

*Les projets de bibliothèques numériques  
de référence en mathématiques :  
un état des lieux*

Thierry Bouche

Cellule MathDoc & institut Fourier,  
Université de Grenoble

*Histoire des journaux mathématiques : problématiques et méthodes*  
CIRM, Luminy, 7-11 septembre 2009

- 1 La documentation mathématique
- 2 La documentation électronique mathématique
- 3 La bibliothèque numérique de mathématiques (WDML)
- 4 Le projet EuDML
- 5 Les services de la Cellule MathDoc
- 6 Quelques références bibliographiques

# La documentation en mathématiques

## Enjeux spécifiques

- La documentation mathématique *validée* ne se périmé pas (Euler 1999)
- Les résultats anciens ne sont pas remplacés par les nouveaux : ils sont leur fondation (Richelot 2004)
- Elle est valide comme un *tout*, qui forme un vaste réseau (Corona bug)
- Elle est utile pour d'autres sciences, de façon *asynchrone* (Weber crypto)
- Elle doit donc être soigneusement validée, rangée, indexée et conservée (GDZ Spr. Zbl MR)
- Elle doit rester accessible sur le très long terme (Galois 1828)

# La documentation en mathématiques

## La bibliothèque de référence

Nous avons donc besoin d'une bibliothèque

- exhaustive
- à jour
- bien rangée
- grande ouverte
- facile d'accès pour les non-mathématiciens

Papier OK ? (bibliothèques, prêt inter., fourniture de documents, catalogues fusionnés, bases de données MR/ZM. . .)

Électronique Un rêve . . . (WDML)

⇒ EuDML (2010-2012) projet pilote des centres de numérisation publics en Europe + quelques éditeurs et ZM

⇒ EVLM (2011-2014) ? infrastructure de recherche européenne intégrée avec le contenu de tous les éditeurs commerciaux

# La documentation en mathématiques

## La bibliothèque de référence

Nous avons donc besoin d'une bibliothèque

- exhaustive
- à jour
- bien rangée
- grande ouverte
- facile d'accès pour les non-mathématiciens

**Papier** OK ? (bibliothèques, prêt inter., fourniture de documents, catalogues fusionnés, bases de données MR/ZM. . .)

**Électronique** Un rêve. . . (WDML)

⇒ EuDML (2010-2012) projet pilote des centres de numérisation publics en Europe + quelques éditeurs et ZM

⇒ EVLM (2011-2014) ? infrastructure de recherche européenne intégrée avec le contenu de tous les éditeurs commerciaux

# La documentation en mathématiques

## La bibliothèque de référence

Nous avons donc besoin d'une bibliothèque

- exhaustive
- à jour
- bien rangée
- grande ouverte
- facile d'accès pour les non-mathématiciens

**Papier** OK ? (bibliothèques, prêt inter., fourniture de documents, catalogues fusionnés, bases de données MR/ZM. . .)

**Électronique** Un rêve . . . (WDML)

⇒ **EuDML (2010-2012)** projet pilote des centres de numérisation publics en Europe + quelques éditeurs et ZM

⇒ **EVLN (2011-2014) ?** infrastructure de recherche européenne intégrée avec le contenu de tous les éditeurs commerciaux

# La documentation en mathématiques

## Échelle de temps

- Prépublications instantanées (labos, arXiv, courriel, pages perso)
- Délais de publication assez longs : 1-2 ans
- Publication à fins de prestige, carrière et d'attribution  
Fournit une version de référence pour les travaux à venir
- Environ 50 % des articles cités aujourd'hui  
sont parus il y a moins de 10 ans
- Environ 25 % des articles cités aujourd'hui  
sont parus il y a plus de 20 ans

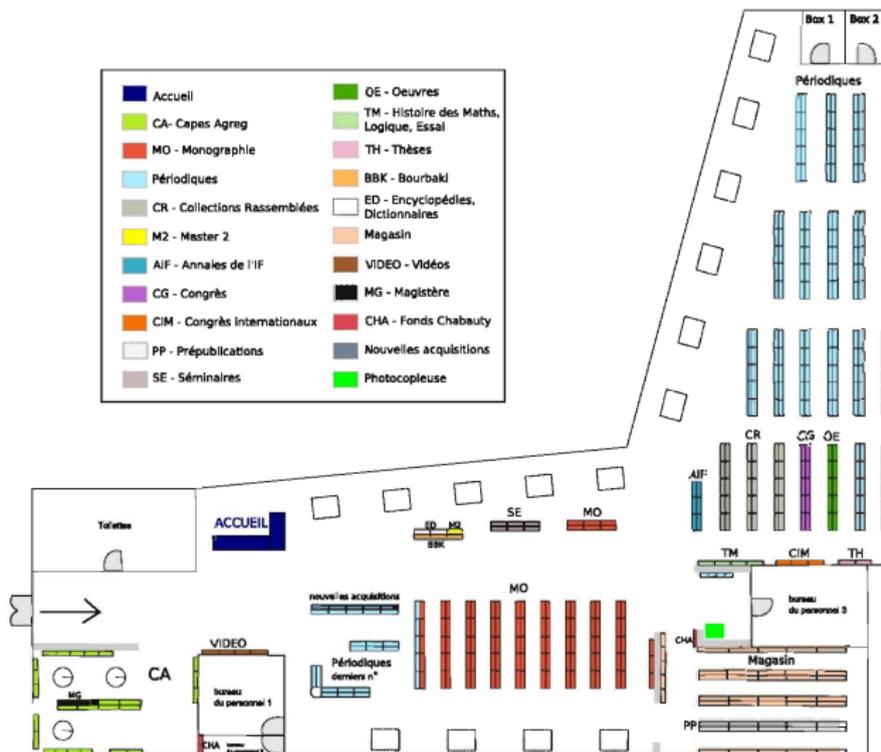
# La documentation en mathématiques

## En chiffres

Une estimation de la taille du corpus mathématique publié dans la tradition occidentale depuis Euclide :

- 3 millions de textes couvrant < 100 millions de pages
- 100 000 nouveaux textes paraissent chaque année
- < 20% parus avant 1900
- > 50% parus après 1950
- 80% articles de revues, 10% chapitres dans des ouvrages collectifs, 10% livres
- 600 journaux vivants dédiés à la recherche mathématique, 2000 journaux comportant des maths
- 15 millions pages numérisées ? (Rehmann WDML)  
65% des journaux « importants » disponibles numériquement ? (DMR)

# Une bibliothèque de mathématiques typique (Institut Fourier, France)



# E-documentation mathématique

## Souhails

- Passer à l'électronique *devrait* être un atout formidable pour ouvrir de nouvelles voies de fouille dans le corpus mathématique et d'accès à celui-ci
- L'infrastructure nécessaire fournirait les fonctions de base de la bibliothèque de référence, enrichie des possibilités offertes par le format numérique
- Soit :
  - Une archive globale (distribuée) stockant ce qui paraît dans le monde au fur et à mesure (par numérisation ou nativement numérique)
  - Un registre à jour de toutes les ressources disponibles
  - Des outils pour résoudre des références ou croiser les contenus de différentes bases de données
  - Une navigation sans frontières dans la totalité du corpus

# *E-documentation mathématique*

## *Les inconvénients du numérique*

Le numérique n'a pas que des avantages pour les chercheurs et les bibliothécaires. . .

- Les grandes plateformes d'édition et les outils courants ne sont pas adaptés au contenu mathématique
- Un babel de « standards » pour la structuration des données, les interfaces utilisateurs, les formats de fichier, etc.
- Des nouvelles barrières d'accès apparaissent (copyright, licences DRM. . . )
- Les coûts augmentent
- On pense pouvoir mesurer la « valeur » d'une publication à partir de chiffres faciles à produire mais dont la signification reste à déterminer (nombre de téléchargements, nombre de liens ou de citations, « impact », etc.)

# E-documentation mathématique

## Désordres

- De nombreux textes parus sur papier n'ont pas d'équivalent numérique, *mais*
  - de nombreux textes numériques sont dupliqués sur plusieurs sites (dans des versions différentes ou non), *tandis que*
  - de nombreuses collections sont découpées entre plusieurs sites, *et*
  - les sites disparaissent ou changent d'adresse, de propriétaire ou de politique commerciale comme de chemise
- ⇒ Gérer un accès exhaustif et à jour au corpus numérique dans son ensemble nécessite des ressources financières et humaines infinies

# *La bibliothèque numérique de mathématiques*

## *Qu'est-ce ?*

- Une liste de documents (REHMANN, UW MATH LIBR.)
  - numérisés ?
  - nativement numériques ?
  - scientifiquement validés ?
  - formellement édités ?
- Une base de données ? (MINI-DML)
- Une liste de bases de données ? (DMR)
- Des rayonnages virtuels ? (MATHGUIDE)
- Une base de données de banques de données ? (EuDML ?)
- Une liste de DML nationales ? (DMLs)

# *La bibliothèque numérique de mathématiques*

## *Risques*

- Une désintermédiation dans l'édition avec perte de qualité (auto-archivage et autres modalités d'auto-édition : pages perso, « personal collected works » . . .), de fiabilité (qui valide ? qui relit ?) et de pérennité
- Une privatisation de la fonction de bibliothèque (projets de numérisation d'Elsevier, Springer, mais aussi bibliothèques numériques universitaires « opérées » par NUMILOG . . .)
- Une très grande visibilité des catalogues, mais un accès aux contenus réservé aux plus riches (Google Book, Google scholar, pay-per-view, etc.)

## Projets WDML

### *Vision (≈ Cornell, 2002)*

« A reference digital mathematics library should assemble as much as possible of the digital mathematical corpus in order to

- **preserve** it over the long term,
- make it **available online**
- at **reasonable cost**,
- in the form of an **authoritative** and **enduring** digital collection,
- **growing** continuously with publisher supplied new content,
- **augmented** with sophisticated search interfaces and interoperability services,
- developed and curated by a network of **institutions** »

# Projets WDML

## Historique

- John Ewing. “Twenty Centuries of Mathematics : Digitizing and Disseminating the Past Mathematical Literature”. *Notices of the AMS*, 49(7) :771–777, August 2002.
- Digital Mathematics Library. NSF planning project (2002-2003, Cornell University Library) “toward the establishment of a comprehensive, international, distributed collection of digital information and published knowledge in mathematics”.
- Mathematical Knowledge Management meetings (2001–) + DML workshops (2008–) : technical challenges.
- EMS’ EoI to the European Commission (2003), supported pilot implementation proposals to EC programmes (2003–2008 : FP6, eContentplus. . .)
- AMS/MSRI proposal to the Moore foundation (2005)
- IMU support (2002–2006 : Vision, Best practices)
- EuDML submitted (06/2009), entering negotiation (08/2009)

# Projets WDML Collections

**Ameriques** JSTOR (260,000 items), project Euclid (100,000), CMS (4,000)

**Asie** DML-JP (30,000 items), China ??

**Europe** EuDML ? (190,000 items)

**Allemagne** ERAM/JFM, GDZ, ELibM (85,000 items)

**France** Gallica-Math, NUMDAM, CEDRAM, TEL (50,000 items)

**Pologne** ICM/BWM (13,000 items)

**Portugal** SPM/BNP (2,000 items)

**Espagne** DML-E (5,000 items)

**Rép. Tchèque** DML-CZ (11,000 items)

**Russie** RusDML (13,000 items)

**Bulgarie** BulDML (2,500 items)

**Serbie** No formalised project (3,700 items)

**Suisse** SwissDML (5,000 items)

**Commercial** 700,000 items ?

**PME** CUP 20 journals, OUP 30, Hindawi 18, WdG 13, Wiley 42, T&F 58. . .

**Elsevier** 4 journals in NUMDAM, 63 in Backfiles, 100 alive (320,000 items)

**Springer** 14 journals in GDZ, 1+2 in NUMDAM, 120 in Online Archives,  
179 alive (300,000 items)

## Projets WDML

### Défis politiques

- Renoncer à la vision initiale, pyramidale et centralisée
- Commencer petit en identifiant des partenaires motivés mais représentant une diversité suffisante pour anticiper l'avenir (cf. EuDML ?)
- Trouver un modèle économique assez souple pour permettre à toutes les catégories d'éditeurs de s'y retrouver (créneau mobile ?)
- Mais sans y perdre son âme !

## Projets WDML

### Défis technologiques

- Trouver un noyau commun de métadonnées qui permette des fonctionnalités riches sans écarter trop de collections existantes
- Développer des chaînes de création automatique de métadonnées pour les collections trop dépouillées (GDZ, DML-CZ)
- Mettre en relation toutes les informations disponibles sur un texte donné
- Contourner le multilinguisme des collections (et le monolinguisme des métadonnées)
- Dépasser les limites des formats graphiques : fournir un accès au contenu scientifique plutôt qu'à des images
- Utiliser le contenu mathématique comme un atout et non comme une punition !



## The European Digital Mathematics Library

### CIP-ICT-PSP.2009.2.4 Open access to scientific information

**Consortium** 13 participants européens (IST, CMD, UB, FIZ/ZM, MU, ICM, CSIC, EDPS, USC, IMI-BAS, IMAS, CSIU, MML)

3 partenaires associés (EMS, SUBG, RusDML)

**Objectifs** Implémentation pilote (orientée utilisateur final) d'un guichet d'accès unique au contenu mathématique fourni par 11 partenaires, avec des fonctions innovantes de recherche, accessibilité, multilinguisme, navigation et interactivité

**Profil** 3 années (2010-2012),

488 PM, coût global : 3,20 M€ (financé pour moitié par la CE).

**Contenu** 170,000 items ; 2 600 000 pages

**Rétronumérisé** NUMDAM, Gallica, BWM, GDZ, SPM/BNP, DML-CZ, DML-E, RusDML.

**Numérique natif** CEDRAM, ICM, EDPS, ELibM, DML-CZ, DML-E



# The European Digital Mathematics Library

## Work plan

- WP1 Project management
- WP2 Policies, exploitation, and dissemination
- WP3 Content aggregation
- WP4 System architecture and design
- WP5 Metadata repository and search engine implementation
- WP6 Web and service interface implementation
- WP7 Metadata enhancer implementation
- WP8 Association analyser implementation
- WP9 Annotation component implementation
- WP10 Accessibility component implementation
- WP11 Assessment and evaluation.

# Cellule MathDoc

## Missions

- La *Cellule de coordination documentaire nationale pour les mathématiques* (MathDoc) est une unité mixte de services CNRS-UJF (UMS 5638) installée à Grenoble depuis 1995
- **Missions**
  - 1995** coordination nationale dans le domaine de la documentation mathématique
    - support et veille technologique (bases de données, documentation électronique)
    - Collaboration avec le *Zentralblatt-MATH*
  - 2000** numérisation des principales revues françaises
  - 2005** aide à l'édition des revues académiques de mathématiques

# Cellule MathDoc

## Actions

- Petite équipe formée d'informaticiens et de documentalistes, dirigée ou animée par des mathématiciens
- **Principales actions**
  - 1996** CFPM : catalogue fusionné des périodiques
  - 1997** EDBM : logiciel d'interrogation en ligne du *Zentralblatt-MATH*
  - 2002** NUMDAM numérisation pilote de 5 revues
  - 2005** CEDRAM, v. 1 : production de 3 revues
    - LiNuM, mini-DML, Gallica-Math ;
  - 2007** CEDRAM, v. 2 : 5 séries, MathML/ $\text{\TeX}$ .

# NUMDAM

## Un survol

*« Programme de numérisation pour archivage et diffusion  
du fonds des revues académiques de mathématiques »*

**Archivage :** Scan intégral de la première à la dernière page de chaque volume  
(TIFF 600 dpi)

**Diffusion :** Un fichier multipage par article, en accès libre après un délai variable  
selon l'éditeur

Recherche et de navigation libres sur  
[www.numdam.org](http://www.numdam.org)

# NUMDAM Collections

## 2000-2004 Phase I

6 séries (4 revues les plus importantes en mathématiques pures, une série de mémoires, une série d'actes de conférences)

**7 500 articles, 210 000 p.**

## 2004-2007 Phase II

17 revues (couvrant un large spectre de la physique mathématique aux statistiques, et de Nîmes en 1810 à Amsterdam, Pise au xx<sup>e</sup> siècle. . . )

28 séminaires (dont les séminaires Bourbaki [I.H.P.], Cartan [É.N.S.], Leray [Collège de France], . . . )

**+ 20 000 articles, + 375 000 p.**

## 2008-2010 Phase III

Journaux anciens (NAM, BSMA), thèses de l'entre-deux-guerres, journaux français de math. appliquées et de statistiques, monographies. . .

**+ 25 000 articles, + 300 000 p. ?**

# NUMDAM

## *Orientations*

- L'objectif premier était de doter les revues académiques de mathématiques (vivantes, de qualité) d'archives pérennes, visibles et faciles d'accès
- Le service est avant tout conçu comme un outil pour la recherche contemporaine en mathématiques
- Si nous numérisons toutes les pages, nous ne diffusons pour l'instant que les articles enregistrés comme tels dans notre base de données (selon une typologie un peu fruste)
- NUMDAM est aussi un réservoir d'articles dotés d'identifiants uniques : il est facile de construire des sites dédiés à des parcours ou des accès enrichis de certaines collections
- Les données produites par le programme sont sans doute très utiles pour d'autres disciplines de recherche (*utilisateurs* de mathématiques ou historiens) : ces communautés sont appelées à se manifester pour obtenir les fonctions qui leur seraient utiles !

# NUMDAM

## Coups de projecteurs

Certaines collections ont des propriétés spécifiques

- Le séminaire de probabilités de Strasbourg est lié avec une base de résumés spécifiques qui préexistait à la numérisation (SPS)
- Quand deux articles se suivent sur une même page, nous « nettoyons » les fichiers associé à chacun (BSMF, Gergonne, NAM)
- Certains journaux publient des tronçons d'articles : nous les regroupons ! (AFST)
- Table des matières imprimée pour les séminaires (GAU)
- Nous étudions la possibilité de fournir les pages numérisées mais non diffusées aujourd'hui
  - Soit en ajoutant au sommaire les pages préliminaires et finales (BSMF)
  - Soit en proposant le volume complet (planches et hors-texte en gris ou couleur le cas échéant) (AMPA, NAM)
- À noter un travail sur les liens entre problèmes posés et solutions dans les *Nouvelles annales de mathématiques* (NAM)

# CEDRAM

## Un survol

« *Centre de diffusion de revues académiques de mathématiques* »

**Soutien** : Ensemble modulaire d'outils d'aide à la production et à la gestion de publications de recherche en mathématiques

**Hébergement** : Édition électronique de qualité supérieure à NUMDAM

**Archivage** Versement des fichiers d'exploitation dans NUMDAM pour accès à plus long terme, des fichiers de production dans une archive pérenne

# CEDRAM Collections

- *Annales de la faculté des sciences de Toulouse, mathématiques* (2003- )
- *Annales de l'institut Fourier* (Grenoble : 2001- )
- *Annales mathématiques Blaise Pascal* (Clermont-Ferrand : 2003- )
- *Journal de théorie des nombres de Bordeaux* (2004- )
- *MathS In Action* (SMAI : 2008- )
- **Actes des rencontres du CIRM** (2009- )
- Séminaires : JEDP (Evian : 2001- ), X-EDP (Paris : 1997- ), TSG (Grenoble : 2005- )

## Fonds spéciaux

**Archives Nicolas Bourbaki** documents privés de l'association, catalogués et numérisés par l'unité « Archives de la création mathématique ». En ligne sur <http://portail.mathdoc.fr/archives-bourbaki/>

**Archives Laurent Schwartz** documents personnels versés à la bibliothèque de Polytechnique (projet).

**Travaux de Laurent Schwartz** numérisés en vue de l'édition d'un livre avec CD-ROM (SMF/X/CMD).

**Videos** conférences à l'EHESS (projet)

## Moteurs de recherche

- Gallica-Math** Catalogage au niveau article de volumes numérisés par Gallica (Gallica-Math)
- Œuvres complètes de nombreux mathématiciens (OC)
  - *Journal de mathématiques pures et appliquées* (1836-1934) (Liouville)
  - *Répertoire bibliographique des sciences mathématiques* (1894-1912) (RBSM)
- LiNum** Catalogue de livres numérisés de mathématiques (Gallica, Michigan U, Cornell, Göttingen, etc.) (LiNuM)
- mini-DML** Catalogue d'articles numériques de mathématiques (mini-DML)

## Quelques références bibliographiques

- Quelques exposés antérieurs sur ma page perso ([Lien](#))
- Je signale également deux articles. Le premier est une présentation plus détaillée de mes propositions pour construire la DML ; le second une étude de ce qui est disponible aujourd'hui du corpus « français », accompagnée de quelques commentaires ([Voir ma page](#)).
  - « Report on the current state of the French DMLs », in *Towards a Digital Mathematics Library*, proceedings of the DML 2009 workshop, Grand Bend, Ontario, Canada, July 8-9 2008 (Petr Sojka ed.) : Masaryk University, Brno, 2009, 10 p. (Preprint)
  - « Digital Mathematics Libraries : The Good, the Bad, the Ugly », to appear in *Mathematics in Computer Science*, special issue on Authoring, Digitalization and Management of Mathematical Knowledge (Serge Autexier, Petr Sojka, and Masakazu Suzuki eds.) 2009, 20 p. (Preprint)