

Bulgarian Digital Mathematical Library BulDML and Czech Digital Mathematical Library DML-CZ as parts of the European Digital Mathematics Library EuDML

Jiri Rakosnik¹, Peter Stanchev^{2,3}, George Simeonov², Radoslav Pavlov²

¹Institute of Mathematics AS CR, Czech Republic
rakosnik@math.cas.cz

²Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria
gsimeo@math.bas.bg, radko@cc.bas.bg

³Kettering University, Flint, USA
pstanche@kettering.edu

Abstract. The paper presents in brief the Bulgarian Digital Mathematical Library BulDML and the Czech Digital Mathematical Library DML-CZ. Both libraries use the open source software DSpace and both are partners in the European Digital Mathematics Library EuDML. We describe their content and metadata schemas; outline the architecture system and overview the statistics of its use.

Keywords: Digital Mathematical Library, Open Access, Institutional Repositories, DSpace, Publishing

1 Introduction

Recently there have emerged several local initiatives building digital mathematics libraries. The Bulgarian Digital Mathematical Library (BulDML) [2] at the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences (IMI-BAS) and the Czech Digital Mathematical Library (DML-CZ) [3] at the Institute of Mathematics of the Academy of Sciences of the Czech Republic are two of them. There are some common characteristic features: both DMLs focus on national past and present production, their content is multilingual, they are using the open source software DSpace and they both are participating in the European network of mathematical libraries associated in the project of European Digital Mathematics Library (EuDML) [15].

The EuDML is a project of 13 partners from 9 European countries partly supported by the EU. Its aim is to provide common framework, standards and services for a unified seamless access to the distributed heterogeneous local digital repositories containing relevant mathematical literature published in Europe including periodicals, selected monographs and conference proceedings from the past as well as the currently produced mathematical publications. The BulDML and DML-CZ are part of this network, provide their content to EuDML and meet requirements and standards

for compliance and interoperability that have been developed in the project. The EuDML pilot release 1.3 has been recently made available to public [8].

2 Content and Metadata

BulDML and DML-CZ include journal articles, reports and papers in conference proceedings in addition, selected books, keeping the content grow up over time. The description of all items is based on the well-known and broadly accepted Fully Qualified Dublin Core (QDC) metadata schema [7]. However, it appeared that the most used Dublin Core metadata fields, i.e. the set title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, does not fit specific requirements for mathematics. For this purpose, a more detailed metadata schema has been analyzed and defined and an initial set of recommended best practices has been developed in the EuDML project [14]. Three levels of metadata have been identified. *Obligatory metadata* is the required minimum of metadata in order to unambiguously identify and handle a relevant mathematical publication in the scope of EuDML: item type, authors, original title, bibliographic reference for this publication with enough structure to enable collection's browsing, unique identifier, URL of full text. *Fundamental metadata* is what satisfies the functional requirements for browsing, searching and reference matching over the collections at item level. It enables basic digital library interaction with the EuDML corpus. *Supplemental metadata* is whatever goes beyond fundamental metadata (e.g. relations to subject ontologies, authority lists, MR/ZM IDs, multilingual, multi-script, bibliographies/references, interlinking, math handling, etc.). They are relevance to the EuDML's corpus specificities and EuDML system functionalities.

The system-to-system interoperability of BulDML and DML-CZ with the EuDML network of digital libraries is implemented through OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) [12]. The OAI-PMH defines a protocol and interface that allows EuDML metadata aggregator to harvest metadata from all EuDML compliant content providers, i.e. digital libraries, repositories and archives. The currently supported metadata format in OAI-PMH is Dublin Core and metadata is delivered to the aggregator in XML via HTTP protocol.

Currently, BulDML contains more than 1500 publications from six journals and two proceedings:

Mathematical Journals

- Serdica Journal of Computing
- Serdica Mathematical Journal
- Mathematica Balkanica New Series
- Fractional Calculus and Applied Analysis

Multidisciplinary Journals

- International Journal Information Theories and Applications – IJ ITA
- International Journal Information Technologies and Knowledge – IJ ITK

Proceedings

- ADIS
- REMIA

The DML-CZ offers more than 32 500 publications comprising 28502 papers in 13 journals, 2030 papers in proceedings of 6 conference series, 2006 chapters in al-most 100 monographs:

Mathematical Journals

- Acta Mathematica Universitatis Ostraviensis / Communications in Mathematics
- Applications of Mathematics
- Archivum Mathematicum
- Časopis pro pěstování matematiky
- Commentationes Mathematicae Universitatis Carolinae
- Czechoslovak Mathematical Journal
- Mathematica Bohemica

Multidisciplinary Journals

- Acta Universitatis Carolinae. Mathematica et Physica
- Acta Universitatis Palackianae Olomucensis. Mathematica
- Časopis pro pěstování matematiky a fyziky
- Kybernetika
- Mathematica Slovaca
- Pokroky matematiky, fyziky a astronomie

Proceedings

- Equadiff
- Nonlinear Analysis, Function Spaces and Applications
- TopoSym
- Winter School on Abstract Analysis
- Winter School "Geometry and Physics"
- DML – Digital Mathematics Library

The books segment contains a valuable collection of Bernard Bolzano's writings, a series of books on history of mathematics and a selection of individual monographs. In addition, the DML-CZ introduced a special section devoted to eminent Czech mathematicians, which currently contains collected works of Otakar Borůvka.

Each item in the digital libraries represents a pair of digital full text in searchable PDF and detailed metadata in XML physically archived in the corresponding DML. In conformity with the general EuDML policy, there is a free access to majority of full texts. There is exception of the most recent publications in some journals, whose publishers implement the moving wall to delay online free access over a certain period.

3 System Overview

A common feature of BuDML and DML-CZ is that they both use the DSpace software in their system implementation and as an interface with respect to the EuDML. However, the ways they do it are different.

DSpace [6] is open source software designed to ease submission of digital content with a very simple user interface. It enables implementation of basic services and functionalities for presentation, archiving, local search, system-to-system interoperability and long-term preservation of digital content of various types and formats. DSpace uses Apache Lucene as a search engine, allowing to search in metadata and full texts and provides a variety of browsing options. The digital objects stored in DSpace have persistent identifiers, based on the Handle.net system [9]. DSpace contains OAI-PMH server and supports OpenURL [13].

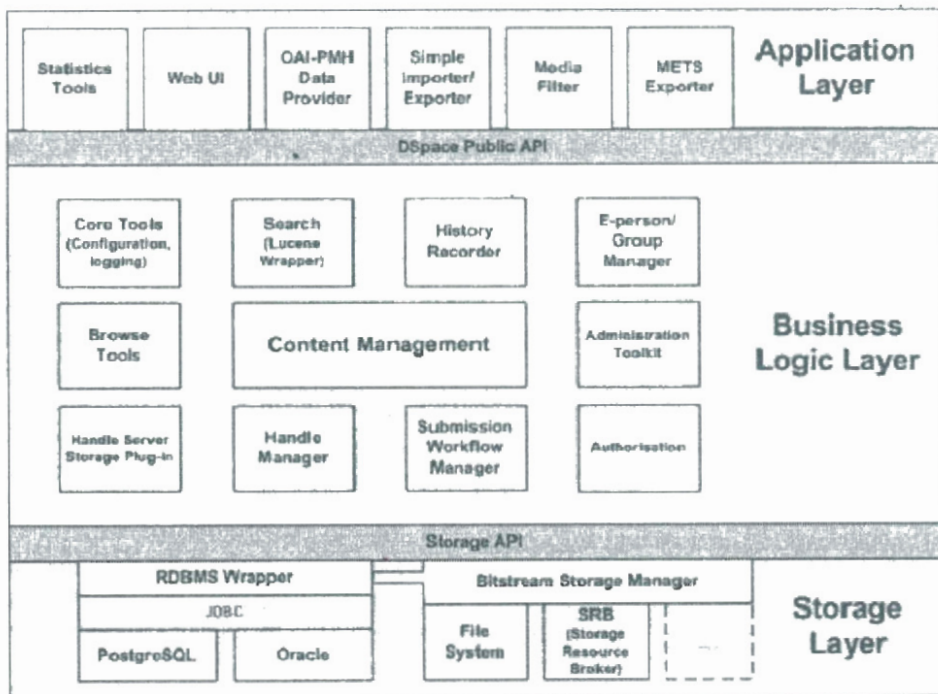


Fig. 1. DSpace layers

DSpace layers are given in Figure 1. The *storage layer* is responsible for physical storage of metadata and content. The *business logic layer* deals with managing the content of the archive, users of the archive (e-people), authorization, and workflow. The *application layer* contains components that communicate with the world outside of the individual DSpace installation, for example the Web user interface and the OAI-PMH for metadata harvesting service. Each layer only invokes the layer below it; the application layer may not use the storage layer directly, for example. Each

component in the storage and business logic layers has a defined public API. The unions of the APIs of those components are referred to as the Storage API (in the case of the storage layer) and the DSpace Public API (in the case of the business logic layer). These APIs are in-process Java classes, objects and methods.

BulDML is based on the DSpace version 1.52, with slightly modified source code for practical reasons to keep consistency with future software releases and make more easily migration from old to a new software versions, because there are still serious issues with not fully implemented DSpace Backup and Restore also known as AIP (Archival Information Package). Mainly for submitting content and metadata in BulDML DSpace native web interface without any difficulties is used. There are some exceptions for importing content. We have used scripts in Python to generate structure of DSpace Simple Archive Format and other administrative tasks with shell scripts and Cron. BulDML has customized web interface, which provides very simple and easy user navigation. DSpace creates collection redundancies if we want to keep journals volumes and numbers organized in different collections. The reason is that in DSpace has no sub-collections and proper navigation for a large number of collections. Currently in DSpace even in the latest versions for almost all changes in web interface we can't avoid to stop server and rebuild entire system. We aim to achieve in future independence of user web interface from other parts of DSpace source code. BulDML has all necessary key features provided by DSpace software platform for building digital library in keeping with well-known standards for interoperability, long-term digital preservation, storage services, etc.

In DML-CZ a customized DSpace version 1.4 is used. For practical reasons it was decided to implement only the reduced Application layer of DSpace for presenting the data to the end users. Most of the work is done in Metadata Editor [1], the specialized complex prototype software developed in the DML-CZ project for storage and handling data, editing and enhancing metadata including their automated verification, bibliographical references processing and creating the final PDFs. The data prepared in the Metadata Editor are periodically imported in the DSpace. For this reason special importing tools had to be developed using Java and DSpace API because the DSpace own import tool was not compatible with the Metadata Editor. Some additional features like Mathematical Subject Classification codes [11], articles interlinking an article counts by authors, which are required for a fully functional digital mathematics library but which DSpace does not provide, had to be implemented. For more details, see [10].

Some key system features in BulDML and DML-CZ are:

– **Permanent Identifiers**

For each published item/article ID or Handle Prefix unique in the scope of entire Internet has been assigned, which is provided by Handle.NET system service. Handle.NET operates in a very similar way to well known Domain Name Service (DNS), all handles are URLs managed and resolved by Handle server. The benefits are that external links pointed to the content in the digital library are permanent and unbreakable, because they provide independence on subsequent changes in physical location of server machine, Internet addresses IP, domain name of the server, or

even changes in filename and/or location of the document when it is moved from one collection to another. Irrespective of all these possible changes the permanent identifiers remain unbreakable and with resolution service they grant instant access to the digital object/resources.

– **Detecting accidental errors in data storage**

The integrity of the data of digital objects is provided by MD5 hash sum (checksum). Digitalized content is stored on physical memory devices, which are prone to failure, and data can be changed, corrupted or even completely lost. This is standard scheduled system routine for generating checksums for the new content and checking all already stored content by re-calculating checksums and comparing them with original stored checksums.

– **Unicode everywhere**

(UTF-8) encoding is applied for all databases, text content storage, queries, web interface. There are no language restrictions for the content of web pages and for the text entered in metadata fields.

– **Embedded document viewer**

DSpace has been modified to provide document viewer - for scanned articles/books it is useful to have preview of separate pages without downloading large files of entire article/book.

4 Statistics of BulDML and DML-CZ Use

The tables in Figure 2 show the number of visit in BulDML and DML-CZ.

BulDML

	2010	2011	Jan 1, 2012 – Jul 30, 2012
Total visits	5 658	8 969	5 708
Unique visitors	4 193	7 105	4 596
Pageviews	18 742	29 005	15 603
Total visits from Bulgaria	2 197	3 266	2 057

DML-CZ

	2008	2009	2010	2011
Total visits	13 960	74 487	124 780	109 913
Unique visitors	10 687	58 782	97 195	82 367
Pageviews	88 421	276 984	441 929	401 914
Total visits from the Czech Republic	2 286	8 554	22 616	18 595

Fig. 2. Survey of visits

Some data about the traffic in BulDML and DML-CZ are given in Figure 3.

BulDML

Source / Medium	Visits
1. google / organic	3 390
2. math.bas.bg / referral	842
3. (direct) / (none)	434
4. scholar.google.com / referral	128
5. nbu.bg / referral	67
6. eudml.org / referral	58
7. roar.eprints.org / referral	28
8. yandex / organic	26
9. oalib.com / referral	20
10. scholar.google.bg / referral	20

DML-CZ

Source / Medium	Visits
1. google / organic	231 163
2. (direct) / (none)	38 457
3. seznam / organic	4 938
4. search / organic	2 870
5. google.com / referral	2 223
6. google.cz / referral	1 149
7. yahoo / organic	1 144
8. scholar.google.com / referral	1 022
9. en.wikipedia.org / referral	850
10. jcmf.cz / referral	792

Fig. 3. Traffic data in BulDML and DML-CZ

5 Conclusions

BulDML and DML-CZ are digital libraries built to preserve and provide free access to national heritage in mathematics. Although developed independently, they have a lot in common with respect to structure and development of content, services to users, access policy of guaranteed eventual free access, and the DSpace software used for presentation. Both libraries fulfill minimal metadata requirements for their content be presented in the unique framework of the European Digital Mathematics Library of which they are partners.

Acknowledgment. This work was supported in part by the EU project "Open Access Infrastructure for Research in Europe" (OpenAIRE), "2nd Generation Open Access Infrastructure for Research in Europe" (OpenAIRE+) and "European Digital Mathematics Library (EuDML)".

References

1. Bartošek M., P. Kovář, and M. Šárky, DML-CZ Metadata Editor. Content creation system for digital libraries, In: DML 2008 – Towards digital mathematics library, Birmingham, UK, July 27th, 2008. Proceedings. Brno: Masaryk University, 2008, 139–151
2. Bulgarian Mathematics Digital Library, <http://sci-gems.math.bas.bg>
3. Czech Digital Mathematics Library, <http://dml.cz>
4. DSpace Resources, <https://wiki.duraspace.org/>
5. DSpace, Example of the new possibilities for DSpace when adding SKOS thesaurus and authority lists management, DSpace User Group Meeting (DSUG) 2009 in Gothenburg
6. DSpace, <http://dspace.org/>
7. Dublin Core Metadata Initiative, <http://dublincore.org/>
8. EuDML beta 1.3, <http://eudml.org>
9. Handle System, <http://www.handle.net/>
10. Krejčíř V., Building the Czech Digital Mathematics Library upon DSpace system, In: DML 2008 - Towards digital mathematics library, Birmingham, UK, July 27th, 2008. Proceedings. Brno: Masaryk University, 2008, 117–126
11. Mathematical Subject Classification 2010, <http://www.zentralblatt-math.org/msc/en/>
12. Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH), <http://www.openarchives.org/pmh/>
13. OpenURL, <http://en.wikipedia.org/wiki/OpenURL/>
14. The EuDML Metadata Schema. Deliverable 3.2 in the project of the EuDML, <http://project.eudml.org/work-package-3-content-aggregation>
15. The Project of the European Digital Mathematics Library, <http://project.eudml.org>