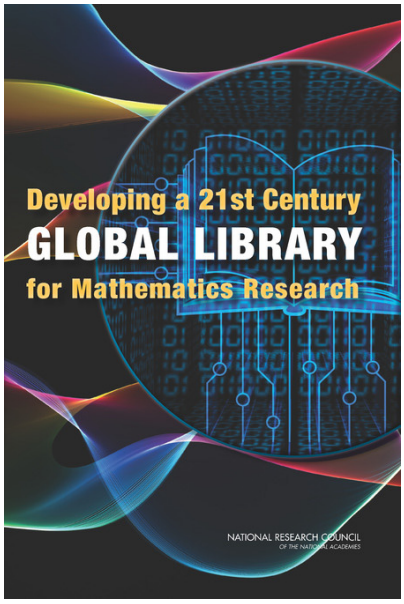
NATIONAL ACADEMY OF SCIENCES
NATIONAL ACADEMY OF ENGINEERING
INSTITUTE OF MEDICINE
NATIONAL RESEARCH COUNCIL

REPORT IN BRIEF

Developing a 21st Century Global Library for Mathematics Research

Board on Mathematical Sciences and Their Applications ·
Division on Engineering & Physical Sciences · March 2014

Mathematics is a cumulative discipline: new research is reliant on well-organized and well-curated literature. Web-based access to that literature is now mature and quite effective, allowing mathematicians to search individual items within the literature using attributes such as subjects, titles, authors, dates, and citations. Yet, while much information is contained in the individual items that comprise the mathematical literature, a greater amount of information is represented by the links between items. This is not just via references but in the interrelation of concepts, insights, and techniques as they are developed and spread between mathematical disciplines. For example, if mathematicians were able to search the literature for instances where a specific equation was used or solved, it would allow them to consider alternative approaches toward solving their own research questions.



This report details how information contained in individual items within the literature could be readily extracted and linked to create a comprehensive Digital Mathematics Library that is more than the sum of its contributing publications. Specifically, the report proposes the establishment of an organization; the development of a set of platforms, tools, and services; the deployment of an ongoing applied research program to complement the development work; and the mobilization and coordination of the mathematical community to take the first steps toward these capabilities.

Background

The idea of a comprehensive digital mathematics library has been around for decades, and there have been several incarnations of the idea with different foci. The first step in this vision was the retrospective digitization of the older parts of the literature that did not already exist in digital form. This step has largely been achieved. Yet, current digital mathematical resources do not allow a user to systematically explore the information captured within all the literature and readily explore connections that may not be obvious from looking at the material alone.

Mathematics is at a pivotal junction. It can either continue to utilize digitized mathematics literature in ways similar to traditional printed literature, or it can take advantage of new and developing technologies to enable new ways of advancing knowledge. In 2011, the Alfred P. Sloan Foundation commissioned this report to evaluate the potential value and future capabilities of a Digital Mathematics Library, as well as the issues to overcome in setting up such a library.

Current Databases

Finding: The construction of mathematical libraries through centralized aggregation of resources has reached a point of diminishing returns, particularly given that much of this construction has been coupled with retrospective digitization efforts.

To be clear, what this report proposes builds on the extensive work—digitizing and aggregating the literature—done by many dedicated individuals under the rubric of the World Digital Mathematical Library and other community initiatives. Necessary incremental improvements will likely continue to occur in these areas, but they do not require an initiative on the scale of what is being called for in this report. Nor is another secondary indexing service that tracks bibliographic information needed at this time.

The Digital Mathematics Library

Currently, the real opportunity is in offering mathematicians new and more direct ways to discover and interact with mathematical objects and mathematical knowledge through the Web. Through a combination of machine-learning methods and a community-based editorial effort, a significantly greater portion of the information in the global mathematical corpus could be made available to researchers as linked open data through a central organizational entity—referred to in this report as the Digital Mathematics Library (DML).

Recommendation: A primary role of the Digital Mathematics Library should be to provide a platform that engages the mathematical community in enriching the library's knowledge base and identifies connections in the data.

The DML would aggregate and make available collections of ontologies, links, and other information created and maintained by human contributors, curators, and specialized machine agents with significant editorial input from the mathematical community. It would include capabilities for annotating, searching, browsing, navigating, linking, computing, and visualizing both copyrighted and openly licensed content. While the DML would store modest amounts of new knowledge structures and indices, it would not generally replicate mathematical literature stored elsewhere. Instead, it would strive to represent the mathematical knowledge presented within a publication and illustrate how it is connected with other resources.

Recommendation: The Digital Mathematics Library should rely on citation indexing, community sourcing, and a combination of other computationally based methods for linking among articles, concepts, authors, and so on.

Development of the DML

Recommendation: A Digital Mathematics Library organization should be created to manage and encourage the creation of a knowledge-based library of mathematical concepts such as theorems and proofs.

The report suggests an incremental development of the DML, starting with the creation of a small nonprofit organization, referred to here as the DML organization. The DML organization will need a small and dedicated paid staff, including a well-respected mathematician in a senior role, to ensure its development and growth. Ideally, the DML would be attached to and draw support from some host institution in order to facilitate sharing of services and to reduce overhead. The DML organization could be governed ultimately by the mathematical sciences community through an existing organization.

The first and foremost challenge that the DML will face is finding a set of primary funding sources that could support its initial development and early operations (a period of between 5 and 10 years). Upon launching the DML effort, there would ideally be a coalition of partners with a commitment to the DML concept. Later, the DML would become a self-sustaining entity once some of its key capabilities are established and a potential sustainable business model is chosen from among options.

Recommendation: The initial DML planning group should set up a task force of suitable experts to produce a realistic plan, timeline, and prioritization of components, using this report as a high-level blueprint, to present to potential funding agencies (both public and private).

Partnerships

A key early issue for the DML organization is how to establish constructive and effective partnerships with existing publishers, Web services, and other resources, both those specific to mathematics and those serving the much broader scholarly community. The report recommends partnering with these current service providers whenever possible rather than replicating the capabilities of existing resources. Some of these partnerships might be challenging because of copyright concerns. However, establishing fruitful partnerships is essential to the success of the DML.

Recommendation: The Digital Mathematics Library should be open and built to cooperate with both researchers and existing services. In particular, the content (knowledge structures) of the library, at least for vocabularies, tags, and links, should also be open, although the library will link to both open and copyright-restricted literature.

Research

Another key issue for the DML will be in establishing the necessary technical, organizational, and community-coordinating capabilities to reach its goals. Some of the technologies required to build the tools and services described in this report do not exist today or are not sufficiently mature. The establishment of the DML needs to be complemented by a long-term (5 to 10 years) commitment to a focused applied research program. This program would encompass both needed technology and independent research to understand how the DML is being used and how well it is working.

Recommendation: Community engagement and the success of community-sourced efforts should be continuously evaluated throughout DML development and operation to ensure that DML missions continue to align with community needs and that community engagement efforts are effective.

Recommendation: The Digital Mathematics Library should serve as a nexus for the coordination of research and research outcomes, including community endorsements, and encourage best practices to facilitate knowledge management in research mathematics.

Conclusion

Like other scientific disciplines, mathematics is now completing a complex multi-decade transition from print to a digital system. Currently, this new system closely emulates print for both authors and readers. The mathematics community is thus at an inflection point where it has the opportunity to think about how its collective knowledge base will be structured and used in the digital world.

Mathematics maintains a healthy and constructive relationship with its past. The literature of the field goes back hundreds of years and some of this literature has a long "shelf life." Initiatives that refresh and restructure the corpus of mathematical literature—abstracting it into a knowledge base for future centuries—are an investment in the future of mathematical scholarship. The Digital Mathematics Library proposed in this report provides a platform and a context to achieve this and also offers a critical point of focus for the mathematical community to engage in discussions about the creation, curation, and management of mathematical knowledge in a genuinely digital environment.

Committee On Planning A Global Library Of The Mathematical Sciences: Ingrid Daubechies, Duke University, *Co-Chair*; **Clifford A. Lynch,** Coalition for Networked Information, *Co-Chair*; **Kathleen M. Carley,** Carnegie Mellon University; **Timothy W. Cole,** University of Illinois at Urbana-Champaign; **Judith L. Klavans,** University of Maryland, College Park; **Yann Lecun,** New York University; **Michael Lesk,** Rutgers University; **Peter Olver,** University of Minnesota, Minneapolis; **Jim Pitman,** University of California, Berkeley; **Zhihong (Jeff) Xia,** Northwestern University

Project Staff: Michelle Schwalbe, Study Director; **Scott Weidman,** Director, Board on Mathematical Sciences and Their Applications; **Barbara Wright,** Administrative Assistant

Support for this project was provided by the Alfred P. Sloan Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organization that provided support for the project, or the National Research Council.

Copies of this report are available free of charge from http://www.nap.edu.

Report issued March 2014.