



*Mathdoc*

## **RAPPORT d'ACTIVITÉ**

Cellule Mathdoc – UMS 5638

2010–2014

Cellule de Coordination Documentaire Nationale pour les Mathématiques  
UMS 5638 UJF-CNRS

Cellule Mathdoc  
Université Joseph Fourier  
B.P. 74  
38402 Saint-Martin d'Hères cedex  
(France)

Tél. 04 76 63 56 36  
Fax. 04 76 63 56 11  
[accueil@mathdoc.ujf-grenoble.fr](mailto:accueil@mathdoc.ujf-grenoble.fr)  
<http://www.mathdoc.fr/>



# SOMMAIRE

<b>INTRODUCTION.....</b>	<b>5</b>
<b>RÉSUMÉ.....</b>	<b>7</b>
I.    HISTORIQUE .....	7
II.   PERSONNEL .....	8
<b>LES ACTIONS DE LA CELLULE MATHDOC.....</b>	<b>9</b>
I.    NUMÉRISATION ET BIBLIOTHÈQUES NUMÉRIQUES .....	9
A. <i>Numérisation</i> .....	9
B. <i>Acquisition</i> .....	16
C. <i>Réindexation et mise en ligne</i> .....	21
D. <i>Agrégation de métadonnées, moteurs de recherche et interfaces</i> .....	22
E. <i>Sauvegarde et archivage</i> .....	25
F. <i>Synthèse</i> .....	26
II.   ÉDITION .....	29
A. <i>Présentation</i> .....	29
B. <i>Activités</i> .....	31
III.  SERVICES DE DOCUMENTATION.....	36
A. <i>Description des activités</i> .....	36
B. <i>Description des services de documentation</i> .....	38
IV.  ACTIVITÉS TRANSVERSALES .....	44
A. <i>Développement logiciel</i> .....	44
B. <i>Site Web de l'unité et identité visuelle</i> .....	46
C. <i>Infrastructures</i> .....	46
V.    COOPÉRATIONS .....	48
A. <i>Locales</i> .....	48
B. <i>Nationales</i> .....	48
C. <i>Internationales</i> .....	51
<b>PROJET.....</b>	<b>52</b>
I.    BIBLIOTHÈQUES NUMÉRIQUES.....	52
II.   ÉDITION .....	54
III.  PORTAIL MATH ET SERVICES DE DOCUMENTATION .....	55
IV.  INFRASTRUCTURES .....	56
<b>ANNEXES.....</b>	<b>57</b>
I.    MISSIONS .....	57
II.   PUBLICATIONS .....	60
<b>GLOSSAIRE.....</b>	<b>62</b>



## Introduction

La Cellule Mathdoc a pour mission de contribuer à la structuration nationale de la documentation mathématique en fournissant des services destinés soit aux professionnels (documentalistes, éditeurs académiques, laboratoires...) soit directement aux chercheurs à travers différents portails d'accès. Partenaire technique de la base de données européenne zbMATH, de plusieurs projets européens, elle a été associée à tous les projets internationaux de bibliothèques numériques de mathématiques en Europe et dans le monde depuis l'an 2000.

Ses activités principales tournent autour de l'informatique documentaire et de la documentation électronique : la numérisation des revues de mathématiques et des ouvrages anciens (programme Numdam), l'acquisition de la production récente d'un certain nombre d'éditeurs pour archivage et accès différé sur le portail Numdam, des contributions à la construction d'une bibliothèque numérique européenne de mathématiques (projet EuDML), le développement d'un pôle de production et de diffusion électronique pour les revues académiques françaises de mathématiques (Cedram), et enfin la maîtrise d'œuvre d'un projet national de portail à destination des mathématiciens des laboratoires français (portail math). Ces actions se font en liaison constante avec les réseaux métiers de l'INSMI : le réseau national des bibliothèques de mathématiques (RNBM, GDS 2755) et le réseau des informaticiens des laboratoires de mathématiques (Mathrice, GDS 2754), mais aussi les réseaux internationaux qu'elle a elle-même contribué à créer comme l'initiative EuDML.

Le présent rapport tente de passer en revue l'ensemble des activités sur une période de cinq ans allant de l'année 2010 à la fin de l'année 2014. Une difficulté est que l'unité a connu un profond renouvellement pendant cette période assez mouvementée : trois personnes sont parties, auxquelles on peut ajouter un contractuel qui est resté 4 ans, tandis que quatre nouveaux agents ont été recrutés, dont trois il y a un an avec parfois des biseaux négatifs de plusieurs années rendant la continuité des services, et la transmission des compétences, acrobatique sinon illusoire. Ces mouvements de personnel concernent le CNRS (à part les personnels sur contrat puisqu'il se trouve que nos contrats récents se sont surtout faits à l'UJF). Nous avons deux ingénieurs UJF à Mathdoc (en fait, le personnel de l'unité était constitué de ces deux ingénieurs en 1995), il n'en reste plus qu'un dont le départ à la retraite est désormais imminent. Il est très important que ce départ ne déséquilibre pas définitivement l'apport des deux tutelles, et ne déstabilise pas l'unité si le remplacement n'est pas anticipé.

Unité de taille extrêmement réduite, dont la liste des activités est plus longue que celle du personnel, Mathdoc est d'une certaine façon un chœur de solistes (une formation resserrée dans laquelle chaque partition est jouée par un seul interprète) : la complémentarité entre les membres de l'unité a permis une très grande efficacité dans des configurations adaptées aux différents objectifs, mais l'absence d'une personne peut mettre un projet en péril. La période qui s'achève a bien illustré ce risque car il a fallu faire face à des congés (maternité, maladie) et des départs non remplacés, ce qui a affecté profondément les temps de réponse pour certaines activités (un an et demi pour mettre une revue numérisée en ligne, ou six mois pour finaliser le site web d'une revue rejoignant le Cedram, par exemple), le personnel prenant sur lui pour éviter que la qualité des

services n'en pâtisse. Un autre exemple est qu'il n'y a toujours pas à Mathdoc de personnel permanent qui pourrait prendre en charge les tâches que je réalise dans l'une des chaînes de production où j'interviens (fabrication des PDF et DjVu utilisateurs de Numdam, développement, maintenance, paramétrisation du système de production LaTeX du Cedram).

Alors qu'un certain nombre de nos projets passent du statut de pilote ou de démonstrateur à celui d'infrastructure, avec changement d'échelle et de périmètre (DML européenne et globale, centre d'excellence pour l'édition de revues scientifiques, portail math pleinement fonctionnel), il est nécessaire d'accompagner ces ambitions par des moyens adéquats et non volatiles.

Je voudrais terminer en remerciant les organisations qui nous soutiennent et toute l'équipe de Mathdoc pour le travail accompli.

Thierry Bouche, directeur de la Cellule Mathdoc

Décembre 2014

## Résumé

### I. HISTORIQUE

**1995** : Création de la Cellule Mathdoc, unité mixte de service 5638 sous la co-tutelle du Centre National de la Recherche Scientifique (CNRS) et de l'Université Joseph Fourier (UJF).

- Directeur : Pierre Bérard, directeur adjoint : Laurent Guillopé
- Personnel permanent : 2 ingénieurs UJF (Elizabeth Cherhal et Claude Goutorbe), 1 secrétaire CNRS.

La Cellule Mathdoc est créée avec pour missions :

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

Les services développés pendant cette période : moteur de recherche et interface Web du Zentralblatt ; index des thèses et des prépublications ; annuaire des laboratoires et des formations (DEA) ; catalogue fusionné des périodiques.

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

**2000** : Début de l'opération de numérisation (Numdam<sup>1</sup>). Thierry Bouche rejoint l'équipe en tant que conseiller scientifique. Ouverture du Portail documentaire mathématique.

**2001-2002** : Recrutement de deux ingénieurs CNRS. Phase préparatoire de Numdam.

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

**2003-2004** : Phase active de Numdam I. Préparation de Numdam II. Lancement de Gallica-Math, mini-DML.

**2005** : Début du programme de soutien à l'édition Cedram (recrutement d'un ingénieur CNRS qui restera un peu plus d'un an, recrutement d'un CDD pour adapter RUCHE aux AIF). Phase active de Numdam II. Début des acquisitions post-numérisation pour Numdam.

**2008** : Une revue est créée sur le Cedram (MathematicS in Action). Le Cedram étend son offre aux séminaires. Deux séries d'actes de rencontres et de cours donnés au CIRM sont créées. Le projet NUMIR des universités Paris Diderot et Pierre et Marie Curie est largement mis en œuvre à Mathdoc.

**2009** : Départ (non remplacé) d'Elizabeth Cherhal.

**2010** : Début du projet EuDML (recrutement d'un ingénieur en CDD UJF).

---

<sup>1</sup> Pour tous les sigles, se reporter au glossaire en dernière page du rapport.

**2012** : Lancement du projet portail math (recrutement de Jean-Luc Archimbaud). Participation active au chantier de la bibliothèque scientifique numérique (MESR, groupes 2, 5, 6, 7). Départ en mobilité d'Hélène Falavard.

**2014** : Ouverture de la première version du portail math. Participation aux plans d'actions partagées de la DIST. Constitution de l'initiative EuDML et participation au groupe de travail GDML de l'IMU. Lancement du Journal de l'École polytechnique et publication d'un premier livre au Cedram. Recrutement d'Hanane Ayadi (gestionnaire CNRS, en remplacement de Myriam Laurens), Sophie Min-Picault (documentaliste CNRS, remplacement d'Hélène Falavard), Olivier Labbe (informaticien CNRS). Départ de Catherine Barbe-Zoppis.

## II. PERSONNEL

- *Personnel permanent au 1<sup>er</sup> janvier 2015*
  - Claude Goutorbe, ingénieur de recherche (info), UJF, depuis octobre 1995 ;
  - Isabelle Costerg, technicienne (doc), CNRS, depuis janvier 2007 ;
  - Franck Lontin, ingénieur d'étude (info), CNRS, depuis janvier 2008 ;
  - Jean-Luc Archimbaud, ingénieur de recherche (info), CNRS, depuis mai 2012 ;
  - Sophie Min-Picault, assistante ingénieur (doc), CNRS, depuis décembre 2013 ;
  - Hanane Ayadi, assistante ingénieur (gestion), CNRS, depuis janvier 2014 ;
  - Olivier Labbe, ingénieur d'étude (info), CNRS, depuis janvier 2014.
  - Simon Chevance, IE info à l'institut Fourier, CNRS, consacre 25 % de son temps à Mathdoc
  
- *Contractuel*
  - Elodie Viavant, technicienne (doc), UJF (BSN5), depuis mai 2014.
  
- *Direction*
  - Yves Laurent, directeur de recherches CNRS, directeur d'octobre 2002 à décembre 2010 ;
  - Thierry Bouche, maître de conférences à l'université Joseph Fourier, responsable du projet Numdam en octobre 2000, responsable du programme Cedram en octobre 2005, directeur adjoint en 2007, directeur depuis janvier 2011.
  
- *Mouvements de personnel intervenus en 2010-2014*
  - Hélène Falavard, IE doc, CNRS, de 2002 à 2012 ;
  - Myriam Charles, TCE gestion, CNRS, de 2005 à 2013 ;
  - Catherine Barbe-Zoppis, IR info, CNRS, de 2001 à 2014.
  
- *Mouvements de personnel contractuel*
  - Michael Lefevre, alternance licence pro (info), 9/2010-9/2011 ;
  - Laurent Pasquali, IE info, UJF, 11/2011-5/2012 ;
  - Nicolas Houillon, IR info, UJF (EuDML), 3/2012-5/2014.

## Les actions de la Cellule Mathdoc

### I. NUMÉRISATION ET BIBLIOTHÈQUES NUMÉRIQUES

#### A. Numérisation

La numérisation est un volet important de l'activité à Mathdoc. Elle est la source du service le plus utilisé par la communauté des mathématiciens dans le monde.

##### a) Descriptions des tâches : processus

Un projet de numérisation comporte trois phases principales : la préparation, la production et la mise en ligne. Elles sont illustrées sur la figure 1.

- *Phase de préparation*

La préparation est réalisée en trois étapes qui commencent par la gestion d'un appel d'offre de marché public (rédaction d'un CCTP – cahier des clauses techniques particulières). C'est une étape qui prépare la phase de production qui est en partie réalisée par un prestataire parce que Mathdoc ne dispose pas d'une chaîne de numérisation en interne.

La seconde étape est la récupération des collections. Mathdoc ne possède pas de fonds documentaires, donc la cellule emprunte des collections à ses partenaires : bibliothèques du RNBM, bibliothèques municipales, éditeurs... Il faut donc repérer les collections et assurer le transport de celles-ci.

La dernière étape est celle du dépouillement. Cette tâche consiste en la description physique des documents (état général, taches, mauvaise impression, présence de tableaux, figures, images en tons continus...) et de la collection. Elle permet d'avoir un état des lieux et de relever les cas particuliers à traiter dans les fichiers de production. C'est une étape importante pour déterminer les responsabilités (Mathdoc, prestataire, bibliothèque) en cas de dommage sur un volume. Elle est réalisée dans un fichier tableur comportant 3 niveaux de données : la description d'un lot, la description des fascicules composant ce lot et la description des articles. Ce fichier est aussi utilisé pour le contrôle qualité, fait lors de la phase de production.

- *Phase de production*

La production est réalisée en partie à Mathdoc et chez un prestataire. Le but est de préparer les fichiers et les métadonnées pour la mise en ligne.

La première étape, le scan, est réalisée chez un prestataire. Elle consiste en la numérisation des pages d'un document. Le résultat attendu est un ensemble de fichiers images TIFF qui reproduit exactement le document original. Donc les pages blanches sont aussi scannées.

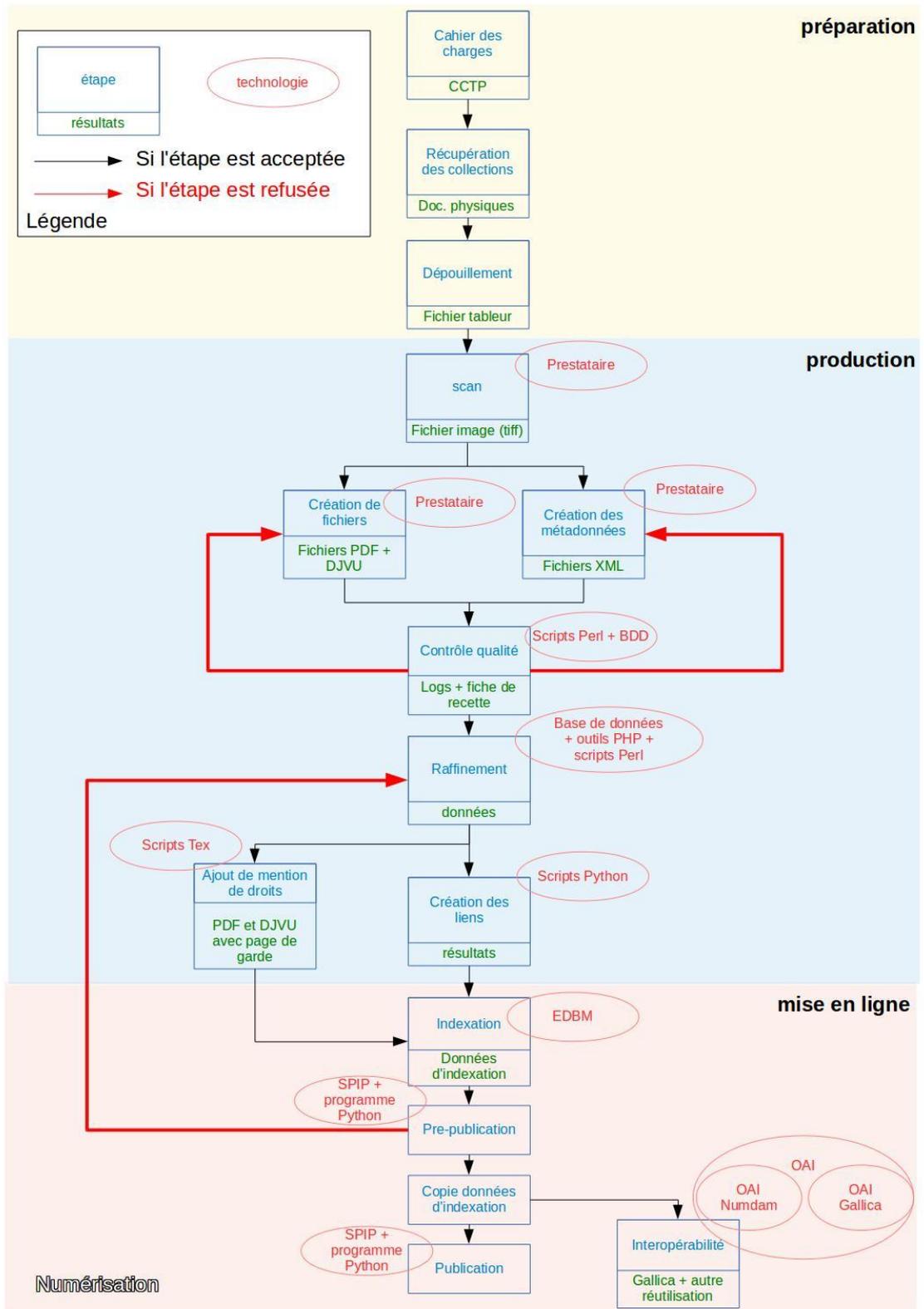


Figure 1 Étapes de la numérisation

Ensuite le prestataire réalise deux phases importantes de la production. C'est-à-dire la création des fichiers « article » : PDF, DJVU et TIFF multipages, et la création de métadonnées au format XML à partir de la DTD Volphys<sup>2</sup>. Elle permet de rassembler des informations physiques, bibliographiques, et descriptives sur les articles et les volumes.

Le travail du prestataire est validé par une fiche de recette. Ce document est le résultat du contrôle qualité réalisé par Mathdoc. Préalablement à cette tâche, il faut bien entendu récupérer les fichiers et les métadonnées produits par le prestataire, soit sur un CD-ROM ou par FTP. Un double contrôle est exécuté : l'un est exhaustif, l'autre sur échantillon.

- Le contrôle exhaustif : une série d'analyses automatiques sont effectuées sur la totalité des fichiers livrés de manière à repérer rapidement toutes les erreurs éventuelles.
- Le contrôle par échantillon : ce contrôle permet de vérifier finement la qualité visuelle des fichiers ainsi que la qualité des bibliographies. L'interface Web utilisée pour cette étape permet de garder trace de tous les fichiers contrôlés et des erreurs détectées. La taille des échantillons de données ainsi que le nombre d'erreurs sont déterminés selon la norme AFNOR X06-021, X06-22, X06-028 (principes du contrôle statistique de lots).

Le raffinement et l'enrichissement des métadonnées sont importants pour augmenter la qualité des données. Cette étape consiste à améliorer et à unifier les métadonnées : à corriger la syntaxe ou l'orthographe des titres, à rajouter les formules en TeX dans les titres, résumés et bibliographies, à corriger les bibliographies, à unifier la base de données d'auteurs de Numdam (dédoublonnage, fusion), à rajouter des relations entre articles de type « suite de », « erratum de », etc.

Pour préparer la mise en ligne des fichiers de production, une page de garde est ajoutée aux fichiers PDF et DJVU des articles pour rappeler l'origine de l'article (référence bibliographique de l'article, logo Numdam et lien sur l'URL pérenne sur le site numdam.org). Elle comporte aussi une mention des usages autorisés.

En parallèle, la création de liens est faite pour réaliser la correspondance entre les articles et les références bibliographiques avec les entrées des bases de données Zentralblatt et Mathscinet, mais aussi à l'intérieur de Numdam.

- *Phase de mise en ligne*

La mise en ligne sur internet est la finalité du projet de numérisation.

La première étape est l'indexation des données pour permettre les fonctionnalités importantes en ligne : feuilletage<sup>3</sup> et recherche<sup>4</sup>. Cette étape est concomitante à la création de liens.

---

2 DTD disponible sur <http://www.numdam.org/dtd/volphys.dtd>

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

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

Ensuite une première épreuve appelée prépublication est mise en ligne pour permettre une dernière vérification des documents et métadonnées présentés. C'est aussi à ce moment que les pages éditoriales (page d'accueil, barrière mobile) sont mises à jour en français et en anglais. Si la prépublication n'est pas validée, il faut revenir à l'étape de raffinement pour parfaire les données.

Si la prépublication est validée une copie des données d'indexation est faite pour permettre au site public de présenter le contenu validé. Ces données d'indexation servent de base pour les serveurs OAI qui assurent l'interopérabilité d'autres projets comme Gallica, BASE ou EuDML.

La dernière étape (la publication) est une simple copie sur le site public de toutes les fonctionnalités et pages de la prépublication validée.

Au cours des années 2010 à 2014, Mathdoc a travaillé sur quatre projets de numérisation. Nous détaillerons ci-dessous les spécificités des projets et les activités entreprises.

## *b) Programme Numdam*

### *1) Présentation*

Le principal projet de numérisation de la Cellule Mathdoc est le programme Numdam (numérisation de documents anciens mathématiques) lancé en l'an 2000. La raison d'être du programme est la nature même de la documentation mathématique qui ne se périmé pas et qui 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'originalité de Numdam repose sur 4 points :

- **Validation scientifique** : ne sont présentés que des textes 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.

Actuellement, la bibliothèque Numdam est composée de 29 revues, 35 séminaires et 3 mémoires, soit un total de 67 collections. Elle est reconnue comme une référence dans le paysage national et international.

La période 2010-2014 correspond à la fin de la phase 3 et au début de la phase 4.

### *2) Activités*

Les activités réalisées dans le cadre du programme Numdam sont toutes celles décrites dans le processus de numérisation.

### *3) Achèvement de Numdam 3*

Le tableau ci-dessous présente les titres numérisés au début de la période couverte par ce rapport, qui correspond à la fin de la phase 3 de Numdam dont le reste des travaux est décrit dans le précédent rapport d'activité. Le marché de numérisation a dû être prolongé de six mois (financés sur le budget de l'unité) pour réussir à traiter la collection complète du *Journal de la société*

*française de statistiques*. Cette prolongation nous a également permis de répondre positivement à la demande de la revue *Philosophia Scientiae*.

<b>Titre de revue (période de publication traitée)</b>	<b>Volumes</b>	<b>Articles</b>	<b>Pages</b>
Cahiers de l'analyse des données (1976-1997)	88	716	11 153
Journal de la société française de statistique (1860-2006)	227	6 210	56 577
Mémorial des sciences mathématiques (1925-1972)	163	163	12 729
Philosophia Scientiae (1996-1999)	14	144	2 800
<b>Total Numdam 3 (mi 2011)</b>	<b>492</b>	<b>6 593</b>	<b>83 260</b>

En 2011, nous avons accumulé un retard de mise en ligne de l'ordre de 100 000 pages. Nous avons donc mis à profit l'arrêt des financements de numérisation pour nous concentrer sur la publication des collections numérisées. C'est en juillet 2012 que nous sommes parvenus à résorber ce retard.

#### 4) Numdam 4

La phase 4 du programme Numdam a été retenue dans le cadre de l'appel d'offre BSN5 de la Bibliothèque Scientifique Numérique (MESR) en 2013. Elle bénéficie en outre du soutien financier du laboratoire de mathématiques de Rennes (IRMAR) ainsi que du Labex associé (Centre Henri Lebesgue) et du soutien moral de la SMF et de la SFdS. Les financements ont été reçus début 2014 pour un calendrier de travaux s'échelonnant sur 18 mois.

La volumétrie prévisionnelle est de 241 volumes, 21 021 articles et 237 171 pages. C'est le plus petit programme depuis le lancement de Numdam.

Les titres prévus pour l'instant sont les suivants :

- Publications des séminaires de mathématiques et informatique de Rennes
- Les Cahiers du Bureau universitaire de recherche opérationnelle (devenu dans l'usage : Les cahiers du BURO)
- Séminaire de théorie des nombres de Grenoble
- Publications du département de mathématiques (université de Lyon 1)
- Recherche Coopérative sur Programme RCP numéro 25 (Physique théorique-Mathématiques, Strasbourg)
- Collection de cours avancés de l'Institut Fourier (université de Grenoble 1).

Les étapes de préparation pour ces titres ont été réalisées. Nous sommes actuellement en phase de production. Il n'y a que les « Publications des séminaires de mathématiques et informatique de Rennes » qui ont été validées pour l'instant.

<b>Titre de revue (période de publication traitée)</b>	<b>Volumes</b>	<b>Articles</b>	<b>Pages</b>
Publications des séminaires de mathématiques et informatique de Rennes (1964-2000)	91	806	17 233
<b>Total Numdam 4 (fin 2014)</b>	<b>91</b>	<b>806</b>	<b>17 233</b>

### *c) Œuvres scientifiques de Laurent Schwartz*

#### *1) Présentation*

Mathdoc a collaboré à l'édition de trois volumes et du DVD des œuvres scientifiques de Laurent Schwartz dans la collection « Documents mathématiques » de la Société mathématique de France. Ce projet qui incubait à la SMF et à l'X dès 2006 a effectivement démarré en 2008 : une bibliographie complète de Laurent Schwartz a été compilée par Alain Guichardet et une sélection des articles à reproduire dans les volumes papier et sur un CD-ROM a été faite sous sa direction.

Mathdoc a utilisé le marché Numdam 3 pour numériser en 2008 une soixantaine de tirés à part provenant des archives personnelles de Laurent Schwartz à Polytechnique. Mathdoc a ensuite effectué un travail sur les métadonnées pour préparer une interface de consultation de l'ensemble des textes sélectionnés (d'origine Numdam, les tirés à part, et quelques textes récupérés ici ou là). Ce travail s'est achevé en 2010. Finalement, les volumes et le DVD ont été publiés en 2011.

L'accord de partenariat prévoyait une mise en ligne en accès libre à l'issue d'un délai raisonnable, ce qui fait donc partie de nos projets, dans le cadre de la nouvelle plateforme de bibliothèque numérique.

#### *2) Activités*

Pour une grande partie des textes reproduits dans les volumes imprimés et sur le DVD, Mathdoc a assuré la phase de préparation et de production. Cela représente 122 documents pour un total de 1391 pages.

Le reste du travail a été de préparer les pages HTML de navigation pour le DVD et les données correspondantes (présentation des bibliographies, liens...). Une partie considérable du temps passé sur ce projet a été consacrée à des discussions parfois houleuses entre les partenaires entrecoupées de périodes d'inactivités parfois très longues.

### *d) NuMIR*

#### *1) Présentation*

NuMIR est un projet de numérisation de monographies des bibliothèques Mathématiques Informatique Recherche (MIR) des Universités Paris Diderot et Pierre et Marie Curie. L'accès aux textes intégraux est réservé aux seuls membres du RNBM. Il a été démarré en 2009 sur des ouvrages épuisés des éditeurs Springer et Birkhäuser.

#### *2) Activités*

La Cellule Mathdoc est partenaire du projet en proposant ses méthodes et son support aux phases de préparation et de production. La phase de mise en ligne est réalisée à Mathdoc.

Dans la phase de préparation, la cellule a collaboré à la rédaction du cahier des charges et des spécifications techniques. Nous avons participé à l'analyse comparative des offres, puis au choix

du prestataire. Auparavant, nous n'avions jamais numérisé des monographies, donc un travail d'adaptation sur la DTD a été nécessaire.

La phase de production, en particulier le contrôle qualité, a été réalisée à la bibliothèque MIR, après formation à Mathdoc et en utilisant nos outils spécifiquement adaptés.

La phase de mise en ligne a été réalisée à Mathdoc sous notre supervision par un contractuel rémunéré par l'université de Paris 6 sur le budget NUMIR.

### 3) Collections

La collection présente sur NuMIR représente 372 ouvrages. Elle concerne les collections suivantes :

- Applications of mathematics
- Applied mathematical sciences
- Ergebnisse der Mathematik und ihrer Grenzgebiete
- Graduate texts in mathematics
- Grundlehren der mathematischen Wissenschaften
- Progress in mathematics
- Progress in probability.

Ce sont 282 titres édités par Springer, 81 titres édités par Birkhäuser. À cela se sont ajoutés ultérieurement les œuvres complètes de Charles Ehresmann, un cours de Bruno Poizat ainsi qu'un livre de Pierre Dolbeault qui doivent en principe être diffusés en libre accès car non soumis au contrat passé avec Springer. Ça n'est pas aujourd'hui le cas car le libre accès ne faisait pas partie du cahier des charges de la plateforme de diffusion créée pour ce projet, qui est actuellement figée.

## e) *Revue d'Écologie Alpine*

### 1) *Présentation*

Mathdoc a reçu en 2012 un financement de la mission patrimoine scientifique et technique de l'UJF pour numériser la Revue d'Écologie Alpine et des documents cartographiques.

### 2) *Activités*

Les activités dans les phases de préparation et de production ont été proches du programme Numdam. Par contre, la phase de mise en ligne a utilisé des outils différents pour présenter les cartes et les liens entre les articles et les cartes. Elle a également permis de tester des technologies différentes de celles de la plateforme numdam.org, avec en toile de fond le projet de refonte de notre plateforme d'hébergement et de diffusion de collections numériques. Le site a donc été développé en Python dans le framework Django avec moteur de recherche SolR.

### 3) Volumétrie

Ce programme de numérisation porte sur 36 volumes et 179 figures.

Titre de revue (période de publication traitée)	Volumes	Articles	Pages	Planches
Documents pour la carte de la Végétation des Alpes (1963-1972)	10	68	1763	73
Documents de Cartographie Écologique (1973-1988)	22	165	2489	76
Revue d'Écologie Alpine (1991-1997)	4	28	435	30
<b>Total</b>	<b>36</b>	<b>261</b>	<b>4687</b>	<b>179</b>

## B. Acquisition

### a) Présentation

L'activité d'acquisition consiste à traiter les données des éditeurs partenaires pour assurer la continuité des collections. Elle se fait uniquement dans le cadre du projet Numdam. L'enjeu est de faire de Numdam une archive vivante de la production des revues. Les acquisitions sont idéalement réalisées l'année suivant la parution d'un volume complet. Les métadonnées sont alors navigables sur le site numdam.org mais les textes des articles sont téléchargeables à l'issue de la barrière mobile.

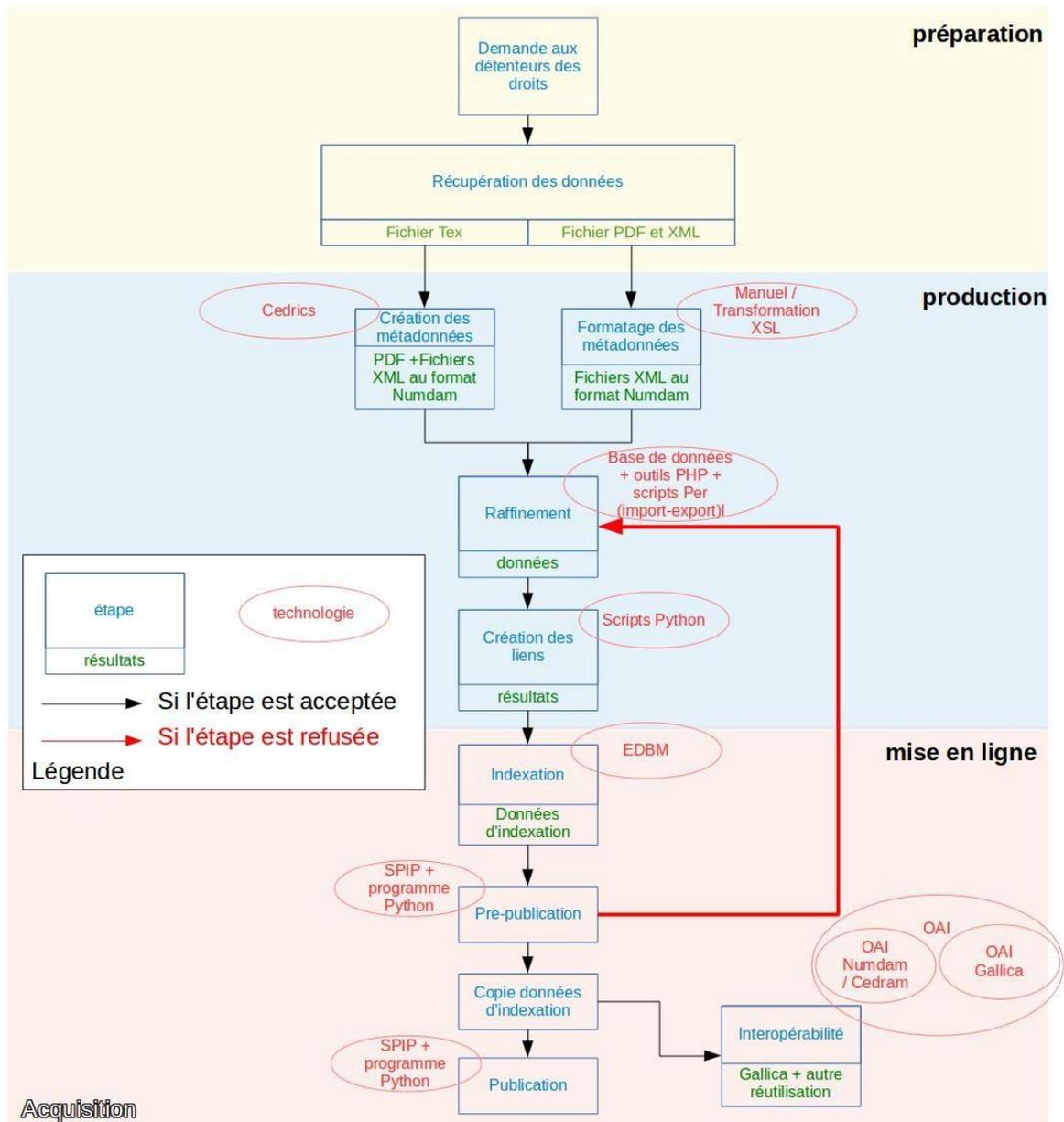
### b) Activités

- *Descriptions des tâches : processus*

Un projet d'acquisition comporte trois phases principales : la préparation, la production et la mise en ligne. Elles sont illustrées sur la figure 2.

- *Phase de préparation*

L'étape, la plus délicate et laborieuse, de la phase de préparation est d'obtenir l'accord des éditeurs pour la récupération des données. Les accords sont fragiles et nombreux sont les éditeurs qui « oublient » de transférer leur production récente. Certains éditeurs ont cependant automatisé cette tâche, parfois avec des résultats surprenants (Springer nous livrant les fascicules d'une revue d'urologie en lieu et place des *Publications de l'IHES...*). Les fichiers numériques collectés sont hétérogènes et peuvent être des sources LaTeX, des fichiers XML de structure et granularité variables. Un fichier PDF par article est impératif. À l'heure actuelle nous diffusons ce fichier tel que reçu, sans ajouter de page de garde comme nous le faisons pour les articles numérisés. Dans une logique d'archivage à long terme il serait intéressant de convertir les fichiers en PDF/A et de les signer numériquement, ce que nous n'avons pas pu faire jusqu'à présent, faute des ressources nécessaires.



**Figure 2 Étapes de l'acquisition numérique**

- *Phase de production*

La première étape de la phase de production est découpée en deux types de chaînes qui dépendent des types de fichiers reçus. Le but étant d'avoir des fichiers PDF et des métadonnées produites selon la DTD Volphys exploitée sur Numdam. Pour des fichiers LaTeX, la chaîne d'acquisition utilise l'environnement de production du Cedram. C'est-à-dire que la revue est virtuellement refabriquée à partir de ces sources de façon à obtenir un XML complet tel qu'il serait exploité par le Cedram, lequel est versé dans Numdam à l'aide du process opérationnel pour les revues du Cedram. Ce travail est en partie scripté (si les données reçues le permettent) mais requiert toujours une mise au point manuelle et de nombreuses vérifications. D'une certaine façon,

la mécanique du Cedram (environnement LaTeX/Talrics, dit Cedrics) peut être vue comme un moyen puissant de produire des métadonnées XML précises à partir de sources LaTeX.

Pour les fichiers XML des transformations XSLT sont appliquées.

Pour récupérer les formats propriétaires des éditeurs, douze chaînes d'acquisition ont été développées : six utilisant des mécanismes XSLT (provenance EDP sciences pour les années récentes, la revue de Padoue, le projet Euclid, Elsevier, Springer, le Cedram), et six autres utilisant du LaTeX avec l'outil Cedrics (provenance SMF, la revue de Pise, *Philosophia Scientæ*, EDP Sciences avant 2009, plus ponctuellement pour les années 2001-2003 des *Publications de l'IHES* et 2007-2008 du *Journal de la société française de statistiques*).

Dans certains cas les métadonnées n'existaient pas ou étaient inexploitable, nous les avons donc créées à la main en XML (revue *Diagrammes*) ou insérées dans une chaîne de numérisation (*Cahiers de topologie et géométrie différentielle catégoriques*).

- *Phase de mise en ligne*

Dans tous les cas, le résultat de la phase précédente est un ensemble de PDF articles et de métadonnées XML au format Numdam. Les fascicules acquis sont donc insérés dans la chaîne de production Numdam où un travail de raffinement et d'enrichissement des métadonnées et de création de liens est réalisé. La mise en ligne est donc identique à celle des collections numérisées de Numdam.

### c) *Titres en acquisition numérique entre 2010 et 2014*

Sur la période concernée, Mathdoc a traité **354 volumes** représentant 2838 articles et 74 500 pages. Le tableau ci-dessous présente les titres acquis sur cette période.

Titre de revue	Années	Volumes	Articles	Pages
Annales scientifiques de l'École Normale Supérieure	2008-2013	36	148	6177
Annales de l'institut Fourier	2009-2012	26	318	10440
Annales de l'institut Henri Poincaré, section B	2008-2013	26	314	7258
Annales de la Faculté des sciences de Toulouse : Mathématiques	2009-2012	19	162	4325
Annales mathématiques Blaise Pascal	2009-2012	8	70	1987
Annali della Scuola Normale Superiore di Pisa - Classe di Scienze	2004-2008	20	13	3704
Bulletin de la Société Mathématique de France	2001-2013	52	273	8005
Cahiers de Topologie et Géométrie Différentielle Catégoriques	2007-2008	8	33	640
Diagrammes	2004-2007	4	4	223
ESAIM: Control, Optimisation and Calculus of Variations	2009-2010	8	97	2101
ESAIM: Mathematical Modelling and Numerical Analysis	2009-2010	12	105	2517

<b>Titre de revue</b>	<b>Années</b>	<b>Volumes</b>	<b>Articles</b>	<b>Pages</b>
ESAIM: Probability and Statistics	2009-2010	2	45	913
Journal de la société française de statistique	2007-2008	8	38	805
Journal de théorie des nombres de Bordeaux	2009-2012	12	168	3032
Journées équations aux dérivées partielles	2007-2011	4	44	823
Philosophia Scientiae	2000-2007	16	172	3716
RAIRO - Operations Research	2009-2010	8	54	835
RAIRO - Theoretical Informatics and Applications	2009-2010	8	64	1305
Rendiconti del Seminario Matematico della Università di Padova	2010-2013	8	103	2181
Revue d'histoire des mathématiques	1995-2012	36	144	5654
Revue de statistique appliquée	2005-2006	8	42	767
Séminaire Bourbaki (Astérisque)	2002-2006	4	58	1664
Séminaire Équations aux dérivées partielles (Polytechnique)	1996-2010	14	311	4309
Séminaire de théorie spectrale et géométrie (Grenoble)	2004-2011	7	58	1098
<b>Total</b>		<b>354</b>	<b>2838</b>	<b>74449</b>

#### *d) Perspectives*

Le travail sur les acquisitions est récurrent sur les 29 publications vivantes de Numdam : 24 revues, 4 séminaires et une collection de mémoires. Mais seulement six sont à jour. Le tableau suivant présente le détail de cet état des lieux. Les revues qui ne sont pas à jour sont surlignées.

<b>Titre complet</b>	<b>Années sur numdam.org</b>	<b>Acquisition post-num</b>	<b>Acquisition à jour</b>
Annales de la faculté des sciences de Toulouse	1887-2012	oui	oui
Annales de l'institut Fourier	1949-2012	oui	oui
Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques	1983-2013	oui	oui
Annales de l'institut Henri Poincaré (C) Analyse non linéaire	1984-2009	oui	non
Annales mathématiques Blaise Pascal	1994-2012	oui	oui

<b>Titre complet</b>	<b>Années sur numdam.org</b>	<b>Acquisition post-num</b>	<b>Acquisition à jour</b>
Annales scientifiques de l'École Normale supérieure	1864-2013	oui	oui <sup>5</sup>
Annali della Scuola Normale Superiore di Pisa - Classe di Scienze	1871-2008	oui	oui <sup>6</sup>
Bulletin de la société mathématique de France	1872-2013	oui	oui
Compositio Mathematica	1935-1996	non	non
ESAIM : Control, Optimisation and Calculus of Variations	1996-2010	oui	non <sup>7</sup>
Cahiers de Topologie et de Géométrie Différentielle Catégoriques	1966-2008	oui	non <sup>8</sup>
Diagrammes	1979-2007	oui	non <sup>9</sup>
RAIRO - Informatique Théorique et Applications	1974-2010	oui	non
Journées Équations aux dérivées partielles	1974-2011	oui	oui
Journal de la Société française de Statistique	1860-2008	oui	non <sup>10</sup>
Journal de théorie des nombres de Bordeaux	1989-2012	oui	oui
ESAIM : Modélisation Mathématique et Analyse Numérique	1967-2010	oui	non
Mathématique et Sciences Humaines	1962-1999	non	non <sup>11</sup>
Mémoires de la société mathématique de France	1964-2000	non	non <sup>12</sup>
Philosophia Scientiæ	1996-2007	oui	non <sup>13</sup>

<sup>5</sup> En 2014, un gros effort a été fait de la part de la SMF et de la Cellule Mathdoc pour mettre à jour leurs revues.

<sup>6</sup> Les annales de Pise implémentent le créneau mobile à leur façon, en nous transférant chaque année leur production cinq ans après parution.

<sup>7</sup> Les revues EDP Sciences ont été tout d'abord converties à partir de sources LaTeX. À partir de 2010 nous avons récupéré leurs XML JATS (format adopté par EuDML dont ils étaient partenaires). Il est probable que le retard sur Numdam provient uniquement d'une surcharge de travail à Mathdoc et qu'il est facile à corriger.

<sup>8</sup> Cette revue un peu atypique n'a pas d'édition électronique. Un processus manuel d'acquisition a été mis en place, qu'il est prévu de renouveler dans le cadre de Numdam 4 pour les volumes récents (que nous avons reçus).

<sup>9</sup> Cette autre revue atypique nous a envoyé des PDF. Des métadonnées XML ont été produites à la main et doivent prochainement être mises en ligne.

<sup>10</sup> Cette revue est désormais publiée en accès libre après un bref passage sur le site de la SMF que nous avons archivé. Les revues en accès libre sont bizarrement moins motivées pour se voir archivées sur Numdam.

<sup>11</sup> Cette revue est publiée sur revues.org. Nous n'avons pas réussi à convaincre les parties prenantes de faire en sorte de faire apparaître les collections numérisées par Numdam sur le site de la revue ni de transférer les articles récents. Seuls des liens croisés globaux permettent d'orienter le lecteur.

<sup>12</sup> Projet 2015 à la SMF.

<b>Titre complet</b>	<b>Années sur numdam.org</b>	<b>Acquisition post-num</b>	<b>Acquisition à jour</b>
Publications Mathématiques de l'Institut des Hautes Études Scientifiques	1959-2007	oui	non
ESAIM : Probability and Statistics	1997-2010	oui	non
Revue d'histoire des mathématiques	1995-2006	oui	oui
RAIRO - Recherche Opérationnelle	1968-2010	oui	non
Rendiconti del Seminario Matematico della Università di Padova	1930-2013	oui	oui
Séminaire Bourbaki	1948-2006	oui	non
Séminaire de probabilités de Strasbourg	1967-2002	non	non
Séminaire de théorie spectrale et géométrie	1982-2011	oui	oui

### C. Réindexation et mise en ligne

Nous mettons sous cette bannière l'activité qui consiste à reprendre des fonds déjà numérisés pour les rendre exploitables avec un niveau d'ergonomie comparable à celui de nos services. L'exemple principal est la réindexation au niveau article de collections de Gallica, qui est décrite dans un précédent rapport d'activité. Nous sommes fréquemment sollicités par des détenteurs de collections numériques qui souhaitent déposer leurs collections chez nous car

- ils nous font confiance ;
- ils pensent que ces collections auront une plus grande visibilité chez nous, en particulier si elles sont intégrées à Numdam.

Nous avons ainsi mis en ligne par le passé les Publications mathématiques d'Orsay (publications orange), numérisées et déjà mises en ligne par la bibliothèque Hadamard, les archives Bourbaki (numérisées ailleurs, nous avons cependant numérisé une petite partie du corpus pour qu'il soit complet). Mais nous ne les avons pas intégrées à Numdam et nous avons donc créé des sites *ad hoc*, ce qui a conduit à un émiettement de notre offre et une perte de lisibilité.

Le problème de l'intégration à Numdam de collections que nous n'avons pas produites est assez semblable à celui de l'acquisition : il nous faut des fichiers articles de qualité acceptable, et des métadonnées précises et complètes de façon à conserver la qualité du service et l'homogénéité des fonctions disponibles. Lorsque nous récupérons des articles il est rare que les métadonnées contiennent les bibliographies. Mais il n'est pas rare non plus que les collections concernées n'aient même pas de métadonnées au niveau article (dans Gallica, il y a les métadonnées au niveau du volume physique comme on pourrait les trouver dans le catalogue d'une bibliothèque avec, dans

---

<sup>13</sup> Cette revue est découpée entre Cairn, revues.org et Numdam.

les bons cas, une table des matières à partir de laquelle on peut déduire approximativement le découpage en articles).

Nous espérons comme résultat du projet EuDML une chaîne de traitement automatisée permettant de préparer ce travail, mais c'est un point qui a été rapidement abandonné et qui n'a donc pas avancé. Pendant la période couverte par ce rapport, nous avons recensé les mouvements sur Gallica (certaines œuvres ont été retirées pour des questions de droits, que nous avons donc retirées également, une dizaine d'années du *Journal de mathématiques pures et appliquées* a été numérisée, ainsi que de nombreuses années du *Bulletin des sciences mathématiques* — à noter que les CRAS sont presque entièrement disponibles jusqu'en 1997 et mériteraient également une indexation au niveau article pour être utilisables par les chercheurs). Mais nous n'avons pas eu les moyens humains de faire avancer ce projet. Il n'empêche que cette activité nous semble importante dans la mesure où elle rend utilisable des collections par ailleurs disponibles mais dotées de métadonnées insuffisantes : c'est un verrou qui limite grandement tous les projets DML dans le monde et qu'il faudra bien un jour faire sauter.

#### *D. Agrégation de métadonnées, moteurs de recherche et interfaces*

Un projet comme Numdam, quels que soient sa réussite ou son caractère exemplaire, ne propose qu'un infime pourcentage des collections utiles au mathématicien actif, pour ne pas parler du patrimoine mathématique en général. Mathdoc a donc toujours cherché des partenariats pour former des bibliothèques de plus grande envergure et promu le protocole OAI-PMH qui est un moyen à la fois simple et efficace d'agrégier des bibliothèques virtuelles à partir de briques dispersées. C'est pourquoi nous nous sommes dotés de serveurs OAI-PMH pour diffuser nos métadonnées qui sont ainsi réutilisées par tous les sites intéressés ; un serveur spécifique a été développé en 2012 dans le cadre de notre partenariat avec la BNF, de telle sorte que Numdam est désormais bibliothèque associée à Gallica. Mais les collections Numdam peuvent aussi être cherchées dans des ensembles plus vastes adaptés à des publics variés : OAIster (l'ancêtre des agrégateurs OAI), Isidore (dédié SHS, où l'on pense toucher les historiens des sciences, p. ex.), BASE (Bielefeld academic search), etc.

Dans l'autre direction, nous n'avons eu de cesse de chercher à faire avancer en parallèle l'idée d'une bibliothèque universelle de mathématiques, qui serait constituée de réservoirs de données de tailles variables produits par des organisations réparties dans le monde mais interopérables. Pour faire avancer les choses, nous avons dès 2006 créé la mini-DML comme démonstrateur de faisabilité : il s'agit d'un agrégateur OAI-PMH qui récupère les métadonnées d'une dizaine de sources et fournit une interface simple de recherche. Pendant de nombreuses années, certaines collections n'ont été cherchables que là (notre JMPA par exemple, mais aussi la revue *Portugalia Mathematicæ*, et dans une certaine mesure une partie des collections numérisées à Göttingen). Dans la même veine, et à la même époque, le service Linum a permis de naviguer dans plusieurs sources de livres numérisés. Le catalogue n'a pas été mis à jour récemment. La stratégie retenue a été de faire avancer les projets européens ou internationaux plutôt que de créer de nouveaux silos de données locaux. Cela dans le but d'y intégrer nos collections isolées (comme par exemple le JMPA qui a été versé dans EuDML), mais aussi de faire développer et tester des technologies à une échelle que nos moyens ne nous permettent pas d'atteindre en vue de nous les réapproprier *in fine* pour fournir un meilleur accès à nos collections.

Au cours de la période, un certain nombre d'avancées majeures ont été obtenues :

1. Le projet EuDML a été financé et a donc réalisé le premier projet concret dépassant les frontières nationales.
2. L'académie des sciences de Washington a décroché un financement de la fondation Sloan pour faire une étude de faisabilité d'une bibliothèque de mathématiques historiques. La fondation Sloan ayant conscience du caractère mondial de la littérature mathématique, a pris contact avec le CEIC de l'IMU pour organiser un symposium et impliquer la communauté internationale dans la réflexion. Il s'en est suivi un rapport du National Research Council en 2014, un certain nombre d'activités à l'occasion du congrès international de Séoul, et la création d'un groupe de travail pour concevoir l'implémentation de ce projet.

## *a) Le projet EuDML*

### *1) Présentation*

Après de nombreuses tentatives infructueuses (qui remontent à 2003) et dont nous avons toujours été partie prenante, un projet de bibliothèque numérique de mathématiques a finalement été financé par la Commission européenne. C'est Thierry Bouche qui a été chargé par la SME en 2007 (il était alors membre du comité EPC) de réunir un consortium et de faire une proposition dans le cadre du programme eContentPlus dans lequel une proposition coordonnée par la bibliothèque universitaire de Göttingen avait déjà échoué. La proposition se fonde sur les principes qu'il a proposés et discutés au cours de nombreuses visites en Europe ; elle se distingue de propositions précédentes par son nom (DML-EU : chapitre européen de la DML, contre *EuDML* où « Eu » se prononce comme un *U* anglais, donc comme la moitié d'un *W*), le principe ferme d'une bibliothèque distribuée de textes *librement accessibles à terme*, l'introduction enfin de techniques de gestion des savoirs mathématiques (ou MKM : *mathematical knowledge management*) et d'une ambition de faire bouger les lignes en matière d'interaction avec le corpus. Du point de vue grenoblois, la principale difficulté a été d'obtenir un support administratif de la part des tutelles ce qui a abouti à proposer au partenaire portugais de coordonner le projet du point de vue administratif, technique et financier, Mathdoc conservant la coordination scientifique uniquement. Ce projet n'a pas été retenu par eContentPlus en 2008, mais l'année suivante dans le programme qui lui faisait suite : CIP ICT PSP Digital Libraries : Open access to scientific information. EuDML a donc été un projet pilote de type B, budget consolidé total de 3,2 M€, financement européen de 1,6 M€, qui aura été actif du premier février 2010 au 31 janvier 2013.

### *2) Actions de Mathdoc*

Mathdoc est donc coordonnateur scientifique du projet, et responsable du 3<sup>e</sup> *Work package* (agrégation de métadonnées). Le 2<sup>e</sup> *Work package* est de fait dirigé aussi pour moitié par Mathdoc (communication avec le conseil consultatif scientifique, relation avec les partenaires potentiels). Par ailleurs, les collections apportées par Mathdoc (Numdam, Cedram, Gallica-Math) sont les deuxièmes par la quantité, après celles du projet allemand GDZ.

Mathdoc a contribué de façon significative aux actions suivantes :

- constitution d'un réseau européen de centres DML ;
- recensement des collections existantes, analyse détaillée sur les plans quantitatif et qualitatif ;

- veille technologique sur les standards de métadonnées adaptées au corpus mathématique ;
- définition du schéma EuDML (basé sur NLM JATS) et des types de documents pris en compte ;
- mise en œuvre des conversions de formats de métadonnées ;
- conseil, soutien, développement à tous les stades requis pour faire rentrer dans le système les collections d'un partenaire ou associé du projet (BDIM a rejoint le projet en cours, une bibliothèque numérique serbe juste après la fin) ;
- outils de conversion à la volée de TeX vers MathML (basé sur Tralics et le savoir-faire du Cedram) ;
- outils d'association de référence bibliographique (basé sur le logiciel de *matching* développé pour Numdam) ;
- rétroconversion des articles scannés de Numdam vers LaTeX et XML/MathML à l'aide d'une version spécifique du logiciel InftyReader développée pour le projet ;
- validation des choix techniques, des outils et de l'interface, notamment en faisant appel à des mathématiciens grenoblois.

### 3) *Données chiffrées*

Comme CIP n'est pas un PCRD, les règles de financements sont un peu différentes (et nettement moins favorables : remboursement de 50 % max. des coûts éligibles, taux d'environnement plafonné à 30 %, pas de tierce partie liée). Elles ont en particulier imposé que Mathdoc participe au projet comme deux entités distinctes pour que les coûts des personnels de chaque tutelle puissent être affectés au projet. La charge administrative a donc été assez lourde pour un bénéfice modeste : pour une raison non entièrement élucidée, le personnel CNRS n'a pas vraiment réussi à consacrer du temps à ce projet, qui a au final reposé pour l'essentiel sur trois personnes, toutes employées par l'UJF. Les comptes finaux font donc apparaître une contribution de Mathdoc en ressources humaines de l'ordre de

- 4 personnes-mois pour le CNRS (16 k€ perçus) ;
- 44 personnes-mois pour l'UJF (environ 200 k€ perçus).

Mathdoc a défini trois types de documents (article de revue, livre, ouvrage en plusieurs volumes) et converti ses collections (y compris le *Journal de math. Pures et appliquées* : en tout 55000 références), ainsi que celles du projet espagnol (DML-E), et russo-allemand (GDZ, RusDML) : 83500 références supplémentaires. En fait, les 225000 références connues de EuDML ont été converties et validées sous la supervision de notre équipe.

### 4) *La suite : EuDML initiative ?*

L'évaluation du projet par la Commission européenne a salué le dépassement des objectifs dans le domaine où Mathdoc était très impliqué (agrégation de métadonnées, intégration des collections, ralliement de nouveaux partenaires), mais a été assez négative sur la façon dont l'évaluation interne du projet avait été conduite, et sur l'étude économique attendue de la part du consortium pour anticiper ses coûts et assurer sa pérennité.

Le plan en vue d'une infrastructure durable passait par la constitution d'une association formelle des partenaires responsable de la vie du site web, de la maintenance des outils et de la bonne gestion des collections. C'est ce qui a été réalisé au cours de l'année 2013 sous le nom *EuDML initiative*. Le temps que les statuts soient signés par les douze partenaires fondateurs a permis de réunir l'assemblée générale à Varsovie seulement en juin 2014. L'initiative est une

association dont le président élu est Thierry Bouche. Elle s'est dotée d'un comité exécutif et d'un comité technique dans lequel Olivier Labbe a été élu aussi. Un plan de travail a été rédigé, comportant des actions urgentes (défauts qui nuisent à l'image du site EuDML), des actions à moyen terme et des objectifs à plus long terme. Il a été présenté au conseil de la SME à San Sebastian, un peu plus tard dans l'année. Malheureusement, l'initiative manque de main d'œuvre et, si le travail à faire a été préparé par certains collègues, l'implémentation qui repose sur la bonne volonté des ingénieurs de Varsovie n'a pas avancé.

### *b) Les projets WDML*

Thierry Bouche étant membre du CEIC, il a participé avec d'autres membres de ce comité à l'organisation du symposium de Washington les 1-3 juin 2012 (conférence invitée sur EuDML, modération de la table ronde sur les bibliothèques DML locales), puis auditionné par la commission des National academies « Planning a Global Library of the Mathematical Sciences » en novembre 2012. Il a fait ensuite partie des referees du rapport produit par ce groupe, été invité à la table ronde du congrès international de Séoul sur ce sujet et, en parallèle à ces activités, nommé par Ingrid Daubechies membre du groupe présidé par Patrick Ion travaillant sur l'implémentation concrète des plans précédents. Le travail de ce groupe s'est pour l'instant essentiellement matérialisé par une conférence téléphonique hebdomadaire et une certaine confusion entre définition des objectifs, des besoins, et recherche de moyens.

### *E. Sauvegarde et archivage*

La Cellule Mathdoc assure une sauvegarde régulière de toutes les informations qu'elle manipule comme tout centre de traitement de l'information. Ceci est décrit dans le chapitre « Infrastructures ».

Mathdoc héberge un volume conséquent de données (articles...) d'une valeur scientifique importante et dont il n'existe souvent pas de copie dans une autre structure. En parallèle, des logiciels spécifiques internes ont été développés et des bases de données sont utilisées pour la gestion et l'accès à ces données. Le volume global est d'environ un Tera-octet et les documents ne sont que du texte ou des images, pas de multimédia.

Dans cette situation il y a deux risques majeurs qu'il faut prendre en compte :

- Une destruction des matériels (machines, disques) entraînant aussi la perte des données sur le site Mathdoc à Grenoble. Pour ce risque une solution est d'assurer une copie régulière des données, bases de données et logiciels sur un autre site non grenoblois de manière à pouvoir remettre en place les services assurés avec les données. Cela correspond à un plan minimal de reprise d'activité.
- L'impossibilité de relire d'anciennes données qu'il est nécessaire de conserver et qui ne sont plus utilisées au quotidien mais qui ont une certaine valeur. Il y a ainsi des données historiques sur d'anciens supports physiques qu'on ne sait déjà plus lire. Il faut donc étudier comment assurer un archivage pérenne des données importantes avec les outils pour les traiter.

Trois actions sont menées pour ces 2 objectifs :

- La mise en œuvre d'une sauvegarde à distance sur un site Mathrice non Grenoblois, actuellement à Lille pour être précis. Le système fonctionne avec les données des services majeurs mais il manque encore les outils logiciels et d'autres données. Il reste donc à faire la liste exhaustive de ce qu'il faut sauvegarder et le mettre en œuvre. C'est en cours.
- La participation au groupe de travail BSN6 sur l'archivage pérenne. Ce type de service est assuré principalement par le CINES (PAC) et BNF (SPAR) pour le patrimoine scientifique. Le déploiement de ces solutions a été achevé et plusieurs centres similaires au nôtre les utilisent déjà (Persée, Cleo, CCSD, SICD...). Il y a actuellement un coût élevé et surtout un modèle opératoire qui nous semble inadapté (pas pérenne, justement !) : la structure qui a des données à préserver (pour l'éternité) doit payer une somme non négligeable au CINES pour l'archivage de ses données, chaque année. Que se passe-t-il en cas de coupe budgétaire ? Mathdoc a décidé de participer aux réflexions sur la mise en œuvre en étant éventuellement volontaire pour devenir site pilote. C'est l'objectif de la participation à ce groupe de travail (où l'on apporte aussi des exemples de besoins concrets).
- La participation à la réflexion sur la mise en place d'un système de copies des données documentaires de maths distribué sur plusieurs sites européens (au moins 3). Nous avons proposé ce projet aux partenaires de l'initiative EuDML, sous la forme de la mise en place d'un réseau privé LOCKSS. Un début d'étude documentaire sur LOCKSS est en cours.

## F. Synthèse

### a) Activités

Voici un tableau par site de diffusion de nos activités dans le secteur des bibliothèques numériques.

<b>Projet : Adresse web</b>	<b>Activités réalisées à Mathdoc</b>	<b>Personnes</b>
<b>Numdam</b> www.numdam.org	numérisation, acquisition numérique, indexation, pages de garde, mise en ligne	IC, HF, SMP, EV, CG, TB, FL
<b>Numir</b> www.numir.org	support à la numérisation, développements contrôle qualité, mise en ligne	CBZ, IC, ML, FL
<b>Publications mathématiques d'Orsay</b> : sites.mathdoc.fr/PMO	réindexation, mise en ligne	EC, TB
<b>Gallica-Math : œuvres complètes</b> sites.mathdoc.fr/OEUVRES	réindexation, pages de garde, mise en ligne	IC, EC, TB, CG
<b>Gallica-Math : articles</b> sites.mathdoc.fr/JMPA	réindexation, pages de garde, mise en ligne	IC, EC, TB, CG

<b>Projet : Adresse web</b>	<b>Activités réalisées à Mathdoc</b>	<b>Personnes</b>
<b>Gallica-Math : répertoire</b> sites.mathdoc.fr/RBSM	mise en ligne	CG
<b>Mini-DML</b> : minidml.mathdoc.fr	agrégation, mise en ligne	CG, HF, TB
<b>LiNum</b> : sites.mathdoc.fr/LiNuM	agrégation, mise en ligne	CG
<b>EuDML</b> : eudml.org	agrégation	CG, TB, NH
<b>SemProba</b> sites.mathdoc.fr/SemProba	indexation, mise en ligne	CG
<b>Revue de géologie alpine</b> geologie-alpine.ujf-grenoble.fr	numérisation, indexation, mise en ligne	IC, HF, CBZ, CG
<b>Revue d'écologie alpine</b> : ecologie-alpine.ujf-grenoble.fr	numérisation, indexation, mise en ligne	IC, CG, CBZ
<b>Archives Bourbaki</b> sites.mathdoc.fr/archives-bourbaki	numérisation, indexation, mise en ligne	HF, IC, CG, EC

### *b) Statistiques Numdam*

- *Volumétrie*

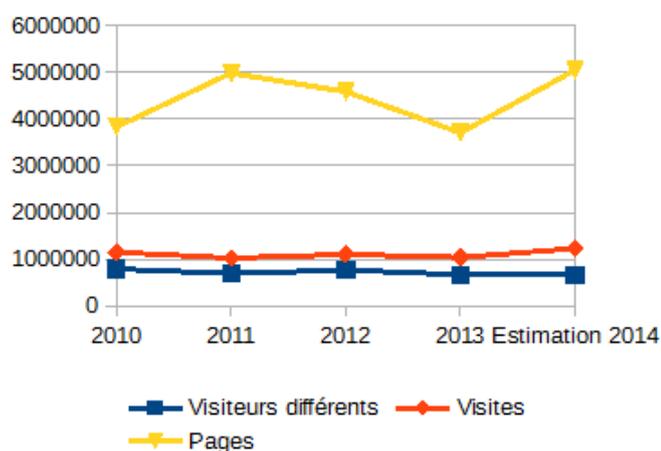
La Bibliothèque Numdam est composée de 29 revues, 35 séminaires et 3 mémoires, soit un total de 67 séries s'étalant de 1810 à 2013. Les acquisitions numériques représentent environ 15 % du volume total.

<b>Volumes</b>	<b>Articles</b>	<b>Pages</b>	<b>Art IDs MR, ZM, JFM</b>	<b>Biblios</b>	<b>Bibitem</b>	<b>Bib IDs MR, ZM, JFM</b>	<b>Liens directs Numdam, DML</b>
5 509	54 416	1 054 130	25 557, 28 036, 5 894	30 067	503 452	321 010, 367 776, 22 270,	37 683, 33 003

- *Consultations*

Le site numdam.org a reçu en moyenne 1 112 549 visites par 720 188 visiteurs différents et 4 431 140 pages visités sur la période 2010-2014.

<b>Année</b>	<b>Visiteurs différents</b>	<b>Visites</b>	<b>Pages</b>
2010	795 333	1 152 760	3 834 012
2011	696 228	1 028 682	4 981 416
2012	773 578	1 106 306	4 585 587
2013	670 591	1 040 890	3 706 861
Estimation 2014	665 208	1 234 108	5 047 826



- *Téléchargements*

Le nombre de téléchargements sur le site numdam.org est en nette augmentation. Il a été sur la période de 2010 à 2014 en moyenne de 1 980 915 PDF et 267 436 DJVU.

	<b>PDF</b>	<b>DJVU</b>
<b>2010</b>	1 793 633	287 729
<b>2011</b>	1 885 488	236 546
<b>2012</b>	1 881 447	318 288
<b>2013</b>	1 673 346	150 333
<b>Estimation 2014</b>	2 670 663	344 284
<b>Moyenne</b>	1 980 915	267 436

## II. ÉDITION

### A. Présentation

Le support à l'édition est à Mathdoc réalisé sous la bannière Cedram, conçue comme un ensemble d'outils qui permettent la fabrication, la diffusion électronique et tout ce qu'il faut pour préparer la diffusion papier de revues académiques de recherche en mathématiques (ainsi que des actes et séminaires).

Cedram (centre de diffusion de revues académiques mathématiques) lancé en 2005 est assumé entièrement par Mathdoc et supervisé par un comité de pilotage<sup>14</sup>. Les revues, en accès libre ou non, sont accessibles via un serveur spécifique à chaque revue et via un portail général [www.cedram.org](http://www.cedram.org).

Le service vise la production et la diffusion aux meilleures normes d'édition électronique (pas seulement sur le plan technique).

#### a) *Éléments principaux du service de fabrication et diffusion*

- Production de fascicules papier et électroniques en pdfLaTeX ;
- extraction automatique et standardisation XML/MathML des métadonnées issues des articles LaTeX ;
- diffusion numérique et référencement des articles sur le site de la revue ;
- référencement des articles dans le portail commun Cedram ;
- référencement des articles dans un serveur OAI ;
- affectation de DOI ;
- intégration des collections rétrospectives complètes depuis Numdam ;
- versement des articles récents dans Numdam à fin d'archivage et d'accès sur le long terme ;
- statistiques d'accès.

#### b) *Autres services*

Le Cedram a aidé les revues et les autres séries qui l'ont rejoint à définir leur maquette (web et papier).

Sur la période concernée, le principal service au-delà de la fabrication et de la diffusion concerne le logiciel RUCHE, qui avait précédemment été adopté et mis en Open source dans le but de le mutualiser entre les revues, et de le faire évoluer vers un outil de gestion des flux éditoriaux mieux intégré avec les autres pièces du puzzle.

Cette activité n'a pas été couronnée de succès :

- si quelques revues (pas que de maths) ont demandé à évaluer le logiciel, aucune ne se l'est approprié ;

---

<sup>14</sup> Présidé par Claude Sabbah, il est constitué de représentants des tutelles, de la SMAI et de la SMF, et de chaque revue.

- les *Annales de l'institut Fourier* ont utilisé le logiciel de façon régulière, mais pas forcément de façon satisfaisante (des erreurs de jeunesse du logiciel nous ont été signalées très tard) ;
- un stagiaire a été encadré pour partie aux AIF et par Mathdoc en 2012 pour corriger un certain nombre de ces erreurs ;
- le soutien technique aux revues (en fait, aux seules AIF) et la maintenance ont été de qualité médiocre par insuffisance de moyens humains.

Au final, il a été décidé d'arrêter le support pour ce logiciel après le départ de Catherine Zoppis, dernière personne à Mathdoc le connaissant un peu, et d'accompagner le mouvement vers OJS.

### *c) Présentation des revues*

Chaque revue est présentée avec une charte graphique spécifique, élaborée avec le responsable de la revue, sur un serveur dédié xxx.cedram.org (xxx étant l'acronyme de la revue).

### *d) Accès aux revues*

La participation au Cedram impose la définition d'une barrière mobile. Une revue peut choisir d'être :

- totalement en accès libre ;
- en accès sur abonnement pour les numéros récents (généralement < 2 ans ou < 5 ans) et en accès libre pour les numéros plus anciens. La technique utilisée pour le contrôle d'accès est le filtrage par adresse IP.

### *e) Identification des articles*

Cedram, adhérent à Crossref, est habilité à délivrer des identifiants DOI pour les articles qu'il publie. Exemple d'identifiant : doi:10.5802/aif.1458.

Cedram utilise aussi une identification interne pour chaque article, comme AIF\_1995\_\_45\_2\_375\_0, dont découlent les URL pérennes de la forme [http://www.numdam.org/item?id=AIF\\_1995\\_\\_45\\_2\\_375\\_0](http://www.numdam.org/item?id=AIF_1995__45_2_375_0).

Il dispose d'un serveur OAI-PMH <http://www.cedram.org/oai> qui permet un partage de ses métadonnées et une meilleure visibilité de ses collections, interrogeable avec le protocole OAI-PMH.

### *f) Visibilité des articles*

L'accès aux articles peut ainsi se faire via le site de la revue, xxx.cedram.org, via le portail [www.cedram.org](http://www.cedram.org) et via tout service interrogeant le serveur OAI-PMH de Cedram comme EuDML.

### *g) Métadonnées*

Chaque article est présenté avec : titre, auteurs, résumé, lien vers le texte en PDF, bibliographie, revue, numéro de revue, pages, identifiant DOI, lien vers la notice zbMATH, lien vers la notice MathSciNet.

Lorsque cela est possible de façon automatique, chaque référence bibliographique est enrichie avec les liens vers les notices zbMATH et MathSciNet, JFM (Jahrbuch), Numdam, article (Cedram, DOI, mini-DML).

Ces notices peuvent être affichées avec les formules mathématiques en TeX ou en MathML (option utilisateur persistant dans un cookie, défaut en fonction du navigateur).

### *h) Moteur de recherche*

Sur les serveurs des revues ou sur le portail Cedram, la recherche peut s'effectuer en plein texte et sur titre, auteur, date, bibliographie. Un feuilletage des revues est aussi possible.

### *i) Les publications actuellement diffusées*

- *Revue*
  - Annales de la faculté des sciences de Toulouse (AFST)
  - Annales de l'institut Fourier (AIF)
  - Annales mathématiques Blaise Pascal (AMBP)
  - Confluentes Mathematici (CML)
  - Journal de l'École polytechnique — Mathématiques (JEP)
  - Journal de théorie des nombres de Bordeaux (JTNB)
  - MathS In Action (MSIA)
  
- *Livre*
  - Le séminaire de mathématiques, par Michèle Audin
  
- *Actes et séminaires*
  - Actes des rencontres du CIRM (ACIRM)
  - Les cours du CIRM (CCIRM)
  - Journées Équations aux dérivées partielles (JEDP)
  - Séminaire Équations aux dérivées partielles (SEDP, mort et remplacé par le suivant)
  - Séminaire Laurent Schwartz — EDP et applications (SLS-EDP)
  - Séminaire de Théorie spectrale et géométrie (Grenoble) (TSG)

Au total, environ 7650 articles pour un total de 172000 pages sont diffusés sur le cedram.org.

## *B. Activités*

L'activité du Cedram se décompose en trois volets principaux :

1. développement de divers composants de la chaîne d'édition, de la sélection des articles à la diffusion en passant par la fabrication ;
2. support aux revues ;
3. production et publication au fil de l'eau.

Dans la première catégorie, on trouve évidemment la plateforme de diffusion, qui est en cours de refonte et dont la mise en production devrait suivre celle de Numdam, mais aussi plus spécifiquement les outils de fabrication (classe LaTeX cedram et ses déclinaisons pour les diverses séries, Tralics et les transformations XML vers les formats utilisés pour la diffusion). Le support, l'installation et la documentation de RUCHE a également mobilisé des moyens sur la période, même s'il n'y a pas eu de développements à proprement parler, l'activité s'étant consacrée

principalement à des correctifs. Mathdoc consacre aussi une certaine énergie à la mise en place et à la maintenance journalière des sites web des revues, dont une grande partie est en Spip.

La seconde catégorie est assez variée et couvre un spectre d'activités assez large, allant d'une veille sur les nouveautés du monde LaTeX à la fourniture d'outils comme RUCHE, voire à la réponse à pas mal de questions au jour le jour, ceci pouvant être en rapport avec une documentation parfois trop elliptique, pour ne pas dire inexistante, de certains de nos outils.

La troisième catégorie passe un peu inaperçue : c'est le labeur quotidien qui consiste à accompagner un fascicule finalisé par une revue vers sa mise en ligne sur le site de diffusion. Ce chemin est à la fois très largement automatisé et plein d'embûches. Ceci peut donc nécessiter une technicité non bornée *a priori*, et des interventions manuelles à chaque étape. Depuis le départ en congé longue durée d'Isabelle Costerg, cette tâche est réalisée sur la base du bon vouloir par deux personnes, dont le directeur de l'unité, pour ne pas pénaliser les revues.

### *a) Chaîne de production*

Le but recherché par le Cedram est de permettre aux revues académiques de s'approprier leur production en mettant à leur disposition des outils performants et facile d'accès. C'est une des raisons pour lesquelles la chaîne de production est entièrement basée sur (et largement écrite en) LaTeX : ce qui est validé par les auteurs, ce sont les sources LaTeX et le PDF qui en résulte. Par conséquent c'est ce qui fait foi et est donc à la source des tous les traitements subséquents.

Le fonctionnement actuel du Cedram suppose que les revues disposent de la structure éditoriale leur permettant de gérer le processus allant de la soumission ou sollicitation d'un article à l'établissement de la version définitive après évaluation par les pairs, relectures, corrections et mise en page au sein du fascicule. À partir de là, Mathdoc (cf. figure 3)

- réceptionne tous les éléments nécessaires à la fabrication du fascicule ;
- lance des scripts et programmes qui produisent simultanément le fichier XML de métadonnées et tous les PDF nécessaires à l'impression comme à l'édition électronique ;
- met en ligne le résultat de ces traitements sur un site de test ayant fonction d'épreuve électronique, à partir duquel la revue décide de faire des corrections, demande des modifications, ou valide pour mise en ligne sur le site public ;
- bascule les fascicules validés sur le site de la revue ;
- retourne à la revue les éléments de production définitifs, de telle sorte qu'elle puisse les archiver, faire imprimer le cas échéant, etc.

L'hypothèse ci-dessus n'étant finalement que rarement vérifiée, Mathdoc a pris en charge au fil du temps de nombreux aspects qui ne sont pas de son ressort (à l'exclusion de décisions scientifiques). Une solution est d'étendre la palette des services proposés par Mathdoc (voir le projet Scientific Edition ci-dessous) pour répondre à tous les cas mais il faudra mettre des moyens humains pour que cela puisse fonctionner !

La classe Cedram a vu son nombre d'options augmenter assez rapidement pour faire face à tous les cas de figure. Il existe aujourd'hui deux modes opératoires radicalement différents :

- un mode natif dans lequel les articles sont composés avec la classe Cedram (c'est-à-dire que les textes des auteurs sont mis aux normes en utilisant la déclinaison de la classe Cedram qui implémente la maquette de la revue, de ce fait la qualité est optimale) ;

- un mode de compatibilité qui permet d'utiliser un PDF déjà composé avec une autre classe (voire avec autre chose que LaTeX) et d'insérer cependant l'article dans le système de production standard du Cedram en préparant un fichier LaTeX paramétrant les métadonnées.

Les *Annales de l'institut Fourier* utilisent des sous-traitants pour la mise aux normes des articles, si bien qu'il y a une qualité de la préparation des fascicules relativement homogène. L'intervention de Mathdoc consiste à régler finement les métadonnées et corriger des erreurs qui peuvent avoir échappé aux étapes précédentes car le XML n'y est pas produit. Quoiqu'organisées un peu différemment, la situation est similaire pour AMBP, JTNB, JEP, CML, TSG, qui sont toutes produites en mode natif.

MSIA et plusieurs séries d'actes sont également produites en mode natif, la plupart du temps par l'équipe de Mathdoc car il n'y a pas de personnel pour prendre en charge la mise aux normes du Cedram dans ces publications. C'est un point préoccupant, surtout dans les circonstances actuelles d'absence de personnel à Mathdoc dédié à ces tâches.

Deux séries sont produites en mode de compatibilité : le séminaire SLS-EDP par décision de son responsable qui ne souhaite pas passer du temps à reprendre les textes des orateurs, et les *Annales de la faculté des sciences de Toulouse* qui font imprimer leurs fascicules avant de nous les envoyer, et pour lesquelles la mise en ligne sur le site du Cedram s'apparente à une acquisition dans Numdam.

De fait, la chaîne d'intégration au Cedram en mode de compatibilité est assez largement utilisée pour les acquisitions de Numdam quand nous pouvons disposer de sources LaTeX de façon à générer des métadonnées XML de qualité.

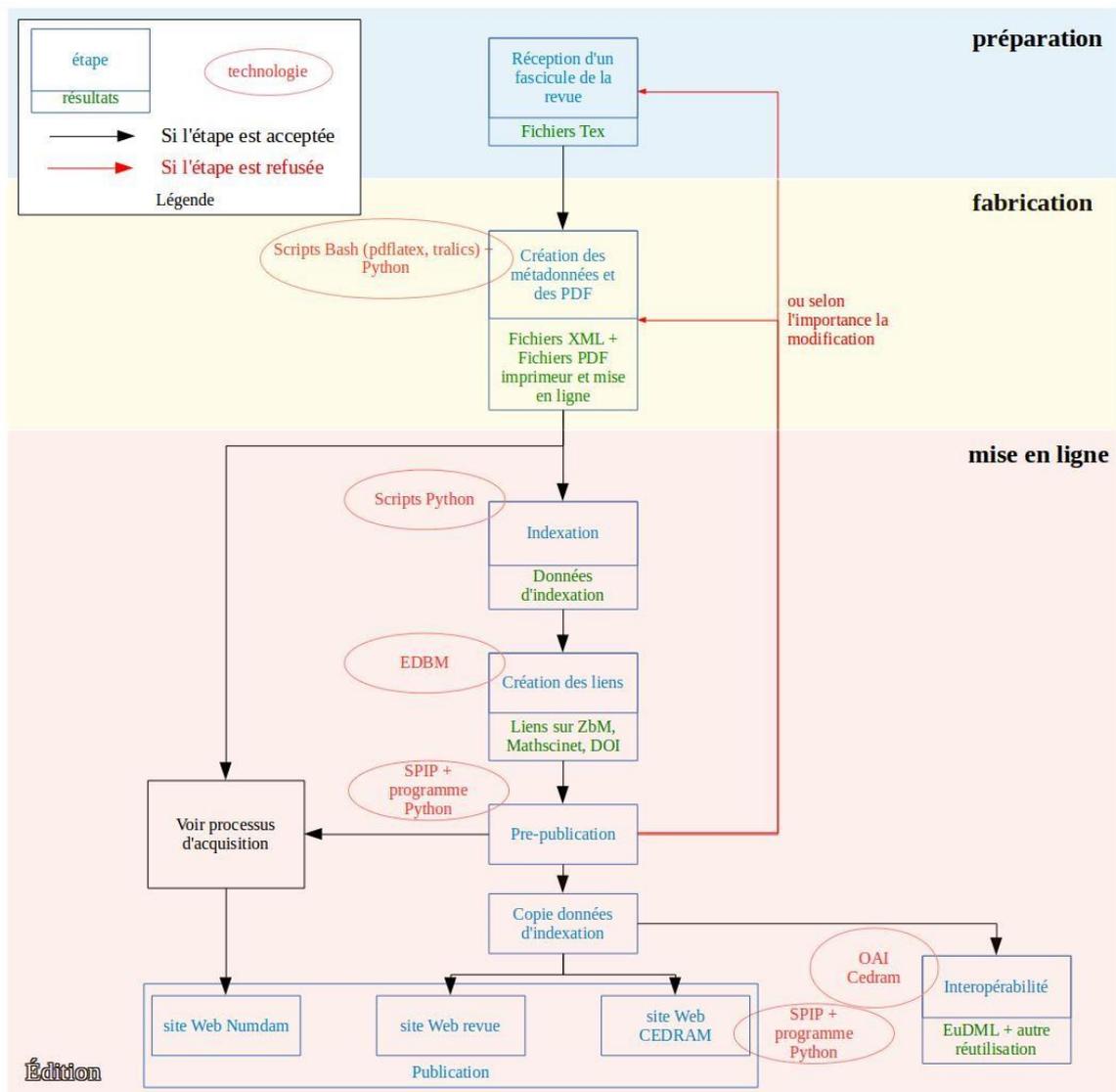
Une part importante de la production actuelle du Cedram est représentée par les AIF (80 articles et 2500 à 3000 pages par an sur les 200 articles et 6000 pages annuelles). Cependant le volume de travail dépend beaucoup de l'organisation de chaque publication, de l'existence ou non d'une personne référente pour les différentes tâches à réaliser en amont ou en aval de notre chaîne de fabrication et diffusion.

### *b) Crossref*

CrossRef est le service d'une association d'éditeurs (Publishers International Linking Association, Inc.) qui développe une infrastructure permettant d'établir des références croisées entre publications. La Cellule Mathdoc est devenue membre de l'association CrossRef en 2011, et est donc habilitée à attribuer des DOI (Digital Object Identifier) aux articles scientifiques qu'elle publie, et à les déclarer dans le système CrossRef. CrossRef fournit également un service permettant d'obtenir un DOI à partir d'une référence bibliographique (par *matching*).

Un système simple d'attribution et d'enregistrement des DOI a été développé en 2011, ainsi qu'un système d'interrogation de la base CrossRef.

Mis en place en une quinzaine de jours, ces systèmes sont peu ou mal intégrés au flux de production Cedram existant. Une des tâches à prévoir dans les années qui viennent pourrait être une amélioration du flux de production intégrant dans sa conception la gestion des DOI.



Notons également que tous les services CrossRef fonctionnent de manière asynchrone, les notifications étant envoyées par courrier électronique. Ceci rend délicate une automatisation plus poussée de leur utilisation.

**Figure 3 Étapes de production du Cedram**

### c) Versement dans Numdam

À l'issue d'un délai d'environ un an après leur publication sur un des sites hébergés par le Cedram, les volumes sont versés dans la bibliothèque numérique Numdam. Pour l'essentiel, les XML Cedram sont convertis (simplifiés) à la DTD Volphys de Numdam. Le seul domaine dans lequel il arrive que les données Cedram soient moins précises que celles de Numdam est le balisage des bibliographies, car les bibliographies répondent à des logiques différentes dans les deux services : dans Numdam, ce sont des chaînes de caractères surbalisées *a posteriori*, tandis que dans Cedram, ce sont soit des données provenant de Bibtex demandant un traitement pour être lues, soit des chaînes de caractères stylées mais non balisées.

Comme pour les autres acquisitions Numdam, les PDF sont versés tel quel.

#### *d) Lancement de séries*

Depuis le démarrage du Cedram les séries diffusées étaient restées relativement stables (une seule création : *Math in Action*, qui n'a jamais atteint un flux d'articles considérable).

En 2011, le « séminaire X-EDP » s'est éteint après quarante ans de service, il a laissé la place au « séminaire Laurent Schwartz — EDP et applications » réalisé en partenariat par Polytechnique et l'IHES, et publié par l'X. Le lancement de cette série s'est fait sans difficulté particulière. C'est la première série utilisant explicitement une licence Creative Commons. Une page de garde a été dessinée, ainsi qu'une vignette de couverture pour l'identifier sur le site web.

En 2012, les travaux ont commencé pour accueillir la revue *Confluentes Mathematici*, précédemment publiée par World Scientific et désormais en accès libre au Cedram, l'éditeur étant l'institut Camille Jordan (Lyon). La conception du site web en Spip a été comme d'habitude un peu compliquée mais largement prise en main par le laboratoire lyonnais, lequel a aussi maîtrisé assez rapidement le système LaTeX.

En 2014, le Cedram a participé au lancement du *Journal de l'école polytechnique*, renaissant dans une nouvelle série consacrée aux mathématiques. Création d'une maquette entièrement nouvelle, devant tenir compte des souhaits des graphistes de l'X et très étroitement pilotée par Claude Sabbah. Quatorze articles sont parus en 2014. Une particularité de cette revue est que les sources TeX des articles sont diffusés à côté des PDF.

Enfin, fin 2014, nous avons publié un ouvrage de Michèle Audin : livre historique principalement consacré au séminaire Julia, dont elle a ressaisi et annoté l'intégralité des textes (dont la numérisation est par ailleurs programmée dans Numdam 4). Cet ouvrage d'un millier de pages à la structure complexe a été entièrement reconçu en vue d'une édition purement électronique avec possibilité d'impression à la demande séparée de la partie historique, voire des tomes annuels. Cet ouvrage était en principe programmé dans une collection historique d'un éditeur académique qui s'est retiré du projet.

En 2014, Mathdoc a aussi participé aux travaux préliminaires de lancement de la revue *SMAI-Journal of Computational Mathematics* (maquette LaTeX, conseil, adhésion à Crosscheck...) et donné son accord pour l'accueil d'une nouvelle série d'actes : Winterbraid Lecture Notes (WBLN).

#### *e) Projet Scientific Edition*

Un grand nombre de réunions ont eu lieu au cours de la période entre les revues du Cedram et avec d'autres partenaires dans le but de faire avancer la mutualisation (bouquet commun, avec la SMF, outils communs partagés par des revues sur différentes plateformes...). Pour l'instant, le seul résultat tangible est constitué par les décisions prises lors de la journée sur les publications organisée à Grenoble par Mathdoc le 24 novembre 2014 : formalisation d'un projet Scientific Edition alliant le Cedram, Epi-math, le volet mathématique d'Epi-Sciences, et mathjournal, un projet d'évolution d'OJS pour mieux servir les revues de math. Avec notamment pour objectif d'accompagner les revues académiques vers un libre accès « diamant » qui suppose un soutien fort et déterminé d'institutions assurant la gratuité de la publication et de l'accès. Mathdoc a commencé à se former sur OJS, organise d'ailleurs une formation à Grenoble le 6 janvier 2015, et a décidé d'accompagner les AIF dans leur migration à partir de RUCHE. Les autres aspects du projet restent largement spéculatifs car dépendant de financement ou de soutien non débloqués à l'heure de boucler ce rapport.

### III. SERVICES DE DOCUMENTATION

Seule ou en coopération avec le RNBM, la Cellule Mathdoc a créé et entretient plusieurs services de documentation, à destination des professionnels de la documentation ou de la communauté mathématique.

#### A. Description des activités

- *Développement de nouveaux systèmes d'informations*

Plusieurs systèmes d'informations ont été conçus ou intégralement refaits pendant la période 2010-2014: le catalogue fusionné des périodiques mathématiques (CFP), le portail math, les services web des bibliothèques et des laboratoires.

- *Maintenance et évolution des systèmes d'information existants*

Afin de conserver un haut niveau de service envers la communauté mathématique, la plupart des systèmes d'information conçus avant 2010 ont été maintenus. Les données ont été régulièrement mises à jour, des anomalies ont été corrigées, les serveurs et sites web ont été mis à niveau pour garantir la sécurité des données.

- *Catalogage des périodiques*

Le cœur du CFP consiste en une liste de périodiques. L'activité de catalogage est donc primordiale pour ce système et comprend la récolte de métadonnées, le nettoyage des données, le formatage pour assurer la cohérence.

- *Gestion de projets*

La Cellule Mathdoc intervient en tant que coordinateur technique sur des projets nationaux tel que le portail math. Elle assure donc notamment les activités de planning, gestion des risques, coordination des tâches, suivi de la performance et de la qualité.

- *Archivage des données*

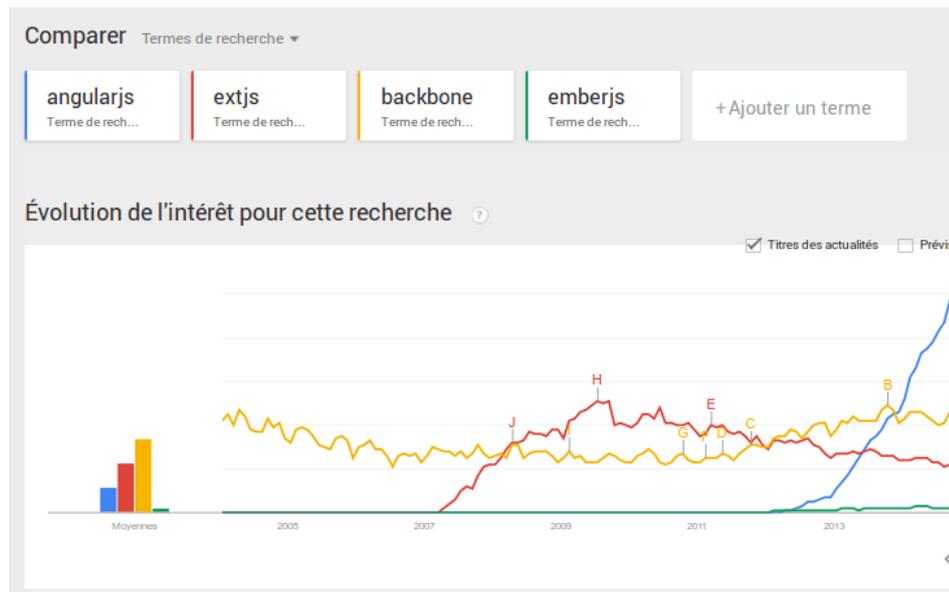
L'archivage des données est nécessaire pour assurer des services pérennes à la communauté mathématique, comme les annuaires des laboratoires et des bibliothèques

- *Coopération RNBM, Mathrice, Inist*

La Cellule Mathdoc participe depuis toujours à plusieurs groupes de travail du RNBM et de Mathrice : groupe PEB sur le prêt entre bibliothèques et groupe PCMaths sur le plan de conservation partagée des bibliothèques de mathématiques, et bien entendu le groupe CFP et le groupe de travail du portail en lien avec le groupe PLM. Une tentative de coopérer plus étroitement avec l'Inist notamment autour des projets ISTEEX et portail math s'est concrétisée par une réunion en 2012 quand les projets n'étaient pas mûrs, et une réunion en décembre 2014 pour étudier les possibilités d'échanges de données et d'expériences sur l'exploitation des ressources ISTEEX sur le portail math.

- *Veille technologique*

La Cellule Mathdoc apporte son expertise technique sur différents projets et veille à étudier régulièrement de nouvelles technologies. AngularJs a ainsi permis la fusion des sites plm.math.cnrs.fr et portail-math.fr ; Swagger la description des services web de api.mathdoc.fr



**Figure 4 Popularité d'AngularJs**

## API.MATHDOC.FR

Description des services web de Mathdoc

### Biblio

Show/Hide | List Operations | Expand Operations

GET	/bibliotheques	Liste des bibliothèques de maths
GET	/bibliotheques/{bibid}	Information d'une bibliothèque
POST	/bibliotheques/biblio	Ajoute une bibliothèque
PUT	/bibliotheques/biblio/{bibid}	Modifie une bibliothèque
DELETE	/bibliotheques/biblio/{bibid}	Supprime une bibliothèque

### Calendar

Show/Hide | List Operations | Expand Operations

GET	/calendars	Liste des séminaires
-----	------------	----------------------

### Event

Show/Hide | List Operations | Expand Operations

GET	/events	Liste des événements
-----	---------	----------------------

### Labo

Show/Hide | List Operations | Expand Operations

GET	/labo	Liste des laboratoires
GET	/labo/{labid}	Information d'un laboratoire

**Figure 5 Documentation des services web sur api.mathdoc.fr avec swagger**

- *Formation*

L'activité de formation découle naturellement de la veille technologique. La Cellule Mathdoc a ainsi participé aux journées Mathrice, et à plusieurs ANF RNBM, Mathrice ou autres réseaux métiers du CNRS pour contribuer à la diffusion des connaissances.

## B. Description des services de documentation

### a) Services à destination du RNBM

#### 1) CFP

Le catalogue fusionné des périodiques contient environ 19 000 périodiques de mathématiques mortes et vivantes, y compris des revues électroniques en accès libre. Pour chaque titre, les informations suivantes sont mentionnées : une notice du périodique provenant de la base de données de l'ISSN (ou d'un catalogage direct dans le cas de journaux très récents ou purement électroniques) ; les localisations de ce périodique dans les différentes bibliothèques du RNBM, et les états de collections précis. Ces données sont obtenues par le retraitement informatique de différents fichiers envoyés par les bibliothèques participantes (toutes les bibliothèques de mathématiques importantes y participent) ; différents URL associés aux titres : site de l'éditeur pour les articles récents, site pour les archives le cas échéant. Des listes de tous les périodiques en ligne sont extraites dynamiquement de la base.

Sur la période 2010-2014, le site web a été intégralement refait (Figures 6 et 7).

**Catalogue fusionné des périodiques de Mathématiques**

Cette base, constituée par la Cellule MathDoc, recense les journaux de mathématiques (y compris des journaux électroniques d'accès libre), obtenus par le retraitement informatique des différents fichiers fournis par les [bibliothèques participantes](#), membres du [RNBM](#).  
Si vous remarquez des problèmes, envoyez nous un courriel (lien "contactez nous" ci dessus)

[Interrogation du catalogue \(périodiques papier et électroniques\)](#) | [Aide à la consultation](#)

Titre contient

ISSN :  Bibliothèque :   
Angers  
Besançon

Si vous n'avez pas choisi de bibliothèque en particulier, vous pouvez afficher:

Tous les journaux de la base répondant à votre question.  
 Seuls les journaux ayant une localisation dans une bibliothèque

[Périodiques en ligne uniquement :](#)

[A](#)/[B](#)/[C](#)/[D](#)/[E](#)/[F](#)/[G](#)/[H](#)/[I](#)/[J](#)/[K](#)/[L](#)/[M](#)/[N](#)/[O](#)/[P](#)/[Q](#)/[R](#)/[S](#)/[T](#)/[U](#)/[V](#)/[W](#)/[Z](#)

Ou mot du titre :

Figure 6 Le CFP en 2009

La structure de la base a été profondément modifiée et un accès en écriture pour les bibliothécaires du réseau a été mis en place. Une interface a été développée, permettant aux membres du RNBM de mieux gérer les informations sur leurs collections, en pouvant renseigner par exemple l'état d'une collection. Des outils internes ont été conçus pour permettre l'import automatique des données, comme l'import d'identifiants au format PPN présents dans le SUDOC.

La base de données a évolué pour supporter les nouvelles fonctionnalités et est régulièrement tenue à jour grâce au contact avec les bibliothèques.

Nom d'utilisateur :   
 Mot de passe :

## Catalogue fusionné des périodiques de mathématiques

RECHERCHE BIBLIOTHÈQUES PORTAIL DOCUMENTAIRE MATHÉMATIQUE CELLULE MATHDOC

### Annals of mathematics [ISSN 0003-486X] [PPN 038665220]

#### Notice :

Identifiant interne : 4381  
 Identifiant ISSN : 0003-486X  
 Identifiant PPN : 038665220  
 Première Année Publication : 1884  
 Dernière Année Publication : Vivant  
 Langue Publication : anglais  
 Pays Publication : United States  
 Fréquence Publication : Bimestriel  
 Editeur Commercial : [Princeton, N.J., etc., Princeton University Press, etc.]  
 Titre Abrégé : Ann. math.  
 Source de la notice : ISSN  
 Dernière mise à jour de la notice : 31/01/2014 10:00:46  
 Format Publication : Imprimé

Voir aussi ce journal dans le [SUDOC](#)

#### Liens :

##### Suite de

- [Analyst \(Des Moines, Iowa\) \(The\)](#) [ISSN 0741-7918] [PPN 093456662]

##### A pour autre édition sur un support différent

- [Annals of mathematics \(Online\)](#) [ISSN 1939-8980] [PPN 11836264X]

#### Collections en bibliothèque :

Bibliothèque	Pôle	Etat de Collection	Lacunes	Dernière mise à jour	Détails
<a href="#">Angers</a>		87 (1968) ->	Aucune		> <a href="#">Détails</a>
<a href="#">BNF</a>		Vol. 133, nr. 1 (1991, Jan.) ->	Aucune		> <a href="#">Détails</a>
<a href="#">Besançon</a>		Vol. 34 n° 2 (1933) ; Vol. 37 n° 3 (1936) ; Vol. 38 (1937) ->	Vol. 59 (1954) ; Vol. 123 n° 1 (1986) ; Vol. 131 n° 2 (1990)	15/01/2014 11:24:33	> <a href="#">Détails</a>
<a href="#">Bordeaux</a>		Vol. 1 (1884) - Vol. 12 (1899) ; Série 2 Vol. 1 (1899) ->	Vol. 165 n° 3 (2007) - Vol. 166 n° 3 (2007) ; Vol. 174 n° 1 (2011)	30/04/2013 17:14:24	> <a href="#">Détails</a>
<a href="#">Cergy</a>		vol. 129 no. 1 (1989) ; vol. 135 no. 2 (1992) ; vol. 147 no. 1 (1998) -....	Aucune		> <a href="#">Détails</a>
<a href="#">Clermont</a>	oui (associé)	Vol. 1 (1900) ->	Aucune	21/11/2014 09:52:27	> <a href="#">Détails</a>
<a href="#">Dijon</a>		(1952) - Vol. 103 n° 2 (1976) ; Vol. 104 n° 3 (1976) ->	Vol. 165 n° 1 (2007) ; Vol. 166 n° 1 (2007)	11/10/2012 16:06:57	> <a href="#">Détails</a>
<a href="#">Grenoble [IF]</a>		Vol. 1 (1899) ->	Aucune	19/06/2012 16:35:14	> <a href="#">Détails</a>
<a href="#">INIST</a>		34,1933-140,1994(2);141,1995	Aucune		> <a href="#">Détails</a>

Figure 7 Le CFP en 2014

## 2) PCMaths

La Cellule Mathdoc participe avec le RNBM au développement le plan national de conservation partagée des périodiques de mathématiques : elle est membre du comité de pilotage et a adapté le CFP pour en faire un outil de gestion de ce plan. Par exemple, les bibliothèques peuvent désormais se déclarer candidates pour être pôle de conservation d'un périodique. L'outil permet aussi au comité de pilotage du PCMath de valider les choix des différents pôles.

## *b) Services à destination de la communauté mathématique*

### *1) Le portail math*

Avant juillet 2012 il y a eu plusieurs essais de création d'un portail pour les maths ainsi que deux propositions d'EQUIPEX 2010 (PURMATH) et 2011 (PRIAM) non retenues dont un volet portait sur ce point. À Mathdoc un portail documentaire avait été développé par Elizabeth Cherhal mais n'avait plus évolué depuis plusieurs années. Malgré l'échec de la recherche de financement, la direction de l'INSMI a décidé à l'été 2012 de lancer avec les moyens disponibles « un projet de création d'un portail mathématique national permettant l'accès aux ressources documentaires et aux services utiles aux chercheurs du domaine en associant tous les acteurs concernés ». Thierry Bouche a rédigé une note de cadrage et une réunion de lancement s'est tenue début septembre 2012. Thierry Bouche a été désigné responsable du projet, Jean-Luc Archimbaud coordinateur technique et 4 partenaires ont été associés : INSMI, Mathrice, Mathdoc, RNBM. La demande de l'INSMI était d'ouvrir ce portail le plus rapidement possible.

La première phase du projet visait à décrire l'existant afin de servir d'inventaire des ressources documentaires à rendre plus facilement accessibles. Le premier objectif était de préparer les phases suivantes et de remplacer le portail documentaire de Mathdoc par un service beaucoup plus général. La seconde phase devait utiliser les résultats de la première pour avancer sur l'accès aux ressources électroniques avec comme premier objectif de donner un accès aux collections disponibles à travers la PLM. Les phases suivantes étant orientées vers le déploiement à plus grande échelle des différentes fonctions réparties entre répertoire d'informations, accès personnalisé à la documentation et aux services numériques pour les mathématiciens travaillant en France.

- *Projet portail math V1 (juillet 2012-avril 2014)*

Le portail math v1, ouvert en janvier 2014 (<http://www.portail-math.fr/>, cf. figure 8), contient un ensemble de fiches descriptives couvrant les ressources documentaires et les moyens d'accès à ces ressources (licence, point d'accès). S'y ajoutent des fiches décrivant les services numériques soit développés par la communauté, soit disponibles sur internet et, pour ce qui ne rentrait pas dans les catégories précédentes mais semblait devoir être porté à l'attention des utilisateurs, une liste de sites utiles.

Ce portail permet plusieurs types de recherches : en texte intégral, avec des facettes, par mots-clés. Par exemple, une recherche sur un éditeur ou un titre de revue combinée avec la facette « libre accès » permet d'obtenir une liste de point d'accès correspondants. Un glossaire et des annuaires sont également disponibles sur le portail.

La Cellule Mathdoc a contribué activement au projet : mise en place de l'organisation, création d'un groupe d'utilisateurs référents, « démarchage » et animation du groupe des concepteurs (informaticiens, mathématiciens, documentalistes), définition d'un processus de publication et du format des fiches descriptives, développement du site web, pilotage de prestataires pour le graphisme du site.

L'ouverture de ce site a permis de fermer l'ancien portail documentaire. Tous les services recensés dans l'ancien portail sont recensés dans celui-ci.

**Ressources documentaires**

- Ensembles de documents
- Bases bibliographiques
- Documents utiles

**Services numériques**

**Sites utiles**

**Annuaire**

- Personnes 
- Laboratoires 
- Bibliothèques 
- CNRS 
- Universités 

**Recherche**

 **NUMDAM** 

BASE : Bielefeld Academic Search  EUDML  Mini-DML  NUMDAM 

Bibliothèque numérique hébergeant et diffusant principalement des périodiques mathématiques publiés en France [En savoir plus](#)

« Annales de Gergonne Annales de l'IHP (A) Physique théorique Annales de l'IHP (B) ... Probabilités et Statistiques Annales de l'IHP (C) Analyse non linéaire Annales de ... Institut Fourier Annales de l'Institut Henri Poincaré Annales de l'université de ... »

---

 **Euclid Direct- archives en accès libre** 

Les archives d'une partie des revues indépendantes diffusées en ligne par le Project Euclid [En savoir plus](#)

accès Reviews Licence libre accès Accessible à tous Avec barrière mobile **Annales ...** de l'Institut Henri Poincaré, Probabilités et Statistiques The **Annals ...** of Applied Probability The **Annals** of Applied Statistics The **Annals** of ...

---

 **De Gruyter Mathematics Journals- en partie en libre accès** 

Revue publiée en ligne par l'éditeur De Gruyter dans le domaine des mathématiques et leurs interactions. [En savoir plus](#)

Analysis and Geometry in Metric Spaces **Annales** UMCS, Mathematica **Annals** of the ... Alexandru Ioan Cuza University- Mathematics **Annals** of West University of Timisoara» ...

---

 **Elsevier Free Archives** 

Archives des 53 revues en mathématiques publiées par Elsevier [En savoir plus](#)

anciennes), Dans le cas des **Annales** de l'IHP (série C), la collection commence ... in Mathematics **Annales** de l'Institut Henri Poincaré (C) **Annals** of ... Mathematical Logic **Annals** of Pure and Applied Logic Applied and Computational Harmonic ...

---

 **CEDRAM partie Libre accès** 

Cedram : centre de diffusion de revues académiques de mathématiques - revues en libre accès [En savoir plus](#)

mathématiques mise en ligne par Mathdoc. Les revues en libre accès sont: **Annales ...** congrès Autres autres = cours Les revues : **Annales** mathématiques Blaise Pascal ...

---

 **Publications de la SMF en accès libre** 

NUMDAM  Publications de la SMF 

Publications de la Société Mathématique de France [En savoir plus](#)

+ accès libre) Descriptif sur ce portail- Accès Les article s des «**Annales** de ... **Annales** de l'ENS:  **Euclid Open Access** 

**Affiner votre recherche**

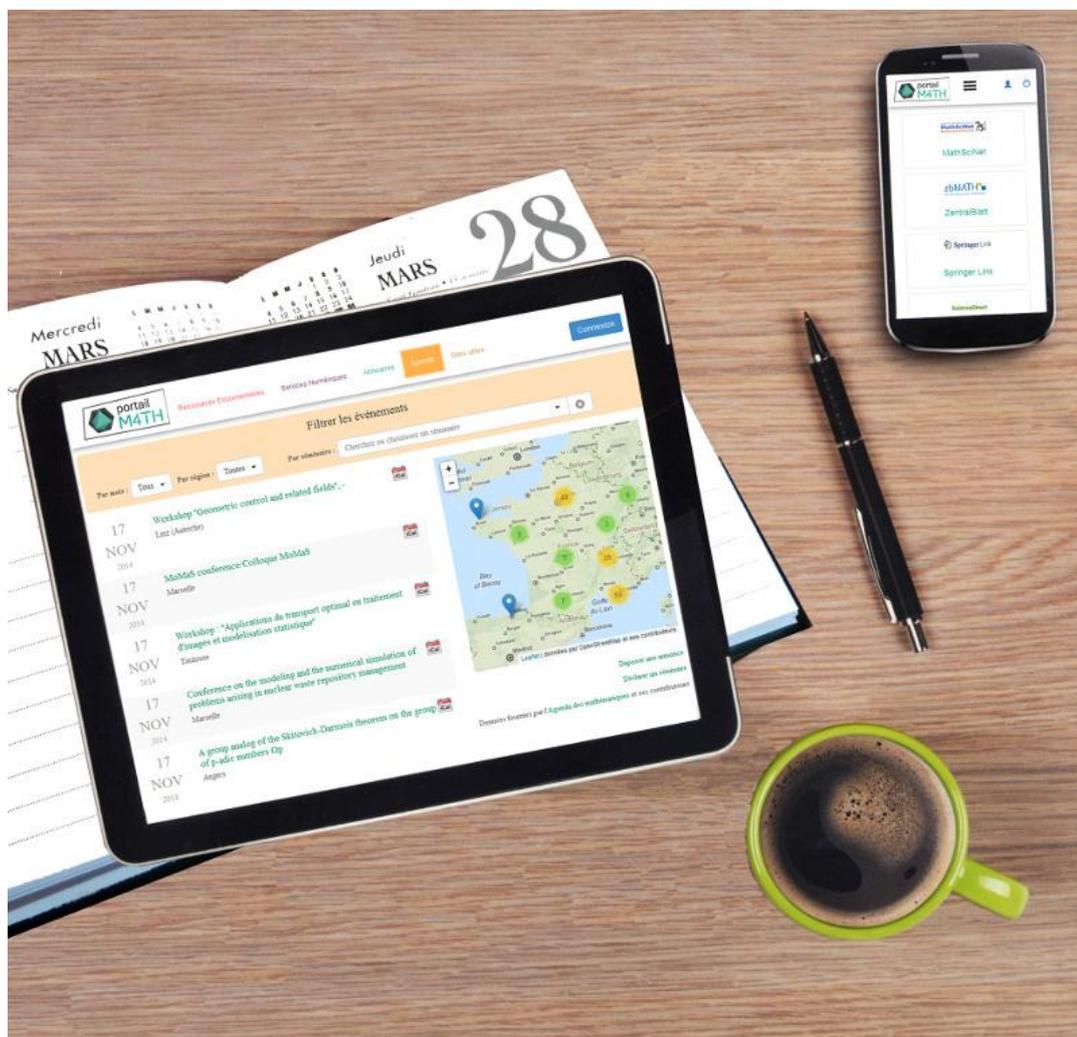
- Descriptif
- ▾ Type de documents
  - x Revues
  - Actes de congrès (5)
  - Livres (2)
  - Autres (1)
  - Thèses (1)
- Editeur
- Gestionnaire/hébergeur
- ▾ Conditions d'accès
  - x Accessible à tous

**Figure 8 Le portail math v1**

- *Projet Portail Math v2 (avril-décembre 2014)*

Le comité de pilotage du projet a précisé la deuxième phase en juin 2014 en définissant comme objectif à court terme l'intégration d'une authentification unique des utilisateurs pour les accès documentaires et les services numériques précédemment hébergés par la PLM, et la fusion des sites portail-math.fr et plm.math.cnrs.fr. Le nouveau site web, portail.math.cnrs.fr est principalement destiné aux chercheurs. Les fonctionnalités prévues à l'ouverture du site en décembre 2014 sont

- authentification, avec mécanisme SSO ;
- intégration de l'agenda des maths (affichage des événements avec une carte graphique et une liste d'événements) ;
- accès direct aux ressources documentaires précédemment gérées par la PLM ;
- annuaires des personnes des laboratoires et des bibliothèques ;
- gestion par les utilisateurs de leurs informations personnelles, notamment l'association de plusieurs adresses mails ;
- dépôt de clés SSH.



**Figure 9 Le portail math V2**

La Cellule Mathdoc intervient en collaboration avec Mathrice et le RNBM dans la définition de l'architecture et des choix technologiques du site, dans la gestion du projet en mode itératif et dans le développement du site web. Une collaboration a été également mise en place avec l'AMIES et l'équipe projet de l'agenda des maths pour développer des services web et permettre l'accès aux données de l'agenda des maths.

Le site [portail.math.cnrs.fr](http://portail.math.cnrs.fr) est développé en Ruby/Sinatra côté serveur et avec Angular/ExtJS côté client.

- *Développement informatique*

Le portail v1 est développé avec le CMS Drupal 7 et le moteur de recherche Solr. Des modules spécifiques ont été développés en PHP pour s'affranchir du système de *hook* de Drupal et proposer un affichage personnalisé des résultats de recherche.

Le site de la PLM était développé jusqu'en juin 2014 en Ruby/Sinatra côté serveur et ExtJS côté client.

Afin de permettre la fusion de ces deux approches diamétralement opposées, le site [portail.math.cnrs.fr](http://portail.math.cnrs.fr) a été développé en Ruby/Sinatra côté serveur et avec AngularJS côté client, tout en gardant ExtJS pour les services numériques, bibliothèque chargée dynamiquement lors de l'appel à l'un de ces services. Une architecture MVC (Modèle Vue Contrôleur, un des patrons de conception de l'orienté objet) est utilisée ; le serveur fournissant principalement les données avec

des API REST. Le paradigme Single Page Application (SPA) est suivi afin de réduire les chargements de page au strict nécessaire. Par exemple, l'utilisateur peut interagir avec les événements de l'agenda (zoomer, filtrer, sélectionner...) sans recharger la page.



Figure 10 L'agenda et son menu interactif

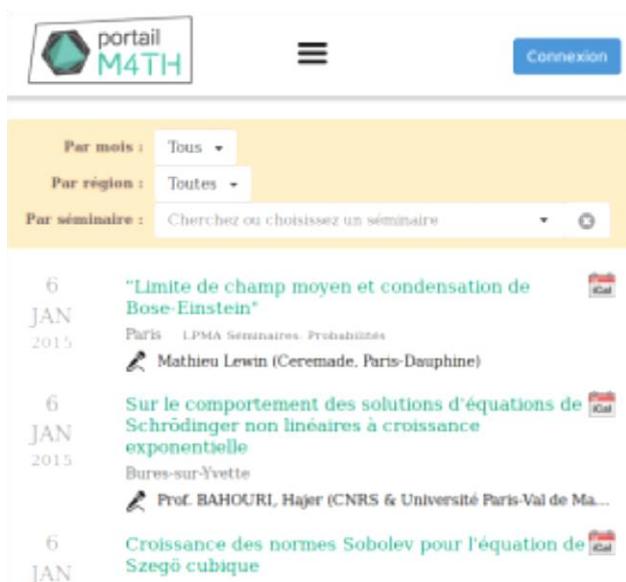


Figure 11

Des services REST ont été développés sur [api.mathdoc.fr](http://api.mathdoc.fr) pour obtenir les coordonnées GPS des laboratoires et améliorer les événements produits par le site de l'agenda des mathématiques. La bibliothèque Bootstrap a été choisie pour le *responsive design* et le portail math offre un affichage optimal quel que soit le matériel utilisé pour consulter le site (smartphone comme sur la figure 11, tablette, ordinateur personnel, écran large comme ci-dessus).

## 2) Les annuaires

La Cellule Mathdoc maintient également un annuaire des laboratoires et bibliothèques de mathématiques (coordonnées, responsables, site web...). La base de données est régulièrement tenue à jour par le contact avec les bibliothèques pour récupérer les nouvelles informations. Des services web, disponibles depuis [api.mathdoc.fr](http://api.mathdoc.fr) ont été conçus pour rendre les données accessibles par d'autres sites web, en particulier le portail math. Le site [api.mathdoc.fr](http://api.mathdoc.fr) décrit les API de ces services web.

## IV. ACTIVITÉS TRANSVERSALES

### A. Développement logiciel

Nous regroupons ici les développements qui n'ont pas été décrits dans une activité, en général parce qu'ils sont utilisés par plusieurs projets.

#### a) EDBM

Développé à la fin des années 90 pour servir de moteur de recherche et interface web pour le Zentralblatt MATH, dans un contexte d'offre logicielle très différent de ce que nous connaissons depuis déjà d'assez nombreuses années, ce système a évolué au tout début des années 2000 en vue d'héberger la première mise en ligne de Numdam. Il est toujours la base d'application telles que Numdam et Cedram, et est toujours maintenu.

Aucun développement n'est prévu, son remplacement à court terme par des logiciels offrant de meilleures fonctionnalités, et dont la pérennité est assurée, doit être envisagé.

#### b) « Matching »

On désigne ici par « matching » (appariement) le processus automatique qui permet de découvrir dans une base de données l'entrée correspondant à des données non structurées. Cela nous concerne typiquement pour la découverte de références croisées entre publications, via les références bibliographiques, ou la création de liens vers les analyses des bases MathSciNet et zbMATH.

Un tel système a été développé par la Cellule Mathdoc dans la période précédant immédiatement celle de ce rapport d'activité. Il permet de générer les liens vers la base zbMATH, pour les articles et leurs références bibliographiques, et donc d'éventuelles références internes par comparaison des identificateurs Zbl. Ce système est utilisé par les applications Numdam et Cedram. Il a été ensuite adapté dans le cadre du projet EuDML, et constitue la base du service dit « Reference lookup » offert par le site EuDML. Il a également été fourni au FIZ dans le cadre de notre partenariat et est utilisé en interne, dans une version possiblement remaniée, pour affecter des identifiants Zbl aux références bibliographiques désormais intégrées au zbMATH (à l'heure actuelle, les sources principales sont Springer et Numdam).

Les liens vers Mathscinet sont quant à eux établis grâce à un service fourni par ce site.

#### c) Transformations XSLT

Étant donnée la nature des activités de l'unité, et les formats de données utilisés par de nombreux partenaires extérieurs aussi bien que de façon interne, l'écriture de transformations XML (XSLT = XML Stylesheet Language Transformations) est une pratique habituelle dans tous les domaines nécessitant un échange ou un affichage de données XML:

- acquisitions de données provenant des éditeurs ;
- export de données vers le site EuDML ou via les serveurs OAI-PMH ;
- affichage HTML (web).

Plusieurs dizaines de transformations ont été développées, de la plus simple à la plus complexe, souvent intégrées dans des programmes nécessitant la manipulation de données XML.

#### *d) Plateforme de diffusion pour Écologie alpine*

La mise en ligne des articles numérisés de trois revues d'écologie alpine, et des cartes associées, a été l'occasion de mettre en œuvre des logiciels autres que ceux habituellement utilisés à la Cellule Mathdoc pour ce type d'application.

Le site web mis en place (<http://ecologie-alpine.ujf-grenoble.fr>) a été développé avec le « framework » Django (langage Python), avec stockage des données dans une base MySQL, associé au moteur SphinxSearch qui assure la recherche textuelle.

Ce « montage » inhabituel s'est révélé facile à mettre en œuvre et robuste, le cœur de l'application ayant nécessité environ 600 lignes de code. L'aspect visuel du site, conçu en collaboration avec le laboratoire d'écologie alpine, a finalement représenté l'essentiel du temps consacré à ce travail.

#### *e) Plateforme de diffusion pour Numir*

Michael Lefevre a été employé (par Paris 6) à Mathdoc pour mettre en place la plateforme de diffusion du projet Numir. Il a installé le logiciel PHPBoost et paramétré le moteur de recherche pour avoir une recherche (séparée) sur les métadonnées des ouvrages et leurs bibliographies. Il a réalisé aussi un certain nombre de développements simples pour obtenir l'ergonomie attendue par les collègues parisiens. Il est peu probable que ce développement sera maintenu très longtemps. Nous prévoyons plutôt de nous doter d'une solution générique pour héberger des collections numériques de toutes natures (articles de revue, livres, thèses...).

#### *f) Plateforme numérique Mathdoc*

Mathdoc a commencé à l'automne 2012 le développement d'une nouvelle plateforme destinée à gérer et diffuser les collections numériques créées ou hébergées par Mathdoc. Elle utilise des logiciels bien établis et éprouvés :

- framework Python Django (avec base de données Postgres) ;
- moteur de recherche Solr pour les fonctionnalités de recherche avancées (recherche textuelle, facettes).

Une version alpha a été mise en ligne à usage interne récemment. Elle propose une partie des fonctionnalités attendues :

- le multilinguisme n'est pas implémenté ;
- pas de contrôles d'accès à l'heure actuelle ;
- l'articulation avec les chaînes de production de données reste à spécifier de façon détaillée (prépublication, publication, mises à jour ponctuelles) ;
- pas encore de serveur OAI-PMH.

Cette version représente environ :

- 8000 lignes de code python, dont environ 30 % pour l'importation des données ;
- 2000 lignes de code de présentation (gabarits pour 3 sites hébergés) ;
- 3500 lignes de transformations XSLT (sans compter les transformations qui produisent les données initiales) ;
- 150 lignes de scripts bash.

Ainsi qu'il était prévisible, il a été constaté que le succès d'un tel projet repose avant tout sur la disponibilité de bons schémas de métadonnées, de métadonnées complètes et fiables, plus que sur un logiciel particulier. Le travail de conversion et d'éventuel enrichissement des métadonnées disponibles, bien que largement commencé à l'occasion du projet EuDML, a dû être adapté et complété, et devra être poursuivi, certaines collections n'ayant pas encore été converties dans les schémas utilisés par la plateforme.

### *B. Site Web de l'unité et identité visuelle*

Dans la période 2011-2012, pour le site Web de l'unité, un changement de CMS (SPIP vers Drupal) et une réorganisation complète du contenu ont été réalisées.

Un IE en CDD avait commencé le travail qu'un ingénieur de l'unité a terminé, avec l'aide de l'ensemble de l'équipe. Comme points à relever :

- utilisation de certains modules de Drupal, pas de développement spécifique ;
- la finalisation du look et un nouveau bandeau ont été réalisés avec l'aide d'un web designer ;
- le même site web public est aussi utilisé comme intranet avec contrôle d'accès pour l'unité ;
- un fil RSS a été installé.

Dans le même temps, un nouveau logo a été conçu et réalisé, avec l'aide d'un graphiste de la DR11, dont au final aucune proposition n'a été retenue.

### *C. Infrastructures*

#### *a) Description*

L'infrastructure et les services informatiques de la cellule Mathdoc s'appuient sur ses ressources humaines et matérielles internes mais aussi sur des ressources de l'institut Fourier et de Mathrice, dans un souci de mutualisation maximale. Ainsi, l'infrastructure réseau, câblage et équipements actifs, le robot de sauvegarde et la salle informatique sont partagées avec l'institut Fourier; les nouveaux services nationaux comme le portail math sont maintenant sur des machines hébergées par Mathrice.

L'architecture choisie, recommandée par Mathrice, repose sur des serveurs virtualisés avec trois serveurs physiques utilisés comme hyperviseurs et un serveur partagé avec Mathrice.

Le pool de stockage est constitué d'un disque NAS, d'un robot de sauvegarde sur bandes (Dell powerwault TL2000), de deux baies de disques Mathdoc et d'une baie de disque partagée avec Mathrice.

Les spécifications des serveurs sont les suivantes :

1. Un Dell 2950, équipé de deux Xeon quadcore avec 16 Go de mémoire vive, de deux disques de 146 Go pour le système et de 4 disques de 300 Go pour les serveurs virtuels.  
L'hyperviseur s'appuie sur le système d'exploitation Debian et les logiciels kvm-libvirt pour la virtualisation. Cet hyperviseur gère le robot de sauvegarde et une baie de disques de 4 To pour les données et de 2 To pour les serveurs virtuels.
2. Deux Dell R710 équipés de deux Xeon quadcore avec 24 Go de mémoire vive, de deux disques de 146 Go pour le système et de quatre disques de 300 Go pour les serveurs

virtuels. Ces hyperviseurs s'appuient sur le système d'exploitation Debian et proxmox-kvm pour la virtualisation. Ces deux serveurs forment un cluster sur lequel est attachée une baie de disque d'une capacité de 2 To pour héberger les serveurs virtuels. Ces hyperviseurs gèrent onze serveurs virtuels, basés sur le système d'exploitation Ubuntu Serveur LTS.

3. Un Dell R420 équipé d'un Xeon six cœurs avec 16 Go de mémoire vive, de deux disques de 300 Go, et exploite le système d'exploitation Debian. Cet équipement utilisé en commun avec Mathrice gère une baie de disques durs d'une capacité de 15 To.
4. Un serveur Nas synology Rs3413xs+ avec une capacité utile de 9 To.

Les principaux logiciels utilisés sont : Hobbit pour la supervision et le monitoring des serveurs, Bacula pour la sauvegarde, LDAP pour annuaire, Apache pour les serveurs Web, MySQL et Postgresql pour les bases de données, SolR comme moteur de recherche, Java Tomcat comme serveur d'applications, Repox logiciel de gestion et d'agrégation de données (OAI-PMH), Spip, PHPboost et Drupal comme CMS, Awstats pour les statistiques d'accès, Redmine pour la gestion des projets (tickets principalement).

#### *b) Historique 2010-2014*

La liste ci-dessous, qui n'inclut pas les mises à jour de l'ensemble des logiciels et matériels, est destinée à donner une idée du travail d'administration systèmes et réseaux, souvent caché mais obligatoire pour la bonne marche et l'évolution des services offerts par une unité comme Mathdoc.

- *2010*
  - Arrivée de Simon Chevance dont le poste d'administration système et réseaux est mutualisé entre l'institut Fourier et Mathdoc.
  - Création d'un serveur NFS, migration de serveurs physiques sur des serveurs virtualisés.
  - Étude et analyse des flux éditoriaux et de numérisation pour centraliser les données des serveurs.
  - Mise en place du robot de sauvegarde sur bande LTO pour remplacer l'ancien système.
  - Migration du service mail de la cellule Mathdoc dans celui de l'Institut Fourier.
- *2011*
  - Mise en place de deux serveurs supplémentaire sous proxmox pour la virtualisation et d'une seconde baie de disques.
  - Installation d'une baie de disques mutualisée avec Mathrice.
  - Migration des sites de la cellule Mathdoc sur des serveurs virtualisés.
  - Mise en place d'un serveur Samba et amélioration du serveur NFS pour centraliser les données et mieux séparer les serveurs de production publics des serveurs de développement ou de test.
- *2012*
  - Mise en place de la supervision et du monitoring des serveurs.
  - Mise en place d'un serveur Drupal pour le Web de l'unité et migration de l'ancien sous SPIP avec création de 2 sites séparés : tests et développement.
  - Mise en place des archives sur la deuxième baie de disques.

- 2013
  - Mise en place d'un logiciel de gestion de projets, principalement utilisé pour la gestion de tickets : Redmine.
  - Mise en place de l'authentification LDAP pour le CMS SPIP.
  - Installation de 3 serveurs pour le projet portail math sous Drupal.
  - Installation de nombreux modules Drupal.
  - Développements spécifiques pour Drupal concernant l'exploitation, la présentation des données, la recherche...
  - Tests d'authentification via Shibboleth sur le CMS Drupal pour le projet portail math.
  - Mise en place du logiciel Repox pour l'agrégation de contenu.
- 2014
  - Ouverture du portail math <https://www.portail-math.fr/> et fermeture du vieux portail documentaire de Mathdoc <http://portail.mathdoc.fr/> (nom de domaine, redirections, migrations des sites et services intégrés à l'ancien portail sur <http://sites.mathdoc.fr/>).
  - Mise en place d'une sauvegarde à distance des données de la Cellule Mathdoc sur un serveur Mathrice situé à Lille.
  - Modification de l'infrastructure de stockage : intégration d'un NAS pour la baie de disques Mathdoc-Mathrice.

## V. COOPÉRATIONS

### A. Locales

Plusieurs réunions se sont tenues avec le SICD 1 à propos d'une politique documentaire de site. Nous sommes de fait solidaires sur les projets de numérisation car nous partageons désormais un marché cadre (dont nous nous servons pour Numdam 4, et sur lequel nous avons donc un plafond de dépenses commun). Un problème récurrent est l'inexistence d'une plateforme de mise en ligne de documents numérisés pour valoriser les collections du site (Annales de l'université de Grenoble, publications de l'école des Houches, de l'OSUG ou de l'institut Fourier...). Les progrès sont lents.

Nous avons numérisé plusieurs collections de l'OSUG, au cours de cette période les revues d'écologie alpine. Nous avons été contactés pour mettre en ligne des revues de glaciologie numérisées par l'INIST mais sans les métadonnées minimales selon nous. Nous n'avons pas répondu positivement à cette demande.

### B. Nationales

#### a) RNBM

Les relations entre Mathdoc et le RNBM ont toujours été étroites et importantes. Elles se traduisent par un soutien au réseau et par la participation à des groupes de travail.

D'une part, le soutien au réseau prend la forme d'une participation à l'organisation et des interventions aux journées du RNBM (Grenoble : 7-9 février 2011, Grenoble, 8-10 avril 2014) et aux ANF<sup>15</sup> (CIRM, 21-24 mai 2013, CIRM : 2015 en cours de préparation).

D'autre part, Mathdoc participe à plusieurs groupes de travail : groupe PEB sur le prêt entre bibliothèque, groupe PCMaths sur le plan de conservation partagée des bibliothèques de mathématiques, groupe CFP sur l'évolution du catalogue fusionné des périodiques.

Les bibliothèques du RNBM participent au programme Numdam par le prêt de leurs collections pour la numérisation.

### *b) Mathrice*

Le réseau Mathrice est un soutien à la recherche en mathématiques et un lieu d'échange et d'entraide pour ses informaticiens. Mathdoc intervient régulièrement lors des journées et a participé à l'organisation des journées Mathrice et de la session commune RNBM-Mathrice à Grenoble en avril 2014.

Les liens se sont resserrés à l'occasion du projet portail math dans la mesure où Mathrice collabore sur les services numériques et une intégration poussée a été recherchée entre la mise à disposition de services numériques et l'aspect plus documentaire du portail. L'objectif de fusionner le portail documentaire et la PLM est actuellement porté par le groupe PLM de Mathrice auquel Olivier Labbe a été intégré.

### *c) INIST (institut de l'IST, unité propre de service du CNRS)*

Dans le cadre du projet portail math, l'INSMI a provoqué en 2012 une réunion entre les acteurs de ce projet et l'INIST pour étudier les opportunités de collaboration principalement sur le projet ISTEEX. Les calendriers des projets n'étant pas vraiment compatibles (et le souhait de l'INSMI de voir des moyens ISTEEX alloués au projet portail math étant rendu impossible par le caractère contractuel du projet), cette collaboration n'a pas démarré. Une nouvelle réunion en décembre 2014 a cependant confirmé l'intérêt de l'INIST pour échanger nos expériences sur le portail math d'une part comme portail à façon pour une communauté disciplinaire offrant un accès aux collections ISTEEX, et d'autre part en relations avec les portails BiblioSciences de l'INIST en cours de réfection.

### *d) DIST (direction de l'IST au CNRS)*

Thierry Bouche a participé au groupe de travail national du projet Medoanet coordonné par la DIST à Lyon en novembre 2012.

En décembre 2013, la DIST a défini une stratégie de l'information scientifique et technique pour « mieux partager les connaissances » et construit quatre plans d'actions partagées (PAP) visant à faire avancer les choses au cours de l'année 2014.

---

15 ANF : action nationale de formation

Mathdoc participe au PAP 2 (Publier) à travers l'action 3 portée par Thierry Bouche sur l'accompagnement de l'édition académique. Mathdoc a également participé aux deux colloques organisés par la DIST à Meudon en mars et novembre 2014 (interventions de Thierry Bouche dans les ateliers).

#### *e) BSN (Bibliothèque scientifique numérique)*

La bibliothèque scientifique numérique est une infrastructure de recherche nationale (IR), inscrite en 2008 sur la feuille de route des IR et activée au cours de l'année 2011 à l'initiative du ministère de l'Enseignement supérieur et de la Recherche sous la forme de groupes de travail et d'un groupe plénier devant produire des recommandations sur la structuration des acteurs de l'IST en France en découpant le travail en 9 chantiers ou segments. Mathdoc a participé à ces travaux en prenant part au groupe plénier et aux segments suivants :

- BSN 2 — Accès (portails) ;
- BSN 5 — Numérisation ;
- BSN 6 — Archivage pérenne ;
- BSN 7 — Édition scientifique publique ;
- BSN 9 — Formation, compétences et usages.

À la fin de la phase de réflexion, le groupe plénier a été remplacé par un comité technique composé des pilotes et copilotes des segments ; les décisions relevant d'un comité de pilotage. La composition des groupes de travail a été redéfinie. Mathdoc fait actuellement partie des groupes 5, 6 et 7. Le groupe 2 a semble-t-il cessé de fonctionner (attendant les résultats d'ISTEX). Le groupe 5 s'est transformé en agence de financement en gérant un appel d'offre de numérisation.

#### *f) BNF (Bibliothèque nationale de France)*

Depuis 2002, Mathdoc est signataire d'une convention du pôle associé avec la BNF sous l'intitulé « *numérisation concertée en mathématiques* ». La dernière convention cadre a été signée en 2012 pour les années 2012-2014 et devrait être reconduite en 2015.

Dans ce cadre, nous avons travaillé à la réindexation de documents numérisés dans le cadre de Gallica et donc à la création des sites Gallica-Math. Notons à ce propos que nous avons versé l'index des articles du JMPA produit par Gallica-Math dans EuDML. Depuis avril 2011, Mathdoc est également bibliothèque partenaire de Gallica. Les collections de Numdam sont ainsi moissonnées dans un but d'une convergence documentaire (complémentarité de l'offre documentaire). Cet événement a donné lieu à la diffusion d'un billet sur le blog Gallica.

À la même époque, Mathdoc a participé à l'enquête sur les usages des bibliothèques numériques en sciences et techniques conduite par la BNF et dont Numdam était l'une des bibliothèques étudiées. Cette étude a été publiée en janvier 2012 dans un rapport déposé sur HAL.

#### *g) Association GUTenberg*

Rappelons pour finir que Mathdoc est le siège de l'association GUTenberg et héberge le site de diffusion des *Cahiers GUTenberg*, publication de référence auprès des utilisateurs de TeX et des

amateurs de typographie numérique, dont le rédacteur en chef est Thierry Bouche (dernier numéro daté d'avril 2012, paru en juillet 2014).

### *C. Internationales*

#### *a) ZbMath*

La coopération avec le FIZ Karlsruhe (bureau de Berlin) se poursuit sous une forme assez légère. Thierry Bouche est membre du comité de coordination du Zentralblatt (désormais zbMATH) qui se réunit tous les 6 mois à Berlin. EDBM a été utilisé par le Zentralblatt jusque très récemment comme moteur de recherche, l'interface web étant gérée autrement. Ils ont finalement cessé de l'utiliser au lancement de la nouvelle interface, basée sur des facettes. En revanche, ils ont repris plusieurs résultats obtenus par Mathdoc dans le projet EuDML dont le service de matching, le format de métadonnées JATS.

#### *b) EuDML initiative*

L'initiative EuDML s'est constituée formellement en 2014. Les membres fondateurs sont au nombre de 12. Elle dispose d'un comité exécutif, d'un comité technique et d'un comité consultatif scientifique. L'assemblée générale du 3 juin 2014 a élu les membres des deux premiers comités et le président de l'initiative. Les membres du comité scientifique doivent être nommés par la SME.

#### *c) Coopérations scientifiques*

Thierry Bouche a été membre du comité Electronic Publishing (EPC) de l'EMS (2005-2012), du committee on Electronic information and communication de l'IMU (2011-2018), du comité scientifique des conférences CICM (Conferences on Intelligent computer mathematics) qui comportent en particulier des sessions sur les thèmes MKM (Mathematical knowledge management) et DML (digital mathematics library) dans lesquels Mathdoc est directement concerné, et d'ailleurs souvent actif. Thierry Bouche a été élu au comité de pilotage CICM.

Thierry Bouche a été élu à plusieurs reprises au conseil de la SME (en tant que représentant des membres institutionnels en 2008, puis en tant qu'adhérent individuel en 2012). Il a présenté rapidement l'initiative EuDML au dernier conseil.

Thierry Bouche est membre du groupe de travail GDML formé à Séoul sous les auspices de l'IMU.

## Projet

Nous reprenons les grands chapitres de ce rapport et donnons les actions qu'il nous semble nécessaire d'entreprendre au cours des années à venir.

### I. BIBLIOTHÈQUES NUMÉRIQUES

- *Plateforme et chaîne de numérisation*

Une des premières actions est de mettre à plat notre chaîne de production et de diffusion des collections numérisées. Il y a là à la fois une opportunité et une certaine urgence car après le départ de Catherine Zoppis, l'évolution de la chaîne actuelle est rendue nécessaire notamment pour accompagner le projet BSN 5 pour lequel nous avons des délais assez stricts de mise en ligne, mais ne peut se faire comme de petites modifications incrémentales. Tandis que Claude Goutorbe travaille avec Olivier Labbe sur la plateforme numérique de Mathdoc qui permettra de remplacer EDBM pour l'hébergement et la diffusion de nos corpus numériques, nous allons utiliser nos reliquats du projet EuDML pour recruter un ingénieur sur CDD pour une mission d'un an environ dont l'objectif sera de créer la nouvelle base de données de production de Numdam, ainsi que les applications satellites permettant de gérer cette production. Le modèle de données sera un sur-ensemble de celui de la base de diffusion, simplifiant ainsi la publication et permettant de remonter certaines tâches à l'étape logique de la chaîne de production, ce qui n'est pas le cas aujourd'hui (un exemple : le matching des références bibliographiques). La maîtrise des technologies de la nouvelle plateforme par plusieurs membres permanents de l'équipe en prévision du départ à la retraite de Claude Goutorbe est également un objectif prioritaire. Cela permettra la mise en service de cette plateforme courant 2015. Un avantage attendu de cette plateforme est de disposer d'une infrastructure de bas niveau assez universelle pour héberger toutes les collections numériques dont nous avons la charge (collections numérisées, acquises, archivées dans le cadre de licences nationales, hébergées pour le compte de partenaires académiques...), voire pour réaliser un miroir de la DML remplaçant les services mini-DML et LiNum qui sont appelés naturellement à être fermés dès que leur contenu se retrouve dans un plus grand réservoir de données (ce qui n'est pas le cas actuellement puisque EuDML est toujours limité à des partenariats européens).

- *Politique d'acquisition documentaire*

Un second objectif est de définir une politique d'acquisition des collections numériques de Mathdoc. La numérisation de livres en aveugle est hors de portée pour une unité de notre taille. En revanche, une numérisation concertée pourrait se faire en lien avec le RNBM sur le principe que tout mathématicien pourrait recommander la numérisation d'un texte rare ou difficile d'accès, de façon à constituer un corpus utile. Il reste également quelques revues qui n'ont toujours pas été numérisées, comme le *Journal de mathématiques pures et appliquées* ou le *Bulletin des sciences mathématiques*. Ces deux dernières revues étant numérisées par Gallica avec une barrière mobile de 70 ans, le problème de la réindexation de ces collections se pose également avec une certaine acuité, ainsi d'ailleurs que celle des CRAS. Toutes ses actions trouvent assez naturellement leur place dans les projets nationaux (BSN, PAP de la DIST), mais nous ne les entreprendrons que si on nous en donne les moyens humains et financiers. De même, l'acquisition de la production récente

des publications archivées dans Numdam reste une priorité scientifique pour Mathdoc. Le projet Scientific Edition prend en compte cette activité et souhaite en faire bénéficier les revues nouvelles et innovantes. Il ne faut pas sous-estimer le caractère chronophage de cette activité.

- *Agrégation et bibliothèques virtuelles*

Nous participons à une proposition de projet européen porté par le réseau Eu-MATHS-IN qui a souhaité intégrer un *work package* créant un lien entre les outils à destination des industriels et de la communauté académique en simulation et modélisation (et plus généralement en mathématiques) et la bibliothèque EuDML. De façon générale, nous ferons tout ce que nos moyens nous permettront de faire pour élargir les collections EuDML à tous les partenaires possibles, et pour clarifier et améliorer le service en ligne. Nous envisageons de faire un miroir des collections comme test pour notre propre plateforme, et comme sécurité dans le cas où le partenaire principal ne maintiendrait pas le service si aucun financement spécifique ne devait apparaître rapidement.

Nous participerons aussi au travail à l'échelle mondiale, avec pour principale stratégie de faire avancer en parallèle ce qui peut être fait dans chaque continent en fonction des opportunités et des communautés locales, en cherchant à faire adopter des standards d'interopérabilité simples et peu contraignants mais permettant au final de combiner les résultats des différents projets et constituer de la sorte une infrastructure globale.

- *Archivage pérenne*

Mettre en place un archivage pérenne de nos collections est un objectif affiché par Mathdoc depuis des années. Nous avons cru l'atteindre en cours de ce quinquennal à la faveur de la BSN dont une recommandation des groupes 2, 5 et 6 était de labelliser des collections (sur critères scientifiques et techniques) et d'offrir en contrepartie un dispositif d'accès et d'archivage pérenne de plein droit. Ces recommandations n'ayant pas survécu à la réalité de la rigueur budgétaire, il nous resterait la possibilité de payer une somme non négligeable (et annuelle, aussi longtemps que nous souhaitons conserver l'archive, donc pour l'éternité ?) au CINES, ce que nous nous sommes refusés à faire pour l'instant. Comme alternative, nous avons proposé aux partenaires de l'initiative EuDML de monter une archive répliquée et dupliquée CLOCKSS au moins pour les collections de Numdam. Nous allons faire pression en faveur de ce projet mais, comme il dépend de moyens dans plusieurs pays européens pour sa mise en place, nous ne nous faisons pas d'illusions sur l'imminence de sa mise en œuvre.

- *Mettre en place un contrôle qualité pour les acquisitions numériques*

En fonction des moyens qui nous seront alloués, nous souhaitons mettre en place une véritable chaîne d'acquisition pour Numdam, qui prenne en compte les formats d'archivage comme PDF/A et opère un véritable contrôle qualité sur les données reçues comme nous le faisons pour la numérisation.

- *Produire des statistiques pertinentes sur l'usage de nos services*

Nous disposons de très peu d'outils de statistiques pour mesurer et surtout mieux comprendre l'usage qui est fait de nos services. Nous ne savons pas par exemple dans quelle mesure la recherche sur les bibliographies est utilisée ni à quelles fins. Numdam en propose une version très élaborée depuis la première mise en ligne ; elle est disponible sous une forme très différente dans Numir et n'est pas proposée dans EuDML par exemple. Nous savons que la grande majorité des utilisateurs font leurs recherches dans des moteurs de recherche généralistes ou dédiés et

atteignent donc directement l'article sans utiliser le moteur de recherche local (ce qui est impossible dans Numir). Nous savons aussi que quelques utilisateurs font un usage très avancé de nos moteurs de recherche (les historiens, notamment). Mais nous n'avons pas de données au-delà de quelques logs que nous n'avons pas le temps d'analyser. Nous avons pu faire un peu avancer ces questions lors de l'enquête de la BNF à laquelle nous avons participé en tant que bibliothèque associée à Gallica. Un partenariat avec des chercheurs et quelques développements à l'occasion de stages de master pourraient être envisagés.

## II. ÉDITION

- *Plateforme et sites de diffusion*

Une fois la nouvelle plateforme éprouvée pour Numdam et la bibliothèque numérique de Mathdoc, nous la déclinons pour refondre les sites du Cedram (un site « portail » et un site par revue). Cela nous permettra de fermer les sites Spip en simplifiant la gestion des pages statiques.

- *Projet Scientific Edition et OJS*

Le projet Scientific Edition propose de créer à Grenoble un centre d'excellence pour l'édition de revues académiques en sciences dures (composées en LaTeX), à travers une offre modulaire de services soutenant en particulier les revues évoluant vers un libre accès « diamant ». Il s'agit d'un changement d'échelle du Cedram dans le périmètre et la palette de services proposés. En particulier, nous voyons ce projet comme une opportunité pour mettre à plat la chaîne de production des revues, y inclure la gestion de DOI de façon propre, ajouter le service antiplagiat crosscheck auquel nous venons d'adhérer à la demande de la SMAI, et automatiser ou simplifier de nombreuses tâches systématiques de la chaîne de production d'une revue. Il s'agira en particulier de déployer plus systématiquement OJS pour gérer les flux de production et créer des passerelles avec les autres outils, y compris hors du strict périmètre du Cedram, puisque Epi-Sciences ou certaines revues utilisant OJS pour la diffusion bénéficieraient de nos services.

Parmi les autres services envisagés, il y a la mise aux normes des fichiers LaTeX, l'impression *a posteriori* ou à la demande, sous une forme qui reste à déterminer : carnet d'adresse mutualisé ou prestation de service.

- *Documenter et former*

Comme nous travaillons toujours dans l'urgence et en sous-effectif, il y a un certain nombre de tâches utiles qui restent en souffrance. Nous avons le projet de trouver les moyens et de réaliser enfin :

- la documentation du système LaTeX du Cedram et de ses applications satellites ;
- de façon générale la documentation et la diffusion en licence libre de plus d'applications développées à Mathdoc ;
- la formation des revues et des sous-traitants à nos outils.

- *Produire des statistiques Counter*

Il existe une norme professionnelle pour la collecte et la présentation de statistiques d'utilisation des sites d'édition électronique : le projet Counter <http://www.projectcounter.org/>. C'est un service qui doit être ajouté à celui de la plateforme de diffusion. De telles statistiques nous sont fréquemment demandées par des bibliothèques abonnées aux revues du Cedram. On peut

supposer qu'elles seront moins demandées pour les revues en libre accès, mais toujours utiles pour les responsables de ces revues.

### III. PORTAIL MATH ET SERVICES DE DOCUMENTATION

- *À court terme*

Le portail math va évoluer pour permettre l'accès direct à plus de ressources documentaires, il faut donc prendre en compte les abonnements locaux. Le CFP devra s'interfacer avec le portail math, et des services web seront développés pour rendre disponible la liste des revues de façon à amener la granularité des objets documentaires disponibles sur le portail à la revue et non plus à l'éditeur ou à la plateforme. De nouvelles fonctionnalités seront apportées au CFP pour pouvoir modéliser les licences documentaires et permettre de répondre à la question « qui a accès à quoi ». La collaboration avec le RNBM sera poursuivie en vue de cataloguer toutes les licences documentaires. Des composants logiciels informatiques jouant le rôle d'intermédiaire en se plaçant entre deux hôtes (proxy) seront mis en place en collaboration avec Mathrice.

Pour répondre aux besoins du RNBM, le CFP sera étendu pour gérer le prêt entre bibliothèques. De plus, le site portail-math.fr sera maintenu ou évoluera, pour continuer le catalogage des licences documentaires dans la mesure où ces fonctionnalités ne rentreraient pas facilement dans le futur CFP.

Les annuaires des laboratoires et des bibliothèques seront complètement intégrés au portail math. Des services et sites web seront proposés pour gérer les données de la base maintenue par les membres du RNBM.

- *À plus long terme*

L'objectif final en matière d'accès à la documentation au travers du portail math est de disposer d'une modélisation complète des accès. Il s'agit en principe de croiser plusieurs informations : connaissance des personnes qui sont identifiées, de leurs rattachements institutionnels (à travers l'annuaire des personnes, des laboratoires et des organismes) donc de leurs droits d'accès (en lien avec l'annuaire des bibliothèques et une base de données des licences qui pourrait être une forme d'état des collections électroniques dans le CFP qui recense déjà l'état des collections papier dans toutes les bibliothèques du RNBM). Ces droits portent en général sur des titres (mais plus souvent, sur des segments de titres, comme par exemple : *Annals of Math*. Disponible pour les archives chez JSTOR et pour le courant chez MSP, avec un segment sur arXiv ou ELibM). L'idéal serait donc de disposer d'une base de données des articles qui soit liée à un CFP amélioré permettant de déterminer les conditions d'accès (il n'y a pas bijection) et les chemins d'accès correspondants pour l'utilisateur connecté. Il s'agit de répondre non seulement à la question « qui a accès à quoi » mais aussi à « comment », et d'implémenter cette action pour créer l'accès d'un individu à un article.

Un écueil sur notre route est qu'il n'existe pas de base de données publique de tous les articles susceptibles d'intéresser les mathématiciens. Une stratégie est d'utiliser les bases de données MSN et ZM, mais on pourra aussi évaluer les *discovery tools*.

## IV. INFRASTRUCTURES

L'architecture actuelle n'est pas complètement satisfaisante et montre des faiblesses en termes de temps de réponse, d'espace de stockage pas entièrement disponible, d'authentification séparée sur chaque serveur.

Il faut donc la faire évoluer, en particulier :

- Modifications d'infrastructure de virtualisation et des pools de stockage, évolution technologique et intégration d'un 4<sup>e</sup> hyperviseur.
- Mise en place d'une authentification centralisée pour les serveurs.
- Centralisation des logs.

Pour les projets décrits ci-avant il sera nécessaire en particulier d'installer :

- plusieurs serveurs OJS avec l'environnement nécessaire et certainement d'autres briques pour le projet Scientific Edition ;
- un serveur LOCKSS pour l'archivage partagé et l'intégrer dans un réseau international.

## Annexes

### I. MISSIONS

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

#### 1) Réunions du projet européen EuDML

- 03-05/02/2010 – Lisbonne (POR) – Yves LAURENT
- 03-05/02/2010 – Portugal – Thierry BOUCHE
- 03-05/02/2010 – Portugal – Claude GOUTORBE
- 05-07/07/2010 – Paris – Nicolas HOUILLON
- 05-08/07/2010 – Paris – Claude GOUTORBE
- 05-09/07/2010 – Paris – Thierry BOUCHE
- 14-17/10/2010 – Prague (RTCH) – Yves LAURENT
- 13-17/10/2010 – Prague (R.TCH) – Claude GOUTORBE
- 14-17/10/2010 – Prague (R.TCH) – Thierry BOUCHE
- 30/01-04/02/2011 – Madrid (ESP) – Thierry BOUCHE
- 30/01-02/02/2011 – Madrid (ESP) – Claude GOUTORBE
- 30/01-02/02/2011 – Madrid (ESP) – Nicolas HOUILLON
- 08-09/09/2011 – Luxembourg – Thierry BOUCHE
- 11-13/12/2011 – Berlin (ALL) – Thierry BOUCHE
- 11-13/12/2011 – Berlin (ALL) – Claude GOUTORBE
- 20-23/06/2012 – Corfu (GRC) – Franck LONTIN
- 20-23/06/2012 – Corfu (GRC) – Thierry BOUCHE
- 20-23/06/2012 – Corfu (GRC) – Claude GOUTORBE
- 20-23/06/2012 – Corfu (GRC) – Nicolas HOUILLON
- 12-14/12/2012 – Birmingham (ANG) – Thierry BOUCHE
- 12-14/12/2012 – Birmingham (ANG) – Claude GOUTORBE
- 12-14/12/2012 – Birmingham (ANG) – Nicolas HOUILLON
- 13-15/03/2013 – Bruxelles (BEL) – Yves LAURENT
- 13-15/03/2013 – Bruxelles (BEL) – Thierry BOUCHE

#### 2) Réunions du comité de coordination du Zentralblatt Math

- 15/11/2010 – Karlsruhe (ALL) – Yves LAURENT
- 15-16/05/2011 – Heidelberg (ALL) – Thierry BOUCHE
- 17-18/11/2011 – Karlsruhe (ALL) – Thierry BOUCHE
- 06-07/05/2012 – Heidelberg (ALL) – Thierry BOUCHE
- 18-19/11/2012 – Berlin (ALL) – Thierry BOUCHE
- 01-02/12/2013 – Berlin (ALL) – Thierry BOUCHE
- 04-05/05/2014 – Berlin (ALL) – Thierry BOUCHE
- 07-12/12/2014 – Berlin (ALL) – Thierry BOUCHE

#### 3) Projet Portail-Math

- 11/05/2011 – Paris – Hélène FALAVARD

- 04/10/2011 – Lyon – Hélène FALAVARD
- 13/03/2012 – Réunion avec G.HENRY sur l’agenda des maths – Marseille – Catherine ZOPPIS
- 18/12/2012 – Paris – Jean-Luc ARCHIMBAUD
- 18/12/2012 – Paris – Franck LONTIN
- 01-03/07/2013 – Bordeaux – Nicolas HOUILLON
- 28-30/01/2014 – Massy-Palaiseau – Olivier LABBE
- 16-17/03/2014 – Dijon – Catherine ZOPPIS
- 16-17/03/2014 – Dijon – Thierry BOUCHE
- 16-17/03/2014 – Dijon – Hélène FALAVARD
- 24-25/03/2014 – Massy-Palaiseau – Olivier LABBE
- 26-28/05/2014 – Massy-Palaiseau – Olivier LABBE
- 02-03/07/2014 – Marseille – Olivier LABBE
- 25-28/11/2014 – Bordeaux – Olivier LABBE

#### 4) *Collaborations nationales*

- 14/04/2011 – Réunion préparatoire GT5 de la BSN – Paris – Thierry BOUCHE
- 26/04/2011 – Réunion BSN5 et BSN6 – Paris – Thierry BOUCHE
- 27/04/2011 – Réunion RNBM – Paris – Thierry BOUCHE
- 11/05/2011 – Réunion groupe de travail BSN2 – Paris – Thierry BOUCHE
- 16-17/06/2011 – BSN2 CP Mathrice – Paris – Thierry BOUCHE
- 22/06/2011 – Réunion BSN2 – Paris – Thierry BOUCHE
- 05/10/2011 – Réunion BSN2 – Paris – Thierry BOUCHE
- 12/01/2012 – Réunion Cedram SMF – Paris – Thierry BOUCHE
- 16-17/01/2012 – Réunion RNBM – Paris – Thierry BOUCHE
- 22/03/2012 – Réunion groupe technique BSN 5 – Paris – Thierry BOUCHE
- 25/04/2012 – Réunion groupes BSN 5 MESR – Paris – Thierry BOUCHE
- 14/06/2012 – Fin 1<sup>ère</sup> phase numérisation bib. MIR – Paris – Catherine ZOPPIS
- 14-15/06/2012 – Numir table ronde SMF – Paris – Thierry BOUCHE
- 20/09/2012 – Réunion ISTEEX – Nancy – Thierry BOUCHE
- 17/12/2012 – Assemblée Générale RNBM – Paris – Jean-Luc ARCHIMBAUD
- 17/12/2012 – Assemblée Générale RNBM – Paris – Thierry BOUCHE
- 04-05/02/2014 – PAP2 + BSN 5 – Paris – Thierry BOUCHE
- 02-03/04/2013 – RNBM PCP + BSN 5 MESR – Paris – Thierry BOUCHE
- 18-19/06/2013 – Comité de pilotage PCPmath – Paris – Thierry BOUCHE
- 04/07/2013 – Rencontre projet de numérisation BSN5 – Lyon – Thierry BOUCHE
- 14/03/2014 – Réunion PCPmath – Paris – Jean-Luc ARCHIMBAUD
- 17-20/03/2014 – Réunion PAP2 INSMI – Paris – Thierry BOUCHE
- 19-20/03/2014 – Réunion BSN 6 – Paris – Jean-Luc ARCHIMBAUD
- 10/07/2014 – Commission d’évaluation BSN5 – Paris – Thierry BOUCHE
- 06/10/2014 – Commission d’évaluation BSN5 – Paris – Thierry BOUCHE
- 15/05/2014 – Réunion PAP2 (C.SORGER, M.DACOS) – Paris – Thierry BOUCHE
- 21/10/2014 – Réunion BSN 6 – Paris – Jean-Luc ARCHIMBAUD

#### 5) *Collaborations internationales*

- 09/07/2010 – Conseil de la SME – Sofia (BUL) – Thierry BOUCHE
- 12-18/09/2011 – Projet « modernisation des bib de maths » - Tunis (TUN) – Franck LONTIN
- 04-09/05/2011 – Réunion CEIC – New York (USA) – Thierry BOUCHE
- 18-20/07/2012 – Réunion CEIC – Berlin (ALL) – Thierry BOUCHE
- 15-16/07/2013 – Réunion CEIC – Bath (UK) – Thierry BOUCHE
- 02-04-06/2014 – Assemblée Générale EuDML initiative – Varsovie (PLG) – Thierry BOUCHE
- 12-13/07/2014 – Réunion CEIC – Berlin (ALL) – Thierry BOUCHE
- 28-30/06/2014 – EMS Council – San Sebastian (ESP) – Thierry BOUCHE

#### 6) *Collaboration avec la Bibliothèque Nationale de France*

- 22/03/2010 – Lyon – Thierry BOUCHE
- 04/02/2011 – Paris – Hélène FALAVARD
- 25/11/2011 – Paris – Hélène FALAVARD
- 25/11/2011 – Paris – Thierry BOUCHE
- 24/09/2014 – Paris – Sophie MIN-PICAULT
- 02/10/2014 – Montpellier – Sophie MIN-PICAULT

#### 7) *Congrès*

- 07-08/04/2011 – 10 ans Isidora – Eveux – Hélène FALAVARD
- 18-22/07/2011 – Conference CICM 2011 – Bertinoro (ITA) – Thierry BOUCHE
- 19-21/07/2011 – Conference CICM 2011 – Bertinoro (ITA) – Yves Laurent
- 19-22/07/2011 – Conference CICM 2011 – Bertinoro (ITA) – Nicolas HOUILLON
- 31/05-04/06/2012 – Colloque WDML IMU – Washington (USA) – Thierry BOUCHE
- 22/11/2012 – Conférence MEDOANET – Villeurbanne – Thierry BOUCHE
- 10-12/07/2013 – Conférence CICM – Bath (ANG) – Thierry BOUCHE
- 14-22/08/2014 – ICM 2014 – Séoul (CDS) – Thierry BOUCHE
- 04-05/11/2014 – Colloque DIST – Paris – Sophie MIN-PICAULT
- 03-05/11/2014 – Colloque DIST – Paris – Thierry BOUCHE
- 25-26/11/2014 – Congrès CARIST – Nancy – Sophie MIN-PICAULT

#### 8) *Journées en France*

- 28/01/2010 – Réunion INSMI – Paris – Yves LAURENT
- 28/01/2010 – Réunion documentation CNRS PMA – Paris – Thierry BOUCHE
- 29/03/2010 – Réunion INSMI – Paris – Yves LAURENT
- 07-09/07/2010 – Réunion INSMI – Paris – Yves LAURENT
- 12-14/10/2010 – ANGD « documentation mathématique » - Marseille – Nicolas HOUILLON
- 20/01/2011 – INSMI Journée Directeur Unité – Paris – Thierry BOUCHE
- 07-09/02/2011 – Séminaire national nouveaux DU – Meudon – Thierry BOUCHE
- 20/03/2011 – Séminaire national nouveau DU – Meudon – Thierry BOUCHE
- 10-12/10/2011 – Journée Frédoc – Bordeaux – Thierry BOUCHE
- 10-13/10/2011 – Journée Frédoc – Bordeaux – Hélène FALAVARD

- 01/06/2013 – Journée GUTenberg – Paris – Thierry BOUCHE
- 08/10/2013 – Journée 30 ans SMAI – Paris – Thierry BOUCHE
- 12-13/02/2014 – INSMI Journée Directeur Unité – Paris – Thierry BOUCHE
- 12/06/2014 – INSMI Journée Gestionnaire – Paris – Hanane AYADI

### 9) Formations

- 11/06/2010 – « La numérisation des collections » - Paris – Catherine ZOPPIS
- 15/10/2010 – « La numérisation des collections » - Paris – Catherine ZOPPIS
- 30/11/2010 – « La numérisation des collections » - Paris – Catherine ZOPPIS
- 24/01/2011 – Journée Fédération RENATER – Paris – Franck LONTIN
- 15-17/03/2011 – Journée Mathrice – Dijon – Franck LONTIN
- 14-15/03/2011 – Formation URFIST – Lyon – Hélène FALAVARD
- 04-06/10/2011 – Journée Mathrice – Lyon – Franck LONTIN
- 09-11/10/2012 – Journée Mathrice – Orléans – Jean-Luc ARCHIMBAUD
- 09-11/10/2012 – Journée Mathrice – Orléans – Franck LONTIN
- 26-28/03/2013 – Journée Mathrice – Caen – Jean-Luc ARCHIMBAUD
- 26-28/03/2013 – Journée Mathrice – Caen – Frank LONTIN
- 19/09/2013 – Réunion Drupal – Lyon – Jean-Luc ARCHIMBAUD
- 30/09-04/10/2013 – Journée Mathrice – Rennes – Franck LONTIN
- 01-04/10/2013 – Journée Mathrice – Rennes – Jean-Luc ARCHIMBAUD
- 14/11/2013 –Drupal – Lyon – Jean-Luc ARCHIMBAUD
- 23/01/2014 – Drupal – Lyon – Jean-Luc ARCHIMBAUD
- 07-10/01/2014 – Drupal – Paris – Sophie MIN-PICAULT
- 20/02/2014 – Drupal – Lyon – Jean-Luc ARCHIMBAUD
- 19-21/05/2014 – Journée ABES – Montpellier – Sophie MIN-PICAULT

### 10) Invité

- 23-25/11/2014 – Journée INSMI sur les publications – Grenoble – Nicolas WICKER (Lille)

## II. PUBLICATIONS

Ci-dessous, les articles parus dans des revues à comité de lecture. Les autres publications sont des présentations et des rapports qui se trouvent sur le site web de Mathdoc dans la rubrique Publications.

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.) **3**#3 (May 2010), p. 227-241.

Thierry Bouche. « Introducing EuDML—The European Digital Mathematics Library »

*EMS newsletter* **76** (June 2010) 11-16.

José Borbinha, Thierry Bouche, Aleksander Nowiński, Petr Sojka & Wojtek Sylwestrzak. « Towards the European Mathematics Digital Library »

in *Towards a Digital Mathematics Library*, proceedings of the DML 2010 workshop, CNAM, Paris, France, July 7-8 2010 (Petr Sojka ed.): Masaryk University, Brno, 2010, p. 11-26.

Thierry Bouche, Claude Goutorbe, Jean-Paul Jorda & Michael Jost. « The EuDML Metadata Schema, version 1.0 »

in *Towards a Digital Mathematics Library*, proceedings of the DML 2011 workshop, Bertinoro, Italy, July 20-21, 2011 (Thierry Bouche & Petr Sojka ed.) : Masaryk University, Brno, 2011, p. 45-61.

José Borbinha, Thierry Bouche, Aleksander Nowinski & Petr Sojka. « Project EuDML—A First Year Demonstration »

in *Intelligent Computer Mathematics*, proceedings of the CICM 2011 conferences, Bertinoro, Italy, July 18-23, 2011 (James H. Davenport, William M. Farmer, Josef Urban, Florian Rabe eds). Lecture Notes in Computer Science, 2011, Volume 6824/2011, p. 281-284.

Thierry Bouche, Olaf Teschke & Krzysztof Wojciechowski. « Time Lag in Mathematical References »  
*EMS newsletter* 86 (December 2012) 54-55.

Thierry Bouche. « Le projet EuDML, une bibliothèque numérique européenne de mathématiques »  
In *L'information scientifique et technique au prisme de l'Europe*, Marielle Jézéquel (coord.)  
actes des journées Frédoc 2011 (Bordeaux). Centre Emile Durkheim (2013), p. 47-56.

Thierry Bouche. « Reviving the Free Public Scientific Library in the Digital Age? The EuDML project »

in *Topics and Issues in Electronic Publishing*, proceedings of the AMS Special Session on Topics and Issues in Electronic Publishing at 2013 Joint Mathematics Meetings, 910 January 2013, San Diego, USA. Klaus Kaiser, Steven Krantz, Bernd Wegner (Ed.), FIZ Karlsruhe (2013), p. 57-80.

Thierry Bouche & Jiri Rakosnik « Report on the EuDML external cooperation model »

in *Topics and Issues in Electronic Publishing*, proceedings of the AMS Special Session on Topics and Issues in Electronic Publishing at 2013 Joint Mathematics Meetings, 910 January 2013, San Diego, USA. Klaus Kaiser, Steven Krantz, Bernd Wegner (Ed.), FIZ Karlsruhe (2013), p. 99-108.

Thierry Bouche. « Update on Recent Global Digital Mathematics Library (GDML) Developments »

*EMS Newsletter* No. 94, December 2014, p. 41

« The Digital Mathematics Library as of 2014 »

*Notices of the AMS* 61(9) (October 2014) p. 1085-1088

## Glossaire

- AMS** American Mathematical Society <http://www.ams.org/>
- BASE** Bielefeld Academic Search Engine <http://www.base-search.net/about/en/>
- BNF** Bibliothèque nationale de France <http://www.bnf.fr/>
- CEIC** Committee on Electronic Information and Communication (IMU)
- CIP** Competitiveness and Innovation framework Programme (EU)
- CCSD** Centre pour la communication scientifique directe (UMS 3668) <http://ccsd.cnrs.fr/>
- Cedram** Centre de diffusion de revues académiques mathématiques <http://www.cedram.org/>
- CNRS** Centre National de la Recherche Scientifique <http://www.cnrs.fr/>
- DIST** Direction de l'information Scientifique et Technique du CNRS <http://www.cnrs.fr/dist/>
- DTD** (XML) Document Type Definition
- DML** Digital Mathematics Library <http://www.mathunion.org/ceic/wdml/>
- EDBM** European database manager for mathematics
- EMS/SME** European Mathematical Society <http://www.euro-math-soc.eu/>
- EuDML** The European Digital Mathematics Library <https://eudml.org/>
- FIZ Karlsruhe** FachInformationsZentrum Karlsruhe <http://www.fiz-karlsruhe.de/>
- GDML** Global Digital Mathematics Library
- GDZ** Göttinger DigitalisierungsZentrum <http://gdz.sub.uni-goettingen.de/gdz/>
- HAL** Hyper article en ligne <https://hal.archives-ouvertes.fr/>
- IMU/UMI** International Mathematical Union <http://www.mathunion.org/>
- INIST** Institut de l'information scientifique et technique (UPS 76) <http://www.inist.fr/>
- INSMI** Institut national des sciences mathématiques et de leurs interactions (CNRS)
- JFM** Jahrbuch über die Fortschritte der Mathematik
- JSTOR** Journal Storage
- LINUM** Livres numérisés mathématiques
- Mathrice** GDS 2754 <http://www.mathrice.org/>
- MESR** Ministère de l'Enseignement supérieur et de la Recherche (avant 2012)
- MENESR** Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche
- MR** Mathematical Reviews
- MSN** MathSciNet: Mathematical Reviews on the web <http://www.ams.org/mathscinet/>
- Numdam** Numérisation de documents anciens mathématiques <http://www.numdam.org/>
- OAI-PMH** Open Archives Initiative—Protocol for Metadata Harvesting
- OSUG** Observatoire des Sciences de l'Univers de Grenoble (structure fédérative)
- PCRD/FP** Programme cadre de recherche et développement (Framework Programme, EU)
- RNBM** Réseau national des bibliothèques de mathématiques (GDS 2755) <http://www.rnbn.org/>
- SMAI** Société de mathématiques appliquées et industrielles <http://smi.emath.fr/>
- SME/EMS** Société mathématique européenne <http://www.euro-math-soc.eu/>
- SMF** Société mathématique de France <http://smf.emath.fr/>
- SUDOC** Système universitaire de documentation <http://www.sudoc.abes.fr/>
- UJF** Université Joseph Fourier Grenoble 1 <https://www.ujf-grenoble.fr/>
- WDML** World Digital Mathematics Library
- ZM** zbMATH (était : Zentralblatt Math) <http://zbmath.org/>