Planning a 21st Century Global Library for Mathematics Research

Jim Pitman and Clifford Lynch

The literature corpus and knowledge base of mathematics is both bedrock and wellspring for future research. How can we make this resource more valuable as it transitions to digital form and is used in conjunction with new network and computationally based services? Over the past two years, a committee established by the National Research Council, a part of the United States National Academies, with funding from the Alfred P. Sloan foundation, has explored these issues. Many scholarly disciplines are considering these questions; the answers for each discipline seem to be different, and they are shaped by the practices of the discipline, the extent to which the historical literature continues to be relevant to current research, and many other factors. The report of this Committee on Planning a Global Library of the Mathematical Sciences has been recently published by the National Academies Press at `http://www.nap.edu/catalog.php?record_id=18619` and is also available at `http://arxiv.org/abs/1404.1905`.

The committee's charge called for an evaluation of the potential value of such a library and some consideration of its appropriate scope; an assessment of the issues and alliances involved in establishing such a library; the identification of a range of desired capabilities for such a system and a consideration of which of these capabilities were likely to be within reach given current and foreseeable technology; and a sketch of resource needs and the way forward with such a project. The committee was specifically asked not to focus on issues such as copyright and open access, but rather to take a pragmatic view of operating within the current diverse landscape of scholarly information. Further, the committee was asked to focus on the needs of researchers in the mathematical sciences, rather than the needs of the vast range of other disciplines that use and rely on mathematics. In its work, the committee took a broad view of the ecosystem of mathematical literature and information services and the extent to which the needs of mathematics researchers (particularly within major research institutions) are currently met.

The committee report included a number of findings and recommendations, indicated here by italicized paragraphs, which are quoted here verbatim:

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.*

Put another way, the committee recognized the potential value of moving toward a broader view, going beyond aggregation alone to create a comprehensive digital mathematics information resource that could be of much greater value than the sum of its contributing publications. Building on the extensive work done by many dedicated individuals under the rubric of the World Digital Mathematical Library, as well as many other

Jim Pitman is a professor in the Departments of Statistics and Mathematics at the University of California, Berkeley. His email address is `pitman@stat.berkeley.edu`.

Clifford Lynch is Director of the Coalition for Networked Information. His email address is `cliff@cni.org`.

Pitman and Lynch were members of the Committee on Planning a Global Library of the Mathematical Sciences, which wrote the report discussed here. Other members were Ingrid Daubechies, Duke University; Kathleen Carley, Carnegie Mellon University; Timothy Cole, University of Illinois at Urbana-Champaign; Judith Klavans, University of Maryland, College Park; Yann LeCun, New York University; Michael Lesk, Rutgers University; Peter Olver, University of Minnesota, Minneapolis; and Zhihong "Jeff" Xia, Northwestern University. Daubechies and Lynch co-chaired the committee.

community initiatives, the committee recommended establishment of an organization, tentatively called the Digital Mathematics Library (DML), to support a wide variety of new functionalities and services over aggregated mathematical information, including dramatically improved capabilities for searching, browsing, navigating, annotating, and linking mathematical concepts. Specifically, the DML organization should

• develop a collection of platforms, tools, and services for curation and navigation of mathematical information;

• mobilize and coordinate the mathematical community to engage with these capabilities;

• support an ongoing applied research program in mathematical information management to complement the development work.

The report envisions a combination of computational machine learning and textual analysis methods along with community-based editorial efforts, in order to make a significantly greater portion of the information and knowledge in the global mathematical corpus available to researchers as linked open data through the DML. The report describes how such a library might operate—discussing development and research needs; its role in facilitating discovery and interaction; and the importance of establishing partnerships with publishers, abstracting and indexing services, and other current players in the mathematical information ecology. Several of the capabilities described—such as the ability to annotate across the full corpus of mathematical literature—seem to be low-hanging fruit. One focus of particular interest, tantalizingly on the fringe of current technology in the committee's view, when combined with community editing, was the extent to which the mathematical literature might be adequately tagged with identifiers for mathematical concepts to facilitate linking and navigation.

Finding: *While fully automated recognition of mathematical concepts and ideas (e.g., theorems, proofs, sequences, groups) is not yet possible, significant benefit can be realized by utilizing existing scalable methods and algorithms to assist human agents in identifying important mathematical concepts contained in the research literature—even while fully automated recognition remains something to aspire to.*

Following are some further recommendations of the report:

Recommendation: *A DML organization should be created to manage and encourage the creation of a knowledge-based library of mathematical concepts such as theorems and proofs… It should be an advocate for the mathematics community and help develop plans for development and funding of open information systems of use to mathematicians.*

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

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

Recommendation: *Community engagement and the success of community-sourced efforts need to 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 DML 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.*

Recommendation: *The DML 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.*

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).*

Recommendation: *The DML needs to build an ongoing relationship with the research communities spanning mathematics, computer science, information science, and related areas concerned with knowledge extraction and structuring in the context of mathematics and to help translation of developments in these areas from research to large-scale application.*

Members of the committee and others hope to offer several articles in coming issues of the *Notices* that look into various facets of the proposed DML in more detail, and there are ongoing discussions about how the work outlined in the report might be advanced. But perhaps the most important objective now is broad vetting of the report within the mathematical sciences community. As one step in this process, there will be a presentation and panel discussion about these recommendations by the Committee on Electronic Information and Communication (CEIC) at the upcoming International Congress of Mathematicians in Seoul, Korea, 6:00–7:30 p.m. on August 20, moderated by Peter Olver of the committee that authored the Academies report. The panel discussion will involve Thierry Bouche, Ingrid Daubechies, Gert-Martin Greuel, Rajeeva L. Karandikar, and June Zhang.