

## Front Matter

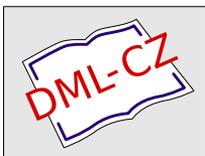
In: Petr Sojka (ed.): Towards a Digital Mathematics Library. Grand Bend, Ontario, Canada, July 8-9th, 2009. Masaryk University Press, Brno, 2009. pp. I--VIII.

Persistent URL: <http://dml.cz/dmlcz/702587>

## Terms of use:

© Masaryk University, 2009

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

**DML 2009**  
**Towards a Digital**  
**Mathematics Library**

**muni**  
PRESS



<http://www.fi.muni.cz/~sojka/dml-2009.html>

**Petr Sojka (editor)**

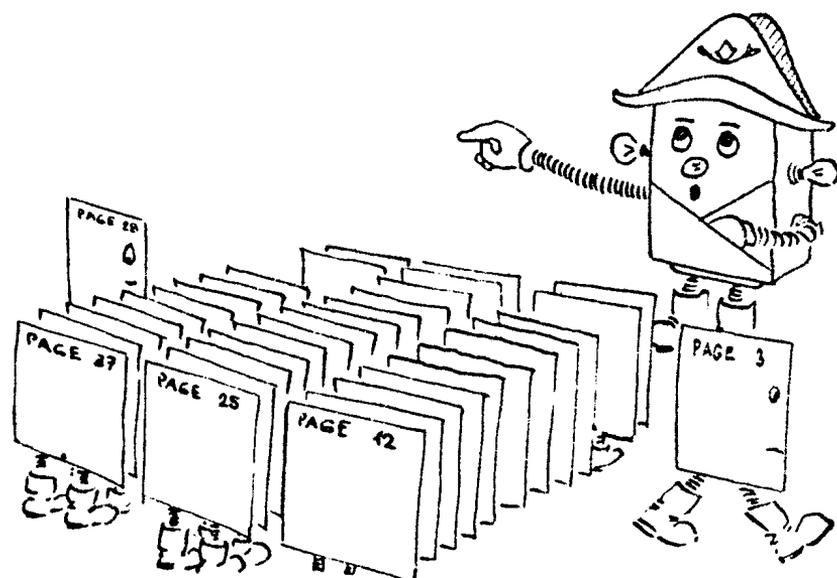
# **DML 2009**

Towards a Digital Mathematics Library

Grand Bend, Ontario, Canada

July 8–9th, 2009

Proceedings



Masaryk University, Brno, 2009

Proceedings Editor

Petr Sojka

Faculty of Informatics, Masaryk University  
Department of Computer Graphics and Design  
Botanická 68a  
CZ-602 00 Brno, Czech Republic  
Email: sojka@fi.muni.cz

CATALOGUING-IN-PUBLICATION – NATIONAL LIBRARY OF THE CZECH  
REPUBLIC

DML 2009 (Grand Bend, Ontario, Canada)

DML 2009 : Towards a Digital Mathematics Library : Grand Bend, Ontario,  
Canada, July 8-9th, 2009 : proceedings / Petr Sojka (editor). - 1st ed. -  
Brno : Masaryk University, 2009. - X+139 p.

ISBN 978-80-210-4781-5

025:004.08 \* 930.25:004.08 \* 51:81'42'373.46 \* 002.2:004 \* 004.91 \*  
004.352.242 \* 004.93'1 \* 004.832.2

- digital libraries
- digital archives
- mathematical texts
- digitization of documents
- data processing
- OCR technology
- pattern recognition
- fulltext search
- proceedings of conferences
- digitální knihovny
- digitální archivy
- matematické texty
- digitalizace dokumentů
- zpracování dat
- technologie OCR
- rozpoznávání vzorů
- fulltextové vyhledávání
- sborníky konferencí

006 - Special computer methods [23]

004.9 - Speciální počítačové metody. Počítačová grafika [23]

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the Czech Copyright Law, in its current version, and permission for use must always be obtained from Masaryk University. Violations are liable for prosecution under the Czech Copyright Law.

© Masaryk University, 2009

ISBN 978-80-210-4781-5

# Preface

Mathematicians dream of a digital archive containing all peer-reviewed mathematical literature ever published, properly linked and validated/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.

The 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 related 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
  - retrodigitized mathematics (from images via OCR to MathML and/or  $\text{\TeX}$ );
  - retro-born-digital mathematics (from existing electronic copy in DVI, PS or PDF to MathML and/or  $\text{\TeX}$ );
  - born-digital mathematics (how to make needed metadata and file formats available as a side effect of publishing workflow [CEDRAM model, Euclid])?

The intention was to have the workshop as a forum for presentation and discussion of the latest developments in the the field of digitization of mathematics, based on the previous bilateral discussions and very successful workshop DML 2008 held as part of CICM multiconference in Birmingham, UK, last year.

*Topics of the Workshop included*

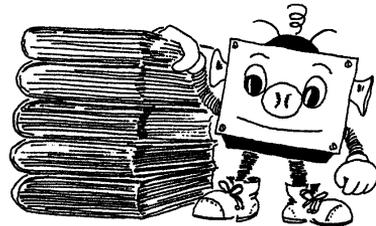
- \* search, indexing and retrieval of mathematical documents
- \* ranking of mathematical papers, similarity of mathematical documents
- \* math OCR with MathML/ $\text{\TeX}$  output
- \* document conversions from/to MathML, OpenMath,  $\text{\LaTeX}$ , PostScript and [tagged] PDF
- \* mathematical document compression
- \* processing of scanned images
- \* algorithms for crosslinking of bibliographical items, intext citations search
- \* mathematical document classification, MSC 2010
- \* mathematical text mining
- \* mathematical documents metadata exchange via OAI-PMH and/or OAI-ORE

- \* long term archiving, data migration
- \* reports and experience from math digitization projects
- \* math publishing with long term archival goal
- \* software engineering aspects of creating, handling MathML, OMDoc, OpenMath documents, and displaying them in web browsers.

This year we took the possibility to compare DML tools, systems and technologies in US and Europe, with invited David Ruddy speaking about Euclid and Thierry Bouche referring about planned European activities and projects EuDML/EVLM.

This volume contains the Proceedings of the Workshop *Towards a Digital Mathematics Library (DML 2009)*, organized by the Faculty of Informatics, Masaryk University with the help of Ontario research Centre for Computer Algebra (University of Western Ontario and University of Waterloo) and held on July 8–9<sup>th</sup>, 2009 in Grand Bend, Ontario, Canada, as a satellite event of CICM 2009 (Conference on Intelligent Computer Mathematics). The Proceedings is divided into five parts:

1. Towards a Digital Mathematics Library,
2. Towards Mathematical OCR and Search,
3. Digitization Reports,
4. Digitization Technologies and Platforms  
and
5. Tools and Techniques.



My very special thanks go to the Program Committee members for their hard work during review periods. Most of the submitted papers were reviewed by three members of the Program Committee. We employed rebuttal phase, where authors were given the possibility to comment on the preliminary review reports and to answer anonymous reviewer's questions. It helped to increase the quality of final paper versions considerably.

I would also like to express my appreciation to the members of the Organizing Committee for their efforts in organizing the Workshop and ensuring its smooth running, and to CICM general chair Stephen Watt.

Last but not least, the cooperation of Masaryk University as a publisher of these Proceedings is gratefully acknowledged.

DML 2009 offered a rich program of presentations, short talks/posters, technical papers and [panel] discussions. I hope that another step towards fulfilling the dream of the world-wide Digital Mathematics Library has been made.

Brno, June 2009

Petr Sojka

# Organization

DML 2009 was organized by Faculty of Informatics, Masaryk University, Brno, Czech Republic with the help of Ontario research Centre for Computer Algebra (University of Western Ontario and University of Waterloo). Web page of the workshop is <http://www.fi.muni.cz/~sojka/dml-2009.html>.

## Program Committee

José Borbinha (Technical University of Lisbon, IST, PT)  
Thierry Bouche (University Grenoble I, Cellule Mathdoc, FR)  
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)  
Janka Chlebíková (Comenius University, MFF, Bratislava, SK)  
Enrique Maciás-Virgós (University of Santiago de Compostela, ES)  
Jiří Rákosník (Academy of Sciences, Institute of Mathematics, Prague, CZ)  
Eugénio Rocha (University of Aveiro, Dept. of Mathematics, PT)  
David Ruddy (Cornell University, Library, US)  
Petr Sojka (Masaryk University, Faculty of Informatics, Brno, CZ) [chair]  
Volker Sorge (University of Birmingham, UK)  
Masakazu Suzuki (Kyushu University, Faculty of Mathematics, JP)  
Bernd Wegner (Technical University Berlin, Mathematical Institute, DE)

## Organizing Committee

Michal Růžička (*technical support and administrative contact*), Michael Doob (*local organization*), and Petr Sojka (*chair, Proceedings*)

## Sponsors and Support

The DML workshop and preparation of the Proceedings was supported by the Masaryk University, Brno, and by project 1ET200190513 (DML-CZ) of the Academy of Sciences of the Czech Republic.

