EuDML: well established prototype of a digital library

Jiří Rákosník
Institute of Mathematics
Czech Academy of Sciences, Praha

Digital Presentation and Preservation of Cultural and Scientific Heritage
Veliko Tarnovo, 27 September 2016
EuDML project

• Partly funded by the EC in the Competitiveness and Innovation Framework Programme, February 2010 to January 2013

• Two challenges:
  – setting up technical infrastructure for a unified access point for the digital math content hosted by a number of different organizations across various countries
  – defining a cooperation model with a variety of stakeholders for building a reliable global reference library, long term and eventually exhaustive

• Project partners: universities, research institutes, scientific service institution, commercial publisher, private digital media agency, large library, learned society
Principles

• EuDML as a distributed system, each partner assumes full responsibility for the corresponding segment
• Unified seamless navigation through the distributed data
• Building on existent environment (zbMATH, MathSciNet, Mathematical Subject System distinguishing ~ 6,000 fields in maths)
• Policies
  – 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 close access later on
  – digital full text of each item contributed to EuDML must be archived physically at one of the EuDML member institutions
System architecture

Web User Interface

Accessibility Component | Annotation Component | Search Engine | External Services Interface

Metadata Repository

Association Analyser | Metadata Enhancer

Local Repositories
Distributed digital library
Interoperability model

• Contributing content to EuDML
  – digitization standards (600 DPI)
  – OAI-PMH server
  – EuDML XML metadata schema based on NLM Journal Archiving and Interchange Tag Suite and created after a detailed study of existing repositories
  – metadata validation tool
  – some other ways also possible (basic OAI-DC, FTP) → additional work!
  – content providers may be external – second tier of the EuDML network (MathDoc for GDZ and DML-E, DML-CZ for EMS-PH)
  – content providers can use the enhanced data back in their collections

• External interoperability devices
  – number of tools for third parties to enrich their services thanks to the availability of collections in the EuDML system
Services and tools

- Metadata enhancements
- On-the-fly conversion from TEX encoding of formulae to the Presentation MathML
- EuDML reference matching, zbMATH matching
- Experimental search for formulae
- Experimental similarity computation
- Experimental production of accessible formats of mathematical texts
- Web 2.0 features and annotation module
- etc.
Specifics of mathematical literature

- Never becomes obsolete
- Ever-growing edifice, every stone is important and remains forever
  - currently approx. 120,000 new scientific publications a year
- Particularly rich metadata
  - catalogue of 6,000 different mathematical fields
  - interlinking
- Considerably structured text
- Mathematical formulae
EuDML content

- **Items**
  - 258,539 articles from 323 journals
  - 2,249 individual contributions from 71 edited books
  - 3,523 monographs
  - 747 volumes from 295 multi-volume works

- **Collections**
  - new: e-Library of the Mathematical Institute SANU, EMS-PH
Number of documents by language

- English (en): 113028
- Russian (ru): 17489
- French (fr): 44871
- Czech (cs): 9834
- German (de): 8084
- Italian (it): 3468
- Others: 5203
- Greek Modern (el): 2697
Number of documents by publishing year

- 2011-2016: 12,497
- 2001-2010: 49,031
- 1991-2000: 38,682
- 1981-1990: 36,319
- 1971-1980: 33,130
- 1961-1970: 20,934
- 1951-1960: 9,340
- 1941-1950: 3,211
- 1931-1940: 5,731
- 1921-1930: 4,853
- 1911-1920: 3,987
- 1901-1910: 4,782
- 1891-1900: 4,508
- 1881-1890: 4,510
- 1871-1880: 4,451
- 1861-1870: 2,912
- 1851-1860: 2,190
- 1841-1850: 2,458
- 1831-1840: 862
- 1821-1830: 380
- 1811-1820: 118
- 1801-1810: 87
- 1701-1800: 598
- 1601-1700: 107
- 1584-1600: 3
# Visibility

Approx. 1000 visitors/day

<table>
<thead>
<tr>
<th>Country</th>
<th>Sessions</th>
<th>% of total 1,149,842</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>159,078</td>
<td>13.8</td>
</tr>
<tr>
<td>France</td>
<td>117,176</td>
<td>10.2</td>
</tr>
<tr>
<td>Germany</td>
<td>84,378</td>
<td>7.4</td>
</tr>
<tr>
<td>China</td>
<td>67,794</td>
<td>5.9</td>
</tr>
<tr>
<td>Italy</td>
<td>44,800</td>
<td>3.9</td>
</tr>
<tr>
<td>India</td>
<td>41,014</td>
<td>3.6</td>
</tr>
<tr>
<td>Japan</td>
<td>39,281</td>
<td>3.4</td>
</tr>
<tr>
<td>Russia</td>
<td>38,650</td>
<td>3.4</td>
</tr>
<tr>
<td>UK</td>
<td>38,557</td>
<td>3.3</td>
</tr>
<tr>
<td>Canada</td>
<td>26,545</td>
<td>2.3</td>
</tr>
</tbody>
</table>
EuDML Initiative

- Association established in 2014
- 12 partners (universities, research institutes, scientific service institution, large library, learned societies)
- Executive Board
- Technical Committee
- Scientific Advisory Board
- General Assembly
Challenges

• Resources for sustainability and development
  – dependence on individuals and their institutions ➔ infrastructure?
• Majority of maths literature is in hands of private for-profit entities
  – how to get it in EuDML?
• Should we insist on all policy principles?
  – *digital full text of each item contributed to EuDML must be archived physically at one of the EuDML member institutions*
• Extension
  – new collections
  – new type of items (software, videos, scientific data sets)
• New developments in publishing and communication, interoperability with other systems, semantic web, …
Conclusions

• EuDML is
  – (the only one) functional verified prototype of a full-fledged DML with a critical mass of documents
  – stable and ever growing
  – capable of essential extensions (if the content provider follows the metadata scheme system)
  – fit for including new types of data
  – open to further partners and collaborators

• There is a need of
  – resources for substantial further development
  – wide support by the community

• EuDML provides a solid base for the future GDML
References


