DELIVERABLE

Project Acronym: EuDML
Grant Agreement number: 250503
Project Title: The European Digital Mathematics Library

D2.2: Workshop Report

Revision: 1.0 as of 11th January 2012

Authors:
Petr Sojka  Masaryk University, MU

Contributors:
Felipe Gago  Univ. Santiago de Compostela, USC
Michal Růžička  Masaryk University, MU

| Project co-funded by the European Comission within the ICT Policy Support Programme |
|-----------------------------------|---------------------------------|
| Dissemination Level               |                                |
| P Public                          | ✓                               |
| C Confidential, only for members of the consortium and the Commission Services |                                |
## Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Author</th>
<th>Organisation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>30th August 2011</td>
<td>Petr Sojka</td>
<td>MU</td>
<td>First notes.</td>
</tr>
<tr>
<td>0.2</td>
<td>14th October 2011</td>
<td>Petr Sojka</td>
<td>MU</td>
<td>First draft.</td>
</tr>
<tr>
<td>0.3</td>
<td>12th December 2011</td>
<td>Petr Sojka</td>
<td>MU</td>
<td>First internal release.</td>
</tr>
<tr>
<td>0.4</td>
<td>13th December 2011</td>
<td>Petr Sojka</td>
<td>MU</td>
<td>Second internal release.</td>
</tr>
<tr>
<td>0.5</td>
<td>14th December 2011</td>
<td>Petr Sojka</td>
<td>MU</td>
<td>Third internal release. Parts moved to Appendix, com-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ments reflected.</td>
</tr>
<tr>
<td>1.0</td>
<td>11th January 2012</td>
<td>Petr Sojka</td>
<td>MU</td>
<td>Final release, added info about EuDML session.</td>
</tr>
</tbody>
</table>

### Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.
Contents

1 Introduction
  1.1 Workshop Objectives ........................................... 2
  1.2 Workshop Topics ............................................... 3

2 Workshop Programme
  2.1 Submission Format, Categories, Reviewing ..................... 4
  2.2 Workshop Programme Overview ................................ 4

3 EuDML Reporting and Discussion Session .......................... 5

4 EuDML Panel Discussion
  4.1 Panelists .......................................................... 8
  4.2 Panel Topics and Questions ..................................... 8
  4.3 EuDML Plenary Discussion: All you wanted to know about EuDML but were afraid to ask ........................................ 8

5 Workshop Proceedings ................................................. 10

6 Summary, Conclusions and Acknowledgement ....................... 10

Executive Summary
This document reports on the Digital Mathematics Library (DML) workshop held as part of Conference on Intelligent Computer Mathematics, CICM 2011 in Bertinoro, Forli, Italy, 18–23 July 2011. Workshop consisted of ten reviewed presentations (out of which seven reported outcomes of EuDML project), panel discussion with working mathematicians and leading researchers in the field, and EuDML reporting and discussion session. It allowed to get important feedback from working mathematicians and helped to establish SAB.
By studying the masters and not their pupils.
Niels H. Abel (Norwegian mathematician, 1802–1829)
in reply to a question about how he got his expertise

1 Introduction


There is a long tradition of Mathematics Knowledge Management (MKM) series of conferences (see http://www.mkm-ig.org/meetings/) since 2001. MKM has merged with AISC and Calculemus conferences into joint Conference on Intelligent Computer Mathematics, CICM in 2008. Since 2008, there has also been a DML workshop organized as a CICM satellite event. CICM has become a major event for mathematicians, librarians and computer scientists working in the area of MKM.

To attract as wide audience as possible for EuDML workshop, we have evaluated the possibilities to hold it either with CICM or with TPDL (Theory and Practice of Digital Libraries, formerly ECDL, European Conference on Digital Libraries) conferences. Even though in general TPDL has wider audience of “digital librarians”, as EuDML will rather serve mathematicians attending CICM, we have opted for CICM and organized EuDML workshop there.

1.1 Workshop Objectives

Mathematicians dream of a digital archive containing all peer-reviewed mathematical literature ever published, properly linked, validated and verified. It is estimated that the entire corpus of mathematical knowledge published over the centuries does not exceed 100,000,000 pages, an amount easily manageable by current information technologies.

Following success of DML 2008 (Birmingham, UK), DML 2009 (Grand Bend, Canada), DML 2010 (Paris, FR) workshop’s objectives were to formulate the strategy and goals of a global mathematical digital library and to summarize the current successes and failures of ongoing technologies and digitization projects, asking such questions as:

- What technologies, standards, algorithms and formats should be used and what metadata should be shared?
- What business models are suitable for publishers of mathematical literature, authors and funders of their projects and institutions?
- Is there a model of sustainable, interoperable, and extensible mathematical library that mathematicians can use in their everyday work?
- What is the best practice for
  - retro-digitized mathematics (from images via OCR to MathML or \TeX);
  - retro-born-digital mathematics (from existing electronic copy in DVI, PS or PDF to MathML or \TeX);
born-digital mathematics (how to make needed metadata and file formats available automatically as part of publishing workflow CEDRAM/Euclid model)?

Figure 1: The audience

1.2 Workshop Topics

Areas related to building [Eu]DML were specified as
- search, indexing and retrieval of mathematical documents
- ranking of mathematical papers, similarity of mathematical documents
- math OCR with MathML or \TeX output
- natural language processing of mathematical content (math NLP), mathematical corpora linguistics, mathematical thesauruses, ontologies, mathematical text mining
- document conversions from and to MathML, OpenMath, \TeX, PostScript and [tagged] PDF
- mathematical document compression
- processing of scanned images
- algorithms for crosslinking of bibliographical items, intext citations search
- mathematical document classification, MSC 2010
- mathematical documents metadata exchange via OAI-PMH and/or OAI-ORE
- long term archiving, data migration
- reports and experience from math digitization projects or EuDML
- math publishing with long term archival goal
- software engineering aspects of creating, handling MathML, OMDoc, OpenMath documents, and displaying them in web browsers
2 Workshop Programme

2.1 Submission Format, Categories, Reviewing

Papers were requested in the Springer LNCS style, preferably using \texttt{LaTeX} and the Springer \texttt{lnsc} class files.

There were two submission categories:

- full paper: 5–15 LNCS pages
- short paper/poster/demo/work in progress report: 2–5 LNCS pages. Paper length was not strict for both categories.

Submissions were collected and reviewed via EasyChair conference system.

Every submission was refereed by at least two (some by 3 or 4) PC members on the basis of technical quality, novelty, potential impact for building DML, and clarity.

There were also two EuDML presentations accepted for the main CICM conference, namely EuDML demonstration paper [3] and MKM paper about EuDML math search [11].

2.2 Workshop Programme Overview

Workshop consisted of ten reviewed presentations (out of which seven reported outcomes of EuDML project), panel discussion with working mathematicians and leading researchers in the field, and EuDML reporting and discussion session. Detailed programme
Figure 3: Petr Sojka and Martin Líška present EuDML math search engine at CICM [11]

is posted on http://www.fi.muni.cz/~sojka/dml-2011-program.html, together with most of the presentation slides, and can also be found in the Appendix.

There were six papers [8, 2, 12, 4, 1, 6] reporting on achievements EuDML has reached during it’s first half. There are detailed notes by James Davenport from the workshop available at http://www.bath.ac.uk/~masjhd/Meetings/CICM2011.pdf, so they are not reproduced there. Solely devoted to EuDML project there was half day “EuDML Session”.

3 EuDML Reporting and Discussion Session

EuDML session started with Thierry Bouche (TB) presenting main goals of the project, e.g. to assemble as much of the digital mathematical corpus as possible, with a view to

• helping preserve it over the long time;
• making it available online (not interested in stuff which will not be eventually free: moving walls permitted);
• being an authoritative and enduring digital collection;
• growing continuously with publisher-supplier new content;
• being augmented with sophisticated search interfaces and interoperability services technique in computational linguistics.
Every day an archival copy (metadata and full text) is checked (via OAI-PMH), transferred, ingested and indexed. Full text links point to the publisher’s site (and access control), until the wall moves past the item, then to the EuDML copy.

TB presented things being distributed. Ultimately, he would like to drop limitation to “European”. Would like a website with personal work spaces, allowing searching and browsing of collections. We plan a (batch) service turning citations into links, and this provides added value for new text, as references turn into links. Note that preservation is important, and is a task that has traditionally been assigned to libraries and publishers. There is little direct profit in preservation!

Current state: 235K items; 185K journal articles, 45K chapters/conference contributions, 2500 books, 300 multi-volume works; 2M pages; both retro-digitised and born-digital material.

Then EuDML Project Manager—José Borbinha (JB) took the floor. He noted that EuDML is not a research project in the EU’s eyes. He noted also that we are in mid-project, with a mid-term evaluation in September. Then followed a discussion:

Patrick Ion: Europeana?

JB: This is funded by the same programme as EuDML. Their aims are to set up new services based on existing technology. Their projects tend to have hundreds of partners/sub-partners/… Note that EuDML could be viewed as an aggregator for Europeana.

Michael Kohlhase: Is there a collections of URLs to access the \texttt{\LaTeX} produced.

TB: We do not produce \texttt{\LaTeX} as such, more \texttt{\LaTeX}-producing tools. As regards the \texttt{\LaTeX} of the papers, this is not necessarily available as a result of the moving wall. The list of resources will be itself available as an EuDML query.

Alan Sexton: The core of the project is aggregation, and the tools are, formally, a by-product.

Jim Pitman: Encouraged to hear that you are committed to machine interfaces as well as human interfaces. RESTful interfaces and JSON are not new technologies any more. Why you, not JSTOR? “You are talking about libraries, but the key is STORage”.

TB: My opening slides showed the benefits of EuDML, against Springer and LMS.

Jim Pitman: You should force European publishers to deposit with EuDML.

James Davenport: When LMS set up the “electronic only” JCM 14 years ago, we printed one paper copy to force the British Library to archive it, since it would not at that time do digital archiving. Legal deposit of digital materials is not even born, let alone in its infancy—you should be pressing the European Parliament, not EuDML.

TB: The latest version of TRALICS has a \texttt{\LaTeX} to MathML converter: install TRALICS and mail me for the incantation. He also demonstrated the linker, and showed it picking up a cited translation (in EuDML) for an article that was not. However, he gets NUMDAM metadata, not EuDML metadata (which does not exist yet).

Michael Kohlhase: You have this TRALICS-based \texttt{\LaTeX}\texttt{\LaTeX}→MathML—can you please wrap it compatibly with the LaTeXXML daemon so that people can do plug-and-play.

Christoph Lange: How do I get the RDF metadata?

TB: There is no RDF metadata currently.
Figure 4: Famous bibliography collector Jim Pitman

JB: The project proposal predated the successful arrival of “linked data”. We need to discuss the whole RDF issue.

Jim Pitman: Is something like the Bernouilli society, an international society based in the Netherlands, “European” enough for you? We currently use Project Euclid. The answer seemed to be yes, and Jim Pitman was willing to go and ask the appropriate authorities.

4 EuDML Panel Discussion

As part of the workshop, also a panel Towards a Digital Mathematics Library: On The Crossroad has been organized.

Panel was chaired by Petr Sojka. He stated that we are now in the middle of the project at the important crossroad [9]. We have to take important decisions now, which will influence EuDML and envisioned WDML significantly.

4.1 Panelists

Panelists that accepted the invitation were:

Jim Pitman (U.C. Berkeley, BibServer project, USA),
James Davenport (University of Bath, CEIC committee, UK),
Volker Sorge (University of Birmingham, EuDML project, UK),
Patrick Ion (Mathematical Reviews, AMS, MathML W3C, USA),
Wolfram Sperber (Zentralblatt Math, Berlin, DE)
4.2 Panel Topics and Questions

Panelists were given two questions:
1) What functionality and incentives would made a working mathematician to login and use a modern DML as EuDML?
2) What are the most decisive factors on this crossroad?

Panelists’ responses were:

Jim Pitman: No login (like in Google Scholar, Microsoft Academic Search, . . . ). Integrate with already available services (ZMath, MR). Open access to metadata encourages cooperation.

James Davenport: Working mathematicians want to get it quickly. Formula search capacity would overcome Google. Social networking for the mathematical world would be a huge step towards collaborative mathematics.

Volker Sorge: Cut and paste formulæ would be a killer app.

Patrick Ion: Login does not matter if the content is good enough to make people come back. Even can make people to feel proud of belonging to a community. Bookshelves. Annotations. Google is not as good/helpful as it used to be (spam, on purpose cheating with lots of keywords, . . . ). High quality search.

Wolfram Sperber: The importance of good metadata.

4.3 EuDML Plenary Discussion: All you wanted to know about EuDML but were afraid to ask

Patrick Ion: Relation with Europeana?
Figure 6: Famous computer scientist and working mathematician Volker Sorge at the panel, calling for formulae cut&paste.

José Borbinha: Founded by the same framework, difference in scale (thousands of partners, infinitely many levels, . . .)

Michael Kohlhase: Is there available a corpus of \LaTeX EuDML?

Thierry Bouche: There is not much \LaTeX in EuDML, only recently.

Alan Sexton: (enhanced) PDF Primary purpose is aggregation. Not a research project.

Jim Pitman: I like the idea of both human and machine interfaces. I’ll be glad to collaborate and be a consumer of these data.

Michael Kohlhase: I am missing from the description more room for editorial boards, publishers, other services (JSTOR, Archiv, . . .).

Thierry Bouche: Eu is to be removed over the years. EuDML is a sort of proof of concept: “this can be done: now let us collaborate”).

Anonymous: Features for allowing publishers, people from other sciences (physics, chemistry, . . .), librarians?

Thierry Bouche: We have approached Springer and LMS with little success so far.

James Davenport: Electronic archiving by libraries.

E. Tsivtsivadze: What can EuDML offer to my Ph.D. student who is writing his/her thesis in relation with references compared to, say, ZMath?

Thierry Bouche: Your chances to get to the full text would increase.

Patrick Ion: What would EuDML’s answer be to a proposal of collaboration?

José Borbinha: This would be a continuation project.

Thierry Bouche: Yes, as soon as we can get rid of the Eu.
Jim Pitman: Would you accept our Probability Electronic Journals? Are they European enough?

Patrick Ion: Do not drop Eu, just the E. Make it u(niversal) DML.

5 Workshop Proceedings

DML proceedings were published by the Masaryk University Press [10] and indexed by Thomson Reuters in Conference Proceedings Citation Index CPCI and Google Scholar and are available in digital form from electronic archive DML-CZ. Best papers might be chosen for a postconference book published by renowned publisher or for a journal special issue as in 2008, cf. MCS Vol 3, issue 3.

6 Summary, Conclusions and Acknowledgement

We have described the preparation and programme of the EuDML workshop. Papers [9, 8, 2, 12, 4, 1, 6] were reporting on progress done during the first half of the project. We have got valuable feedback during the panel discussion and during the EuDML discussion session.

Immediately following the workshop, EuDML general meeting has been organized, with several CICM attendees and experts invited to attend. Based on these joint discussions, many decisions on EuDML has been done, routing EuDML on a crossroad.
Several people expressed interest in EuDML and finally become members of Scientific Advisory Board (SAB), namely James Davenport and Masakazu Suzuki.

We would like to thank all who has made the workshop success, and to James Davenport for his detailed notes he has made during the CICM.

References


Figure 9: James Davenport telling how math search is important


DELIVERABLE

Project Acronym: EuDML
Grant Agreement number: 250503
Project Title: The European Digital Mathematics Library

Annexes to Deliverable D2.2
Workshop Report
Contents

A Participants 2
   A.1 Workshop Programme Committee 2
   A.2 Organizing Committee 3
   A.3 EuDML Session Participants 3

B Workshop Programme 4
A Participants

A.1 Workshop Programme Committee

An international Programme Committee has been formed:

- José Borbinha (Technical University of Lisbon, IST, PT)
- Thierry Bouche (University Grenoble, Cellule Mathdoc, FR) [co-chair]
- Michael Doob (University of Manitoba, Winnipeg, CA)
- Thomas Fischer (Goettingen University, Digitization Center, DE)
- Yannis Haralambous (Télécom Bretagne, FR)
- Václav Hlaváč (Czech Technical University, Faculty of Engineering, Prague, CZ)
- Michael Kohlhase (Jacobs University Bremen, DE)
- Janka Chlebíková (Portsmouth University, School of Computing, UK)
- Enrique Maciá-Virgós (University of Santiago de Compostela, ES)
- Jiří Rákosník (Academy of Sciences, Mathematical Institute, Prague, CZ)
- Eugenio Rocha (University of Aveiro, Dept. of Mathematics, PT)
- David Ruddy (Cornell University, Library, US)
- Volker Sorge (University of Birmingham, UK)
- Petr Sojka (Masaryk University, Faculty of Informatics, Brno, CZ) [co-chair]
- Masakazu Suzuki (Kyushu University, Faculty of Mathematics, JP)

![Figure 1: Thierry Bouche argues about EuDML.](image_url)
A.2 Organizing Committee
Workshop has been organized by Petr Sojka, Michal Růžička, in addition to the CICM local chair Andrea Asperti and CICM programme chair James Davenport.

A.3 EuDML Session Participants
We were able to identify the following people participating in the EuDML session.
1. Alan Sexton (University Birmingham, UK)
2. Aleksander Nowiński (ICM, University Warsaw, PL)
3. Beatriz Martinez (USC, ES)
4. Christoph Lange (Jacobs University, Bremen, Germany)
5. Deyan Ginev (Jacobs University, Bremen, Germany)
6. Dominika Tkaczyk (ICM, University Warsaw, PL)
7. Enrique Macias (USC, ES)
8. Felipe Gago (USC, ES)
9. Gilberto Pedrosa (IST, PT)
10. Jim Pittman (U.C. Berkeley, USA)
11. Josef Urban (Radboud University, NL)
12. James Davenport (University Bath, UK)
13. Jiří Rákosník (Institute of Maths ASCR, Prague, CZ)
14. José Borbính (IST Lisbon, PT)
15. Josef Baker (University Birmingham, UK)
16. Krzysz Wojciechowski (ICM, University Warsaw, PL)
17. Lídia Rodrigues (IST Lisbon, PT)
18. Łukasz Bolikowski (ICM, University Warsaw, PL)
19. Masakazu Suzuki (Kyushu University, Japan)
20. Michael Kolhase (Jacobs University, Bremen, Germany)
21. Michał Politowski (ICM, University Warsaw, PL)
22. Michal Růžička (Masaryk University, Brno, CZ)
23. Nicolas Houillon (UJE, CMD Grenoble, FR)
24. Panayiotis Vlamos (Ionian University, GR)
25. Petr Sojka (Masaryk University, Brno, CZ)
26. Piotr Dendek (ICM, University Warsaw, PL)
27. Serge Autexier (DFKI, Bremen, Germany)
28. Takao Namiki (Hokkaido University, Japan)
29. Thierry Bouche (University Grenoble, FR)
30. Thomas Fischer (Göttingen university Library, DE)
31. Thomas Hales (Pittsburgh, USA)
32. Vittorio Coti Zelati (University Napoli, Italy)
33. Volker Sorge (University Birmingham, UK)
34. Wolfram Sperber (FIZ, Karlsruhe, DE)
B Workshop Programme

Wednesday, July 20th, 2011

Towards a Digital Mathematics Library – Reports (Session Chair: Thierry Bouche)  
14:30–14:40 * Towards a Digital Mathematics Library: On the Crossroad (paper)  
by Petr Sojka (Masaryk University, Brno, Czech Republic)  
14:40–15:10 * Recent Development of the DML-CZ and Its Current State (paper, presentation)  
by Jiří Rákosník (Institute of Mathematics AS CR, Prague, Czech Republic)

![Image of a presentation slide]

Figure 2: Jiří Rákosník presents DML-CZ

15:10–15:35 * An Update on bdim: the Italian Digital Mathematical Library (paper)  
by Vittorio Coti Zelati (Universita degli Studi di Napoli “Federico II”, Napoli, Italy)  
15:35–16:00 * Time Stamping Preprint and Electronic Journal Server Environment (paper, presentation)  
by Takao Namiki (Hokkaido University, Japan), Kazutsuna Yamaji, Toshiyuki Kataoka, Noboru Sonehara (National Institute of Informatics, Japan)

Digitization Workflows and Standards (Session Chair: Petr Sojka)  
16:30–17:00 * Towards a Robust Author Name Disambiguation Framework (paper, presentation)  
by Łukasz Bolikowski and Piotr Dendek (University of Warsaw, Warsaw, Poland)
17:00–17:30  * Workflow of Metadata Extraction from Retro-Born Digital Documents (paper, presentation) by Dominika Tkaczyk and Łukasz Bolikowski (University of Warsaw, Warsaw, Poland)

Figure 3: Dominika Tkaczyk presents EuDML Workflow of Metadata Extraction from Retro-Born Digital Documents [1].

17:30–18:00  * The EuDML Metadata Schema, Version 1.0 (paper, presentation) by Thierry Bouche, Claude Goutorbe (CMD UJF, Grenoble, France), Jean-Paul Jorda (EDP Sciences, Les Ulis, France) and Michael Jost (Zentralblatt MATH, Berlin, Germany)

Thursday, July 21st, 2011

DML Building Technologies (Session Chair: José Borbinha)  09:00–09:30  * Towards Faithful Reverse Engineering of PDF Documents (paper, presentation) by Josef Baker, Alan Sexton and Volker Sorge (University of Birmingham, UK)

09:30–10:00  * Web Interface and Collection for Mathematical Retrieval – WebMIaS and MREC (paper, presentation) by Martin Liška, Petr Sojka, Michal Růžička and Peter Mravec (Masaryk University, Brno, CZ)

10:00–10:30  * Using Discourse Context to Interpret Object-denoting Mathematical Expressions (paper, presentation) by Magdalena Wolska (Universität des Saarlandes, Saarbrücken, Germany), Mihai Grigore (Goethe University, Frankfurt am Main, Germany), Michael Kohlhase (Jacobs University, Bremen, Germany)
Figure 4: Josef Baker presents his research

EuDML session (Session Chair: Quique Macias)

11:00–11:30 * The EuDML status quo (presentation)
by José Borbinha, Thierry Bouche (EuDML project)

11:30–12:15 * Plenary discussion (All you wanted to know about EuDML, but were afraid to ask.)

12:15 * DML Bussiness Meeting, CICM charter approval, Closing (minutes)

References

Figure 5: Thierry Bouche presenting EuDML at workshop’s EuDML session

Figure 6: EuDML panelists