Trusted data and services from the GDML

International Mathematical Union

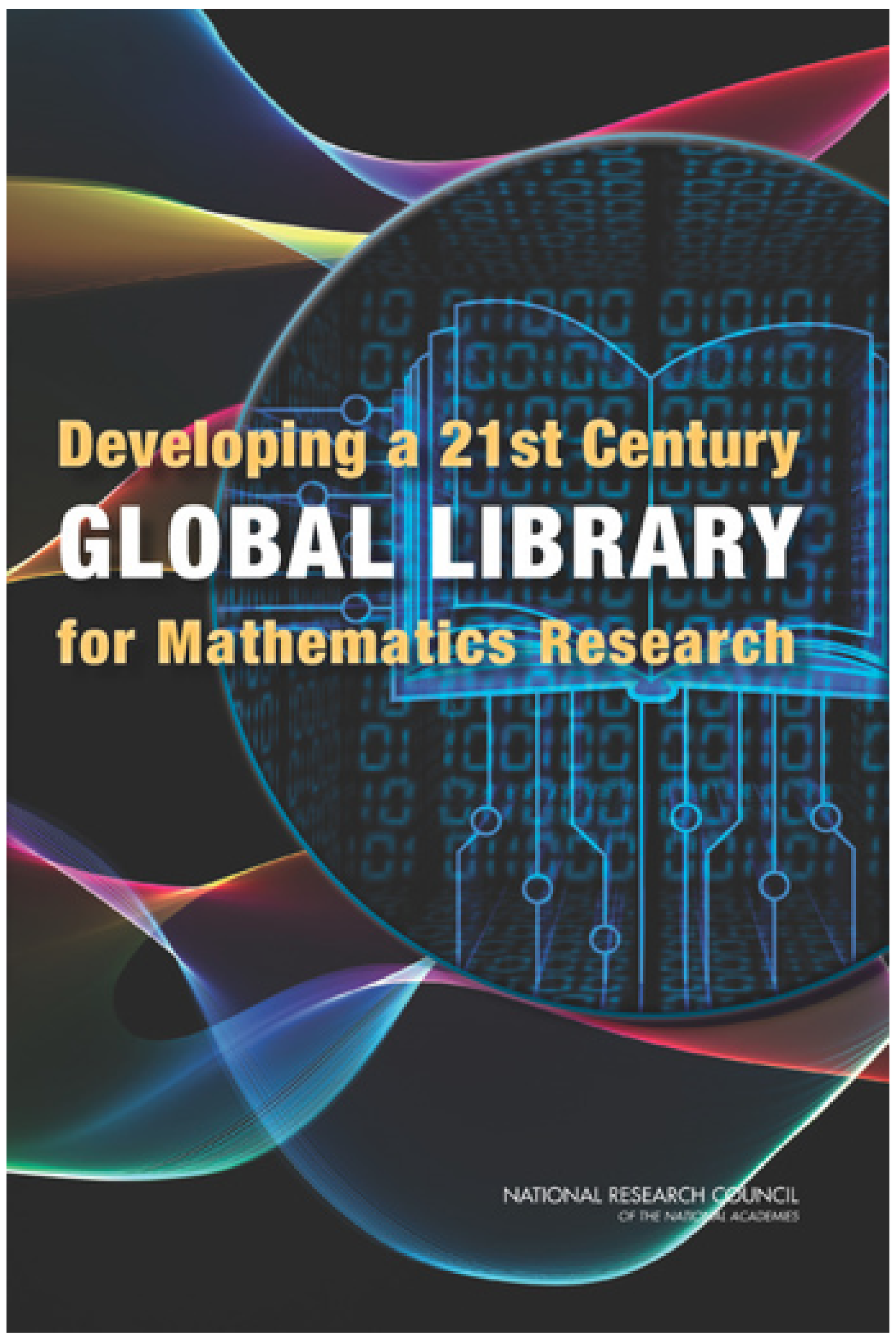
Committee on Electronic Information and Communication

History

The International Mathematical Union (#IMU) strongly supports the establishment of a Global Digital Mathematical Library (#GDML). It realises that mathematicians can either continue to utilise digital literature in ways similar to traditional printed literature, or can profit from emerging technology to advance knowledge.

IMU President Ingrid Daubechies and Chair Peter Olver of the IMU's Committee on Electronic Information and Communication (#CEIC) took the initiative to work toward a GDML through consultations with a broad expert group. The GDML vision was codified by the General Assembly of the IMU who, back in 2006, endorsed a statement, "Digital Mathematics Library: A Vision for the Future", by CEIC that "… endorses this vision of a distributed collection of past mathematical scholarship that serves the needs of all science, and encourages mathematicians and publishers of mathematics to join together in implementing this vision."

In 2012, CEIC organised the International Symposium on *The Future World Heritage Digital Mathematics Library*. The conclusions of the meeting explore practical mechanisms, challenges, and capabilities required for the realisation of the GDML [1].



During the International Congress of Mathematicians in 2014 (#ICM2014), the newly established GDML WG, chaired by Patrick Ion, and under the sponsorship of the IMU, was charged with designing a roadmap towards the GDML, determining its organisational structure, prioritising requirements for its implementation, estimating a budget, including start-up and sustaining funds, and fostering the writing of proposals to funding agencies. The panel discussion, that was begun during ICM2014, now continues on the newly established GDML blog [2].

The IMU supports the statement on Open Access, endorsed by ICSU at the General Assembly in 2014, which contributes significant blocks to the infrastructure needed for the concretisation of the GDML. John Ball, former CEIC Chair, now member of ICSU's Executive Board, who chaired the Open Access WG, inaugurated the "ICSU News" column of the bi-monthly bulletin IMU-Net.

The GDML

The vision for the GDML is that of a growing corpus of public-domain and openly licensed mathematical information, Web services, and software agents, coexisting with present mathematical publishing and indexing services. Imagine being able to search the literature for instances where a specific bit of math was used or solved: it would allow you to consider alternative approaches toward solving your own research questions. This type of search capability could be facilitated through the use of a database of machine-generated and human-cultivated information about the mathematical literature and allow for a variety of other capabilities to be built.

Today we have the opportunity to expand and redefine the way in which mathematical knowledge is represented and used, the character of mathematical literature and how it evolves, and empower mathematicians by new possibilities of interacting with knowledge. This future relationship with the literature and the mathematical knowledge corpus will go beyond new forms of access and analytical tools; it will also include tools and services to accommodate the creation, sharing, and curation of new kinds of knowledge structures.

Which mathematical data?

Important data for the working mathematician, and in general for the scientist, is previous knowledge. Digitisation efforts for mathematical knowledge have been undertaken in the last decades with the European Digital Mathematics Library, EuDML, and the National Science Digital Library, NSDL, being prominent examples of metadata-driven centralised services. The current digital corpus of bibliographic information for mathematical literature is extensive:

- MathSciNet (1940-present) holds over 3 million items and indexes over 2000 journals.
- Zentralblatt, zbMATH (1868-present, incl. Jahrbuch), contains more than 3 million entries and currently indexes more than 3000 journals and serials.

Looking at the increase in mathematical papers added over the past 5 years to arXiv, it seems that more and more mathematical literature will be in digital form, some with high-quality markup, specifically those "born" digital or retro-digitized to be in a machine readable format such as LaTeX or MathML.

Lists and tables have always been essential for the working mathematician. The most basic are numerical tables (e.g., values of logarithms, trigonometric functions, special functions, zeros of the zeta function, integer sequences). More sophisticated are lists of mathematical objects (e.g., indefinite and definite integrals, finite simple groups, Fourier transforms, partial differential equations and their solutions), or even lists of definitions, axioms, and theorems. Example digital collections include:

- LOCOMAT, the LORIA Collection of Mathematical Tables is a library of recomputed historical tables.
- NIST Digital Library of Mathematical Functions (DLMF), a digital revision of Abramowitz and Stegun's Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, likely the most widely distributed NBS/NIST technical publication of all time.
- On-Line Encyclopedia of Integer Sequences (OEIS) celebrating 50 years of identifying sequences and still receiving over 100 new sequences and updates per day, all handled by volunteer editors.
- Wolfram Functions Site, the world's most encyclopaedic collection of information about mathematical functions.

Which mathematical services?

Modelling and simulation have now been applied in many areas of science from Aerodynamics, to Economics, and Medical Imaging. Notable software for the computational scientist relies on mathematical software such as:

- Portable, Extensible Toolkit for Scientific Computation, PETSC, used extensively in modelling and simulation.
- SWMath is an extensive database of information on mathematical software.
- TIDES permits to numerically integrating ODE problems using Taylor Series with multiple precision, allowing to solve ODE problems up to any precision.
- SuperLU is a high-level library for the direct solution of large, sparse, non-symmetric systems of linear equations on high performance machines.

Simulations for climate modelling or for planetary motion, both long-time and large-scale, are examples of situations that may require higher-precision arithmetic, beyond 64-bit accuracy. Recently developed software packages have tackled this challenge of scientific computing making numerical precision a required component as important to the program design as are the algorithms and data structures.

In connection with reliability of science, reproducibility of scientific results has received attention and requires dealing with the entire working environment of the scientist doing the computation:

- Recomputation.org, a first repository for experiments in computational science achieves reproducibility of results by recomputation and by archiving virtual machines alongside code and data.

A big step forward in the discovery of mathematical services will also result from improvements to the documentation of mathematical software which is still basically human oriented and largely uncategorized or lacking any kind of machine readable metadata.

Future perspectives

Expanding the range of computer software that is math-aware will enable moving from text mining to math mining. This work will cover aspects of classification and representation of mathematical knowledge, computational linguistics, aggregation and analysis of corpora, tools for [meta]data and full-text processing, OCR and document analysis, information retrieval developments, and document processing workflows. The needs of large user communities will drive this process towards the goals of the GDML and this will eventually improve access for all.

References

[1] National Research Council, Developing a 21st Century Global Library for Mathematics Research. Washington, D.C., 2014.*ISSN:978-0-309-29848-3*

[2] 'The World Digital Mathematics Library' (WDML). url:http://blog.wias-berlin.de/imu-icmpanel-wdml.