Wojtek Sylwestrzak; José Borbinha; Thierry Bouche; Aleksander Nowiński; Petr Sojka
EuDML—Towards the European Digital Mathematics Library

In: Petr Sojka (ed.): Towards a Digital Mathematics Library. Paris, France, July 7-8th, 2010. Masaryk University Press, Brno, Czech Republic, 2010. pp. 11--26.

Persistent URL: http://dml.cz/dmlcz/702569

EuDML—Towards the European Digital Mathematics Library
Architecture and Design

Wojtek Sylwestrzak[1], José Borbinha[2], Thierry Bouche[3], Aleksander Nowiński[1], and Petr Sojka[4]

[1] Interdisciplinary Center for Mathematical and Computational Modelling
University of Warsaw, ul. Pawińskiego 5A, 02-106 Warsaw, Poland
W.Sylwestrzak@icm.edu.pl, A.Nowinski@icm.edu.pl
[2] Instituto Superior Técnico: Computer Science Department, Lisbon, Portugal
jlb@ist.utl.pt
[3] Institut Fourier (UMR 5582) & Cellule Mathdoc (UMS 5638), Université Joseph-Fourier (Grenoble 1), B.P. 74, 38402 Saint-Martin d'Hères, France
thierry.bouche@ujf-grenoble.fr
[4] Masaryk University, Faculty of Informatics
Botanická 68a, 602 00 Brno, Czech Republic
sojka@fi.muni.cz

Abstract. The paper describes the background, the expected functionalities, and the architecture design goals of the European Digital Mathematics Library (Eu-DML), an infrastructure system aimed to integrate the mathematical contents available online throughout Europe, allowing for both extensive and specialized mathematics resource discovery. The three years long project to build the EuDML, partially funded by the European Commission, started in February 2010.

Key words: bibliography crosslinking, document ranking, digital libraries, DML, EuDML, information systems, information retrieval, citations discovery and extraction, mathematical content search, mathematical metadata, mathematics indexing, REPOX, similarity analysis, text mining, YADDA, Web 2.0

1 Introduction

Mathematics is a specific discipline in many respects. One often hears that all a mathematician needs to work is a pencil and a piece of paper. This is not entirely the truth. One of the unique qualities of mathematics is its intrinsic dependence on previous works—one of the basic tool of a mathematician is thus a library. Similarly, many other disciplines of science and research depend on mathematical knowledge, and for them, too, access to a mathematical library is a requisite. In modern times, this translates into the requirement of online availability of the mathematical content. This is why the numerous local initiatives first to digitize and then to make local mathematical content available spontaneously started and are still active in many countries. However,

Petr Sojka (editor): DML 2010, Towards a Digital Mathematics Library, pp. 11–26.
 ISBN 978-80-210-5242-0

for a number of reasons, it is necessary to provide an integrated access to all the accumulated material, especially because in mathematics, unlike in many other disciplines, the language of a publication is to a lesser extent a barrier. A mathematician looking for a certain theorem proof is still concerned about the place, date or language of the publication, but to a lesser extent than researchers in other, less universal disciplines. Therefore it is essential that all the worldwide mathematical content, although distributed and heterogeneous in nature, is presented in a seamless way, through a unified interface suitable for the mathematics search purposes, as well as future reliable reference.

This claim already impulsed a large array of international activities since year 2001, when the DML concept emerged from pioneering work by John Ewing, Philippe Tondeur and Keith Dennis, who initiated a meeting in San Diego (California) in January 2002, where mathematical societies, digitisation projects, academic publishers reckoned the need for this infrastructure [22,18]. Since then, however, no actual progress towards integration of heterogeneous collections has been achieved. A one-year (2002–2003) planning project [13] coordinated by Cornell University Library was funded by the U.S. National Science Foundation (NSF) "toward the establishment of a comprehensive, international, distributed collection of digital information and published knowledge in mathematics", with a steering committee that happened to be mostly European. Most of the conclusions from that group are still valid today: the need for standardization and coordination, the identification of intellectual property rights and conflict of interests among stakeholders as principal inhibitors. In 2005, the Moore Foundation was approached by the American Mathematical Society (AMS) and considered funding a gigantic World DML, but faced the same inhibiting factors. However, while no consensus could be reached in the areas of ownership of the mathematical heritage, or governance of the DML, and no global funding scheme seemed realistic, numerous digitisation projects were running or launched, most of which were based in Europe. The International Mathematical Union endorsed in 2006 a generous text [23] written by its Committee on Electronic Information and Communication (CEIC), but although this probably served as an incentive to launch new local DML projects, nothing happened in the area of integration although the EMANI project was a pioneering attempt in this direction [32].

The European Mathematical Society (EMS) was a driving force in many of these efforts: it wrote an Expression of Interest to the European Union to support an application for the European chapter of the DML as an integrated infrastructure in the 5th Framework programme. In 2009, it contacted the European Science Foundation which considered supporting a foresight study on a European DML, which showed that almost a decade after the first attempts, we were more or less back to the NSF planning project situation.

One of the reasons for this situation is that commercial publishers had considered for a while using public funding to get their backfiles digitised, and were open to new business models (like JSTOR's moving wall, or even NUMDAM's moving wall) if that happened. But no funding source of the size

needed ever surfaced, and each publisher found its own resources, some of them public (usually at a national level, often associated with some mandate for eventual open access in NUMDAM's line), but many publishers invested their own money and thus expected some return, and started to view DML projects as competitors.

We believe that this situation will not be resolved by itself, and therefore that scientists will still lack the necessary infrastructure for handling the reference mathematical corpus, unless a core group of stakeholders takes up the challenge to set up an actual system going beyond the current individual projects' boundaries. This core group must have a critical mass in content, know-how, and a sufficient organisational diversity. They must take a pro-active approach and set the networking infrastructure, standards and policies so that we can further build on the current state of the art and aim at comprehensiveness in content while expanding geographically. We formed a consortium, and claim that, together with its associated partners (EMS and Göttingen University Library), it does form such a core group in position of making the first paradigmatic shift in this area, thanks to the support it obtained from the European Commission.

2 Overview

Despite the lack of success in building a global Digital Mathematics Library (DML) infrastructure [5,7,6,8], many local initiatives continue their development, and many new DML activities started [2,4,30]. It is estimated that over 200 thousand items are already openly available online in European national projects, of which Germany provides around 85,000, France 60,000, Czech Republic 27,000, Russia 18,000, Poland: 13,000, Spain 6,000, Switzerland 5,000, Serbia 4,500, and Portugal 2,000. A lot of mathematical contents is also ready in digital form owned by commercial publishers like Springer, or Elsevier, these alas are not yet freely available, mostly due to the restricted access polices assumed by the publishers, who in most cases are also the copyright owners. Finally, a still unidentified number of publications may be available in electronic form in institutional repositories, archived there by their authors (with unknown scientific validation status).

Despite the impressive volume, the currently available digital mathematics content is often difficult to access, through a number of isolated interfaces, not registered, difficult to find or virtually inaccessible. Often not adequately curated, some of the content may be available in volatile collections and at the risk of perishing.

A group of European stakeholders in DML joined their efforts to draft a project to build a common DML infrastructure for mathematical knowledge access in Europe. The three years long project, named EuDML, officially started in February 2010 [9]. The EuDML will establish a pilot implementation of an integrated mathematics digital library system allowing for seamless access for otherwise dispersed digital material of the partners and associates. A comprehensive partner institutions list is available at the project's web

site, `http://www.eudml.eu/`. The partners include national DML operators and mathematical content providers (among contributed collections are: NUMDAM [21], CEDRAM [11], Zentralblatt MATH [20], EMIS' ELibM [12], DML-CZ [1], DML-PL [37], DML-E [19], DML-PT [3], EDP Sciences' math journals, Bulgarian and Hellenic DMLs [10,15]), digital library technology developers (IST[5] developed REPOX [28], ICM[6] YADDA [36] and MU[7] several other tools [27,1]), a scientific publisher (EDP Sciences), and experienced service providers (ICM providing large scale content services for years, FIZ[8] and CMD[9] as well, in the area of mathematical literature). A professional company (MML[10]) will design the user interface. The European Mathematical Society, will assess the usefulness of the system and the University Library of Göttingen will contribute ERAM and RusDML books and journals [14,33,29].

The fact that the available digital maths corpus is already considerable provides an opportunity to reach the critical mass needed for wide acceptance of the EuDML infrastructure. The quality and scientific relevance of the freely available content, especially in Europe, is a strong asset. Many EuDML partners have been active for years, making digitally available a substantial portion of the mathematical treasuries produced or published mainly in Europe since the XVIIth century. This is of course only one of the necessary conditions, the others being the overall quality of the proposed tools and solutions, and their acceptance by users.

3 Design Goals

The primary goal of the system is to provide a "one-stop-shop" unified access gateway to the distributed mathematical content with innovative services tailored for a wide range of users. At the same time, the system is expected to play the role of an infrastructure solution, forming the basis of the future mathematical knowledge management and provisioning platform. To this end, the system design has to allow for future seamless integration with relevant mathematics tools as well as with existing and planned knowledge infrastructures. While it is envisaged that the future of the knowledge management will be based on open access principles, and the initial versions of the system will focus on open content handling, it should also be capable of dealing with restricted licenses, and cater for selected intermediate and hybrid solutions, including the moving wall model. EuDML basic policy is to handle only content for which some kind of moving wall license has been obtained:

[5] Instituto Superior Técnico, Lisbon, Portugal.
[6] Interdisciplinary Center for Mathematical and Computational Modelling, Warsaw, Poland.
[7] Masaryk University, Brno, Czech Republic.
[8] The Berlin editorial office of Fachinformationszentrum Karlsruhe (FIZ) is producing the Zentralblatt MATH Database and the digital library of journals ELibM.
[9] Cellule Mathdoc, Grenoble, France.
[10] Made Media Ltd. is a digital media agency based in Birmingham, UK.

this secures eventual open access to the content that is curated in the project, which appears to be a rather popular and efficient model for mathematical literature, given its life span. Besides the metadata, the system will also have the possibility to store local fulltext copies, both for speedy text mining and fulltext index rebuild purposes. This might also help setting up eventual archiving of the curated corpus.

The key features of the EuDML platform are its extensibility, allowing easy addition of new services (and content), and its scalability in many dimensions, including the content's volume, content's structure, number of services, number of concurrent users, etc., without performance or reliability degradation. To this end, the system will be designed in a modular, distributed architecture, allowing to replace or provide alternative modules realizing the same or similar functions in the future.

The EuDML system will enable seamless access to the DML resources distributed through the heterogeneous repositories of the partners and other potential content contributors. A number of specific functional requirements are being defined, including mathematics-specific content support for consistent mathematical data presentation (of various provenance), or mathematical formulae search.

While the initial version of the system may be centralized, the requirement for reliability results in a need for a subsequent migration towards a fully distributed architecture that would allow for a redundancy and limiting single points of failure and at the same time would provide a better overall scalability. For this reason, the architecture design has to assume a distributed operations model from the very beginning, even if the first prototypes are expected to be deployed in a simplified environment. Nevertheless, it is absolutely essential that the system in all its versions is able to work in an environment of distributed document and metadata sources, supporting heterogeneous content providers. OAI-PMH is a primary means of content harvesting but other methods are going to be considered and implemented throughout the system's lifetime.

Besides the import capabilities, EuDML will provide access to its stored contents to external services, through specified access interfaces. A secure integration with commercial content providers services, and possibly also with selected federated authentication systems is also envisaged. While security may not seem to be the critical feature of a DML system, certain aspects require careful consideration. As all distributed systems relying on the open Internet for inter-process communication, EuDML by its nature may be prone to service impersonification and denial of service (DoS) type attacks. While service-to-service authentication and authorisation mechanisms will have to be implemented to guarantee the system's integrity, service and data backup-restore and redundant load balancing functions will guarantee its adequate availability. In order to be able to handle non-free content, special care will be paid to access control and access accounting functions. Finally, user authentication will allow end users to customize the work environments attached to their accounts, annotate content, and use the system's community

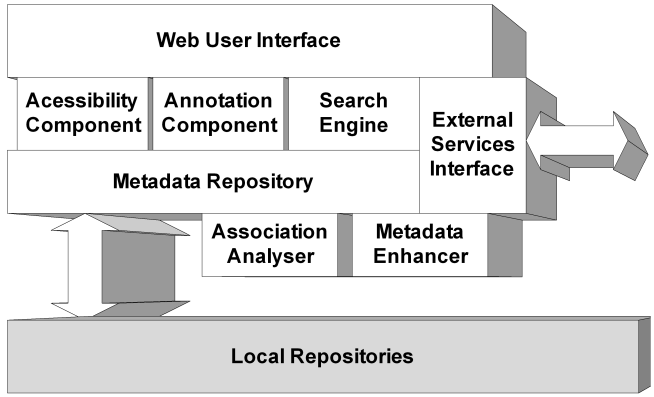

Fig. 1. EuDML functional vision

cooperation and other Web 2.0 functions. The registered authors should be able to claim their works and create their own bibliography portfolios.

The end user features include an efficient way of content presentation through a custom designed web user interface integrated with Web 2.0 content enrichment features. Besides the common search/browse/display scenarios, EuDML will offer the user a number of personalisation features, community and collaboration support services, and content enrichment tools (annotations, personal keywords, personal ranking etc.).

The initial functional expectations of the system, presented in the diagram in Figure 1, conceptually consist of a metadata repository, a search engine, a metadata enhancer, an association analyser, annotation and accessibility functions and of course the interfaces. Each of these abstract concepts will actually materialize as a number of functional components in the functional specification, and may eventually consist of a range of different tools and services, that will be improved and extended over time, and that will be able to handle different aspects of the expected functions. The metadata repository will provide the central point of reference for all the managed contents. It will work directly with an OAI-PMH harvester to ingest repositories' content descriptions, will be able to map the metadata into the internal EuDML schema, will provide item identifier resolution facilities, and store the metadata, and a copy of the actual fulltext content when appropriate. The performance and the quality of responses of the search service will directly influence user experience. Therefore, particularly this service has to be reliable, scalable and customized to fulfil user expectations. Apart from the common search functionality, innovative solutions for searching in mathematical contents will be sought. The metadata enhancer function will consist in a collection of tools that each will contribute to expand or complete the existing items' metadata, depending on the improvements needed. These will 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 analyser will be able to detect, analyse and record relations between individual items. The annotation component will provide mechanisms to attach new material to individual items in the repositories and maintain this new material. The accessibility component will provide support for enhanced accessibility of items, if required, before presentation to end users. Finally, the user and system interfaces will provide access to the collected resources on different levels both to human and machine users, respectively. They will also provide interfaces for integration with other knowledge infrastructure and third party services.

The authors encourage the community to discuss the presented system's features and provide suggestions for additional functionalities, considered to be vital for the broadly defined platform's usability.

4 System Architecture

Based on a detailed functional requirements specification, the EuDML system's architecture is going to be designed. The extensibility, scalability and reliability requirements lead to an eventually fully distributed, platform independent (Java based) solution. While the general architecture design will follow the Service Oriented Architecture (SOA) paradigm, the communication layer will remain abstract, so that individual services will be able to communicate through different adequate means, including possible direct connection when deployed in a single location (host). On the other hand, universal SOAP communication will be maintained as the default for flexibility and compatibility reasons, and more lightweight remote communication (e.g. REST or other content-driven) will be also possible, where required. For this purpose a layered service structure will be adopted, presented in more detail at the end of this section on page 20.

The performance requirement results in the need of a careful selection of critical processing components, and where possible, mature and proven technologies will be used, with which the partners have adequate experience. The additional benefit of using the partner developed technologies will be the system's sustainability after the end of the project, when the partners responsible for individual services will be able to continue curating them by further developing the relevant software code, while keeping EuDML's compatibility and other specific requirements in mind.

4.1 Modular Design

The modular design principle will pertain not only to the backend services but to the user interface as well, allowing for existence of alternative user interfaces, or embedding EuDML portlets in third party services, where required.

Each of the functional modules presented in the previous section shall be realized by means of a separate service or a group of services acting

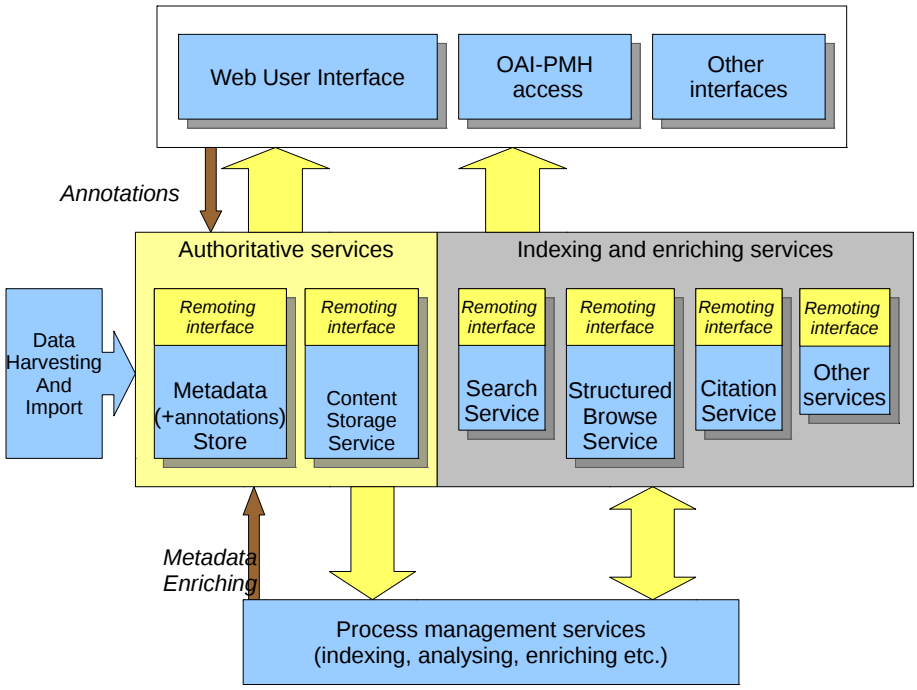

Fig. 2. EuDML core and extension services

together. At the design stage each service will be characterized by its formal contract definition, which will be subsequently used for testing purposes (and particularly for the regression tests in a continuous integration environment). At the same time, a service's contract will define its behaviour and ascertain the fulfilment of the desired functional requirements.

Many technologies and solutions that are required in EuDML already exist either developed by the project partners or elsewhere, and in many areas new developments will not be required. However, a careful design and evaluation of the alternatives will always have to be carried out. Also, in order to assure a proper follow-up, the evolving environment will have to be taken into account, and EuDML will not be limited to its initial specific functional requirements only, but will also conceive other requirements for wider interoperability. Where possible, the system backend modules will be based on existing partner-developed code of proven and deployed services, in order to economize the development effort, capitalize on partners' experience and secure the future sustainable development.

The system, based on Service Oriented Architecture, will consist of a set of core services and a number of extension or enriching services. The core services are defined as a set of services required and sufficient for the basic

system operation. They include the publication metadata store, the indexing and the search services, the content storage system and structured publication browsing services. The architecture concept outline is presented in Figure 2 on the preceding page. The core services realize the functions of the metadata repository, and search engine, described in the previous section.

4.2 The Metadata Store

The Metadata Store will be composed of several separate services acting together: a Metadata Repository Manager; a Storage Manager; and a Metadata Registry. The Metadata Repository Manager REPOX has been chosen. REPOX [28] is a framework developed in the scope of the project TELplus and already deployed in the TEL central service and in several TEL data provider libraries (it also is being redesigned in order to be used by Europeana project). The YADDA Storage Service is our metadata storage tool of choice. YADDA [35] is a service oriented distributed digital content management and provisioning platform, originally developed for Polish national Virtual Library of Science project, with its core components successfully deployed in several large European production content infrastructures (e.g. DRIVER [25], or OpenAIRE [24]). REPOX and YADDA are capable of managing a large number of data sources and storing large quantities of heterogeneous records, with additional version control support, tagging and other required control features. This will allow storing not only bibliographical records in various schemas but also user created content referring to the custom work. An important requirement also will be the assignment of a persistent identifier to each entity (metadata record or document) and the related resolution service able to point to a local copy and back to the original resource's locations.

Despite the fact that EuDML will be using its internal common metadata schema, it is anticipated that multiple different metadata patterns will be used by different content providers and data sources, and these will have to be reversibly mapped onto the internal data structure. That implies the Metadata Store will have to support adapting the different forms of metadata that each provider has about their items to the common format required by the EuDML. For that purpose, the Metadata Repository also will include, in its internal architecture, a Metadata Registry (MDR). Besides the traditional references for this work [16,17], the results of the XMDR project [34] will be also considered. For the purpose of the EuDML Metadata Registry, the MDR technology already developed in the TELplus project and in deployment in TEL will be evaluated, it is being made more generic and robust in the scope of Europeana project. For the Search Service and Structured Browse Service relevant YADDA services will be considered, which proved to be mature, stable and performant, and are already successfully used in a number of European large scale content infrastructures. The YADDA Search Service, based on the Apache Lucene search system is well prepared to be installed and integrated within the whole system. Subsequently, a Solr [31] based YADDA Search Service version will be

evaluated, and solutions that enhance the indexing system with support for math formulae will be sought.

4.3 Extensions

The other expected functions, including annotation, accessibility, metadata enhancer or association analyser will be realised as separate enriching services, that while following the same service design principles, will be considered extension services. Examples of such services include the Citation Service, responsible for citation resolving and indexing or Similarity Service [26,27], which would be able to return similar objects based on a predefined metrics and criteria. Similarly, additional extension services are hoped to be developed in the future by third parties or by the involved partners.

4.4 Interfaces

One or more Web User Interface Services will be developed, based on user requirements criteria. Functional interfaces and widgets may also be prepared to make it possible to include an "EuDML Search Box" in other local systems and portals. A widget configurator may be developed, making it easy for users to create tailored search interfaces for their own websites. Other functional interfaces will be also designed and implemented for services related to interoperability, based on common standards such as OAI-PMH, or on Web services, following the SOAP or REST paradigms where reasonable. All the system's operations will be managed through a process management service, which will be responsible for operation scheduling, synchronisation and timely execution, and for the overall system level integrity of the services.

On a different level, the authorization and authentication services will ensure secure service-to-service operations, and users authentication required to personalize their accounts environment and to access any restricted contents. A YADDA AAS2 service, playing a similar role in European Repository DRIVER infrastructure will be considered for this purpose.

4.5 Layered Services

An important feature of the EuDML system architecture is its layered service structure, as seen in Figure 3 on the facing page. Each service has a pluggable layer for service remote access, allowing to avoid the necessity to select a single service access method and then to support only it throughout the system's lifetime. In EuDML, each service will be defined using a common pattern of a separate API for the client interface (service facade) operating on a common service interface suitable for remote access. A proper definition of the service interface layer, with error encapsulation and strict request-response pattern makes it ideally suited for enabling remote access with any of the existing technologies, including SOAP Web Services, RESTful services, RMI,

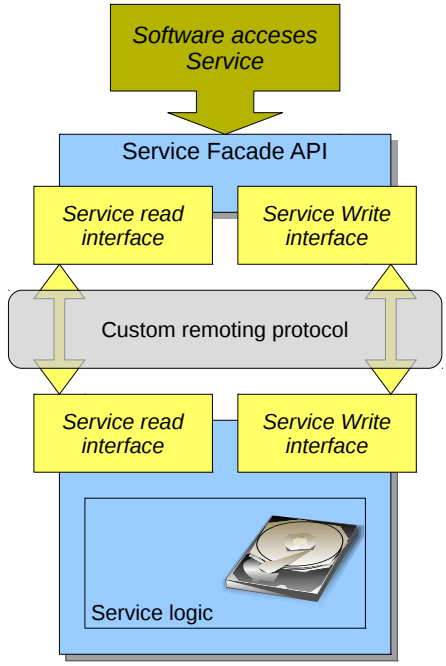

Fig. 3. EuDML service layers

HTTP invoke or no remoting (direct invocation) at all. At the same time the facade interface can be easily extended with any method required. Using well established enterprise frameworks, like Spring, will allow to define remote operation method independently of the source code. In a centralized, high performance scenario, local services will be able to talk to each other directly through their service interfaces, while in a rich, distributed environment, well documented and stable Service Facade APIs shall be exposed for other services to use. This added level of flexibility will allow the infrastructure or subsets thereof to be installed and used in various deployment scenarios, from a high performance single SMP machine based centralised system, through a multiple low-end machines deployment to eventually a global distributed infrastructure.

It is expected that after completion of the initial design phases, where the detailed functional specification and system architecture, internal metadata format and metadata mappings will be defined, a prototype system will be implemented. The fully functional system is expected to be deployed at the end of 2012.

5 Main Challenges

One of the typical challenges of any large scale digital library system dealing with heterogeneous, distributed content is the optimal metadata harmonisation. While all the typical issues related to content multilingualism, different data provenance and quality, versioning, duplicate content, or need to merge controlled vocabularies are present in EuDML, it is hoped that through the use of adequate flexible tools it will be possible to reiterate and optimize the chosen solution. Additionally, exploiting the fact that all the content in question is either mathematical by its nature or closely related to mathematics, should allow to apply mathematical knowledge management techniques to overcome the barriers and find the mathematics-specific relations between individual objects and their groups. To this end, provisions for extensive text mining will be supported, and adequate data structures, and appropriate analytic tools will be designed.

Besides the metadata harmonisation issues, it is apparent that in order to be adequately comprehensive, a mathematical knowledge environment has to be able to manage both open access and restricted licensed content. On the technical level, this will be handled by the AA service developed at ICM and successfully used in DRIVER and YADDA systems (where open content has to seamlessly co-exist with millions of restricted licensed articles and books from commercial publishers). However, the challenge lies in the careful design of license usage scenarios, so that the system would be considered trustworthy by the commercial mathematical content providers.

An obvious challenge will be not only implementing the mathematics specific user interface functions (such as formulae presentation, which parsing is still often a performance bottleneck) but also supporting mathematics specific functions by the backend services (such as formulae search). In the future, close integration with existing mathematics tools and environments is planned.

An important challenge, probably rudimentary for the EuDML's success, although not of technical nature will be to position the service as a recognized authoritative information source not only in the mathematicians environment, but also for the services and users of all other disciplines, searching for mathematical knowledge. One of the prerequisites for this will be to select and accumulate a critical mass of quality content, trusted and relevant to our users (this is assured by the content contributed by the partners) and services (which is the reachable goal described in this paper) taking advantage of the potential and the resources of the project partners. Another issue, not to be neglected, is maintaining high level of true interoperability with other subject specific and general content infrastructures not only on European but also international level.

6 Future Directions

One of the important issues not being addressed directly at the present state is the long-term preservation of the mathematical content. It is feared that

without proper preservation and curation activities some of the publications eventually will be, and some probably already are lost to the community. While the proposed DML system currently does not encompass the complete long-term preservation procedures, it will provide the infrastructural support for their future implementation. Moreover, providing the centralized registry view and control mechanisms, it will make it possible to detect the endangered items in advance. Besides, EuDML may keep electronic copies not only of the metadata but also of the original content fulltexts for its internal processing purposes, that can be also used for emergency recovery. It is essential that these opportunities are exploited and proper long-time preservation strategies are designed and implemented as soon as possible.

Another issue to be considered is the eventual full distribution of the service. While a complete distribution may not be viable for performance reasons, it should be considered as an additional service reliability measure. The system is being designed with a distributed model of operation in mind from the very beginning. However it is assumed, that for the first several years, it will operate from ICM's redundant servers in Warsaw. As ICM operates three separate datacentres, it is planned that servers in two distinct locations in Warsaw will be used. Only after the systems fault tolerance and failover mechanisms are implemented and verified, would a larger scale distribution be reasonable.

A larger long-term organizational and technological challenge will be encompassing all the mathematical contents beyond our first day partners', first by cooperating with more European collections, publicly or privately owned, freely or non-freely accessible, then by extending geographically beyond the continent's boundaries. One of the important future issues is to establish a model for cooperation with the publishers. Finally, it is planned to expand the EuDML infrastructure to integrate external tools and environments, both mathematics related and other relevant. Particularly valuable will be EuDML content and services integration with other international knowledge and content management infrastructures.

7 Conclusions

After many years of efforts, finally the dream of many European mathematicians to have a common digital library of mathematical contents comes true. This is the first breakthrough towards a universal DML since the advent of isolated digitisation programmes around the world at the turn of the 21^{th} century. We thus hope it will help shape and drive the forthcoming efforts towards a more comprehensive worldwide digital mathematics library. Special attention has to be paid in order not to waste this chance.

While the current consortium, building the EuDML system, forms a strong group of reliable partners, the system's sustainability and further development after the project ends will be challenge. It is identified as such and will be specifically addressed as one of the activities scheduled in the project. However, some of the provisions allowing us to trust in the true EuDML sustainability

include the design of gradual development through modular upgrades and extensions of the technologies already developed by individual partners, relatively wide community support from the very beginning of the project, with many stakeholders directly involved and the EMS chairing the advisory committee, and the fact that the infrastructure will be directly maintained and operated by the partners with many years experience in quality service provisioning.

The open questions remaining include subsequent funding scheme, need to design and implement archival policies, the way to cooperate most efficiently with commercial publishers, and the way towards eventually opening access to all mathematical contents.

Acknowledgement

EuDML project 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. 250,503).

References

1. Miroslav Bartošek, Martin Lhoták, Jiří Rákosník, Petr Sojka, and Martin Šárfy. DML-CZ: The Objectives and the First Steps. In: Borwein et al. [4], pages 69–79.
2. Hans Becker, Kari Stange, and Bernd Wegner, editors. *New Developments in Electronic publishing*. FIZ Karlsruhe, 2004. `http://www.emis.de/proceedings/Stockholm2004/`.
3. José Borbinha. Digital Libraries and the Rebirth of Printed Journals. In: Borwein et al. [4], pages 97–110.
4. Jonathan Borwein, Eugénio M. Rocha, and José Francisco Rodrigues, editors. *Communicating Mathematics in the Digital Era*, MA, USA, 2008. A. K. Peters.
5. Thierry Bouche. Introducing the mini-DML Project. In: Becker et al. [2], pages 19–29. `http://www.emis.de/proceedings/Stockholm2004/bouche.pdf`.
6. Thierry Bouche. Some Thoughts on the Near-Future Digital Mathematics Library. In: Sojka [30], pages 3–15. `http://dml.cz/dmlcz/702540`.
7. Thierry Bouche. Towards a Digital Mathematics Library? In: Borwein et al. [4], pages 43–68.
8. Thierry Bouche. Digital Mathematics Libraries: The Good, the Bad, the Ugly. *Mathematics in Computer Science*, 3(3):227–241, May 2010. `http://dx.doi.org/10.1007/s11786-010-0029-2`.
9. Thierry Bouche. Introducing EuDML—The European Digital Mathematics Library. *EMS newsletter*, 76:11–16, June 2010. `http://www.ems-ph.org/journals/journal.php?jrn=news`.
10. BulDML. Bulgarian DML at the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences. `http://sci-gems.math.bas.bg/`.
11. CEDRAM. Centre de diffusion de revues académiques mathématiques. `http://www.cedram.org/`.
12. ELibM. The Electronic Library of Mathematics of the European Mathematical Information Service. `http://www.emis.de/journals/`.

13. Sarah E. Thomas et al. Digital Mathematics Library. Final Report. Technical Report NSF Award # DUE-02066-40, Cornell University, 2004. http://www.library.cornell.edu/dmlib/DMLreport_final.pdf.
14. Thomas Fischer. The Digitization Registry at SUB Göttingen. A Step towards a DML Registry. In: Becker et al. [2], pages 55–63. http://www.emis.de/proceedings/Stockholm2004/fischer.pdf.
15. HDML. Hellenic DML at Ionian University. http://dspace.eap.gr/dspace/handle/123456789/46.
16. ISO/IEC JTC1 SC32 WG2. ISO 11179—Metadata Registries (MDR), 2009. http://metadata-stds.org/11179.
17. ISO/IEC JTC1 SC32 WG2. ISO/IEC 20944—Metadata Registry Interoperability and Bindings (MDRIB), 2009. http://metadata-stds.org/20944.
18. Allyn Jackson. The Digital Mathematics Library. *Notices of Am. Math. Soc.*, 50(4):918–923, 2003. http://www.ams.org/notices/200308/comm-jackson.pdf.
19. Enrique Macias-Virgós. Some Digitization Initiatives in Spain. In: Becker et al. [2], pages 137–142. http://www.emis.de/proceedings/Stockholm2004/virgos.pdf.
20. Zentralblatt MATH. A reviewing database. http://www.zentralblatt-math.org/zmath/.
21. NUMDAM. Numérisation de documents anciens de mathématiques. http://www.numdam.org/.
22. Committee on Electronic Information Communication of the International Mathematical Union. Best current practices: Recommendations on electronic information communication. *Notices Am. Math. Soc.*, 49(8):922–925, 2002. http://www.ams.org/notices/200208/comm-practices.pdf.
23. Committee on Electronic Information Communication of the International Mathematical Union. Digital Mathematics Library: A Vision for the Future, 2006. Endorsed on August 20, 2006 by the General Assembly of the International Mathematical Union. http://www.mathunion.org/ceic/Publications/dml_vision.pdf.
24. OpenAIRE. Open Access Infrastructure for Research in Europe. http://www.openaire.eu/.
25. DRIVER project. Networking European Scientific Repositories. http://www.driver-repository.eu/.
26. Radim Řehůřek and Petr Sojka. Automated Classification and Categorization of Mathematical Knowledge. In: Serge Autexier, John Campbell, Julio Rubio, Volker Sorge, Masakazu Suzuki, and Freek Wiedijk, editors, *Intelligent Computer Mathematics—Proceedings of 7th International Conference on Mathematical Knowledge Management MKM 2008*, volume 5144 of *Lecture Notes in Computer Science LNCS/LNAI*, pages 543–557, Berlin, Heidelberg, July 2008. Springer-Verlag.
27. Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In: Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks. pp. 45–50. Valletta, Malta (2010), software available at http://nlp.fi.muni.cz/projekty/gensim.
28. REPOX. A Metadata Repository Manager. http://repox.ist.utl.pt/.
29. RusDML. Russian DML. http://www.rusdml.de/.
30. Petr Sojka, editor. *Towards a Digital Mathematics Library*, Birmingham, UK, July 2008. Masaryk University. http://www.fi.muni.cz/~sojka/dml-2008-program.xhtml.
31. Solr. An open source search platform. http://lucene.apache.org/solr/.
32. Bernd Wegner. EMANI—Leader and Follower for the WDML. In: Becker et al. [2], pages 161–169. http://www.emis.de/proceedings/Stockholm2004/wegner.pdf.

33. Bernd Wegner. RusDML 2008: Current Facilities of the Core Archive of Digitized Russian Publications in Mathematics. In: Sojka [30], pages 83–86. http://dml.cz/dmlcz/702547.
34. XMDR. eXtended MetaData Registry. https://xmdr.org/.
35. YADDA. A digital content management and provisioning platform. http://yaddainfo.icm.edu.pl/.
36. Katarzyna Zamłyńska, Łukasz Bolikowski, and Tomasz Rosiek. Migration of the Mathematical Collection of Polish Virtual Library of Science to the YADDA platform. In: Sojka [30], pages 127–130. http://dml.cz/dmlcz/702538.
37. Katarzyna Zamłyńska, Alek Tarkowski, and Tomasz Rosiek. Evolution of the Mathematical Collection of the Polish Virtual Library of Science. *Mathematics in Computer Science*, 3(3):265–278, May 2010. http://dx.doi.org/10.1007/s11786-010-0022-9.