Digital Mathematical Libraries: Overview of Implementations and Content Management Services

© A.M. Elizarov © E.K. Lipachev © D.S. Zuev

Volga Region Federal University,
Kazan, Russia

amelizarov@gmail.com elipachev@gmail.com dzuev11@gmail.com

Abstract. The paper gives a review of existing projects of implementation of digital mathematical libraries. An analysis of existing information systems of digital mathematical libraries is performed using the evaluation criteria embedded in the DELOS DLRM model, emphasis is placed to the methods of managing mathematical content on the basis of semantic technologies. All projects are in different degrees of completeness, the range of services provided is different. We found that most of digital mathematical libraries are concentrated on the transfer of the resources to the electronic form and their preservation, rather than on the development of semantic services.

Keywords: digital publishing, library automation, machine-actionable digital library, digital mathematics library, DML, WDML.

1 Introduction

The Digital Era has changed crucially as the methods of research, and the ways in which scientists search, produce, publish, and disseminate their scientific work. A digital library, a collection of information which is both digitized and organized, gives us power we never had with traditional libraries. Information and communication technologies are actively implemented in research and development. Therefore, it became possible to use the entire volume of accumulated scientific knowledge in conducting new research. This requires creation of complex of technologies that ensure management of available knowledge, the organization has effective access to this knowledge, as well as sharing and multiple use of new kinds of knowledge structures. In mathematics also accumulated considerable experience in using of electronic mathematical content within the various projects on creation of mathematical digital libraries (see, e. g., [1]).

Since inception of the first scientific information systems, mathematicians have been involved in the full cycle of software product development, from idea to implementation. Well-known examples are an open source system TEX and commercial systems Wolfram Mathematica and Wolfram Alpha, led by Stephen Wolfram according to his principles of computational knowledge theory [2, 3]. Tools for mathematical content management are developed with the help of communities of mathematicians, e.g. MathJax by American Mathematical Society, information system Math-Net.Ru is developed at the Steklov Mathematical Institute of the Russian Academy of Sciences [4] and the collection of publicly available preprints arXiv.org (https://arxiv.org/).

Main challenges of mathematical knowledge management (MKM) are discussed in [5–9], the most urgent tasks are outlined. Such tasks are: modeling representations of mathematical knowledge; presentation formats; authoring languages and tools; creating repositories of formalized mathematics, and mathematical digital libraries; mathematical search and retrieval; implementing math assistants, tutoring and assessment systems; developing collaboration tools for mathematics; creating new tools for detecting re-purposing material, including plagiarism of others' work and self-plagiarism; creation of interactive documents; developing deduction systems. The solution of this task requires formalization of mathematical statements and proofs [9].

At present, research activities in the field of mathematics are associated with the use of modern information technology (cloud, semantic, etc.). These technologies are used in research of distributed scientific teams, preparation and dissemination of mathematical knowledge in an electronic form. At present, a new type of digital library is being formed, connected with the integration of mathematical knowledge into the scientific information space, see. [1,10,11]. This type of information system is called Digital Mathematical Library (DML), a number of global projects are implemented, such as European Digital Mathematical Library or World Digital Mathematical Library [12–14]. More details about goals, functions and current results are listed below, in Section 3.

Implementation of digital mathematical libraries involves the development of special tools and continuous improvement of their functionality. An example is the Open Journal Systems (OJS, https://pkp.sfu.ca/ojs/). The platform is used in many projects, particularly in Lobachevskii Journal of Mathematics (http://ljm.kpfu.ru/), one of the first digital mathematical journals [15].

In our work, we try to look more deeply into world largest DML to outline current status of described projects and to investigate services and functions that provide these digital mathematical libraries.

Proceedings of the XIX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2017), Moscow, Russia, October 10–13, 2017

2 Mathematical Libraries and DELOS Digital Library Reference Model

2.1 Criteria for investigation

Firstly, we need to establish common criteria and main features and functions that we will look at.

In DELOS Digital Library Reference Model [16, 17] three basic concepts are distinguished for defining what is called a digital library (DL):

• DL – a (potentially virtual) organization that comprehensively collects, manages, and preserves for the long term rich digital content and offers to its user communities specialized functionality on that content, of measurable quality, and according to prescribed policies;

• DL system – a software system that is based on an architecture and provides all functionality that is required by a particular Digital Library. Users interact with a Digital Library through the corresponding DL system;

• DL management system (DLMS) – a generic software system that provides the appropriate software infrastructure to both produce a basic DL system that incorporates all functionality that is considered foundational for Digital Libraries and integrate additional software offering more refined, specialized, or advanced functionality. An intrinsic part of DLMS functionality is related to administrative services that are used to choose the appropriate subset of its functionality, e.g., through relevant parameters of its components, and then install, deploy, and (re)configure a DL system.

A DLMSis "system software". As in several other domains (e.g., operating systems, databases, user interfaces), such kernel software may be used as a foundation to produce Digital Library systems.

While the concept of DL is intended to capture an abstract system that consists of both physical and virtual components, the remaining two capture concrete software systems. For every DL, there is a unique DL system in operation (possibly consisting of many interconnected smaller DL systems in the most general case), where as all DL systems are based on a handful of DLMSs.

In the role-based aspect, the DELOS DLRM model consider following types of users: the end user of the DL; the developer of DL; the system administrator of DL and the developer of applications for DL and, four levels of user views and expectations are formed. In addition, the model identifies six key areas, each of which introduces and defines its own entities and their properties: architecture, information space, functionality, users, policy and quality of services provided. These areas can be considered as evaluation criteria and, by virtue of their universality, can be used to analyze almost any information system.

We will carry out an analysis of existing digital mathematical libraries, performed using the evaluation criteria embedded in the DELOS DLRM model.

2.2 Differences between approaches

It is interesting to stop at the discussion at the approaches of the definition of elementary objects with which digital library works. In particular, an interesting approach to information objects organization lies in the ideology of WDML. We use the same approach in creating a digital mathematical library Lobachevskii DML, which is based on mathematical collections of the Kazan Federal University [18].

Usually digital library consists of collections, and collections in turn from documents or information resources (objects). In 1990–2000 there was a large number of studies carried out on the definition, architectural and technical aspects of DL systems. Finally, it is necessary to mention the creation of the DL manifesto in the DELOS project, which resulted in the creation of a reference model for DL [16, 17].With the development of Semantic Web technologies, it became interesting to investigate the semantics of resources and their links placed in libraries, see, for example, [18]. In this case, an information object can already be considered not only as a document, but as its certain parts – abstract, keywords, bibliography, citations, comments of authors or readers.

From the end user's point of view, DL must satisfy the user's expectations. The document itself as an elementary information object may not be interesting at all. It is much in demand to search for information on a particular entity or subject mentioned in the document. At the same time, much more interesting to find all possible resources where different versions of mentioned subjects, especially in cases when various interpretations and definitions are possible. For example, there are a number of definitions of the concept of "digital library" and the user studying this topic will certainly be interested in all references to the definition of the digital library from different sources. Thus, we observe a change in the elementary information object. The electronic document fragmented into smaller information objects and all services of a library deal with such objects and manage the relationships between them. In mathematics, such elementary objects can be, for example, theorems, lemmas, definitions or formulas, research of which is much more informative on a number of sources. The services of any DML should provide such an opportunity. All this functionality lies in WDML architecture. Its implementation became possible only with the development of semantic technologies and the transfer of library content into digital form with metadata. Now, there is no technical problems in maintaining such approach to the organization of DL.

During our research, we will take into account this transformation of the approach to the organization of information objects. Note that the change in the approach to the organization of DL does not affect the selected criteria for investigation.

3 Functionality of Digital Mathematical Libraries

Below is a brief review of existing digital mathematical libraries. The largest projects are "All-Russian Mathematical Portal Math-Net.RU","Centre de diffusion de revues académiques mathématiques", "Czech Digital Mathematics Library", "The Polish Digital Mathematics Library", "Göttinger

Digitalisierungs Zentrum", "Numérisation de documents anciens mathématiques", Zentralblatt MATH, "Bulgarian Digital Mathematics Library" and "The European Digital Mathematical Library". It should be outlined, that all projects are in different degrees of completeness, the range of services provided is also different.

3.1 Math-Net.ru

All-Russian Mathematical Portal Math-Net.ru [4, 20–22] combines both a digital mathematical library and a publishing system for mathematical texts. It is a web portal developed by the V. A. Steklov Institute of Mathematics, Russian Academy of Sciences.

The key component of the portal – the "Journals" section links Russian periodicals in the field of mathematical sciences to a single information system. Currently contains more than 120 journals with nearly 200 thousand publications. Information about the article includes a bibliographic description, an annotation, lists of literature and a file with the full text of the article. The portal presented in two languages – Russian and English.

The most interesting part is the functionality of the portal. The portal provides the ability to search for publications and links on the bibliographic description and keywords in the title, annotation or text. As result of the search, an abstract, article IDs (DOI, resource references in abstract databases, URIs), a citation pattern, classifier values are issued. There are no recommender service, in fact all semantic services work with a bibliographic description of the resource. MiRef module is used to form correctly the description and links to resources. The module is designed to automatically place links to various publications databases in the literature list. The format of the links must satisfy the rules of the amsbib package and should be entered in the LaTeX format.

Registered users can create personal pages, manage personal collections of publications, authors get access to the full texts of their articles, authors can send the manuscript to the editorial office of the journal electronically, and track the process of its workflow in the editorial office.

Statistics on popular authors and resources are maintained, infometric indicators for resources located on the portal are calculated.

The policy for accessing the full texts of articles is determined by the publisher of the paper. Access for any other information is free.

3.2 CEDRAM

The center for diffusion of academic mathematical journals (CEDRAM, Centre de Diffusion de Revues Académiques Mathématiques) is a web portal for common access to a set of mathematical journals [23], available in French and English. CEDRAM's mission is to provide a large distribution of their current volumes, and range from help for producing journals according to the best standards for electronic publishing to long lasting archiving. CEDRAM is a service of the Cellule MathDoc (UMS 5638 of CNRS and Université Joseph Fourier) which completes its important offer in mathematical documentation. This DML is not so large – contains 9 French math journals, 1 book and 7 proceedings of seminars and conferences.

The CEDRAM websites offer two ways of consulting the hosted articles: quick and advanced search. Search functions provide search by keywords, author, title, bibliography and full text search. Quick search searches in all fields except full text. Advanced search interface offers several types of research, more or less complicated. The full entry of articles produced for CEDRAM contains abstracts and bibliographical references.

All online records exist in two formats, which are only different by the way they display mathematical formulas in titles, abstracts, keywords or references: MathML or TeX and have stable url link.

XHTML+MathML display is best for reading and browsing, but there are some problems with viewing in browsers, that need to be pre-configured to work correctly with MathML. The HTML+TeX version used for compatibility for users who do not have an environment capable of displaying MathML. Now CEDRAM provide following services [12, 13, 23]:

- production workflow of journals;
- dedicated web site for each journal;
- provides creation and maintenance of LATEX styles (using a specific class);
- production of PDFs for print and web with XML/MathML metadata;
- DOI registration (Crossref), reference linking (MSN, ZBM, mini-DML, Crossref);
- provides publishing platform for mathematical articles based on Open Journal System (Public Knowledge Project, https://pkp.sfu.ca/ojs/);
- all resources archived in partner project - the French digital math library NUMDAM.

Policy and quality of services. Starting 2017 all CEDRAM journals are open access. Access to the database containing the bibliographical references of all the articles of all participating journals is totally free. The database itself is the property of Cellule Mathdoc, and contains elements covered by copyright. CEDRAM has OAI-PMH server, which can be used for systematic download of metadata in various schemas. Files of the full texts are the property of the journals and it is necessary to refer to the policy of each of them. Also there are some restrictions of full copying and indexing by web robots.

3.3 Numerisation de Documents Anciens Mathematiques (NUMDAM)

The French digital math library NUMDAM [12–14, 24] started as a digitisation program for a pilot of 6 journals. Now it contains more than 57000 articles in 76 periodicals, 373 books in 4 collections, 263 theses.

The NUMDAM is the reference French digital mathematics library set up by Cellule MathDoc with the assistance of a network of partners.

From 2007 onwards, publishers send digital born articles into DML. Collections are normally indexed

within one year of publication, and full texts are freely downloadable at the end of a period of time set by agreement upon each title.

The NUMDAM program is designed to support academic publishers and provide the research community with a sustainable, reliable and easy-to-use library. The research and dissemination platform was completely redesigned in 2016. Now portal is available on two languages –English and French, formulas can displayed in TeX or in graphical form using MathJax.

System provide following functions: search and navigation by title, author, references or in full text of resources. During search all statistics, related to the search topic is displayed – co-authors, journals and years of publication. Browse functions provide navigation through sorted list of resources (authors, journals etc.).

Full texts available in PDF and DJVU formats. Each article in NUMDAM is available via a stable URL. This URL is a compact address, designed to remain valid in the long term. It is displayed in the web page of the article, on the first page of PDF or DjVu files and by the OAI-PMH server.

There is no any user registration. All functions have open access. NUMDAM only disseminate resources that already published in journals, books or theses but submission process of resources is not clear. Metadata extraction made only for bibliography. Any additional services like formula search or recommender system are absent.

The full text of most recent articles is generally not available. The journals whose archives are on this portal have accepted the principle of "a moving wall". This is a time interval between the publication of a volume (in paper or electronic form, delivered to subscribers) and the availability of the full text on the NUMDAM server. Generally, moving-wall for most of journal in NUMDAM is equal to 5 years.

3.4 The Czech Digital Mathematics Library (DML-CZ)

The Czech Digital Mathematics Library (DML-CZ) [25, 26] has been developed in order to preserve 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 resulted from the project no. 1ET200190513 supported by the Czech Academy of Sciences (CAS) in the R&D programme Information Society, and operated by the Institute of Mathematics CAS. Project seems to be finished in 2010 and now is in stable form.

Functionality. Editors of all journals included in DML-CZ are using tools and work flows 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.

There are some automated procedures for validation of data of new journal issues but all of them are archived in DML-CZ for internal use and development. Based on limiting the name space of allowed TEX macros, validation service get all metadata including abstracts, keywords and references transformed into representation using MathML [27].

End-users cannot submit any resource, everything can be submitted only through editorial board of journals, also there is no any personal area for users.

Search and navigation. As others DMLs DML-CZ allows to search by title, author of publications. Also avaliable search by language or by Zentrablatt MATH and MathSciNet identifiers. Browse functions provide navigation through sorted list of resources (authors, journals etc.).

The most interesting function is search of related articles (finding similarities between papers). This service tries to find similar papers using three methods: "Term frequency–Inverse document frequency" (TF-IDF, see, e. g. [28]), the "Random Projections" or method that is built on TF-IDF and simplifies the computations by projecting vectors onto a subspace of lower dimensionality [28] and with using "Latent Semantic Indexing" (LSI, [29]). Last method gives the most accurate results up to 90%.

Policies and quality of service. The digitized journal and proceedings papers are displayed with the agreement of the publisher who owns the digital data. The digitized monographs are displayed with the agreement of the author and/or the publisher while the digital data are property of the Institute of Mathematics CAS. The database itself, in particular the bibliographic data, are property of the Institute of Mathematics CAS.DML-CZ presents full texts articles and book chapters in PDF format, equipped with enhanced metadata including bibliographical references linked to Zentrablatt MATH and MathSciNet. The digital born documents are being obtained from the original sources provided by publishers. The presented page content and format corresponds to the original one. Journals are presented and accessed according to the terms of a contract with the publisher. The digital documents displayed in the DML-CZ are authorized with electronic stamps.

3.5 The Polish Digital Mathematical Library

The Polish Digital Mathematical Library (DML-PL, [30]) has existed since 2002. The library holds full texts of polish mathematical journals and books. The major part of the collection are archive issues of mathematical journals published before World War II. Library consists of 550 books and 36 journals, but only 3 journals provide access to full text of articles. Portal of DML-PL provide search by attributes and navigation through sorted lists of authors, books and journals.

Brief explanation of the project is given in [31], but nowadays it seems that project is already finished. On the web portal of library there is no additional information about current status. Any information about semantic functions or metadata extraction from resources is missing.

3.6 GDZ–Gottingen Digitization Centre

The task of the GDZ [32, 33] is to record data such as prints, manuscripts and illustrations and to preserve them. Main aim of the project is conversion of resources

into digital form. This is multidisciplinary library, that contains not only mathematical collections but also history of Law, history of the Humanities and the Sciences, travel and North American literature and other collections. Mathematical collections have about 7000 resources and also have some Russian resources. Library contains more than 15 million digitized pages.

Portal provides search in metadata and full text of resources and browse functions. Many resources are historical, not modern, main aim of the project is to digitize and preserve resources. All resources have full texts and can be viewed page by page or in structured mode. Metadata of any resource contain stable URL of resource, metadata can be downloaded in METS format.

3.7 Zentralblatt MATH

Zentralblatt MATH (zbMATH, [34]) is abstracting and reviewing service in pure and applied mathematics. It is hosted by the Berlin office ofFIZ Karlsruhe – Leibniz Institute for Information Infrastructure GmbH (FIZ Karlsruhe) and distributed by Springer. The zbMATH database contains more than 3.5 million bibliographic entries with reviews or abstracts currently drawn from more than 3,000 journals and serials, and 170000 books. zbMATH is not a digital library itself, it is an indexing service and provides easy access to bibliographic data, reviews and abstracts from all areas of pure mathematics as well as applications, in particular to the natural sciences, computer science, economics and engineering.

Search functions provide search for documents, authors and journals. Search can be done in one line, or in structured form using attributes such as title, author, subject, source, keywords etc. Service also provide full-text formula search for indexed arXiv documents [35].The zbMATH formula search uses the MathWebSearchsystem, which is a content-based search engine for MathML formula based on substitution tree indexing.

Portal offer three ways of displaying mathematical formulas – MathML, MathJax and LaTeX. The XML-based MathML is the solution recommended by W3C for displaying mathematical content on the web and is set as default within zbMATH. Mathematical Reviews and zbMATH maintain the Mathematics Subject Classification (MSC), a classification scheme for mathematics. It is used by reviewing services to categorize items in the mathematical sciences literature. The database of service contains about 2.1 million direct links to electronic versions of the indexed publications, to the publishers' websites and/or to electronic libraries with open access to the full texts.

3.8 Bulgarian Digital Mathematics Library

Bulgarian Digital Mathematics Library, BulDML is a digital repository at Institute of Mathematics and Informatics of Bulgarian Academy of Sciences. Library has 7 mathematical journals, 4 book series and proceedings in its repository. In fact, BulDML is an institutional repository and is built on open-source DSpace software [36]. As known, DSpace preserves and enables open access to all types of digital content including text, images, moving images, mpegs and data sets. All functionality of DSpace software is clear and we will not describe it in this paper. For example, additional information about DSpace can be found in [17, 37].

3.9 European Digital Mathematics Library

The European Digital Library (EuDML) was a project partly funded by the European Commission. EuDML [12–14, 38, 39] is an aggregation and indexing services with was established under The EuDML Initiative and promoted by European Mathematical Society. EuDML assemble as much as possible of the digital mathematical corpus in order to make it available online, with eventual open access, in the form of an authoritative and enduring digital collection, growing continuously with publisher supplied new content, augmented with sophisticated search interfaces and interoperability services, developed and curated by a network of institutions.

The system, presented in the diagram in Figure 1, conceptually consists of a metadata repository, a search engine, a metadata enhancer, an association analyser, annotation and accessibility functions and of course the interfaces [38].

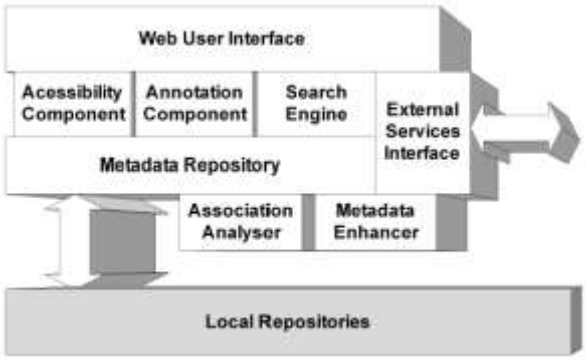

Figure 1 EuDML architecture

The metadata repository provides the central point of reference for all the managed contents. It will work with an OAI-PMH harvester to ingest repositories' content descriptions, maps the metadata into the internal EuDML schema. The performance and the quality of responses of the search service directly influence user experience. Therefore, particularly this service has to be reliable, scalable and customized to fulfill user expectations.

The metadata enhancer function consist in a collection of tools that each contribute to expand or complete the existing items' metadata, depending on the improvements needed. These range from applying OCR over full texts, adding key words or multilingual metadata by merging information from different databases when an item happens to have such non-redundant description, generating MathML for mathematical expressions, etc. The association analyzer detects, analyses and records relations between individual items. The annotation component provides mechanisms to attach new material to individual items in the repositories and maintain this new material. The accessibility component provides support for enhanced accessibility of items, if required, before presentation to end users. Finally, the user and system interfaces provide

access to the collected resources on different levels both to human and machine users. Now EuDML offers several service interfaces that allow other applications to connect with the service. These are OAI-PMH server, REST services, OpenSearch service, which allow to query library index in machine way and annotation retrieval services in JSON.

EuDML aims to be an open source of trusted mathematical knowledge. That is why it has some policies:

- All texts must have been scientifically validated and formally published;
- All items must be open access after a finite embargo period. Once documents contributed to the library are made open access due to this policy, they cannot revert to close access later on;
- The digital full text of each item contributed to library must be archived physically at one of the EuDML member institutions.

All DMLs, described above except All-Russian Mathematical Portal Math-Net.RU are partners of EuDML.

4 Conclusion

In order to outline all differences of observed projects we created comparison Table 1 listed below. Note that, we excluded from table two DMLs due to following. BulDML is and built on open-source DSpace software, so all functionality of it is clear, for DML-PL we could not find any working portal in order to study it more deeply.

In all the projects studied, emphasis is placed on the transfer of the resources themselves to the electronic form, rather than on the development of semantic services. Only a few portals have a mathematical formula search, and only one has a recommender service.

After the analysis done it is clear that there are only two types of repository systems: the first is actually DML, which preserve the resources themselves, the second is indexing and aggregating services that do not have their own database of electronic documents, but provide a wide range of convenient search capabilities.

This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, grant agreement no. 1.2368.2017) and with partial financial support of the Russian Foundation for Basic Research and the Government of the Republic of Tatarstan, within the framework of scientific projects Nos. 15-07-08522, 15-47-02472.

References

[1] Borwein, J.M., Rocha, E.M., Rodrigues, J.F. Communicating Mathematics in the Digital Era, pp. 3-21. A K Peters, Ltd. MKM-IG. Mathematical Knowledge Management (2008). http://www.mkm-ig.org/

[2] Wolfram, S.: A New Kind of Science. Wolfram Media, Inc. (2002)

[3] Wolfram, S.: An elementary introduction to the Wolfram Language. Wolfram Media, Inc. (2015)

[4] Chebukov, D.E., Izaak, A.D., Misyurina, O.G., Pupyrev, Yu.A., Zhizhchenko, A.B.: Math-Net.Ru as a Digital Archive of the Russian Mathematical Knowledge from the XIX Century to Today. Intelligent Computer Mathematics, Lecture Notes in Comput. Sci., 7961, pp. 344-348, Springer (2013), doi: 10.1007/978-3-642-39320-4_26

[5] Carette, J., Farmer, W.M.: A Review of Mathematical Knowledge Management. In Intelligent Computer Mathematics. Lecture Notes in Computer Science, 5625. pp. 233-246 (2009)

[6] Ion, P.D.F.: Mathematics and the World Wide Web. In Intelligent Computer Mathematics. Lecture Notes in Computer Science, 7961, pp. 230-245 (2013)

[7] Lange, C.: Enabling Collaboration on Semiformal Mathematical Knowledge by Semantic Web Integration. Ph. D. Thesis, Jacobs University Bremen (2011)

[8] Elizarov, A.M., Lipachev, E.K., Nevzorova, O.A., Solov'ev, V.D.: Methods and Means for Semantic Structuring of Electronic Mathematical Documents. Doklady Mathematics, 90 (1), pp. 521-524 (2014), doi: 10.1134/S1064562414050275

[9] Elizarov, A., Kirillovich, A., Lipachev, E., Nevzorova, O., Solovyev, V., and Zhiltsov N.: Mathematical Knowledge Representation: Semantic Models and Formalisms. Lobachevskii J. of Mathematics, 35 (4), pp. 347-353 (2014), doi:10.1134/S1995080214040143

[10] Elizarov A., Kirillovich A., Lipachev E., Nevzorova O. (2017) Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management. In: Kalinichenko L., Kuznetsov S., Manolopoulos Y. (eds) Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2016. Communications in Computer and Information Science, 706, pp. 33-46 (2017), doi: 10.1007/978-3-319-57135-5_3

[11] Elizarov, A.M., Kirilovich, A.V., Lipachev, E.K., Nevzorova, O.A.: Mathematical Knowledge Management: Ontological Models and Digital Technology. CEUR Workshop Proceedings, 1752, pp. 44-50 (2016), http://ceur-ws.org/Vol-1752/paper08.pdf

[12] Bouche, T.: Towards a World Digital Library: Mathdoc, Numdam and EuDML Experiences. UMI, La Sapienza, Roma (2016), http://www.mat.uniroma1.it/sites/default/import-files/biblioteca/SEMINARIO2016/bouche.pdf

[13] Bouche, T.: Digital Mathematics Libraries: The good, the bad, the ugly. Mathematics in Computer Science, (3), pp. 227-241 (2010), doi: 10.1007/s11786-010-0029-2

[14] Bouche, T.: Reviving the Free Public Scientific Library in the Digital Age? The EuDML Project. In: Kaiser, K., Krantz, S., Wegner, B. (eds.): Topics and Issues in Electronic Publishing, JMM, Special

Session, San Diego, January 2013, pp. 57-80 (2013), http://www.emis.de/ proceedings/TIEP2013/05bouche.pdf

[15] Elizarov, A.M., Zuev, D.S., Lipachev, E.K.: Mathematical Content Semantic Markup Methods and Open Scientific E-Journals Management Systems. In: Klinov, P., Mouromtsev, D. (eds.) KESW 2014. CCIS, 468, pp. 242-251 (2014), doi: 10.1007/978-3-319-11716-4 22 29

[16] Candela, L., Athanasopoulos, G., Castelli, D., El Raheb, K., Innocenti, P., Ioannidis, Y., Katifori, A., Nika, A., Vullo, G., Ross, S.: The Digital Library Reference Model. FP7-ICT-2007-3. Cultural Heritage and Technology Enhanced Learning (2011)

[17] Candela, L., Castelli, D., Fuhr, N., Ioannidis, Y., Klas, C.-P., Pagano, P., Ross, S., Saidis, C., Schek, H.-J., Schuldt, H., Springmann, M.: Current Digital Library Systems: User Requirements vs Provided Functionality. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage (2006)

[18] Elizarov, A.M., Lipachev, E.K.: Lobachevskii DML: Towards a Semantic Digital Mathema-tical Library of Kazan University, 2017 (in press), DAMDID-2017 proceedings

[19] Kogalovskiy, M.R., Parinov, S.I.: Klassifikatsiya i ispol'zovaniye semanticheskikh svyazey mezh-du informatsionnymi ob'yektami v nauchnykh elektronnykh bibliotekakh. Inform. i yee primen., 3 (6), pp. 32-42 (2012)

[20] All-Russian Mathematical Portal Math-Net.Ru. http://www.mathnet.ru/

[21] Zhizhchenko, A.B., Izaak, A.D.: The Information System Math-Net.Ru. Application of Contemporary Technologies in the Scientific Work of Mathematicians. Russian Math. Surveys, 62 (5), pp. 943-966 (2007), http://dx. doi.org/10.1070/ RM2007v062n05ABEH004455

[22] Zhizhchenko, A.B., Izaak, A.D.: The Information System Math-Net.Ru. Current State and Prospects. The Impact Factors of Russian Mathematics Journals. Russian Math. Surveys, 64 (4), pp. 775-784 (2009), http://dx.doi.org/ 10.1070/ RM2009v064n04ABEH004638

[23] CEDRAM. www.cedram.org

[24] NUMDAM. www.numdam.org

[25] The Czech Digital Mathematics Library (DML-CZ), http://www.dml.cz/

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

[27] Rákosník, J.: Recent Development of the DML-CZ and Its Current State. In Proc. of DML 2011: Towards a Digital Mathematics Library. Bertinoro, Italy, July 20–21st (2011)

[28] Rajaraman, A.; Ullman, J. D.: Data Mining (2011). doi:10.1017/CBO9781139058452.002

[29] Deerwester, S., Dumais, S., Landauer, T., Furnas, G., Beck, L.: Improving Information Retrieval with Latent Semantic Indexing. Proc. of the 51st Annual Meeting of the American Society for Information Science, 25, pp. 36-40 (1988)

[30] The Polish Digital Mathematics Library, http://pldml.icm.edu.pl/

[31] Zamlynska, K., Tarkowski, A., Rosiek, T.: Evolution of the Mathematical Collection of the Polish Virtual Library of Science. Mathematics in computer Science, (3), pp. 265-278 (2010), doi: 10.1007/s11786-010-0029-2

[32] Gottingen Digitalisierungs Zentrum. http://gdz. sub.uni-goettingen.de/gdz/

[33] Gottingen digitization Centre https://www.sub. uni-goettingen.de/en/copying-digitising/ goettingen-digitisation-centre/

[34] Zentralblatt MATH. https://zbmath.org/

[35] Muller, F., Teschke, O.: Full Text Formula Search in zbMATH, EMS Newsletter (2016)

[36] Bulgarian Digital Mathematics Library. http://sci-gems.math.bas.bg/jspui/

[37] DSpace, www.dspace.org

[38] Sylwestrzak, W., Borbinha, J., Bouche, T., Nowinski, A., Sojka P.: EuDML – Towards the European Digital Mathematics Library. In: Sojka, P. (ed.) Towards a Digital Mathematics Library. Paris, July 7–8th, 2010, pp. 11-26. Masaryk University Press, Brno (2010), http://dml.cz/bitstream/handle/10338.dmlcz/702569/DML_003-2010-1_5.pdf

[39] EuDML, www.eudml.org

Table 1 Comparison table of DML projects

DML Criteria	Math-Net.ru	CEDRAM	NUMDAM	DML-CZ	GDZ	zbMATH	EuDML
Information space	There is an object hierarchy. Collections split into journals, issues, articles and so on. Currently contains more than 120 journals with nearly 200 thousand publications. Information about the article includes a bibliographic description, an annotation, lists of literature and a file with the full text of the article.	DML contains 9 French math journals, 1 book and 7 proceedings of seminars and conferences. All CEDRAM journals are open access. Access to the database containing the bibliographical references of all the articles of all participating journals is totally free. The full entry of articles contains abstracts and bibliographical references.	Contains more than 57000 articles in 76 periodicals, 373 books in 4 collections, 263 theses. Full texts available in PDF and DJVU formats. Each article in NUMDAM is available via a stable URL.	The digitized journal and proceedings papers are displayed with the agreement of the publisher who owns the digital data. DML-CZ presents full texts articles and book chapters in PDF format, equipped with enhanced metadata including bibliographical references. The digital born documents are being obtained from the original sources provided by publishers.	This is multidisciplinary library, that contains not only mathematical collections but also history of Law, history of the Humanities and the Sciences, travel and North American literature and other collections. Mathematical collections have about 7000 resources and also have some Russian resources. Library contains more than 15 million digitized pages.	The database contains more than 3.5 million bibliographic entries with reviews or abstracts currently drawn from more than 3,000 journals and serials, and 170,000 books. The database of service contains about 2.1 million direct links to electronic versions of the indexed publications, to the publishers' websites and/or to electronic libraries with open access to the full texts.	This is an aggregation and indexing service. EuDML assemble the digital mathematical corpus in order to make it available online.
Functionality	The portal provides the ability to search for publications and links on the bibliographic description and keywords in the title, annotation or text. As result of the search, an abstract, article IDs (DOI, resource references in abstract databases, URIs), a citation pattern, classifier values are issued. There are no recommender service, in fact all semantic services work with a bibliographic description of the resource	CEDRAM has OAI-PMH server, which can be used for systematic download of metadata in various schemas. Search functions provide search by keywords, author, title, bibliography and full text search. Quick search searches in all fields except full text. Advanced search interface offers several types of research, more or less complicated. The full entry of articles produced for CEDRAM contains abstracts and bibliographical references.	NUMDAM has an OAI-PMH server, thus allowing sharing of metadata and better visibility of collections. System provide following functions: search and navigation by title, author, references or in full text of resources. During search all statistics, related to the search topic is displayed – co-authors, journals and years of publication. Browse functions provide navigation through sorted list of resources. Metadata extraction made only for bibliography. Any additional services like formula search or recommender system are absent.	Editors of all journals are using tools and workflows that enable them to produce inputs 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. There are some automated procedures for validation of data of new journal issues but all of them are for internal use and development. DML-CZ allows to search by title, author of publications, by language or by zbMATH and MathSciNet identifiers. Browse functions provide navigation through sorted list of resources. There is search of related articles.	Portal provides search in metadata and full text of resources and browse functions. All resources have full texts and can be viewed page by page or in structured mode. Metadata of any resource contain stable URL of resource, metadata can be downloaded in METS format.	Search functions provide search for documents, authors and journals. Search can be done in one line, or in structured form using attributes. Service also provide full-text formula search for indexed arXiv documents. The zbMATH formula search uses the MathWebSearch system. zbMATH maintain a classification scheme for mathematics.	EuDML offers several service interfaces that allow other applications to connect with the service. These are OAI-PMH server, REST services, OpenSearch service, which allow to query library index in machine way and annotation retrieval services in JSON.

Users	There are role model of users, everybody can register and create own personal area. Registered users can create personal pages, manage personal collections of publications, authors get access to the full texts of their articles.	No any user registration	No any user registration	No any user registration. End-users cannot submit any resource, everything can be submitted only through editorial board of journals, also there is no any personal area for users.	No any user registration.	There is a personal area for users – for reviewers, publishers etc.	No any user registration.
Quality of service	System is available in two languages. The policy for accessing the full texts of articles is determined by the publisher of the paper. Access for any other information is free.	Portal is available in English and French. Files of the full texts are the property of the journals. All online records exist in two formats, which are only different by the way they display mathematical formulas: MathML or TeX and have stable url link.	Portal available on English and French. Formulas can be viewed in TeX source code or in compiled, graphical way. NUMDAM only disseminate resources that were already published in journals, books or theses but submission process of resources is not clear.	Project was finished in 2010 and now it is in a stable form. Portal is available only in English.	Portal is available in German and English. But main aim of the project is to digitize and preserve resources.	Portal offers three ways of displaying mathematical formulas – MathML, MathJax and LaTeX. MathML is set as default. Not all services of the system are free, some of them need to be purchased.	EuDML has some policies: all texts must be scientifically validated and formally published; all items must be open access after a finite embargo period. Once documents contributed to the library are made open access due to this policy, they cannot revert to close access later on; the digital full text of each item contributed to library must be archived physically at one of the member institutions.