transmis une collection de notes de cours et de séminaires à numériser, qui est en cours d'expertise scientifique et pourrait être en grande partie incluse dans Numdam.

Programme de numérisation de la bibliothèque Mathématiques Recherche Paris Centre

L'ensemble des traitements des volumes numérisés sera réalisé sur nos serveurs et les données gérées dans une base de données dédiée. Enfin la Cellule MathDoc devrait être chargée de l'exploitation et notamment de la diffusion en ligne des ouvrages aux bibliothèques du RNBM, dans le cadre d'une future convention signée entre nos deux établissements.

Collaboration avec la Bibliothèque Nationale de France

Cette collaboration concerne des projets d'échanges de métadonnées et de liens, via soit un travail d'indexation de documents mathématiques mis en ligne par la bibliothèque, soit un travail de fourniture de nos métadonnées via le protocole d'échange OAI-PMH. Ainsi cela permettra par exemple au serveur GALLICA2 d'indexer les revues de Numdam et d'étoffer son offre mathématique.

Europe et international

La participation aux différentes instances internationales se prolongera assez naturellement : comité EPC de l'EMS, comité de coordination du Zentralblatt, comité scientifique de la série de conférences DML, etc.

Du fait de notre visibilité, nous avons été approchés par de nombreuses structures à travers le monde (notamment, la bibliothèque américaine JSTOR et le centre de recherche allemand DFKI). Nous concentrerons nos forces sur le projet EuDML et ses suites à court terme, mais resterons ouverts à des collaborations fructueuses tant que nous aurons les moyens de les conduire dans de bonnes conditions.

II. Publications

Articles :

- Thierry Bouche. — « Introducing the mini-DML project », in *New Developments in Electronic Publishing* AMS/SMM Special Session, Houston, May 2004 & ECM4 Satellite Conference, Stockholm, June 2004 (2005) 19-29.
- Y. Laurent “Documentation mathématique, Spécificités et évolutions”, *Arabesques* (2005), no. 37, p. 8-9.
 - T. Bouche, Y. Laurent & C. Sabbah “L’éditoin sans drame”, *Gazette des mathmaticiens* 108 (2006), p. 86-88.
- Thierry Bouche. — « A pdfLaTeX-based automated journal production system », *TUGboat* 27 (2006) 45-50.
- Thierry Bouche. — « Toward a digital mathematics library? », in *Communicating mathematics in the digital era*, Jon Borwein, Eugénio Rocha, and José Rodrigues (eds): AK Peters Ltd, 2008, p. 47-73.
- Thierry Bouche. — « Digital Mathematics Libraries: The Good, the Bad, the Ugly », in *Intelligent Computer Mathematics*, AISC/Calculemus/MKM 2008 proceedings (S. Autexier et al., eds.), LNAI 5144, p. 331-332, Springer-Verlag Berlin Heidelberg 2008.
- Thierry Bouche. — « Some thoughts on the near-future Digital Mathematics Library », in *Towards Digital Mathematics Library*, proceedings of the DML 2008 workshop, Birmingham, UK, July 27th 2008 (Petr Sojka ed.): Masaryk University, Brno, 2008, p. 3-15.
- Thierry Bouche. — « CEDRICS: When Cedram meets Tralics », in *Towards Digital Mathematics Library*, proceedings of the DML 2008 workshop, Birmingham, UK, July 27th 2008 (Petr Sojka ed.): Masaryk University, Brno, 2008, p. 153-165.
- Thierry Bouche. — « Report on the current state of the French DMLs », in *Towards a Digital Mathematics Library*, proceedings of the DML 2009 workshop, Grand Bend, Ontario, Canada, July 8-9 2009 (Petr Sojka ed.): Masaryk University, Brno, 2009, 10 p.
- C.Goutorbe, *Document interlinking in a digital library*. in *Towards a Digital Mathematics Library*, proceedings of the DML 2009 workshop, Grand Bend, Ontario, Canada, July 8-9 2009 (Petr Sojka ed.): Masaryk University, Brno, 2009.
- Thierry Bouche. — « Bibliothèques numériques de mathématiques en France et dans le monde », in *Images des mathématiques*: CNRS, Paris, Oct. 2009.
<http://images.math.cnrs.fr/Bibliotheques-numeriques-de.html>
- Thierry Bouche. — « Digital Mathematics Libraries: The Good, the Bad, the Ugly », *Mathematics in Computer Science*, special issue on Authoring, Digitalization and Management of Mathematical Knowledge (Serge Autexier, Petr Sojka, and Masakazu Suzuki eds.) 2010 (3), p. 227-241. DOI: 10.1007/s11786-010-0029-2
- Thierry Bouche. — «Démarrage du projet EuDML — La bibliothèque numérique européenne de mathématiques», *Gazette des mathématiciens & MATAPLI* (2010), 11 p.
- Thierry Bouche. — «Introducing EuDML—The European Digital Mathematics Library», EMS newsletter (2010), 8 p.

Logiciel libre :

L’ensemble logiciel CEDRICS (11000 lignes de code en syntaxe TeX) est un outil de production de journaux mathématiques qui permet de gérer la production des revues du

Cedram (création des PDF mais surtout des métadonnées XML/MathML), mais sert aussi à produire des métadonnées pour Numdam. Il a permis à ce jour de traiter en souplesse environ 3000 articles d'origines assez variées. Il est diffusé en logiciel libre (licences LPPL pour le code LaTeX, Cecill pour le code Tralics, cf. <http://www.cedram.org/> rubrique « Boite à outils »). Il est utilisé en République Tchèque, notamment pour la production de la revue *Archivum Mathematicum*, ce qui a donné lieu à plusieurs publications de la part de l'équipe qui l'a fait.