





# Towards a world digital library:

### Mathdoc, Numdam and EuDML experiences

## Thierry Bouche

Cellule Mathdoc, Université Grenoble Alpes

Biblioteche di matematica UMI, La Sapienza Roma, Venerdi 23 settembre 2016









Towards a world digital math library?

## A quotation of Jean-Pierre Serre (according to Michel Broué)

« Les mathématiciens se contentent de mettre leur production à la disposition de tous, comme sur des étagères où l'on peut venir se servir. »

> "Mathematicians just make their results available to everyone as if they were on shelves, waiting to be fetched."

My understanding

- Mathematics has its own agenda
- Once a proof is completed, it's written up, checked, and contributed to the library of all mathematical knowledge so far
- Maybe somewhat later, someone will look up the shelf and do something with that knowledge
- ⇒ The mathematical accumulated knowledge is a scientific commons that should be preserved and made easily accessible for eternity

Towards a world digital math library?

## A quotation of Jean-Pierre Serre (according to Michel Broué)

« Les mathématiciens se contentent de mettre leur production à la disposition de tous, comme sur des étagères où l'on peut venir se servir. »

"Mathematicians just make their results available to everyone as if they were on shelves, waiting to be fetched."

## My understanding

- Mathematics has its own agenda
- Once a proof is completed, it's written up, checked, and contributed to the library of all mathematical knowledge so far
- Maybe somewhat later, someone will look up the shelf and do something with that knowledge
- $\implies$  The mathematical accumulated knowledge is a scientific commons that should be preserved and made easily accessible for eternity



- Libraries have up to now played a central role by providing archiving and preservation of print material
- Highlight of some interesting features:
  - Content selection
  - Permanent acquisition
  - Multiple copies preserved worldwide
- Who will in the future be responsible for acquiring, archiving and preserving the digital mathematical heritage?
- How reliable are the potential candidates?



- Libraries have up to now played a central role by providing archiving and preservation of print material
- Highlight of some interesting features:
  - Content selection
  - Permanent acquisition
  - Multiple copies preserved worldwide
- Who will in the future be responsible for acquiring, archiving and preserving the digital mathematical heritage?
- How reliable are the potential candidates?



- Libraries have not only cared for preservation and archiving: They have provided long-term access for their patrons and, more widely, for the general public
- Highlight of some interesting features:
  - Permanent reading right granted to any visitor, and remote users (ILL!)
  - Mostly supported by public funding and overheads
  - Typically quite nice buildings central to the universities
- Will the mathematical production remain part of our common, freely accessible, heritage or will it be gradually locked by private interests?
- What about the opportunities in our digital world to exchange and consume mathematics through open standards?



- Libraries have not only cared for preservation and archiving: They have provided long-term access for their patrons and, more widely, for the general public
- Highlight of some interesting features:
  - Permanent reading right granted to any visitor, and remote users (ILL!)
  - Mostly supported by public funding and overheads
  - Typically quite nice buildings central to the universities
- Will the mathematical production remain part of our common, freely accessible, heritage or will it be gradually locked by private interests?
- What about the opportunities in our digital world to exchange and consume mathematics through open standards?

Intro: DML	Mathdoc	Portail Math	CEDRAM	NUMDAM	EuDML	GDML
	000	00000	000	00000	0000000	00
				Still	a long way	

A remark from a colleague to whom I showed a preview of the first online posting of Numdam (Dec. 2002, 2 journals):

Why isn't it all there?



A remark from a colleague to whom I showed a preview of the first online posting of Numdam (Dec. 2002, 2 journals):

Why isn't it all there?



Numdam 2003: 2,600 art.



EuDML 2016: 252,000 art.



Numdam 2016: 57,500 art.

DML

DML ready: 1,000,000 art.?

Full corpus: 4,500,000 items?

Towards DML



Accueil » Services documentaires



#### Services documentaires

				Fortalis principaux
Services	Projets	Mots clés	Sites	• Numdam
	NUMDAM	périodiques, numérisation, acquisition numérique, indexation	www.numdam.org	<ul> <li>Cedram</li> <li>Portail math</li> </ul>
	Archives Bourbaki	numérisation, indexation	sites.mathdoc.fr/archives-bourbaki	Coopération europée
	Numir	livres, numérisation	www.numir.org	
Fonds	Publications Mathématiques d'Orsay	numérisation, réindexation	sites.mathdoc.fr/PMO	EuDML initiative
documentaires	Bibliographie de Laurent Schwartz	numérisation	sites.mathdoc.fr/OCLS	Projets nationaux     Numdam 4
	Gallica-Math	oeuvres complètes, réindexation	sites.mathdoc.fr/OEUVRES	<ul> <li>Portail math</li> <li>Scientific edition</li> </ul>
		articles, réindexation	sites.mathdoc.fr/JMPA	
		répertoire bibliographique, restructuration	sites.mathdoc.fr/RBSM	
	Mini-DML	articles	minidml.mathdoc.fr	
Bibliothèques numériques	LiNum	livres	sites.mathdoc.fr/LiNuM	
virtuelles	EuDML	bibliothèque virtuelle européenne	eudml.org	
Edition numérique	Cedram	revues, séminaires	www.cedram.org	
Base de résumés	SemProba	séminaire de probabilités, indexation	sites.mathdoc.fr/SemProba	
Coopération documentaire	Catalogue fusionné des périodiques de mathématique	issn, conservation partagée	cfp.mathdoc.fr	
Annuaires	Annuaire des bibliothèques	mbm	www.portail-math.fr/bibliotheques	
	Annuaire des	insmi	www.portail-math.fr/laboratoires	

9



- The *Cellule de coordination documentaire nationale pour les mathématiques* (Mathdoc) is a joint unit of CNRS-UGA located in Grenoble since 1995
- Small team (~ 8 permanent positions) with
  - 4-5 computer scientists
  - 0-2 information specialist
  - + administrative and temporary staff

lead by mathematicians

• Developing services towards (French, math) librarians and research mathematicians (worldwide)





#### Support for French math libraries

- 1995 CFP: Union catalogue of French math libraries
  - Creation of the search engine and the first Web interface of Zentralblatt MATH
  - A lot of obsolete stuff, including EU projects with Zentralblatt





Digitisation & dissemination of the literature, access portals

- 2000 Digitisation of main French serials
  - $\implies \mathcal{N}$ umdam
- 2005 Technological support for academic publishers  $\implies$  cedram
- 2010 Larger integration projects
  - $\implies$  Gallica-Math mini-DML  $\mathcal{EU}$ DML
- 2012 "Math portal" project launched  $\Rightarrow \boxed{\Phi_{MATH}^{\text{portal}}}$
- 2016 MISO: "Centre Mersenne pour l'information scientifique ouverte"
  - $\implies$  Cedram reloaded (more services, more sciences, more open)

[M]





Bibliothèque numérique Annuaires Agenda Services numériques Sites utiles



Biervenue sur le portail math, site destiné à la communauté mathématique. Ce site vous permet d'accéder aux ressources documentaires ainsi qu'à des services numériques.

#### Agenda des Maths

22	Propriété de Multiplicité un Shimura	à niveau de congruence principal pour les courbes de	Cal
2016	Saint-Martin-d'Hères	🗶 Stefano Morra	
22 SEP 2016	La solitude des groupes Villeurbanne	🗶 Melleray, Julien (UCBL)	ical
22 SEP 2016	Maýlis Limouzineau (Paris) fonctions génératrices Nantes	- Constructions autour des noeuds legendriens et	Kal



Liste Complète 👩

Foire aux questions Contact Mentions légales Le projet





## A one stop shop

- To math literature and associated services (databases, videos, preprint archives and search engines) Open and restricted access
- To other productivity services to ease nomadic and cooperative working
- To some professional informations

Technically

- Single authentication for every mathematician working in France
- A knowledge base on available resources and entitled users

Still in its infancy!





## INSMI

Institut national des sciences mathématiques et de leurs interactions, CNRS 5000 researchers

Mathdoc

UMS CNRS (INSMI)-UGA

RNBM

National math libraries network, GDS CNRS 63 libraries

Mathrice

Network of computer scientists from math labs, GDS CNRS 120 members

• SMF, SMAI, SFdS

Learned societies









cedram.org		
	Accuell Collections Rechercher un article A propos Bolite à outils Contact	
Accueil	Accès aux collections	Nouvellement en ligne
Abornemeria et Locardenaria gendra alea gendra alea Mantena.	Recharche rapide         >> Recharche avancée    En ligne sur CEDRAM : 9 Revues - 1 Livre - 7 Actes et séminaires En ligne sur CEDRAM : 9 Revues - 1 Livre - 7 Actes et séminaires Revues Annales de Infaculté des sciences de Toulouse Annales de Infaculté outer - Rapide - National Actes et séminaires Distribution - National Actes - Nationa	2 nouvelles publications in revue publications in revue Matchematiques de Besançon* et les actes Winter Braids Lecture Notes* Autres sereurs Matboc Cellule Mathdoc Portali mat Nimi-dimi
	Revues de mathématiques soutenues par le CNRS	



## Support for French math academic journals

- Production and dissemination of 9 journals and 6 proceedings series
- 8000 articles, 180000 pages online (including archives from Numdam)
- 200 articles (6 000 pages) produced and published each year All these will be Open access in 2017

New

- CML: Confluentes Mathematici (left World Scientific in 2013)
- JEP: Le journal de l'École polytechnique (since 2014)
- SMAI-JCM: SMAI Journal of Computational Mathematics (since 2015)

One book! (2014)

## cedram



### Current

- Production workflow with journals
- Each journal has a web site with dedicated look & feel
- Creation and maintenance of LATEX styles (using a specific class)
- Production of PDFs for print and web with XML/MathML metadata
- DOI registration (Crossref), reference linking (MSN, ZBM, mini-DML, Crossref) Now also Crosscheck
- OJS for managing refereeing, etc. (3-4 journals)
- ⊮T<sub>E</sub>X typesetting + structuring
- Archiving in NUMDAM





### Soon (Mersenne)

- New platform, new workflow
- Print-on-demand?
- A business model for non-French, non-math journals!

## cedram

Numdam	() III () (III (III () (III) (III) (IIII (IIII () (IIII) (IIII))))))))
Naviguer par •	andredit X AI+ + Q.Germ
Authors (65)  Authors (65) Real-model, Autor (19) Real-model (4) Struther (4) Struther (4) Topold, Autors (4) Veens (6) Veens (6) Veens (6) Veens (6) Veens (6) Veens (7) Tobo 1000 (8) Tobo 1000 (8) Tobo 1000 (9)	B cecula Page indexes Conception and the page indexes developed and the field of the page index Second Conception and Second Advanced Advanced Advanced and the Conception and Second Type of Conception (Page index Page in the Advanced of the Conception and Second Advanced and advanced for the Conception and Second Advanced and advanced advanced andvanced and advanced advanced
1960-1969 (27) C Colections Annual delta Scuola Normale Superiore di Pisa - Casse di Scienze (85) Document types Article (85)	and not all at a given z, one of the case
Classifications	Terrorbiger des geans de Runge Les Sonan des Annal dels Social Tomale Superfore d'Plas - Classe d'Electes: Walnes 17 (1963) no. 4 p. 319 - 322. Pren tente :tembrage des geans de orgen 2 Sana C) asel () que , l'état un aveil de himosphajement compilie (not a éco aveil de Dine), m art (y, -, -, e Spar ) > e. A. Andread el R. Tensambra () orf ansage les game de Runge, (V, -, -, -, -, -, -, -, -, -, -, -, -, -,
	Cons. Starting: Cons. Starting: Annual delisional Jonnes Experiore al Plan-Classed Blocetta: Munice 20 (1960) (no. 2 p. 200 - 200) Perti tatasis: 1, 1, 2, 1, 2, 2, 1, 2, 3, 2, 1, 2, 3, 2, 1, 2, 3, 2, 1, 2, 3, 2, 2, 2, 0, -2, 0, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
¢	Asymptotic score inequalities for periodoconcare manifolds (L) Namescu, George



This all started as a digitisation programme (initial budget from Ministry of research for a pilot of 6 journals  $\rightarrow$  50 series in 6 years)

- Still growing 8 new series 2013-2016, Astérisque scheduled 2017
- Updated 20 % of our content is born digital direct from publishers
- Extending Digitised content from other sources (Gallica/BNF, Orsay...)
- Not only journals Proceedings, Theses, monographs





- Digitisation of 37 serials, 34 seminar proceedings, 4 memoir series **51 000 articles**, **800 000 pages**
- New material post-digitisation from 6 publishers (21 journals) 6 500 articles, 200 000 pages
- Including 2 Italian journals (Pisa, Padova)
- We *might* start an on-demand digitisation service for books or rare documents ("One mathematician, one text") if resources permit





- Full backrun, no editorial choice
- Scan of every page at high resolution (600 dpi) black & white (text), grey or colour when applicable
- Page format reproduced
- XML Unicode metadata (now JATS, with TeX/MathML)
- Clean separation between faithful articles' image and (un)corrected metadata (PDF/DjVu article vs. navigation HTML stuff)
- Bibliographic references captured and linked





- Electronic version under the title's owner control
- Authors are asked to transfer their electronic copyright to them if still alive (preferred: academic institution)
- We contract with the copyright owner to allow access at www.numdam.org
- The publisher can use our digitised files on its own platform
- Clear identification of the originating journal (logo, links, cover page)
- Moving wall agreed by both parties (mean: 5, min: 0, max: 10);
- Free access after some moving wall as a counterpart to public funding and work done by public servants



- Initially we wanted to support academic (independant, society, university or lab ran) journals so that they'd be competitive with more aggressively commercial ones
- The more general goal is to balance publishers' power
- New and old content should be eventually archived and served by an independant organization (public or at least non-profit)
- We try to revive the public library where every knowledge eventually belongs to everyone, and can be accessed without a licence to pay
- We view Numdam as one of the local DMLs that should cover the mathematical corpus as a whole, with similar policies





English (en) ✓ Login | Register | (Why Register?)

Search





The European distributed digital math libray

#### Two ambitious & intertwined objectives

- A one-stop shop for the mathematical corpus
- A cooperation model for a network of institutions and publishers to archive and open access to that corpus



Intro: DML 0000	Mathdoc 000	Portail Math	CEDRAM 000	NUMDAM 00000	EuDML OOOOOO	GDML 00
				The (Eu)	DML Vision	
		8	UDML	4		

The EUROPEAN DIGITAL MATHEMATICS LIBRARY

The Digital Mathematics Library should assemble as much as possible of the digital mathematical corpus in order to

- make it available online
- with eventual open access,
- 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





- Optimistic conjecture!
- Project funded by CIP ICT PSP "Open access to scientific information" 3 years (2/2010-1/2013); budget 1.6 M€; workload: 487 person-month
- Followed-up by the EuDML initiative founded in 2014







- Optimistic conjecture!
- Project funded by CIP ICT PSP "Open access to scientific information" 3 years (2/2010-1/2013); budget 1.6 M€; workload: 487 person-month
- Followed-up by the EuDML initiative founded in 2014



- EMS The European Mathematical Society
  - FIZ FIZ Karlsruhe/zbMATH
- ICM Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw
- UGA Université Grenoble Alpes
  - UB University of Birmingham
- IMI-BAS Institute of Mathematics and Informatics BAS, Sofia
- IM-CAS Institute of Mathematics CAS, Praha
  - IU Ionian University, Corfu
  - SIMAI Società Italiana per la Matematica Applicata e Industriale
    - UMI Unione Matematica Italiana
    - SUB Niedersächsische Staats- und Universitätsbibliothek Göttingen
    - MU Masaryk University, Brno





## The European distributed digital math libray

### Main results

- A critical mass of contents 245 000 texts of all envisionned types
- A partners network *slowly growing*
- A digital library dedicated to math *MathML metadata, math formulæ search, MSC, links to reviewing databases*
- A web site with an unexpected (fresh) look
- A number of tools behind the website some production-ready, others more experimental





## A distributed archive

- EuDML local libraries are public or non-profit institutions with a long-term interest in preservation and delivery of some part of the digital mathematical corpus
- Publishers should provide content through local DMLs unless they are themselves of the kind above

## A content provider needs to agree on these

- The texts in EuDML must have been scientifically validated and formally published.
- EuDML items must be open access after a finite embargo period. Once documents contributed to the library are made open access due to this policy, they cannot revert to closed access later on.
- The digital full text of each item contributed to EuDML must be archived physically at one of the EuDML member institutions. More at https://initiative.eudml.org/policies-annex

Th. Bouche (UGA)

Towards DML



- The initiative is alive and will stay so for at least 5 years
- It is running with pro-bono resources from partners Enough for maintenance, small developments (without planning)
- The content will grow with new collections ingested (agreement from SwissDML, math-net.ru, e.g.)
- Anything more substantial needs specific funding



## "The World Digital Mathematics Library" (WDML)

HOME

## The Global Digital Mathematical Library Working Group (GDML WG)

Leave a reply

At the first ICM in 1897 there was a session under the chairmanship of Peano concerned with questions of how to encode mathematical knowledge; see the article by E. Schröder in the ICM1897 Proceedings.

Indeed it was in connection with such efforts that Peano developed his axioms for the natural numbers. At the 1928 ICM in Bologna there was active discussion of how to provide comprehensive bibliographic resources for mathematics to everyone: see the article by R.C. Archibald in the I<u>CM1928 Proceedings</u>. Now the I<u>MU</u> sees the possibility of realizing the current fream of a Global Digital Mathematical Library or World Digital Mathematical Library.

The adjective digital is important here as it is the new digital technologies that allow better access to the resources of mathematical knowledge than ever before. We are in the presence of a transformative technology and we can capitalize on it to everyone's benefit. I can imagine that in 16th century Europe, or even earlier in <u>14th century Korea</u>, when printing from metal type was a brand-new technology, people saw the possibilities of the new forms of book for the recording and dissemination of knowledge. That they were right we all now know. That sort of opportunity is open to us again now.

The adjective global is important too. We all think the truths of our subject to be global. Independent of focation in this world. We think of ourselves as a world-wide community. This is well demonstrated by our being gathered at the [CM201] from over 120 countries. A GDML can have a global reach as a result of the digital technology methode, particularly the internet. It will be a shared global good. We can hope for global support for the idea and expect that there can be contributions to a GDML from all over the world. It will provide benefits all over the world. The aerliest mathematical artifact some think to be from <u>ishango in Congo</u>, commerorated in a <u>Tm-meter ligh replicar</u> perhaps this GDML can be a help in Africa. But as the <u>leelayati lecture</u> emphasized the main goal has to be to offer solutions to problems that there enous you serve want solved.

IMU President Ingrid Daubechies and Chair Peter Olver of the IMU's Committee on Electronic

	Search
	Scaren
	RECENT POSTS
0	Report of the Panel Discussion at ICM2014
0	The Global Digital Mathematical Library Working Group (GDML WG)
0	Welcome to the World Digital Mathematics Libra
	(WDML) Blog
	RECENT COMMENTS
	ARCHIVES
0	October 2014
0	August 2014
	CATEGORIES
0	GDML WG
0	ICM Panel

## Global Digital Mathematics Library

Activities supervised by the *Committee on Electronic Information and Communication* (CEIC) of the *International Mathematical Union* (IMU)

- The Future World Heritage Digital Mathematics Library: Plans and Prospects Symposium organized by the CEIC June 1-3, 2012, at the National Academy of Sciences, Washington DC
- Developing a 21st Century Global Library for Mathematics Research Committee on Planning a Global Library of the Mathematical Sciences; Board on Mathematical Sciences and Their Applications; Division on Engineering and Physical Sciences; National Research Council (Washington, DC)
- WDML panel, ICM 2014, Seoul
- Business meeting, Seoul Aug. 2014
- ⇒ GDML Working group (9/2014-?) : Thierry Bouche, Bruno Buchberger, Patrick Ion, Michael Kohlhase, Jim Pitman, Olaf Teschke, Stephen Watt, Eric Weisstein





- Coordinate worlwide actions towards a Global Digital Library (understood as an interoperable repository of mathematical knowledge [linked?] open data)
- ⇒ Semantic Representation of Mathematical Knowledge Workshop (Fields inst. Toronto, Feb. 2016)
  - Create an umbrella organization (such as W3C, UNESCO: IMKT or IMPAKT?) to
    - Establish standards for formats, interoperability, exchange, etc.
    - Foster, label projects
    - Hunt for funding
    - Distribute work over continents



## The International Mathematical Knowledge Trust

International Mathematical Digital Library

International Mathematical Knowledge Base

International Mathematical Knowledge Engine

Semantic Representation of Mathematical Knowledge Workshop

Under Construction: This is a place-holder page for the IMKT, forming under the aegis of the IMU working group for digitization of the world's mathematics. Nothing on this mock-up page should be taken as an indication of consent or commitment of any individual or organization.

and a standard and any and there are a standard and the s

When the Property of the Party of the Party

States 13. Octometer

and a summer of a garget and good

The second se

-6-53

and in

The second states and the second second

UNPAS LE

In construction of

a strangener finder og

and the particular



## Thanks !

## Thierry BOUCHE

## Cellule Mathdoc (UMS 5638 CNRS/UJF)

Domaine Universitaire Bâtiment Institut Fourier 100 rue des Maths 38400 Saint-Martin-d'Hères

#### www.mathdoc.fr