EuDML

The EuDML schema, v 1.0

Thierry Bouche
Cellule MathDoc,
Université de Grenoble 1

Towards a Digital Mathematics Library
DML 2011 workshop
CICM 2011, Bertinoro, July 20th 2011

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**

$\implies$ **EuDML**, pilot implementation with content from 12 European partners

The EuDML schema
What for?

A public well-specified EuDML schema is needed:

1. For content providers: *which metadata* is *required*, *appreciated*, *encouraged*

2. For the search engine: *where to look for*

3. User interface: *what and how to display*,

4. Metadata enhancers toolsets: *what* they can start from and *where* and *how* they store their results

5. Content dump, *interoperability*

The EuDML schema
Supported items

Supported item types include:

- A multivolume work;
- A book, namely
  - a single volume from a multivolume work,
  - a monograph (which might be a doctoral dissertation, a memoir…),
  - an edited book (a book that contains chapters or articles that have been written by different authors and collated by scientific editors, which might be a conference proceedings volume);
- A part of a book such as a chapter, or a contribution in a proceedings volume;
- A journal article.

The EuDML schema

Out of scope

- Any text not persistently and formally published (preprints, personal web pages…).
- Papers not generally accessible online
- Version control for documents
- Complicated author/contributor structures for documents
- Licensing, access barriers, digital rights management issues

The EuDML schema
Based on NLM JATS

The Journal Archiving and Interchange Tag Suite (JATS),

- Internal format of one of our partners (EDP Sciences)
- Vastly deployed: PubMed Central, JSTOR, Portico…
- Highly customisable
- Full text oriented (ultimate metadata!)
- Allows parallel versions of the same content encoded differently
- MathML-ready

The EuDML schema
NLM JATS DTDs

JATS provides three DTDs that we will adapt for describing our three main content types:

- **The Journal Archiving and Interchange Tag Set** implements article3.dtd for journal articles (cf. http://dtd.nlm.nih.gov/archiving/)
- **The NCBI Book Tag Set** implements book3.dtd for books and bookcollection3.dtd for collections of books (cf. http://dtd.nlm.nih.gov/book/)

The EuDML schema
EuDML in NLM JATS?

JATS needed more work to suit our needs, in two opposite directions:

1. The item types currently supported out-of-the-box by JATS published DTDs are:
   journal article, **book**, and **book collection**
   EuDML "first class citizens" are more diversified. . .

2. Guidelines, unambiguous encoding
   ("EuDML application profile" in JATS, best practices)

The EuDML schema

Items available

A summary of the EuDML metadata, converted to NLM

Item type	Number
Journal article	185 046 items
Proceedings contribution	3 211 items
Book chapter	41 145 items
Book: monograph	1 590 items
Book: conference	119 items
Book: volume	748 items
Multiple volume work	296 items
Total	232 155 items

206,775 formulae generated from 40,937 input strings from all harvested items.
(Only 3,219 conversion errors were encountered.)

The EuDML schema
References

EuDML http://www.eudml.eu/

EuDML schema v1.0 ressource page http://www.eudml.eu/

JATS tag suite http://dtd.nlm.nih.gov/