

Final report

From dml_wiki

The Future World Heritage Digital Mathematics Library Symposium was held at the National Academy of Sciences (NAS) in Washington, DC, on June 1-3, 2012, organized by the International Mathematical Union (IMU)'s Committee on Electronic Information and Communication (CEIC). The organizers gratefully acknowledge the support of the Alfred P. Sloan Foundation.

Contents

- 1 About the Symposium
- 2 Proposal Development
- 3 Scope of the World Digital Mathematics Library
- 4 Administration
- 5 Building on Existing Repositories
- 6 Search and Computability
- 7 Involving the Community
- 8 Summary

About the Symposium

The purpose of the symposium was to bring together a select group of experts to consult on the ongoing goal of a World Digital Mathematics Library (WDML), assess the requirements for its realization, discuss the progress made to date, and pinpoint significant challenges that remain to be overcome. A total of 52 participants attended in person and one remotely by Skype, coming primarily from the US and Europe and also China, India, Japan, and Russia. Among the participants, 29 were supported through the grant from the Sloan Foundation, of whom 10 were domestic and 19 international. The remaining participants were either local, or supported using AMS funds, or used their own funds.

The conference schedule ran over 2 1/2 days, and consisted of 11 keynote talks, interspersed by 8 panel discussions, and several break out sessions, some organized in advance, and some spontaneously during the meeting. Due to time constraints, the panel discussions were held in parallel pairs. A complete program, list of participants, slides of all keynote talks and many panel presentations, position papers by speakers and other participants, and further discussions of issues can be found on the Symposium wiki

<http://ada00.math.uni-bielefeld.de/dml>

In addition to the workshop, the Sloan Foundation has separately funded a project for the NAS to assemble a committee of experts to conduct a study on the potential of a future digital mathematics library. Thus, an

important goal of the meeting was to provide the recently formed NAS committee with a foundation upon which to base its forthcoming deliberations.

Proposal Development

The first and most important conclusion shared by all participants is the need to maintain momentum now that there is renewed interest and enthusiasm for the WDML. Participants left the meeting mostly convinced that some version of a World Digital Mathematics Library can be achieved within a reasonable time frame. Right now, there exists an exceptional window of opportunity, with several key people prepared to volunteer their time and expertise to furthering the project. But it is not clear how long such an opportunity will last, and hence there is a pressing need to act quickly and decisively. However, to move forward, a variety of critical and challenging issues need to be addressed, with secure funding being a primary consideration. One danger is that the decision making processes of many funding agencies, particularly in the US, are so ponderous that momentum is lost as each new phase is being preplanned and justified. Thus, support from a variety of sources, including government granting agencies, foundations, the UN, professional societies, as well as institutions and libraries, should be explored.

An important initial step is to come up with realistic (rough) estimates of the overall size of the task. These will, of course, depend on the envisioned scope of the project, and include estimates of the overall size of the literature, of what has already been digitized, of the projected expenses, and of the required number of paid staff and volunteers. The funding requirements of both phases of the project — establishing the WDML and then sustained maintenance and upgrading of the system — need to be realistically assessed.

Scope of the World Digital Mathematics Library

Perhaps the most critical topic of discussion, and the biggest cause of disagreement, is in specifying the overall scope of the WDML. At its most expansive, such a library should contain "all" of the world's mathematics in an open, freely accessible, and searchable form. Beyond the question of what "all mathematics" actually encompasses, assembling a comprehensive collection immediately runs into tricky, unresolved copyright issues involving more recently published material, along with the current controversies and dramatic changes resulting from the internet revolution in publishing. In contrast, a less visionary, but more realistic and achievable goal would be to develop a World Heritage Digital Mathematics Library (WHDML), which concentrates on the freely available classic literature, incorporating material that is out-of-copyright or copyrighted but freely accessible, as well as, possibly, orphaned works. Such a WHDML would then serve as a prototype, a testing ground, and a foundation for any eventual broader WDML project.

Thus, the principal challenge is to reconcile development of an initial project of manageable size and scope with the grand vision of an all-encompassing WDML. A realistic strategy would be to formulate an overall plan that starts with the classical mathematical corpus, but includes mechanisms and capabilities for expanding the scope as required. Decisions will have to be made as to what to include, i.e. what qualifies as "mathematics", but these can be left up to individual content providers, perhaps softly regulated by setting good practices. (The same issue is continually dealt with by Math Reviews and Zentralblatt MATH (zbMATH), and the WDML can take advantage of their experience.) In this vein, one question is whether statistics, mechanics, and other close physical applications should be included, especially in view of how these subjects were so closely intertwined

during the classical period of development of mathematics. For the initial project, the main emphasis should be on assembling archival journal articles, say by designating a set of primary historical mathematical journals. Books that are not copyrighted should also be included, since they focus and condense math knowledge; estimates are that these form around 5% to 10% of the total corpus. A potential challenge to the heritage library is how to prevent re-copyrighting of digitized versions of publicly available material.

While the core is being identified and assembled into a usable form, the content growth process must also be addressed. Indeed, it is essential that any initial WHDML be designed so that it can be extended and broadened, allowing additional mathematical material to be successively and straightforwardly incorporated into the corpus. As the library expands through further digitization and incorporation, it should also include heritage mathematics from other traditions and regions (e.g. India, China, Egypt, Arab/Persian, etc.) During this development phase, it will be a continual challenge to resolve the tension between the use of the heritage literature as a limited testing ground for a prototype WDML, versus the need to also include more recent mathematics to maximize the utility of the library to working mathematicians, whose references can include classic literature from the nineteenth century and earlier, modern journal articles, books and text books, newly posted arXiv preprints, conference proceedings, theses, historiography, personalia, collected works, and even modern forums such as math blogs, MathOverflow, social media, etc.

Administration

The symposium debated the type of administrative structure that would be required to bring the WDML (or WHDML) into existence. Most came to the conclusion that the optimum would be a consortium controlled by the international mathematical community, with a full time salaried executive to oversee the project, along with a dedicated staff. While many of the tasks can be delegated to committees consisting of community volunteers and smaller stakeholders, it is essential that the project be headed by a director whose job depends directly upon its success. As the premier international mathematics organization, the role of the IMU as a potential umbrella organization needs to be clarified. One option would be that the IMU appoint a governing board, containing academic, non-commercial, and commercial participants (such as Google, Microsoft, Wolfram, mathematics publishers, etc.), which in turn appoints the WDML director and staff, and to whom they report. One issue that needs to be addressed is to what extent the board would also be responsible for soliciting funding, setting priorities, and allocating resources among the different repositories, versus adopting a more decentralized model for funding and resource development. Further, while there is a need to be aware of the legal implications of a WDML, these must not be allowed to impede current progress. As part of the charge to the WDML consortium, there must be an agreed upon and realistic set of milestones and a timeline for achieving them. While much of the work can be done locally and through online collaboration, there will also be a need for regular general symposia, specialized workshops, and technical meetings. The proposed infrastructure will require significant start-up funding, along with a modest, but absolutely indispensable, long-term income stream for maintaining, sustaining, and upgrading the system.

There was a strong consensus that the underlying structure of the WDML be distributed and decentralized, and easy for the community to contribute to. Curating and maintaining material will depend on volunteers taking ownership of pieces of the library, similar to people "owning" Wikipedia pages, while making their content available with an open license for reuse. Archiving and support for long term preservation of the electronic files is essential, as hardware, software, computing and storage capabilities, and protocols will evolve: in other words, the WDML must be "future-proof". Individual DMLs and repositories should be interconnected through

seamless navigation, efficient search, and additional features, based on the needs and desires of the community. In this manner, the consortium can concentrate its efforts on two main activities:

1. formulating and encouraging the adoption of detailed "best practice" recommendations for distributed archiving and publication of digitized material with both primary text data and metadata that is machine-readable;
2. creating and maintaining a central index for all digitized mathematics material and applications (search, linking, computability, classification, etc.) that are developed by the community in order to access and make use of the corpus.

While the mathematics community must be responsible for the eventual WDML, it will be essential for the consortium to make use of the expertise, experiences, and advice of professional librarians, archivists, information technologists, and commercial publishers and companies involved in building other forms of digital repositories and libraries, while in turn offering mathematical perspectives and tools that can be adapted to other digital libraries.

The consortium will need to develop a basic service platform based on the aggregated content, that is simply and transparently designed so that non-technophiles can easily access and use the material. This in turn will allow other parties, including academic, open source, and commercial developers to design software and applications that will enhance and extend the basic foundation maintained by the WDML consortium. For this and many other reasons, the entire infrastructure needs to be completely open, including the digital data, the metadata, the search algorithms, and all code that forms the core of the WDML, and also open in its extensibility, thereby inspiring the mathematical community, as well as software engineers and programmers, to develop additional applications and services. To foster and accelerate the development, there would be great advantage from having the key review journals (e.g., Math Reviews and Zentralblatt MATH) open their nonproprietary bibliographic metadata to the WDML to incorporate and build on what has already been accomplished. This should be accomplished by appropriate data exchange between the databases of the reviewing journals and the central WDML index. While both commercial and non-commercial agents should be able to develop their own enhancements and applications over WDML data, the licensing of this data must enable the corpus to be freely accessed and available to everyone, especially academic and non-profit users, while preserving the rights, if any, of content owners. It is extremely important that the project be a truly World DML, implemented in such a manner that the developing world will benefit from it at least as much as countries with established scientific, mathematical and computing infrastructure. Services, applications, and enhancements should be offered which make it attractive for institutions, libraries, and even commercial publishers to provide their digital content to the WDML under well defined conditions, that includes a commitment to open access.

Building on Existing Repositories

Since the initial calls in the late 1990's for the establishment of a WDML, a large fraction of the historical literature has now been digitized, and is in principle available in a large variety of locations. The WDML consortium will need to motivate even more organizations to digitize and publish the mathematical material in their possession, so that the community can access it, index it (initially through general purpose search engines), process it, evaluate it, etc. However, even with what is currently available, there are still a host of obstacles and barriers that hinder finding, accessing, interlinking, and searching the extant material. There is a critical need to

assemble content that is currently scattered and hard to locate, and then develop basic search, interlinking, and referencing capabilities that builds on the existing online corpus of heritage material. A significant challenge is that the existing digital repositories employ a variety of file formats, and different standards for data and metadata, which are, moreover, not stable over time. Then a principal task of the WDML consortium will be to develop a set of standards, protocols, and application programming interfaces (APIs) that will allow distributed math content providers to interoperate.

Thus, prior to the establishment of a WDML, an initial step will be to set up and maintain a comprehensive registry of all the existing mathematics literature (books, journals, preprints, etc.) that is available online. (A prototype is Ulf Rehmann's page on "Retrodigitized Mathematics Journals and Monographs", http://www.math.uni-bielefeld.de/~rehmann/DML/dml_links.html, which currently contains links to 4608 digitized books and 576 digitized journals/seminars.) Such a comprehensive registry would, at the very least, enable the user to easily access the material that is already available online, which, even today, presents a significant challenge. The registry can also point to copyrighted material that is available on the web, thereby at least letting people know how (i.e., under what conditions) they can access such material, irrespective of whether such material is eventually incorporated into an open access WDML. The registry would also serve as a focal point for digitizing and collecting additional material, for demonstrating how to integrate the available corpus, and thus for starting to build the WDML. The following step is to assemble a comprehensive open list of article metadata, thereby fostering the design of search and indexing applications. Openness is key throughout, in order that the quality and scope can, at each stage, be controlled by the math community.

During the symposium, the existing European Digital Mathematics Library (EuDML) was often cited as a prototype for building a truly global DML. Consequently, the question of whether such a WDML is feasible is, in principle, resolved. The eventual WDML can capitalize on the accumulated expertise and accomplishments of the project EuDML. One potential strategy would be to merge and extend the EuDML with an as yet unrealized US-DML, as well as include heritage material from other countries, such as Australia, Canada, China, India, Japan, Korea, Russia, etc., in order to form the core of the WDML.

Search and Computability

An essential feature of a WDML is that it include basic search capabilities, which can then be supplemented by more sophisticated search algorithms. Computability, which was discussed in depth by the participants, remains controversial, with some enthusiastic proponents and others quite skeptical that a system (e.g. one modeled on Wolfram Alpha) can be of genuine use to the working researcher. On the other hand, search and computability are not independent since powerful search tools must carry out computations in order to find/select relevant material.

For the heritage literature in a WHDML, sophisticated search processes will require multilingual and multicultural mathematics dictionaries and thesauri, including mechanisms for dealing with the dynamic and culturally-specific changes in terminology, concepts, notation, rigor, etc. This will require extending the existing MSC classification scheme, keywords, and possibly reviews to historical texts, perhaps eventually leading to an advanced automatic classification of mathematics. Currently, the field of mathematical search and computability is progressing rapidly. While the development of sophisticated search algorithms may require input from computer scientists, to be of value to the WDML, they must ultimately be tailored to mathematics, and constitute a significant improvement on existing general-purpose search engines.

Two other important issues:

1. Develop an open system of Name Authority Files — that is, assigning unique author identifiers, so an author's list of publications can be created automatically and correctly from metadata maintained by diverse curators; such a system would integrate and extend the systems already created for this purpose by MathSciNet, zbMath, OCLC, Microsoft, and the broader ORCID initiative.
2. Equip the central WDML index with some form of citation resolver, to form the basis for interlinking of the literature. This tool should integrate with CrossRef, the DOI resolver founded and directed by publishers, with the MRef tool provided by MathSciNet and a similar lookup tool of zbMATH, with indexing and search services provided Google Scholar, Microsoft Academic Search, and with whatever other services may enhance the value of WDML content by linking to related information.

Mathematical document collections can also be enhanced by automated added-value services like definition lookup, prerequisites explanation, or notation adaption, once the deep semantic structure of documents is explicitly marked up. Work is under way in the "Mathematical Knowledge Management" community for developing markup systems, algorithms, and tools that can provide information access at a level that previously only humans could provide (after understanding the documents), leading to "active documents" in semantic document collections. The assembly of a historical corpus of semantically marked-up documents holds the promise of providing access to mathematical knowledge beyond the "one-brain-barrier", e.g. discovering long-range correspondences that remain undiscovered, since no human has studied both subjects to notice structural similarities. The driving vision is to explore to what extent technology can facilitate new mathematical research, in a manner that goes well beyond merely providing linking and pdf delivery of material.

Involving the Community

Ultimately, while certainly of interest to archivists, librarians, and historians, the most important measure of success will be that the WDML/WHDML be actively used, appreciated, and publicized by working research mathematicians, as well as researchers in other scientific disciplines. The WDML consortium, along with the IMU and other involved parties, need to rouse the broader math community, its professional societies and the academies, to enthusiastically support and promote the establishment of a WDML. The realization of a WDML or WHDML has the potential to serve as an exemplar of what can be accomplished for science and for the world at large. Publicity from leading mathematicians, as well as favorable postings on key mathematics blogs, will play an important role in this campaign. The IMU and the WDML consortium should investigate the opportunities and obstacles to building a supporting community, thereby identifying those with time, energy, and financial resources to devote to the project.

Summary

In summary, there is a tremendous opportunity right now to make significant progress on achieving the highly anticipated goal of a WDML within a reasonable time frame. There is a pressing need to enthusiastically and without delay build on the renewed momentum sparked by the IMU/NAS symposium and the support of the Sloan Foundation. Progress will require understanding and addressing the important issues and challenges, which include:

- Organization structure

- Funding
- Scope and focus
- Expanding the online corpus
- Openness of data and metadata
- Intelligent search
- Extensibility
- Building enthusiasm

It is the hope of the participants and organizers that this symposium and the upcoming NAS study committee will, in the future, be regarded as a significant milestone in the achievement of the overall goal of a WDML.

Retrieved from "http://ada00.math.uni-bielefeld.de/mediawiki-1.18.1/index.php?title=Final_report&oldid=894"

- This page was last modified on 23 August 2012, at 18:25.
- This page has been accessed once.