



EuDML project's outline

Content, metadata, interoperability, and partnering policies

Thierry Bouche

Cellule MathDoc & institut Fourier,
Université de Grenoble

EuDML session – The EuDML status quo

Towards a Digital Mathematics Library

DML 2011 workshop

CICM 2011, Bertinoro, July 21st 2011

Outline

- 1 EuDML policies & strategies
- 2 EuDML content: state-of-the-art
- 3 Metadata
- 4 How to join?

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

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

The European Digital Mathematics Library

A global gateway

One access point

For users A website with personal work spaces, allowing to search and browse the collections

For systems A batch lookup for turning citations into links

Benefits

- Easier discovery of mathematical items
- Higher visibility
- Unique lookup for reference linking
- More value to new material as references lead *somewhere*

The European Digital Mathematics Library

A distributed archive

A network of institutions

- Leverage the existing network of local (European) DML projects
- Kind of opt-in legal deposit of mathematical knowledge
- A distributed archive for full texts
- Reliable not-for-profit scientific institutions for long term preservation

Benefits

- Third party curation of full texts
- Long term availability of the corpus
- No preservation hassle on each content producer
- Publishers concentrate on creating new products
(EuDML takes care of the published content to maintain it over the long term)

The European Digital Mathematics Library

Eventual open access

The moving wall approach

- Once a publisher has completed a year's output, an archival copy (metadata and full text) is transferred to the relevant institution
- It is ingested and indexed
- New items appear in EuDML central database
- Full texts links point to the publisher's platform where access is controlled
- After a while, the local copy of the full text is readily available as well

Benefits

- Full text secured for preservation and long term access
- More texts with low market value available as open access
- Basic mathematical knowledge, foundation of current science and enabling new research, much more widely available

The European Digital Mathematics Library

Innovation

We are investigating new technology

- MathML metadata (extracted from images by math OCR, converted from \TeX sources, or extracted from PDF)
- Math search
- Accessible math
- Deep interlinking
- Compute items categorisation and similarity

Benefits

- Advances in Mathematical knowledge management
- Innovative discovery mechanisms
- Testbed for new user interaction with the corpus
- Deliver readily deployable technological solutions
- Deliver enhanced metadata to providers

EuDML content

Current content overview

Collections 235,000 items, 2,600,000 pages

- Germany** ERAM/JFM, GDZ, ELibM (120,000 items)
- France** Gallica-Math, NUMDAM, CEDRAM, TEL (50,000 items)
- Czech Rep.** DML-CZ (27,000 items)
- Russia** RusDML (17,000 items)
- Poland** DML-PL (13,000 items)
- Greece** HDML (2,400 items)
- Spain** DML-E (6,400 items)
- Portugal** SPM/BNP (2,000 items)
- Bulgaria** BulDML (450 items)

Retrodigitised BNP/SPM/IST, DML-CZ, DML-E, DML-PL, Gallica, GDZ, HDML, NUMDAM, RusDML

Born digital BulDML, CEDRAM, DML-CZ, DML-E, DML-PL, EDPS, ELibM, NUMDAM

EuDML content

Selection

Process Cascading: the project selects the partnering institutions, each institution selects contributed collections.

Criteria **Published** texts holding mathematical knowledge that has been **validated** through a scientific editorial process, so that they can serve for further **reference** in future works based on this mathematical knowledge.

Items A EuDML *item* is the relevant logical unit to be ultimately delivered to our users.

A monograph, a journal article, each individual contribution in a proceedings volume or an edited book, as well as the whole book

Concretely, it is a pair

(digital full text [PDF], metadata [XML])

archived at one of the partnering institutions

Summary Currently harvested: **235,000 items in 12 collections**
(185,000 articles, 45,000 chapters and contributions in books, 2500 books, 300 multiple volume works)

EuDML content

Collections (details 1/2)

EuDML partners: digital libraries

BuIDML 3 journals, 1 book series

DML-GZ 12 journals, 6 proceeding series, 35 books

DML-E 22 journals

DML-PL 11 journals, 4 series of books

HDML 5 journals, 1 proceedings series, 6 books

NUMDAM 34 journals, 29 seminars, 3 memoir series

SPM/BNP 1 journal

EuDML partners: publishing platforms

CEDRAM 10 periodicals

ELibM 91 journals

EDP Sciences 6 journals

EuDML content

Collections (details 2/2)

EuDML partner: reviewing database

ZMATH 3 million reviews

EuDML associated partners and collections

Gallica-Math 1 journal, 98 books

GDZ-Math 42 journals, 1549 monographs, 296 multi-volume works

RusDML 11 journals

EuDML content

Next collections?

EuDML future partners?

BDIM *processing* 1 journal

eLib SANU 9 Journals

eLib MATF 150 books, 354 Doct. Th.

SwissDML 4 journals

TEL 1996 Ph. D. Th.

Gallica 800 books

IMU ICM proceedings

EuDML metadata

Metadata formats babel

Metadata notations are of highly varying granularity and shape

SQL Internal relational database (no XML): DML-E

In-house DTD MathDoc, FIZ, IST, HDML

Standard DTD DC, Dspace, minidml, METS, NLM

We chose the

NLM Journal Archiving and Interchange Tag Suite

as basis for EuDML metadata storage and exchange

EuDML metadata

EuDML schema

NLM Journal Archiving and Interchange Tag Suite

Pros

- Widely deployed and tested (EDPS, PubMed Central, JSTOR...)
- NISO standard
- Precise and flexible (structured *and* flat models)
- MathML (and *alternatives*) ready
- Covers periodical content, books, book collections
- Has room for all foreseen metadata elements, yet extensible

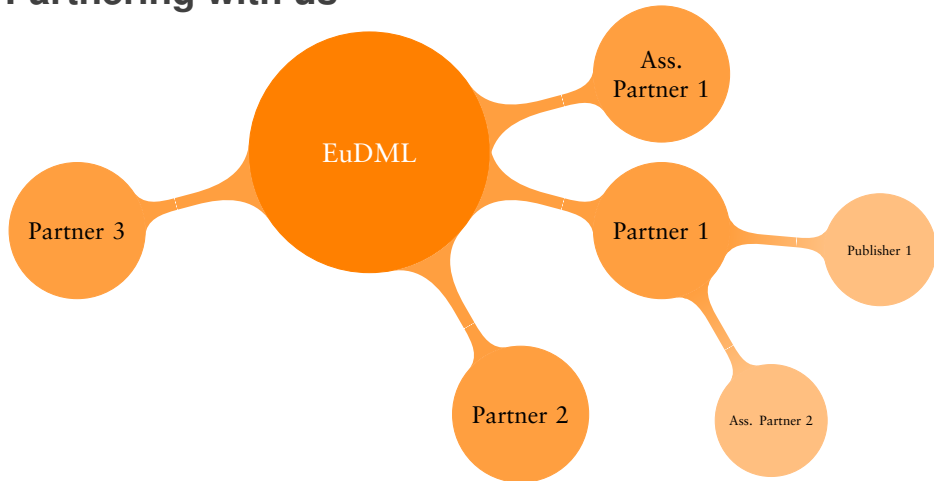
Cons

- Needs “application profile”
- Geared towards item’s full text
- Not meant for all foreseen item types (chapter in edited book...)

⇒ EuDML schema, v. 1.0

- Defines **article**, **book**, **mbook** as transport items
- Minimally deviates from NLM DTDs
- Best practices recommendation

Partnering with us



Partnering with us

Scientific institution

Institution charter

- Be aligned with the project's goals
- Keep committed over the long term
- Develop a preservation policy for the full texts
- Acquire new items (retrodigitisation or direct from publishers)
- Sort out rights and licences
- Quality insurance for metadata
- Manage communication with central registry

Partnering with us

Publisher

Publisher charter

- Be aligned with the project's goals
- Select one local partner as entry point to EuDML
- Set up transfer and update mechanism for new items
- Licence local DML to store transferred files for ever
- Determine moving wall duration

We will *deliver*
a truly open,
sustainable
and *innovative*
framework
for *access* and
exploitation of
Europe's rich
heritage of
mathematics.