DML/CZ: Czech Digital Mathematics Library
Project No. 1ET200190513
R&D Programme Information Society
National Research Programme
Academy of Sciences of the Czech Republic
2005–2009

Final Report

The Czech Digital Mathematics Library DML-CZ resulted from cooperation of five partners during execution of the project No. 1ET200190513 supported by the Academy of Sciences of the Czech Republic in the period 2005–2009 in the frames of the R&D Programme Information Society.

DML-CZ constitutes a full-fledged digital library that is freely accessible on http://dml.cz and by the end of 2009 comprises 275 000 pages of scholarly texts (26 000 articles and book chapters) from almost 10 000 authors. Being endowed with a number of functionalities it enables users in particular

– to browse and download detailed article metadata of all contained documents including bibliographic references,
– to browse and download full-texts of articles and book chapters, which are not blocked by publishers conditions (moving wall),
– to search in author, title and Mathematics Subject Classification indexes,
– to use a simple and advanced search in metadata and in full-texts,
– to exploit the supply of content related papers set up according to a similarity degree calculated on the basis of machine analysis of texts.

Journal papers are provided with structured lists of bibliographic references making altogether about 160 000 items.

The articles and bibliographic references are linked with the databases Zentralblatt MATH and MathSciNet.

The full-texts are presented in the DML-CZ as PDF documents digitally signed to certify their authenticity. Each article is provided with a fly-leaf containing basic metadata and a copyright clause indicating the copyright ownership and warning the user that the document is available strictly for personal use.
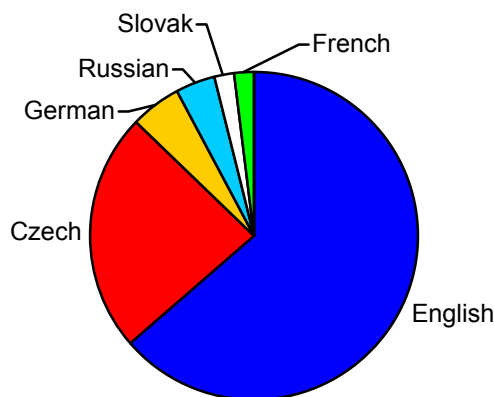
DML-CZ Content

By the end of 2009, the DML-CZ comprises virtually all Czech mathematics journals output (8 scholarly journals including their forerunners, the eldest being *Časopis pro pěstování matematiky a fysiky* published by the Union of Czech Mathematicians and Physicists since 1872), the Slovak journal *Mathematica Slovaca*, proceedings series of five important international conferences organized by Czech and Slovak institutions, and 28 monographs including the collection of 25 works by Bernad Bolzano (or about him).

Table. Survey of DML-CZ content on the project end date 31th December 2009.

Source	From-to	Issues/ volumes	Pages	Articles/ chapters
-- journals				
Acta Mathematica Universitatis Ostraviensis	1993-2008	16	1 711	169
Acta Universitatis Palackianae Olomucensis. Mathematica	1960-2008	47	9 684	769
Applications of Mathematics	1956-2008	313	27 258	2 259
Archivum Mathematicum	1965-2009	175	13 645	1 334
Časopis pro pěstování matematiky a fysiky	1872-1950	353	35 511	4 060
Časopis pro pěstování matematiky	1951-1990	160	19 186	2 119
Commentationes Mathematicae Universitatis Carolinae	1960-2009	201	36 276	3 248
Czechoslovak Mathematical Journal	1951-2008	230	42 947	3 510
Kybernetika	1965-2008	293	30 038	2 413
Mathematica Bohemica	1991-2008	69	8 556	745
Mathematica Slovaca	1951-2007	238	24 922	2 423
Journals total		2 095	249 734	23 049
-- proceedings				
EQUADIFF	1962-2005	15	5 212	646
Nonlinear Analysis, Function Spaces and Applications	1978-2006	8	1 958	85
TopoSym	1961-1976	6	2 362	459
Winter School on Abstract Analysis	1976-2009	36	4 699	693
Winter School "Geometry and Physics"	1981-2009	29	4 245	396
DML – Digital Mathematics Library	2008-2009	2	307	29
Proceedings total		96	18 783	2 308
-- monographs				
Bolzano – book collection	1804-1963	25	5 287	335
History of Mathematics - book collection	1994-1996	4	786	55
Other monographs	1874-1979	3	630	37
Monographs total		32	6 703	427
DML-CZ total		**2 223**	**275 220**	**25 784**

DML-CZ contains texts in 12 different languages, whereas documents including texts in two to four different languages including Russian are not rare.

Such polylingualism brought in a series of interesting problems to be solved (machine-aided recognition of languages, translation of titles into English, Roman transcription of author's names etc.).

Back journal issues generally have a compound structure including not only scholarly mathematical papers but also papers devoted to physics, chemistry and astronomy, reviews, news, information for society members as well as papers focused on history or political events.
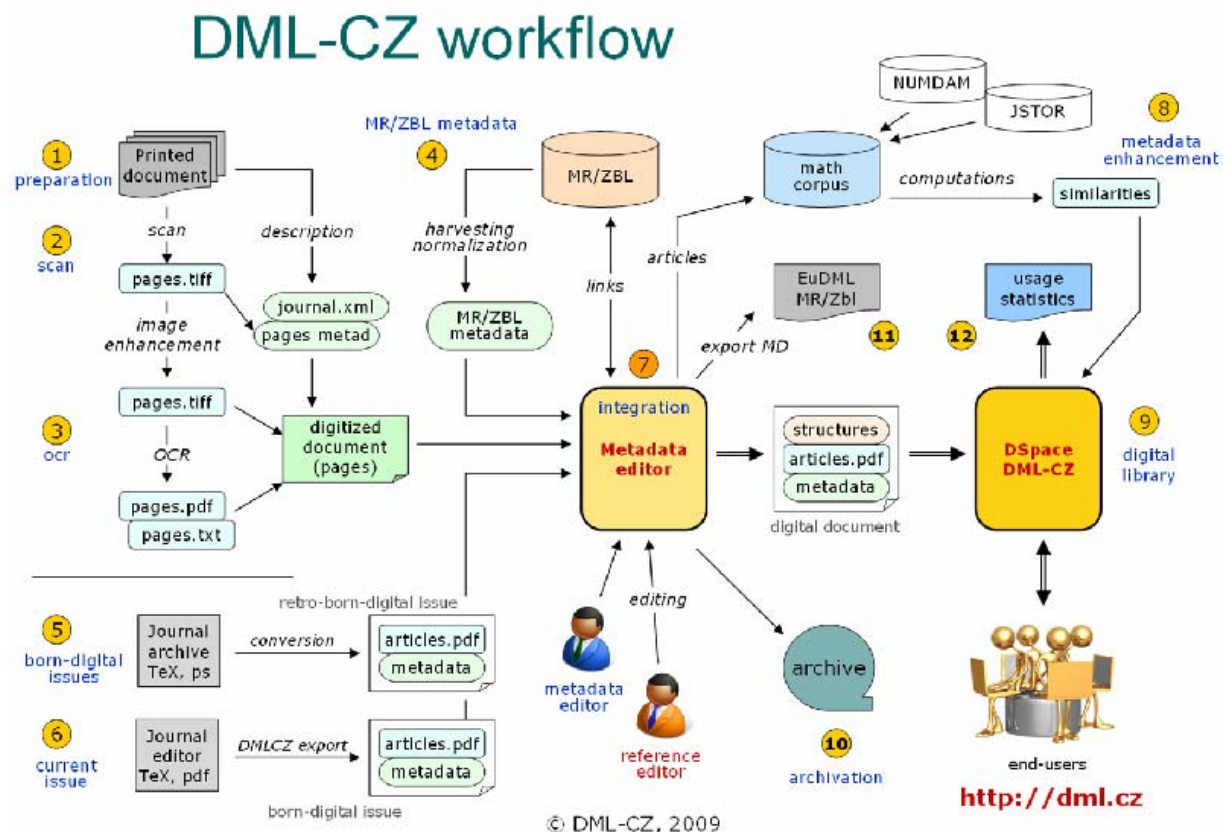
The Main Outputs of the Project

The most important results of the project are seven applications of the type of certified technology, which altogether form the operational core of the Czech Digital Mathematics Library, enable its continuous upgrading and further development. At the same time they represent efficient tools that can be used for creation of other digital repositories. Theoretical as well as practical aspects of these tools have been presented in series of publications emerging during the project.

 - *The Czech Digital Mathematics Library DML-CZ* represents a complex solution, which comprises the six following elements each of them being an independent certified technology.

 - *Metadata Editor* is a complex system supporting all operations carried out during creation of an article oriented digital library through a distributed structured remote access: integration of digitized pages into hierarchical structures, article building, creation, adjustment and machine verification of detailed metadata up to the level of articles and book chapters, processing and interlinking bibliographic references, name authority management, integration of born-digital material and generation of the resulting PDF papers.

 - *Presentation system DML-CZ* is a complex solution for exposing the digital library and its full exploitation through a remote access, built on the basis of the open source system DSpace.

 - *DML-CZ OCR* is a technology developed for optical recognition of mathematical texts, based on the commercial product ABBYY FineReader 8 SDK and on the InftyReader system for recognition of mathematical formulas. The text recognition is optimized for languages in individual paragraphs.

 - *DML-CZ Born-Digital Import* is a complex system based on the public domain software TeX and Tralics-CEDRICS that enables technical editors of mathematical journals typeset by means of the typographical system TeX to create optimized and validated outputs for the digital library in the form of PDF files, article metadata and bibliographic references in an automated way.

 - *DML-CZ Similarity&Classification* is a robust technology based on methods of machine learning on a corpus of author- and/or reviewer-classified article full-texts enabling to train an automated classifier to sort texts according to the Mathematics Subject Classification and to compute their semantic similarity.

- Recompression of scanned bitmaps in PDF is a technology for automation of lossy compressing scanned bitmaps in the JBIG2 standard with an adaptive optimization of degree of lossiness and compression.

The State-of-the-art

The following schema presents a general depiction of activities, relationships and roles in the frame of project realization.

The DML-CZ system including all data has been committed to the Institute of Mathematics AS CR in Prague. The Institute will be in charge of general administration, maintenance and development of this digital library including fund-raising based on financial contributions of content providers. The excellent cooperation of all partners developed during the project will continue. The Institute of Computer Science of the Masaryk University and Faculty of Informatics of the Masaryk University in Brno will provide the technological maintenance and development of DML-CZ according to contracts fixed with the Institute of Mathematics AS CR. The necessary scanning and graphical adjustment of documents will be provided by the Digitization Centre of the Library AS CR. The team of the Faculty of Mathematics and Physics, Charles University will cooperate with the Institute of Mathematics AS CR in terms of scholarly mathematical aspects of the content development. A scientific supervision and promotion will be provided by the Czech Mathematical Society.

A systematic publicity since launching the pilot part of the project in 2008 together with ever growing content and range of services results in increasing interest of the public. It is also supported by the fact that DML-CZ is being indexed by Google where links to items in DML-CZ use to have one of the highest ranking. The Google Analytics registered over 73 thousands accesses from 177 countries to the DML-CZ during 2009 (most frequently from the Czech Republic, USA, Germany, India, Slovakia, China, France, UK, Poland and Russia). The daily visit rate is permanently growing, amounting around 500 in the last weeks of 2009. We recognize also increasing interest of publishers in getting their journals and conference proceedings in DML-CZ.

Further Development

The content of DML-CZ is not finished. In spite of the fact that the original project has been accomplished the digital library DML-CZ will be continuously developed and complemented using the created tools and processes. During 2010 the journal *Pokroky matematiky, fyziky a astronomie* published by the Union of Czech Mathematicians and Physicists will be added. In the next period it is planned to include two specialised journals for mathematics and physics teachers, the book series *History of Mathematics* and – depending on the attitude of copyright holders – some more important monographs. On the proposal of the Otakar Borůvka Society a new section devoted to legacy of prominent Czech mathematicians has been prepared. In agreement with the policy declared by the European Commission insertion of doctoral theses will be considered.

It became evident that DML-CZ is beneficial not only for the community of mathematical scholars but also for teachers who may use the easy access to literature for a more efficient work with students. Last but not least DML-CZ will be conductive to librarians.

The Masaryk University and the Institute of Mathematics AS CR along with 12 further foreign partners take part in the project EuDML with the objective to create a multifunctional digital library which will combine the digital content of existing distributed mathematics repositories in different countries. Integration of the DML-CZ in the European Digital Mathematics Library in frames of this project will undoubtedly increase interest in the Czech mathematics and in scholarly texts published in the Czech Republic.