DML-CZ: The Experience of a Medium-Sized Digital Mathematics Library

Miroslav Bartošek and Jiří Rákosník

Introduction

The Czech Digital Mathematics Library (DML-CZ) [6] is one of the national initiatives that emerged in the past decade when technical and social circumstances allowed making mathematicians' dreams of a global digital mathematics library (DML) come true. Following the vision approved by the International Mathematical Union [3], taking lessons from the successful French project *NUMDAM* [11], and using the financial support from the Academy of Sciences of the Czech Republic, we started the DML-CZ project [4] in 2005, which after five years ended up in the fully functional DML-CZ [5].

The original goal was to build up a sound basis for a digital archive comprising the relevant mathematical literature published in the territory of today's Czech Republic, endowed with all conceivable features and services, making it a comprehensive, topical, and living DML, generally respected and used by the local as well as the global mathematical community. From the very beginning we had in mind that the DML-CZ should constitute a building block for the envisioned global DML.

Miroslav Bartošek is Head of the Library and Information Centre at Masaryk University. His email address is bartosek@ics.muni.cz.

Jiří Rákosník is researcher and deputy director at the Institute of Mathematics of the Academy of Sciences of the Czech Republic. His email address is rakosnik@math.cas.cz.

In this note we want to look back upon the original goals compared to actual achievements, share the experience, and demonstrate how such a medium-scale project can serve the community and contribute to the global vision. We shall discuss various features of the DML-CZ: its content, purpose and usage, system, functionality, access and sustainability, readiness for integration into global structures, and possible further development.

Content

Mathematical literature that has been published by Czech publishers is fairly varied. Our ambition was to build an open comprehensive library for a wide range of users rather than a mere archive of specialized literature for researchers in mathematics. Of course, the research journals, selected conference proceedings, and monographs form the core of the library with emphasis put on validity, topicality, and currency.

The DML-CZ contains virtually all research journals that have been published by Czech publishers since the nineteenth century. Ten of them are still publishing and contributing their new content to the DML-CZ on a regular basis. Research journals represent not only the most important section of research literature but also the easiest one to handle in a DML.

Conference proceedings constitute a little more complex issue requiring more careful selection and laborious processing. Every year there are

many different meetings, and the proceedings have different technical and content quality. Therefore only six series have been included so far.

Digitizing research monographs and providing free access to them requires a particularly careful assessment of the copyright. This is why only a very few recent books have been included so far. The monographs section primarily contains a valuable collection of twenty-five mathematical works by the famous Bernard Bolzano. Then there is the ever-growing number of books devoted to the history of mathematics, both Czech and international. The rest ranges over a selection of historical scripts from the collection of the Royal Czech Society of Sciences, some popular undergraduate textbooks, and a couple of specialized monographs for which the copyright was easy to obtain.

At the public's request, a special section called *Eminent Czech Mathematicians* was established two years ago. At the moment it contains the private archive of Otakar Borůvka; additional ones (of Vojtěch Jarník, Matyáš Lerch, and Eduard Čech) are in preparation. The peculiarity of such collections is in the large variability of the content: research works scattered in a variety of journals, most of which are not included in the DML-CZ; monographs; textbooks; lecture notes and even manuscripts in multiple editions; newspaper articles; and other people's works about a mathematician. Even though the extreme heterogeneity of such personal collections goes far beyond the standard DML structure, their value is unquestionable. The same idea has been independently adopted by the Biblioteca Digitale Italiana di Matematica [2], which presents the collected works of the eminent Italian mathematician Salvatore Pincherle.

Certain technical complications might be connected with other relevant literature, notably theses, which for various reasons (extremely diverse form and technical quality, copyright, necessity of cooperation by universities) still remains in the realm of contemplation.

DML-CZ content growth in the post-project phase

	December 2009	June 2011	December 2012
Journals	11	12	13
Conference series	6	6	6
Monographs	32	65	108
Pages	275,220	313,707	349,988
Articles/Chapters	25,784	30,475	33,179
Issues/Volumes	2,223	2,619	2,846

Purpose and Usage

The DML-CZ has been built to serve a variety of users. The needs of the main group of users—the community of researchers—are met by the access provided to scientific journals, proceedings, and monographs. The general policy guarantees that these exclusively represent validated items published by authorized publishers and that all of them can be freely accessed after a certain fixed period elapses. This is secured by formal contracts with publishers, all of which are not-for-profit entities: universities, research institutes, and learned societies.

Researchers particularly take advantage of the set of specialized services which in addition to standard browsing and searching comprise browsing by MSC codes, links to Zentralblatt and MathSciNet, searching for similar articles, and recently—through the European Digital Mathematics Library (EuDML) [8]—also the search for mathematical formulae. The response from users shows that the rare possibility to access all back volumes of the journals and conference proceedings from one place is highly valued.

Another important target group is historians of mathematics and science. They benefit from the centralized access to the retrodigitized serials (the oldest one, *Časopis pro pěstování matematiky a fysiky*, started in 1872), the specialized monograph series *History of Mathematics* (published mostly in Czech), the already-mentioned Bolzano collection and other old scripts, the series of almanacs documenting the 150-years history of the Union of Czech Mathematicians and Physicists, and the section of collected works of eminent Czech mathematicians.

Teachers and students will find not only textbooks and scientific and historical articles in the DML-CZ, but also a wealth of instructive texts, collections of solved problems, and educational materials that were published regularly in special appendices to *Časopis pro pěstování matematiky a fysiky* until the 1950s and which would be forgotten otherwise.

The library is also of merit for physicists as four of the periodicals also used to publish or still publish articles about physics and astronomy as well as mathematics.

An important kind of user is publishers. For them DML-CZ means a reliable archive independently preserving their collections, enhancing the content with special tools, services, and upgraded metadata. Publishers and their products gain better visibility through the DML-CZ and, subsequently, increased citation rates. The DML-CZ set up a workflow for each publisher which complies with the typesetting system used and enables a more-or-less automatic production of inputs to the DML-CZ [12].

System

What turns a mere heap of documents into a library is metadata and services. One of the crucial tasks in the DML-CZ project was to create a comprehensive software system which would facilitate the production of the DML's content

(production system) and provide users with access to the DML (presentation system). It is made of a combination of specialized tools developed in the project and freely available open-source systems adapted to the needs of the DML-CZ.

In the core of the *production system* is Metadata Editor [1]—the extensive Web application integrating all activities related to processing the digital content, creating metadata, and interconnecting information sources. Generally, materials to be incorporated in a DML are of three types and have to be processed in three different ways:

(1) *Printed* documents pass through standard digitization workflow involving scanning the printed originals, recognition of characters considering specifics of mathematical texts, grouping page images into digital documents, creating and adjusting descriptive, structural, and administrative metadata, etc.

(2) *Partly digital* documents whose digital form is incomplete or unsatisfactory (for instance, if only the final presentation form is available without the digital source files) are processed in a series of semiautomatic transformations and manually completed into a full digital form according to DML standards.

(3) *Born-digital* documents often also have to be converted into the standardized form, but this can usually be done automatically, including creation of all necessary metadata. This highly effective procedure is typically applied to new-born journal issues generated by suitably adapted editorial workflows for direct ingestion of digital data with a minimum of manual work.

Metadata Editor in combination with other tools including various validation procedures and the module for disambiguation of authors proved to be a very useful and efficient device that significantly reduces the work and facilitates creation of the digital library. It is the fundamental production tool for populating the DML with quality data.

For *presentation* of the content and the DML-CZ interface, DSpace software [7] was chosen, a general open-source tool for implementation of digital libraries and documentary repositories. Being largely adapted for the DML-CZ needs [10], DSpace offers most of the basic functions of digital libraries and services for the end users, which thus need not be carried out by the DML designers: user interface, indexing, document search, browsing information sources, persistent document identification, providing metadata for harvesting through the OAI-PMH protocol, support for long-term preservation of digital data, etc.

Separation of the production and presentation systems into two independent parts enabled designing and implementing the system in a more efficient way.

Features and Functionality

As mentioned above, end users employ the DML-CZ through the suitably modified DSpace repository software. The vision was to not only offer basic functions which belong to the standard repertoire of today's digital libraries but also to provide a certain added value in view of the possibilities, foreseen needs, and expectations of the mathematical community. The basic functionality is grounded in the native services of DSpace: searching, browsing, interlinking, communication with users, system functions. The unique services with added value implemented by the project team are represented by the search for similar articles based on machine text analysis and by experimental search for mathematical expressions using the TEX recording or MathML representation.

Among the basic functions, the key one is searching. The user can choose either a simple Google-style search in both the metadata and the full texts or an advanced search in metadata by selected fields and/or by a specified scope of data. Browsing offers the possibility of information "discovery" by going through available collections, hierarchically structured documents (e.g., journal volumes and issues), or specific indexes (author index, title index, Mathematical Subject Classification).

In addition to a simple intuitive access to chosen segments of scholarly literature, the browse functionality provides an illustrative insight into the overall content and structuring of a digital library.

A particular challenge is represented by the multilinguality of the DML-CZ content—a typical feature for the Central European region and for the literature fifty years old and older. The DML-CZ comprises documents in no less than twelve different languages, including such less-known ones as Czech, Slovak, Polish, Croatian, Latin, and Greek. There are no feasible tools for reasonable translations into English, especially if mathematical rigor is required. The DML-CZ's general policy is to provide translations of all titles in English and in many cases also in German or French. Providing higher-level multilingual services (e.g., multilingual key words) is a difficult problem which should be solved on a larger platform than the small national DML.

The important advantage from which the DML-CZ users benefit is the rich interconnection of available information both inside the DML-CZ and to the outside environment. The inner linkage

Czech Digital Mathematics Library

Search

[ ] Go

Advanced Search

Browse

⇒ Collections
⇒ Titles
⇒ Authors
⇒ MSC

About DML-CZ

DML-CZ Home >

Welcome to **DML-CZ** (Czech Digital Mathematics Library), the website offering an open access to the metadata and fulltext of mathematical journals, proceedings and books published throughout history in the Czech lands.

Search DML-CZ

Enter the searched text in the box below

[ ] Go

Browse collections

Journals

Acta Universitatis Carolinae. Mathematica et Physica
Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica
Applications of Mathematics
Archivum Mathematicum
Časopis pro pěstování matematiky
Časopis pro pěstování matematiky a fysiky
Commentationes Mathematicae Universitatis Carolinae
Communications in Mathematics
Czechoslovak Mathematical Journal
Kybernetika
Mathematica Bohemica
Mathematica Slovaca
Pokroky matematiky, fyziky a astronomie

Proceedings

DML
EQUADIFF
NAFSA
Toposym
WSAA
WSGP

Monographs

Bolzano Collection
Dějiny matematiky
Single Books
Z dějin Jednoty českých matematiků a fyziků

Eminent Czech mathematicians

Borůvka, Otakar

The DML-CZ main page.

allows the browsing user to easily move from one entity to another in the DML-CZ, for instance from an article to the author's record with the list of all his/her works, to a list of articles with the identical MSC code, or to a possible continuation of that article. The external links offer connections from an article to its records in the reference databases Zentralblatt and MathSciNet, and this holds even on the level of individual bibliographical references. The persistent identifiers based on the

Handle system [9] enable easy building of reliable connections in the opposite direction too—from the outside into the DML-CZ, e.g., from an EuDML record or from the reference databases to the parent record and corresponding full text in the DML-CZ. As the vast majority of the DML-CZ content is open access, this is a step towards the envisioned global broadly interconnected network of mathematical information resources. In 2012 we started implementing interconnection with

CrossRef through DOI identifiers, the pilot journal being *Archivum Mathematicum.*

In this way, the DML-CZ plays an important role in the national community as a pioneer and intermediary of global infrastructure and modern technologies which would otherwise be much harder for local publishers, universities, and research institutes to adopt individually.

Access and Sustainability

The DML-CZ has been built with public support to form a specialized open access digital library. Its content ranges from immediate open access content to materials which would eventually be freely accessed after a certain period that depends on a publisher's policy, with the maximum being twenty-four months.

The general rule is that the DML-CZ owns the created metadata, while the ownership of the digital content (full texts) remains with the content provider or IPR owner, and the DML-CZ (the Institute of Mathematics of the Academy of Sciences of the Czech Republic (AS CR)) is licensed to archive and possibly enhance the content and to make it freely accessible online for strictly noncommercial purposes. This includes the license for re-serving the content through higher structures like the European Digital Mathematics Library. The Institute, as the public research institution, guarantees that the DML-CZ remains a service representing the public interest.

Access to the DML-CZ content

	Articles/Chapters Total	Open Access	Open Access (%)
Journals	28,759	28,349	98.6
Conference series	2,452	2,452	100.0
Monographs	1,968	1,968	100.0
Total	33,179	32,769	98.8

Building the digital library was exciting work for a team encouraged by the project task and necessary funding. At the project's end, there is the DML-CZ as a functional prototype with a critical core content and an open future. However, to assure its long-term viability and further development, it was necessary to take measures in advance, during the project phase. The core of the project team expressed their explicit interest in continuing. The Institute of Mathematics AS CR as the project coordinator and DML-CZ owner assumed the overall responsibility for the maintenance and resources, acquisitions, intellectual property rights issues, and publicity. Mathematicians' expert supervision is provided in collaboration with the Charles University in Prague and the Czech Mathematical Society. The team from the Masaryk University in Brno, where the DML-CZ core technology was developed, maintains both the digital library and system and secures further

technological development. Retrodigitization is carried out in the Digitization Centre of the Academy Library with the use of workflow and tools developed during the project phase.

This is the DML-CZ sustainability model. In the future, we shall concentrate on continuous content growth and data improvement while for further technological development we shall mostly rely on joint efforts within the EuDML structure to which the DML-CZ wants to actively contribute. For instance, the search for semantically similar documents which was originally developed in the DML-CZ can be enhanced using the much larger reference corpus associated with the EuDML. Some features like mathematical formulae search will be better left on the common EuDML portal.

Even though DML-CZ is a not-for-profit activity, there occur certain expenses which require a sustainable plan. The basic costs related to system maintenance and ingestion of new-born journal issues do not exceed the equivalent of one IT developer's FTE and are proportionally shared by the main and regular content providers—journal publishers. Additional costs connected with special development or integration of irregular or new-type content are covered by the Institute of Mathematics AS CR or by the ordering party. Needless to say, all these obligations are based on long-term contracts which regulate operational and financial conditions as well as the IPR issues.

Conclusions and Future Work

The DML-CZ technology, experience, and expertise already found their application in various external activities. They helped to develop the Digital Library of the Faculty of Arts at Masaryk University. Some tools developed in the project were used to enhance the library system Kramerius, widely used by the Czech public libraries. Preparatory works are ongoing to apply Metadata Editor and other tools for several EuDML partners. This includes enhancement of the Bulgarian Digital Mathematics Library (Bul-DML) and upgrading the metadata of journals in the Spanish Digital Mathematics Library (DML-E) and *Journal of the EMS* published by the European Mathematical Society Publishing House. For the latter, still subject to further negotiation, the DML-CZ may serve as a mediator content provider to the EuDML.

During the two years of independent operation following the project phase, the DML-CZ proved beyond any doubt its viability and general utility. This is well documented by the stable rate of more than 100,000 visitors per year with a day peak of 600. Most of the visitors come from the Czech Republic (19.7 %), followed by the USA (8.5 %), India (6.1 %), Germany (5.4 %), and China (4.8 %).
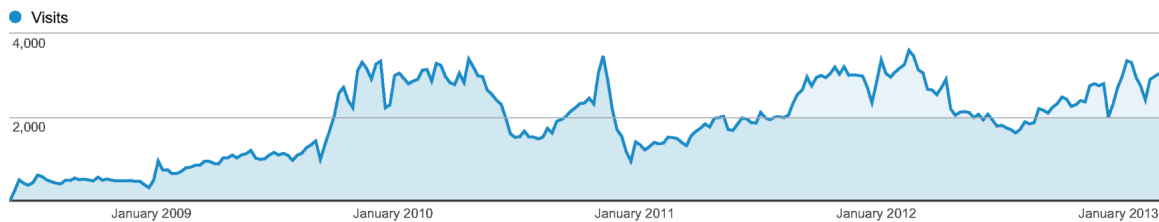
Visits per week to the DML-CZ (source: Google Analytics).

Through the Institute of Mathematics AS CR and Masaryk University, the DML-CZ became a partner in the large consortium for a three-year project aiming at the EuDML, supported by the European Commission.

Cooperation with the EuDML partners and integration of the DML-CZ into the EuDML portal provided the occasion for enhancement of the data, substantial DSpace OAI-PMH system reconfiguration, and—last but not least—interlinking with a large number of documents distributed in a network of other DMLs and thus increasing the visibility and utility of the DML-CZ content. This provides a great challenge as well as opportunity to contribute to the further development of the EuDML, which has been recognized as a basis for the envisioned global DML.

Acknowledgments

References

[1] Miroslav Bartošek, Petr Kovář, and Martin Šárfy, DML-CZ Metadata Editor. Content creation system for digital libraries, in *DML 2008—Towards Digital Mathematics Library* (P. Sojka, ed.), Proceedings of the workshop held in Birmingham, UK, July 27, 2008, Brno: Masaryk University, 2008, pp. 139–151. `http://dml.cz/handle/10338.dmlcz/702537`.

[2] bdim: Biblioteca Digitale Italiana di Matematica, `http://bdim.eu`.

[3] Committee on Electronic Information Communication of the International Mathematical Union, Best current practices: Recommendations on electronic information communication, *Notices of the AMS* **49**(8) (September 2002), pp. 922–925. `http://ams.org/notices/200208/comm-practices.pdf`.

[4] The Czech Digital Mathematics Library, Project funded by the Academy of Sciences of the Czech Republic, 2005–2009. `http://project.dml.cz`.

[5] The Czech Digital Mathematics Library. Final Report, `http://project.dml.cz/docs/dmlcz_finalreport_2010.pdf`.

[6] DML-CZ: The Czech Digital Mathematics Library, `http://dml.cz`.

[7] DSpace, `http://dspace.org`.

[8] The European Digital Mathematics Library, `http://eudml.org`.

[9] Handle System, `http://handle.net`.

[10] Vlastimil Krejčíř, Building the Czech Digital Mathematics Library upon DSpace system, in DML 2008—*Towards Digital Mathematics Library* (P. Sojka, ed.), Birmingham, United Kingdom, July 27, 2008, Masaryk University, Brno, 2008, pp. 117–126, `http://dml.cz/handle/10338.dmlcz/702539`.

[11] NUMDAM: Numérisation de documents anciens mathématiques, `http://numdam.org`.

[12] Michal Růžička, Automated processing of TEX-typeset articles for a digital library, in DML 2008—Towards Digital Mathematics Library, (P. Sojka, ed.), pp. 167–176 (2008), Birmingham, UK, July 27, 2008, `http://dml.cz/handle/10338.dmlcz/702533`.