



*Mathdoc*



# *Work resumed on the Tower of Babel: The Mathdoc Digital Mathematics Library...*

Alexandre Bouquet, **Thierry Bouche**

Cellule Mathdoc, Université Grenoble Alpes

*CICM 2019—WiP*

Prague

July 10th, 2019

NUMDAM

MDML

*Eu*DML

# Part I

## *A recap on the Digital Mathematics Library*

## *Intro: A mathematician's dream*

### *The original vision*

“ In light of mathematicians' reliance on their discipline's rich published heritage and the key role of mathematics in enabling other scientific disciplines, the Digital Mathematics Library strives to make the entirety of past mathematics scholarship available online, at reasonable cost, in the form of an authoritative and enduring digital collection, developed and curated by a network of institutions. ”

(Cornell NSF project 2002,  
endorsed by IMU 2006)

## Towards a world digital math library?

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

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

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

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? ””

## Towards a world digital math library?

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

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

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

**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?””

*Full corpus: 4,500,000 items?*



Numdam 2003: 2,600 items



EuDML 2019: 265,000 items



MDML 2019: 600,000+ items  
(including preprints)



mini-DML 2015: 400,000 items  
(including preprints)



Numdam 2019: 60,000 items

# DML

DML ready: 1,500,000 items?

# The Digital Mathematics Library

## A Timeline (1/2)

- 2001 Emergence of “local DMLs” (GDZ, ERAM, Numdam. . . )
- 2002 John Ewing. “Twenty Centuries of Mathematics: Digitizing and Disseminating the Past Mathematical Literature”. *Notices of the AMS*, 49(7):771–777, August 2002.
- 2001– Mathematical Knowledge Management meetings
  - Reviewing databases start adding links to full texts: Is this the way to go?
- 2002-2003 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”.
- 2003–2010 EMS’ EoI to the European Commission (2003), support of pilot implementation proposals to EC or ESF programmes (FP6, eContentplus, Forward look. . . )
- 2004 **mini-DML**: a proof-of-concept of content aggregation

# The Digital Mathematics Library

## A Timeline (2/2)

- 2006 IMU support (Vision, Best practices)
- 2005 AMS/MSRI proposal to the Moore foundation (others will follow...)
- 2008–2018 DML workshops, then track: technical challenges.
- 2010-2013 **EuDML**: The first medium-scale, professionally-built, supra-national effort
- 2012 Future World Heritage Digital Mathematics Library workshop/Planning a Global Library of the Mathematical Sciences report (NAS, Washington DC)
- 2014 IMU creates the GDML working group
- 2016 The GDML working group fosters the creation of the The International Mathematical Knowledge Trust (IMKT)
- 2019 **Mathdoc DML**: Let's go bigger!



*Introducing the*  
**mini-DML**  
*project*

Thierry BOUCHE  
Université Joseph Fourier (Grenoble)

WDML workshop  
Stockholm  
June 27th 2004

## 6. THE MINI-DML PROJECT

**Synopsis** Collate in *one* place basic bibliographical data for any kind of mathematical digital article and make them accessible to the users through simple search or metadata retrieval. Where...

- *basic bibliographical data* means

- Author
- Title
- Date of publication
- Full bibliographical reference
- Unique identifier
- URL (preferably calculated from the above)

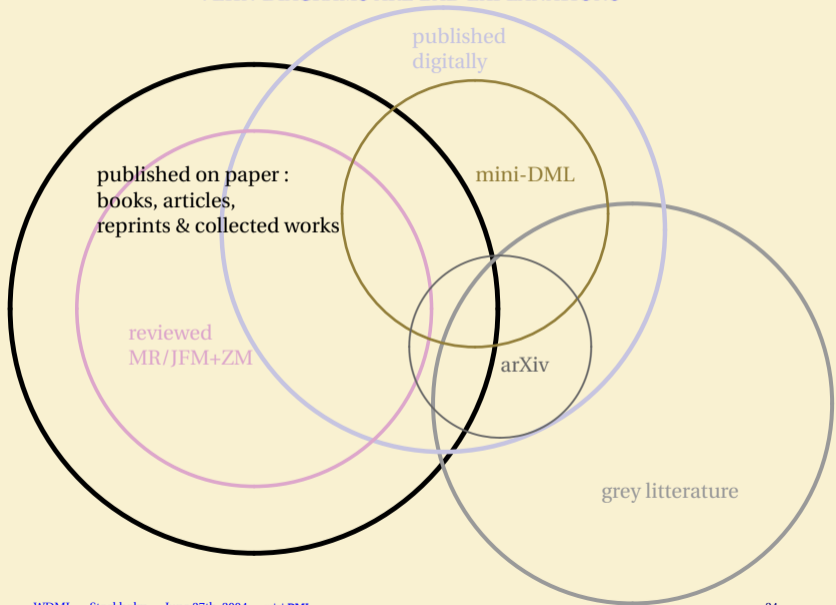
- *mathematical digital article* means

- Journal article
- Proceedings, seminar article
- Reprint, collected works
- Preprint
- Memoirs, books, slides... (not yet)

- *accessible to the users* means

- Simple unified search interface
- Easy retrieval of the used metadata for further processing or different exposure to the public.

# VEHN DIAGRAMS ARE BAD EXPLANATIONS



# The European Digital Mathematics Library

## *EuDML Vision (2008)*

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

- help **preserving** it over the long term,
- make it **available online**
- possibly after some embargo period (**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**

⇒ **EuDML**, pilot implementation with content from 12 European partners

# EuDML initiative

## Policies

### A distributed archive

- 1 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
- 2 Publishers should provide content through local DMLs unless they are themselves of the kind above

### A content provider needs to agree on these

- 1 The texts in EuDML must have been scientifically validated and formally published.
- 2 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.
- 3 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>

# EuDML : The Conjecture

Hence the name

$$U = \frac{W}{2}$$



$$\text{Europe} = \text{W(orld)}/2$$



EuDML is half the effort to build the WDML

- Optimistic conjecture!
  - Project followed-up by the EuDML initiative founded in 2014  
But the content aggregation is stalled
- ⇒ We need to go ahead

# EuDML : The Conjecture

Hence the name

$$U = \frac{W}{2}$$



$$\text{Europe} = \text{W(orld)}/2$$



EuDML is half the effort to build the WDML

- Optimistic conjecture!
  - Project followed-up by the EuDML initiative founded in 2014  
But the content aggregation is stalled
- ⇒ We need to go ahead

## Part II

*Work in progress:  
The Mathdoc DML*





# MDML: *The Objectives*

## Just like EuDML

- A one-stop shop for math content hosted on trusted platforms
- All content types
- As high quality metadata as available
- Automated ingest and updates
- Only index full texts that are or will eventually be Open Access
- Not a broker for publisher content (e.g. Bronze OA)

## With different priorities

- Make it as large as possible
- Accept e-prints of any kind, videotaped conferences. . .
- But keep the distinction clear whether refereed/validated or not



## Why now?

- We are technologically ready (cf. Olivier Labbe, Thierry Bouche: “The New Numdam platform,” CICM 2010, Edinburgh)
- We had the opportunity of Alexandre’s internship that could be dedicated to this goal
- If we don’t do it, who will?

## Done so far

- Harvesting metadata (system backend)
  - Choose content sources, harvesting protocol (OAI-PMH)
  - Set-up the import mechanism
  - Set-up the automated updates
- User interface (platform frontend)
  - Categorize items, and try to offer a decent browsing interface
  - Adapt the search engine, add a switch to include or exclude e-prints
  - Adapt the web display: style, navigation, full-text links, missing metadata. . .



## Backend development

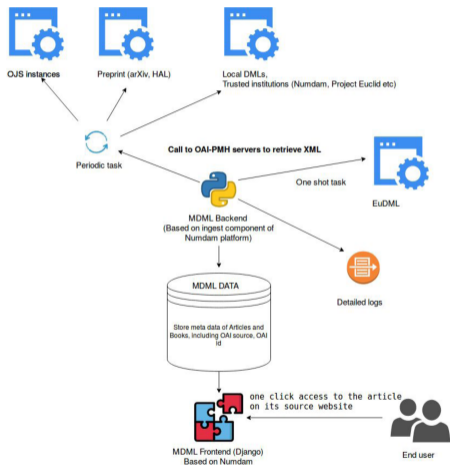
- Analysis of sources and what they offer (arXiv, OJS journals, EuDML, local DMLs, EuDML. . .)
- Create an XML multiplexer:
  - Harvest source  $X$  with XML schema  $Y$ ,
  - Normalize  $Y$  to our model through JATS or BITS (or direct from Dublin Core)
- Cron-like scheduler: Celery

```
CELERY_BEAT_SCHEDULER = "django_celery_beat.schedulers:DatabaseScheduler"
CELERY_BEAT_SCHEDULE = {
    'task-one': {
        'task': 'import_oai',
        'schedule': crontab(hour=13, minute=55),
    },
}
```

## *Frontend development*

- Framework from the new Numdam
- Quite a lot of tweaks and quirks
- Storing original ID and displaying relevant URLs

# Frontend development



# The prototype

Journals

Books

Sources



The Mathdoc Digital Mathematics Library (MDML) is currently indexing 598821 items from 30 sources



All ▾



Search

# The prototype

MDML Journals Books Sources DE -

pl  All

Include e-prints in search results (arXiv, HAL)

**Authors (100+)** 5567 results

Finkel, Olivier (28)  
 Colonna, Jean-François (20)  
 Majumdar, Satya N. (19)  
 Neeb, Karl-Hermann (16)  
 Djakov, Plamen (14)

**Years (18)**

2010-2019 (3479)  
 2000-2009 (1358)  
 1990-1999 (273)  
 1980-1989 (138)  
 1970-1979 (153)

**Collections (100+)**

arXiv (3850)  
 HAL (537)  
 Annales de l'Institut Fourier (114)  
 Ann. Math. Statist. (89)  
 The Electronic Journal of Combinatorics (88)

**Document types**

Article (5550)  
 e-print (945)  
 Book part (17)

**Classifications (100+)**

1 2 3 4 5 6 7 8 9 ... 279

**Size of dot product sets determined by pairs of subsets of vector spaces over finite fields**  
 Koh, Doowon, Pi, Youngjin  
 arXiv, Volume 2014 (2014) p. -  
 Abstract : ... a finite field  $F, q$  with  $q$  elements. As a new result, we prove that if  $E$  and  $F$  are subsets of the paraboloid and  $|E|F| \geq C \cdot q^d$  for some large  $C > 1$ , then  $|P(E, F)| \geq c \cdot q$  for some  $c > 0$

**Modules satisfying ACC on a certain type of colons.**  
 Lu, Chin-Pi  
 Pacific J. Math., Volume 135 (1988) no. 2 p. 303-318

**Factorial Domains**  
 Lu, Chin-Pi  
 Rocky Mountain J. Math., Volume 7 (1977) no. 4 p. 125-140

**Rates of Convergence in Empirical Bayes Estimation Problems: Continuous Case**  
 Lin, Pi-Erh  
 Ann. Statist., Volume 3 (1975) no. 1 p. 165-164

**Rings of analytic functions on any subset of the complex plane.**  
 Su, Li Pi  
 Pacific J. Math., Volume 43 (1972) no. 3 p. 535-538

**Homomorphisms of near-rings of continuous functions.**  
 Su, Li Pi  
 Pacific J. Math., Volume 99 (1971) no. 3 p. 261-266





## What next?

### Technically

- Still a lot of details to fix
- The Mathdoc team will work on it

### Contentwise

- Explore ways to enlarge the content beyond usable OAI-PMH servers
- Systematic source selection process (with the help of EuDML Scientific advisory board?)
- Support more metadata formats
- Investigate other ingest methods (web pages crawling, or human assisted)



*Thanks !*

**Thierry BOUCHE**  
**Alexandre BOUQUET**

**Cellule Mathdoc** (UMS 5638 CNRS/UJF)

Domaine Universitaire

Bâtiment CETA

150 rue de la Chimie

38400 Saint-Martin-d'Hères

[dml.mathdoc.fr](http://dml.mathdoc.fr)