Jiří Rákosník

Recent Development of the DML-CZ and Its Current State

In: Petr Sojka and Thierry Bouche (eds.): Towards a Digital Mathematics Library. Bertinoro, Italy, July 20-21st, 2011. Masaryk University Press, Brno, Czech Republic, 2011. pp. 9--14.

Persistent URL: http://dml.cz/dmlcz/702597

Recent Development of the DML-CZ and Its Current State

Jiří Rákosník

Institute of Mathematics AS CR, Žitná 25, 115 67 Praha 1, Czech Republic
rakosnik@math.cas.cz

Abstract. The project *DML-CZ: The Czech Digital Mathematics Library* has been implemented since 2005 and in 2010 switched over to routine operation. This report describes progress, growth and usage of the DML-CZ, the experience from cooperation with content providers in the designed editorial workflow, some newly implemented features, adjustments of the workflow following from both the ongoing practical experience and the requirements of the advancing EuDML project, the general public acceptance and attendance and the suggested economic model for sustainable development.

Keywords: digital mathematics library, retrodigitization, DML-CZ

1 The Original Project

The Czech Digital Mathematics Library DML-CZ [1] has been developed in frames of the project supported by the Academy of Sciences of the Czech Republic in the period 2005–2009 [2]. The project resulted in a full-featured digital library providing a free access to more than 275,000 pages of digital content in 11 journal titles, 6 conference proceedings series and 32 monographs including a collection of 25 books by the famous Czech mathematician Bernard Bolzano (1781–1848). Structured lists of references contained in metadata of journal articles represented about 160,000 items.

2 Routine Operation and Sustainability

When the funding of the project ended, the project coordinator, the Institute of Mathematics AS CR in Prague, undertook the responsibility for the maintenance and development of the DML-CZ cooperating with other project partners on a non-profit basis. The teams in the Masaryk University in Brno ensure the operation of the DML-CZ on their servers and provide technical service, process new additions and develop further tools. Colleagues from the Charles University in Prague effectively cooperate on provision and enhancement of metadata and the Digitization Centre in the Library of the Academy of Sciences employs the developed workflow for acquisition of new retrodigitized content.

The toolsets and workflow that have been implemented and verified in the project provide technical conditions for regular semiautomatic supply of new journal issues. Validation procedures enable editors to check completeness

Petr Sojka, Thierry Bouche (editors): DML 2011, Towards a Digital Mathematics Library, pp. 9–14.
 ISBN 978-80-210-5542-1

and integrity of data produced for the DML-CZ. The cooperation is based on formal contracts concluded between the Institute of Mathematics and journal publishers. Running costs in the first year of routine operation were covered by the Institute of Mathematics and starting with 2011 they will be shared by all journal publishers proportionally to the volume of delivered material.

The smooth passage to the routine operation of the DML-CZ confirms how important and forethoughtful was the timely decision to devote essential part of the project to development of efficient tools and implementation of a reliable workflow.

3 New Content

The DML-CZ content is growing faster than it has been expected. The journal *Pokroky matematiky, fyziky a astronomie* (*Advances in Mathematics, Physics and Astronomy*) published by the Union of Czech Mathematicians and Physicists since 1956 has been included. This brought new questions into consideration as the journal has rather heterogeneous content comprising original papers for general audience, translations from other journals, discussions on actual problems of mathematics and physics education, announcements, news etc. Major part of the journal published in the pre-TEX era had to be retrodigitized. Including such not purely scientific journal to DML-CZ has an important impact on increasing awareness of mathematics and of the library itself among teachers and students.

The scientific journal *Acta Universitatis Carolinae, Mathematica et Physica* has been recently processed and will be accessed soon. In fact, part of the journal has already been displayed during the project before 2009 in frames of the proceedings series of the *Winter School on Abstract Analysis*. Successful development and public acceptance of the DML-CZ encouraged the publisher to ask for enlistment of the whole journal in the DML-CZ.

The recent growth of DML-CZ is seen from the comparison of its content at the end of project in 2009 and its current content in Table 1.

Table 1. DML-CZ content

	December 2009	June 2011
Journals	11	12
Conference series	6	6
Monographs	32	65
Pages	275 220	313 707
Articles/Chapters	25 784	30 475
Issues/Volumes	2 223	2 619

4 Collected Works

A series of brand new questions emerged with the decision to include a special section devoted to collected works of eminent personalities of Czech mathematics. The first collection (see Figure 1) will be devoted to private archive of late Professor Otakar Borůvka (1899–1995), one of the most important Czech mathematicians in the 20th century who coincidentally worked at two affiliations of DML-CZ partners: the Masaryk University and the Institute of Mathematics AS CR.

Otakar Borůvka's archive contains respectable amount of 209 items representing 3,983 pages. Some of them have already been captured in the DML-CZ but the major part had to be retrodigitized. The collection consists of three main types of works: research works (9 monographs and 81 papers), other works (2 university textbooks and 51 journal and newspaper articles) and works about him (1 monograph, 1 thesis, 64 articles). This structure exceeds the scope of a standard DML focused mainly on research works. However, it meets the public demand and we believe that it belongs to the general mathematical heritage. Collections of further personalities are under consideration.

Investigating the structure of Borůvka's archive we learned that we are facing new problems like treatment of different editions of the same book, different manifestations of the same work (offprint vs. preprint, conference proceedings paper vs. working papers etc.). It came out that it is necessary to switch to the FRBR model (*Functional Requirements for Bibliographic Records*, see [3]) and to work with "creations" rather than with "publications" only. The copyright issues became more complex as well because many items in the archive come from sources which have not been treated in the DML-CZ so far.

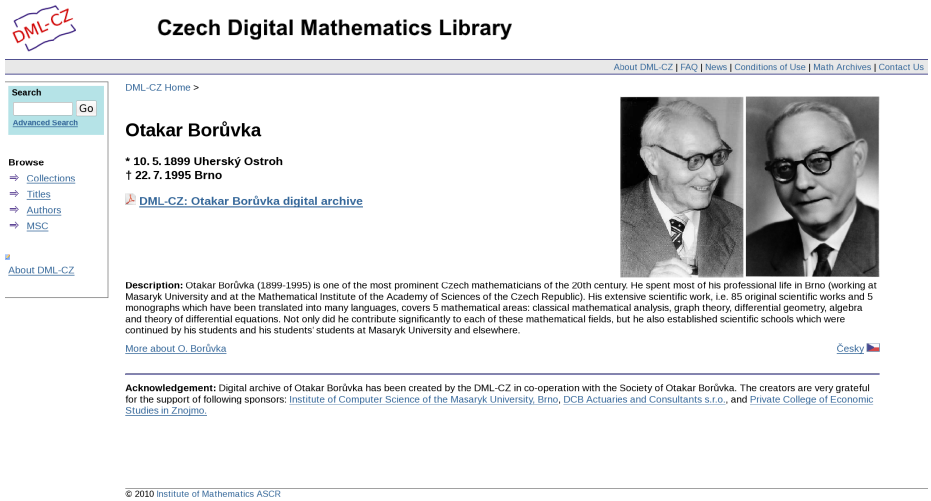

Fig. 1. Otakar Borůvka in DML-CZ

5 Technical Issues

The viability of a digital library rests above all with new acquisitions emerging mainly in the form of born-digital publications. Therefore, the DML-CZ project has been experimenting with automatic born-digital workflows since 2008 [7].

Editors of all journals included in DML-CZ are using tools and workflows that have been tailored to their individual publishing practice and that enable them to produce inputs for DML-CZ in a semiautomatic way. The formal consistency and integrity of the data are controlled by several validating procedures that have been developed in the project. This eliminates the majority of possible defects and decreases the demandingness of the final visual control.

Changes to the original editorial workflow were generally minimal. Automated procedures for validation of data of new journal issues are being gradually improved to catch irregularities, and full primary sources are archived in DML-CZ for internal use and development. Based on limiting the namespace of allowed TEX macros to those supported by our Tralics configuration, the recent improvements aimed at getting all metadata including abstracts, keywords and references transformed into representation using MathML. This was motivated by recent EuDML developments, namely by the possibility of math indexing [8].

Procedures developed during the project phase also helped in adding new journals and monographs or even complex constructs like collected works. Of course, it is less straightforward because many particular problems have to be tackled from the beginning — the structure, individual editorial practice, copyright issues etc.

The careful verification of data preceding their presentation in DML-CZ is done on different levels by different people using the Metadata Editor — a complex web-based system supporting all essential steps in the development of the library, see [4]. It appeared practical to implement a working copy of the DML-CZ presentation in which all the changes are first realized and checked. The working copy is being regularly transferred to the public DML-CZ after the final approval. This arrangement virtually prevents introducing errors to the public version and improves the stability of the public version.

The DML-CZ is one of content providers in the project of the European Digital Mathematics Library [6] and the Institute of Mathematics AS CR and the Masaryk University are partners in the project. Technical requirements stipulated in the project, developed tools and the close cooperation with other project partners have important impact on further development of DML-CZ itself. For the transfer of our metadata to the EuDML core via OAI-PMH a detailed metadata format in the form of tagged Dublin Core has recently been created. However, we find it rather limiting solution as we want to provide much more data than usually available via OAI-PMH, for instance references or even the full texts for indexing. Thus we will switch to providing data in the NLM format [5] in the future, and establish bidirectional secure channels for data exchange between DML-CZ and EuDML.

The Metadata Editor has been recently internationalized so that we can put it at the disposal of those EuDML partners who do not have such tool yet. It has been made more portable and further configuration support has been added. It is available as open source application at `http://sourceforge.net/projects/dme`.

6 Public Acceptance

Thanks to various promotion actions, lectures for mathematicians, librarians, students and teachers, journal articles, radio and TV interviews, the DML-CZ became well-known and exploited in the Czech Republic. Activities of the team in the EuDML project [6] and the fact that DML-CZ is highly ranked by Google Scholar due to the negotiated metadata interface increases its awareness abroad. According to Google Analytics there is a rather stable visit rate around 400 accesses per day. Most of them are naturally from the Czech Republic followed by the USA, Germany, India, Slovakia, China, Iran, France, Poland and the United Kingdom. Approximately 70 % of web traffic to DML-CZ is generated by Google itself. In one month, the site gets about 7,000 unique visitors.

The increasing awareness of DML-CZ is valued not only by the public but also by the cooperating journal editors. Even though it is yet too early for credible conclusions there are indications that the presence of journals in the DML-CZ helps to improve their publicity, increasing access rate to their papers and consequently number of citations.

7 Conclusion

DML-CZ is a living digital library which is growing up from its infancy to maturity while increasing its content as well as the extent and quality of services. Procedures implemented during the project phase proved efficient for inclusion of new material. The structure of DML-CZ has been recently extended with the section of collected works of eminent Czech mathematicians. This required some new arrangements, especially in the metadata scheme. The DML-CZ is profiting from its partnership in the EuDML project and preparing provision of its digital content for the upcoming EuDML.

Acknowledgement. This work is partly financed by the European Union through its Competitiveness and Innovation Programme (Information and Communications Technologies Policy Support Programme, “Open access to scientific information”, Grant Agreement No. 250503).

References

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

3. *Functional Requirements for Bibliographic Records.* http://en.wikipedia.org/wiki/Functional_Requirements_for_Bibliographic_Records.
4. Miroslav Bartošek, Petr Kovář and Martin Šárfy: *DML-CZ Metadata Editor. Content Creation System for Digital Libraries.* In: Sojka, P. (ed.) DML 2008 – Towards Digital Mathematics Library. Proceedings of the workshop held in Birmingham, UK, July 27th, 2008. Brno: Masaryk University, 2008, pp. 139–151.
5. *National Library of Medicine Journal Archiving and Interchange Tag Suite.* http://dtd.nlm.nih.gov/
6. *The European Digital Mathematics Library.* CIP-ICT-PSP project No. 250503. http://eudml.eu.
7. Michal Růžička: *Automated Processing of TEX-Typeset Articles for a Digital Library*. In: Sojka, P. (ed.) DML 2008 – Towards Digital Mathematics Library. pp. 167–176 (2008), Birmingham, UK, July 27th, 2008.
8. Sojka, P., Líška, M.: Indexing and Searching Mathematics in Digital Libraries – Architecture, Design and Scalability Issues. In: Davenport, J.H., Farmer, W., Rabe, F., Urban, J. (eds.) Proceedings of CICM Conference 2011 (Calculemus/MKM). Volume 6824 of Lecture Notes in Artificial Intelligence, LNAI, Berlin, Germany, Springer-Verlag (2011) 228–243.