The Geodesic project

Yet Another Digital Mathematics Library. . .

Thierry Bouche

Cellule Mathdoc & Institut Fourier, Université Grenoble Alpes (France)

9ECM—MIDAS Session
Sevilla
July 18th, 2024

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)

GDML

Towards a (comprehensive) 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?”

GDML

Full corpus: 5,300,000 items?

- It is estimated that about 500,000 items were published before the advent of systematic reviewing
- The math databases hold about 5 million references
- and grow by 100,000+ each year
- This makes the past scholarship an ever diminishing part of the lot...
- but there are essential building blocks of current knowledge there!
- I will focus on rare items, difficult to find
- Open access gaining momentum in the 21th Century doesn't mean the DML problem is solved for recent years, but issues are different (getting the OA content sustainably archived *out* of publishers platforms, e.g.)

GDML

Full corpus: a long distance target

GDZ 2001: 60,000 items

Numdam 2003: 2,600 items

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

EuDML 2024: 271,000 items

Numdam 2024: 72,600 items

Geodesic 2025: 1,000,000+ items?

GDML

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 inception: 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

GDML

The Digital Mathematics Library

A Timeline (2/2)

2006	Official IMU support (Vision, Best practices)
2005	AMS/MSRI proposal to the Moore foundation (others will follow. . .)
2008–2018	DML workshops, then DML track whithin CICM: 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 International Mathematical Knowledge Trust (IMKT)
2021	zbMath database becomes Open
2023	Geodesic project funded
2024	"MPS Workshop on Perspectives on Electronic Information and Communication in the Mathematical Sciences" (NYC) meeting organized by IMU's CEIC

GDML

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**

$\Longrightarrow$ **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

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

$$\Downarrow$$

$$\mathcal{E}u\mathrm{rope} = \mathrm{W}(\mathrm{orld})/2$$

$$\Downarrow$$

$\mathcal{E}u\mathrm{DML}$ 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

$\Longrightarrow$ We need to go ahead

GDML

Part II

Work in progress: The Geodesic project

The Geodesic Project

Abstract

- To develop a DML (Digitial Mathematics Library) to make it easier to find the existing Open Access mathematical literature
- Make 'F' in FAIR a reality

Profile

- Funded by FNSO (French National fund for Open Science, ministry of Research)
- Duration: 2 years
- Staff hired:
  - 1 librarian
  - 1 developer
  + permanent staff part-time
  + advisory committee with math and library expertise

The Geodesic Project

The goal

Standing on the shoulders of EuDML...

- A one-stop shop for math content hosted on trusted platforms
- All relevant content types
- As high quality metadata as available
- Open Access is mandatory (eventual OA is a minimum)
- Not a broker for publisher content (e.g. Bronze OA)

...while building a full replacement

- EuDML content ingested to bootstrap in a different platform
- Automated updates direct from the EuDML provider
  Numdam: Astérisque, CR Math; BDIM; DML-CZ; ElibM...
- Metadata improvements through zbMath Open API

The Geodesic Project

A larger scope

- EuDML borders:
  - Partner's collections out of European Union
  - Metadata through OAI-PMH harvesting
- Geodesic's scope is worldwide, e.g.
  - project Euclid
  - Math-net.ru
  - J-stage?
  - Math. Soc. publications like AustMS, CMS...
- Sources: reliable, reasonably sustainable, not-for-profit but not necessarily institutional partners
  - Hand-made web sites (like TAC, Erdős' Œuvres...)
  - OJS powered journals (Ars Math....)
  - Overlay journals (episciences, scholastica...)

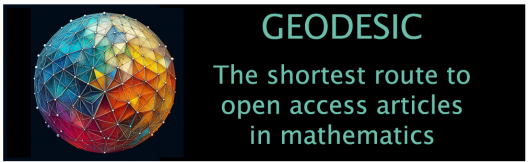
The Geodesic Project
The methods

Pragmatic content acquisition

- Publication websites crawling/scraping
- \+ Calls to source's APIs (including OAI-PMH)
- \+ Metadata improvements through zbMath Open API, Crossref API, etc.
- Automated periodic updates from trusted sources
- A local copy of PDFs is stored in certain cases, to secure the collection, but not exposed publicly unless a "triggering event" occurs

The prototype

Home screen



Geodesic, the Mathdoc digital mathematics library, is currently indexing 261139 articles from 12 sources

Browse by

Journals Seminars Books Conferences

Search

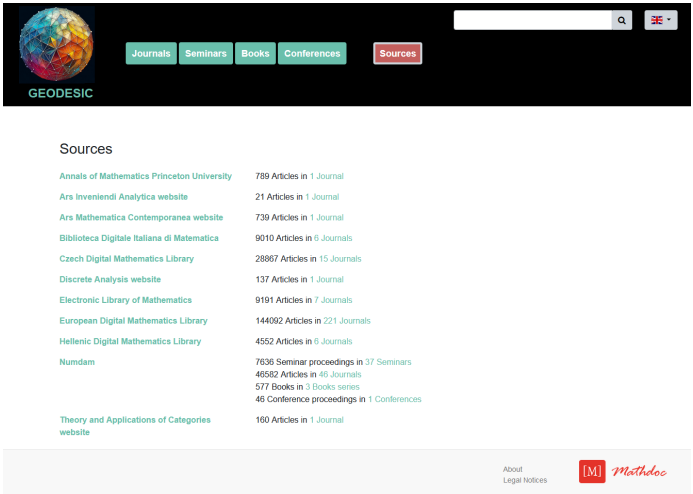
Search articles, authors...

About
Legal Notices

Mathdoc

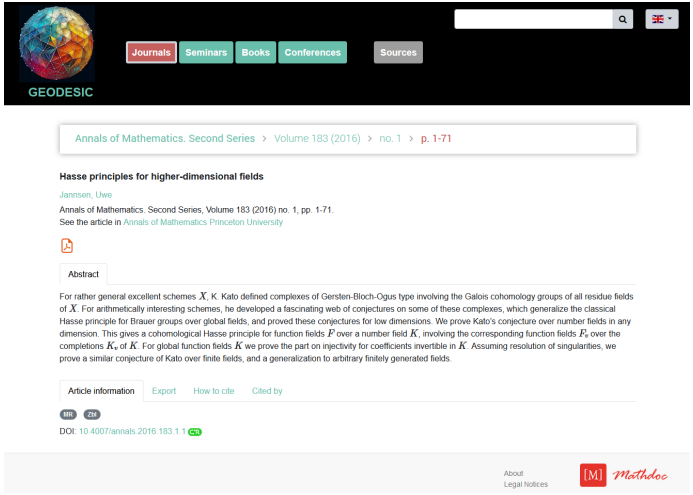
The prototype

Sources



The prototype

An item



What next?

Technically

- Not all metadata yet exploited
  (bibliographic references, ZB/MR numbers. . .)
- More relevant links expected

Contentwise

- Much more sources
- More content types
  (in)complete works, videotaped conferences, preprints. . .
- An API such that other service can retreive the metadata
  and add full text links to their catalogs
  like zbMath enhanced with 170,000 links to EuDML

Thanks !

Thierry BOUCHE

Cellule Mathdoc (URA 5638 CNRS/UGA)
Domaine Universitaire
Bâtiment CETA
150 rue de la Chimie
38400 Saint-Martin-d'Hères

www.mathdoc.fr