



DML-CZ

Czech Digital Mathematics Library

<http://dml.cz/>



DML-CZ

DML-CZ has been developed in order to preserve and present in a digital form the content of major part of mathematical literature that has ever been published in the Czech lands, and to provide a free access to the digital content and bibliographical data.

DML-CZ houses old digitized materials (1872–1990) as well as retro-born (~ 1990–1996) and born digital content (~ 1997 and further).

DML-CZ is alive and ever growing: the newly published content has been regularly included in cooperation with publishers and retro-digitization continues.

DML-CZ content statistics

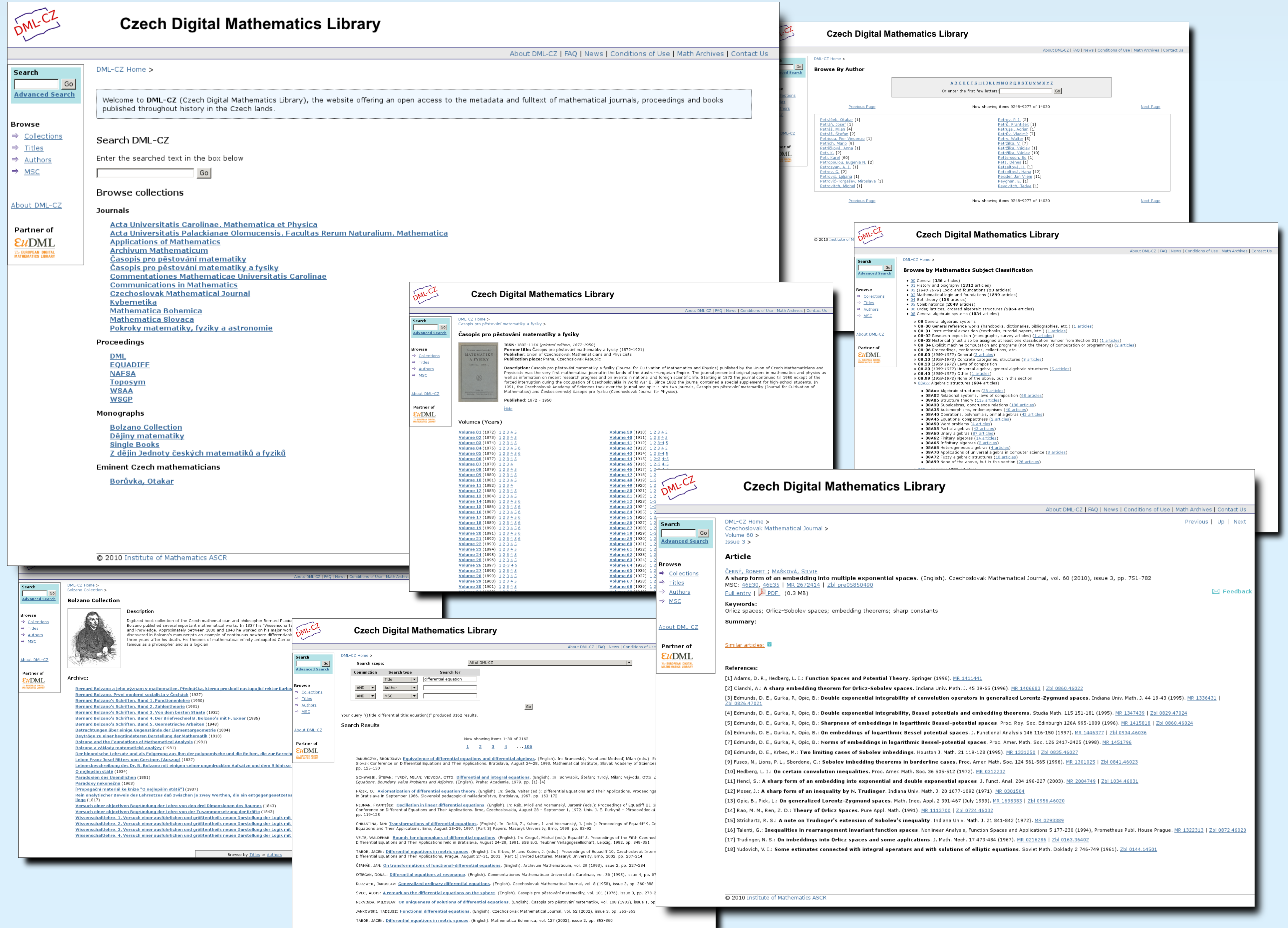
- ▶ 13 journals (~ 28 300 articles, 296 300 pages, 2 610 issues)
- ▶ 6 proceedings series (~ 2 400 articles, 19 200 pages)
- ▶ 4 monograph collections (~ 89 books, 1 560 chapters, 21 350 pages)
- ▶ Otakar Borůvka collection (~ 3 940 pages in various materials)

DML-CZ services for end users

- ▶ browsing and searching (metadata + full text)
- ▶ PDF full text download
- ▶ persistent URL for the content (articles, PDFs, journals, ...)
- ▶ Mathematical Subject Classification (MSC) browse and search
- ▶ authors browse
- ▶ references including links to the mathematical bibliographic databases
- ▶ lists of similar articles (machine computed)
- ▶ partner of and content provider to EuDML (<http://eudml.org>)

DML-CZ partners

- ▶ Institute of Mathematics AS CR, Prague
- ▶ Masaryk University, Brno
- ▶ Charles University, Prague



DML-CZ Software and Technologies

New software and technologies have been created and a lot of open-source software tools have been used for DML-CZ development and routine operation.

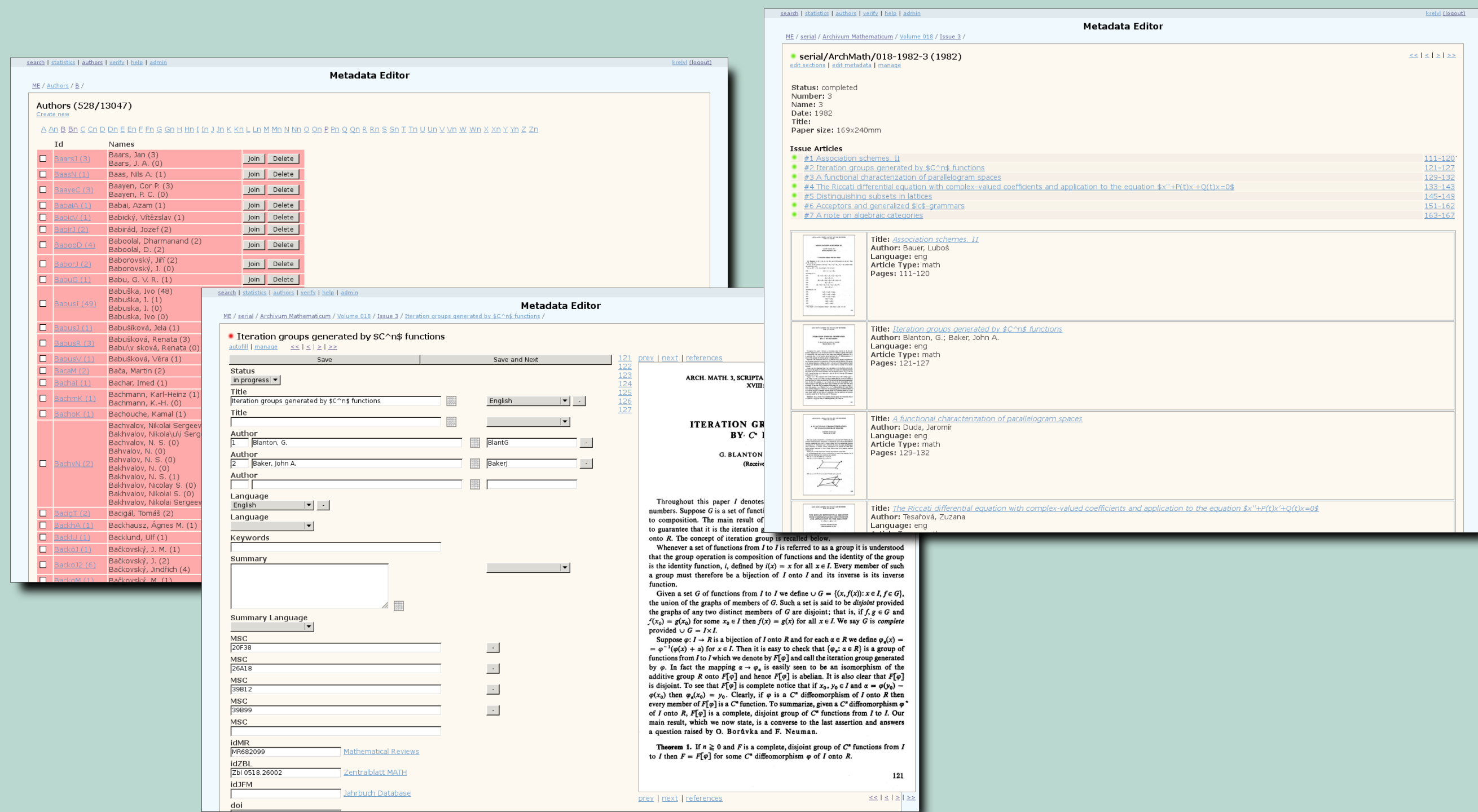
DSpace

The DSpace software is used as a public repository for presentation of the DML-CZ content to the end users. Web interface is customized using the internal DSpace XMLUI tool. Integrated and configurable OAI-PMH server provides the metadata to third-party harvesters. The most important one is the European Digital Mathematics Library (EuDML). We also provide a necessary metadata to cooperate with Google Scholar.

Metadata Editor

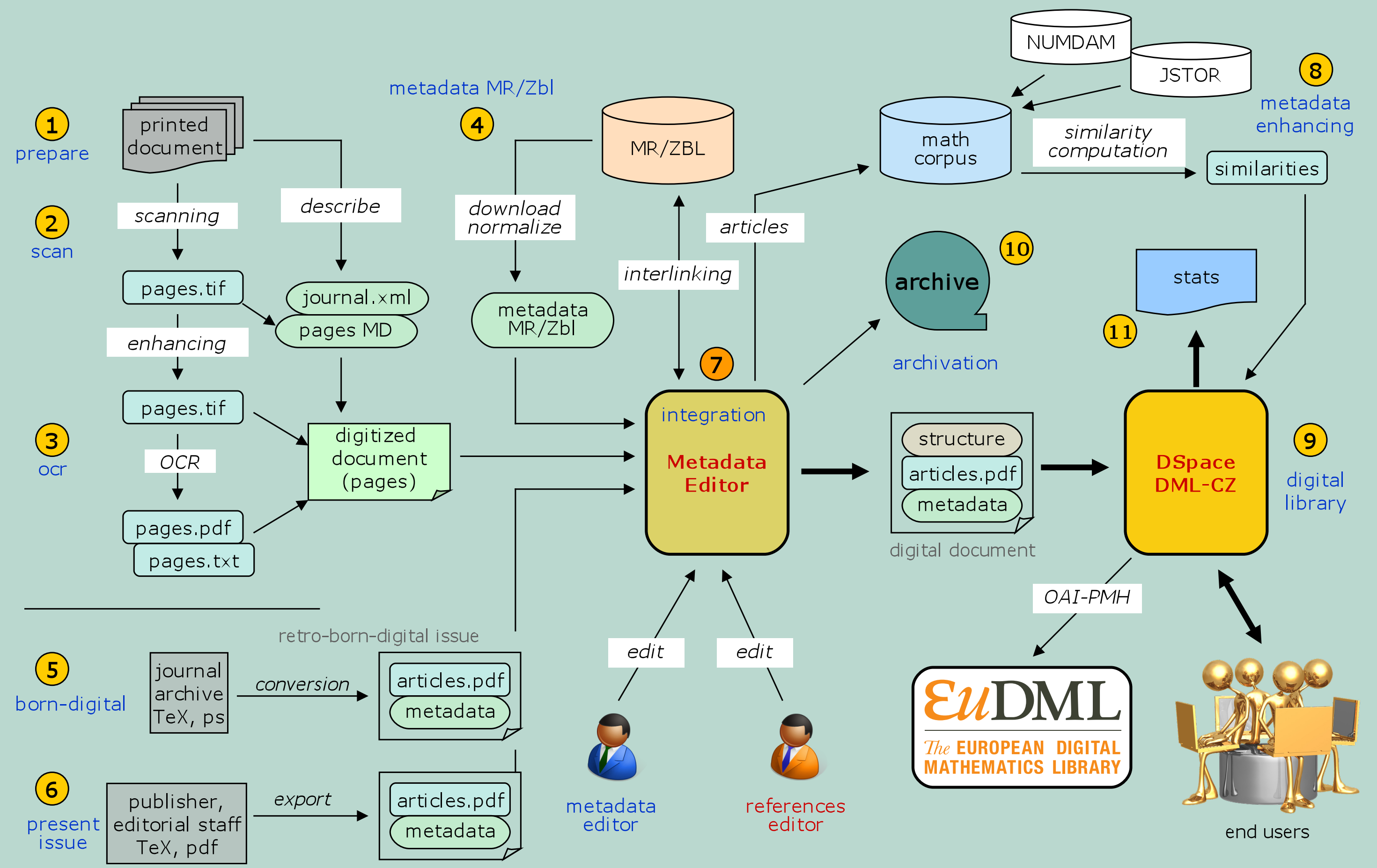
The Metadata Editor (ME) serves as a general tool for managing the content of DML-CZ. Once a new material (e.g. journal issue) comes out it is loaded into ME for further processing. The result is complete data structures with metadata and full texts ready for publishing. ME supports:

- ▶ building the data structures (e.g. journal → volume → issue → article)
- ▶ creation and enhancement of metadata for journals, issues, articles, monographs, chapters, references, ...
- ▶ drag & drop tool for tailoring scanned pages into articles
- ▶ managing name authorities and Mathematics Subject Classification (MSC)
- ▶ interlinking related articles
- ▶ linking to the bibliographic databases Zentralblatt MATH and Mathematical Reviews
- ▶ access control and administration

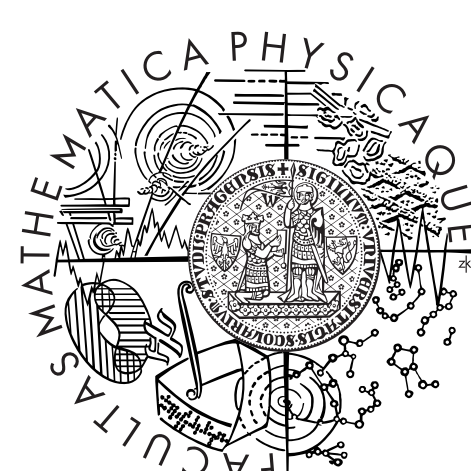


DML-CZ Workflow: from Scanned Images to Final Articles

1. a printed document is prepared for scanning
2. scanning (the method depends on the type of the publication) and enhancing (Book Restorer)
3. OCR is performed on the scans (FineReader, InftyReader) and the outputs are loaded into ME to be processed further
4. metadata found in Zentralblatt MATH and Mathematical Reviews are automatically connected to the articles
- 5, 6. born-digital materials are simply loaded into ME, retro-born-digital materials are preprocessed before loading
7. ME processing: build up the structures, assemble the article, fill in the metadata, check name authorities, create PDFs
8. compute similarities (three different algorithms – LSI, RP and TFIDF)
9. import the content into DSpace, index it for browsing and searching (including full texts) and expose it via OAI-PMH
10. backup and archive content
11. generate statistics



<http://project.dml.cz/>



DML-CZ partners:
Institute of Mathematics, Academy of Science of the Czech Republic, Prague
Institute of Computer Science, Masaryk University, Brno
Faculty of Informatics, Masaryk University, Brno
Faculty of Mathematics and Physics, Charles University, Prague



Authors:
Vlastimil Krejčíř & Miroslav Bartošek
Institute of Computer Science
Masaryk University