

# International Mathematical Knowledge Trust - IMKT: an Update on the Global Digital Mathematics Library

Patrick D. F. Ion<sup>1</sup>, Thierry Bouche<sup>2</sup>, Gadadhar Misra<sup>3</sup>, Alf A. Onshuus<sup>4</sup>, Stephen M. Watt<sup>5</sup>, and Liu Zheng<sup>6</sup>

<sup>1</sup> GDML WG & University of Michigan, MI USA

<sup>2</sup> Université Grenoble Alpes, Grenoble, France

<sup>3</sup> Indian Institute of Science, Bangalore, India

<sup>4</sup> Universidad de los Andes, Bogotá, Columbia

<sup>5</sup> University of Waterloo, Waterloo ON, Canada

<sup>6</sup> National Science Library, Chinese Academy of Sciences, Beijing, China

**Abstract.** A Panel at the 2018 International Congress of Mathematicians concerning the efforts to realize the dream of a Global Digital Mathematics Library consisted of Thierry Bouche, Gadadhar Misra, Alf A. Onshuus, Stephen M. Watt and Liu Zheng and was moderated by the author of the lines below. This report contains a description setting the stage for the panel, summaries of the panelists' statements and of some questions and answers at the session as well as at a later open opportunity for further discussion.

## 1 Intent of the Panel

This panel was arranged by the Committee on Electronic Information and Communication (CEIC) of the International Mathematical Union (IMU) through its Global Digital Mathematics Library Working Group (GDML WG). This panel follows up on a previous panel in Seoul at ICM 2014 with the title “The World Digital Mathematics Library” which coincided with the inception of the GDML WG. That working group came to the conclusion that a significant step to realizing to dream of a World Digital Mathematics Library would be to set up a legal entity, the International Mathematical Knowledge Trust, initially based in Canada, and hence the name of this panel.

A recording of the panel session may be able to be seen in [the YouTube collection of ICM 2018](#) like the [Seoul 2014 panel](#).

The objective of the session was to provide the community with information about further strengthening of access to our mathematical literature and its preservation as a global public good. This is part of building a World Digital Mathematics Library, as endorsed by the IMU as long ago as 2006.

It is necessary to increase the commitment of our community to what it being done and what needs to be done to preserve our mathematical heritage, and it was hoped the variety of relevant issues that can be discussed could improve both support for WDML and the effectiveness of the movement setting it up.

Issues that might have been discussed include:

- The structure of a GDML and its services — why we set up an International Mathematical Knowledge Trust
- Opportunities for worldwide collaboration
- The state of mathematical publishing in the digital age
- Approaches to getting more out of the presently available digital literature by data mining and statistical methods
- The design of a semantic markup language for mathematics
- The present tools for the digital handling of mathematics and how they interact: e.g.  $\text{\TeX}$ , MathML, MathJax, Word, computer algebra systems and proof assistants,
- Issues of reproducibility of results when computing is involved
- The need for authority trails in a reliable literature

A difficulty with developing a GDML is the breadth of relevant issues as shown in the list above, itself by no means an exhaustive one. The actual content of the panel emphasized, rather than technical discussions the global nature of the effort involved by the diversity of panelist origins and experience, and, in particular, by the representation of the largest populations centres of the world.

It seems worth repeating some historical references mentioned at the 2014 panel on GDML efforts. At the first ICM in 1897, in Zürich, there was a session including pasigraphy, under the chairmanship of G. Peano, concerned with questions of how to encode mathematical knowledge. (It was in connection with such efforts that Peano developed his axioms for the natural numbers.) Schröder said there he was sure that pasigraphy would take its rightful place on the agenda of all succeeding ICMs. (Pasigraphy is the study of universal languages of symbols intended to encapsulate semantics and to provide a basis for calculational ratiocination.) This clearly did not happen.

At the 1928 ICM, in Bologna, there was active discussion of how to provide comprehensive bibliographic resources for mathematics to everyone. They were to be based on publishing the exhaustive collection of about a quarter of a million catalogue slips assembled by Georg Valentin. These slips were eventually destroyed in 1942 by fires resulting from bombs dropped on the Prussian State Library on Unter den Linden in Berlin.

Earlier, in the decades from about 1900 on, the Belgians Paul Otlet and Henri La Fontaine planned, and started to create, a master bibliography of all the worlds published knowledge. La Fontaine used his Nobel Prize money for their cause. In 1934, Otlet sketched plans for a global network of “electric telescopes”, as he called them, that would allow people to search and browse through millions of interlinked documents, images, audio and video recordings. He described how people would use the devices to send messages to one another, to ask queries of the central repository and receive answers, to share information and even to get together electrically connected social networks. The whole system was termed a “rseau”. The Second World War killed this effort as well, though there is a revived Mundaneum today to visit in Mons.

The point of these glances to the past is to emphasize that a GDML effort must not give up, even if it takes significant time to mobilize the resources needed.

## 1.1 Panelists

The panelists, listed alphabetically below, were chosen to represent the global intent of the effort to realize access to the legacy literature of mathematics as well as for their expertise with modern digital technology. Their present affiliations are given in parentheses,

- Thierry Bouche (Université Grenoble Alpes, Grenoble, France)
- Patrick D. F. Ion (GDML WG, University of Michigan, Ann Arbor MI, USA) — moderator
- Gadadhar Misra (Indian Institute of Science, Bangalore, India)
- Alf A. Onshuus (Universidad de los Andes, Bogotá, Columbia)
- Stephen M. Watt (University of Waterloo, Waterloo ON, Canada)
- Liu Zheng (National Science Library, Chinese Academy of Sciences, Beijing, China)

## 2 Statements from the panel

The order of speaking was not alphabetical. The initial paragraphs provide short panelist biographies just mentioning details relevant here rather than their much larger CVs. The bracketed initials show the handles used in reporting discussion. Slide sets from the panelists, which are more extensive than those which could be show during the panel session itself, due to time pressures can be found at the [IMKT Projects site](#).

**Stephen M. Watt [SMW]** is now Dean of [Mathematics at the University of Waterloo](#), Ontario, Canada. The Waterloo Faculty of Mathematics includes four Mathematics Departments and Computer Science with more than 8,000 graduate and undergraduate students, 240 full-time professors, and 300 courses in mathematics, statistics and computer science. Previously Watt held positions at Western University (Ontario), IBM T.J. Watson Research Center, INRIA and the University of Nice. Among his main interests have been computer algebra. In particular, he has been a co-author of the CA languages Maple and Axiom, and Aldor. Aside from research publications, he has been a co-author of the World Wide Web Consortium (W3C) markup language standards MathML and InkML. He was Co-Principal Investigator with Ingrid Daubechies on the Alfred P. Sloan Grant that enabled setting up the IMKT based in Waterloo ON.

Stephen Watt had been asked to address particularly IMKT and GDML efforts. He began by noting that Mathematical Knowledge *can* be of permanent relevance, represented exactly, and treated mechanically. This led to envisioning a Global Digital Math Library. Such a notion has been a longstanding vision with many advocates over time, likely including some in the present room. It had been often thought of, in general terms, as linked digitization of (much or) all of past mathematical literature in enhanced, open article repositories. The scope would then be of the order of  $10^6$  to  $10^7$  items. It was the task of the IMU GDML Working Group whose activity was initiated by Ingrid Daubechies at the Seoul ICM to begin to turn this into some sort of reality. The group membership was

- Thierry Bouche (U Grenoble, France)

- Bruno Buchberger (RISC Linz, Austria)
- Patrick Ion (Math Reviews, USA) chair
- Michael Kohlhase (FAU Erlangen-Nrnberg, Germany)
- Jim Pitman (UC Berkeley, USA)
- Olaf Teschke (zbMATH, Germany)
- Eric Weisstein (Wolfram Research, USA)

The GDML WG was charged with the tasks of

- designing a road map for the practical next steps towards the GDML
- determining its organizational structure
- prioritizing the different requirements for its implementation
- estimating an incremental budget, both start-up and sustaining funds
- fostering the writing of proposals to funding organizations

To carry out its remit the WG engaged in the following activities

- extensive regular tele-conferences
- organizing workshops, sessions and panels (such as this)
- specific presentations and conference presentations
- gaining a seed funding grant
- founding a not-for-profit, the International Mathematical Knowledge Trust (IMKT) in Canada
- encouraging seed projects

More about this history can be found at the new [IMKT web presence](#). There can also be found an explanation of the governance of the IMKT, its Board Membership and legal status.

Watt next passed to “What the Future Was”. In an extreme form this could be summed up in the simple question “A linked repository of articles is nice, but don’t we have this already?”, which suggests there’s nothing more to be done.

In the present day there are already all sorts of (conventional) article repositories such as publisher archives — e.g. [Springer](#), [JSTOR](#), [arXiv](#), [EuDML](#), [NUMDAM](#), [Göttinger Digitalisierungszentrum](#), [RusDML](#), valuable private lists like [Ulf Rehmann’s](#), and many more.

In addition there are indexes, review collections, and author databases such as [Inspec](#), [zbMATH](#), [MathReviews/MathSciNet](#), the [Math Genealogy Project](#), and the broad access method [Google Scholar](#).

Furthermore, there are General Summary Works such as the [Wikipedia Mathematics Project](#), [MathWorld](#), the [Wolfram Functions Site](#), the [NIST Digital Library of Mathematical Functions](#), and the [Digital Repository of Math Formulae](#).

There are Specialized Tools and Databases such as the [Online Encyclopedia of Integer Sequences](#), the [Dynamic Dictionary of Mathematical Functions](#), the [Online Integral Calculator](#), the [Inverse Symbolic Calculator](#), [Atlas of Finite Simple Groups](#), the [L-functions and modular forms database](#), the [Combinatorial Statistic Finder](#), [A Catalogue of Lattices](#), and the [Encyclopedia of Triangle Centers](#).

There are Proof Libraries, such as those devoted to Big Proofs of the Four Colour Theorem, the Feit-Thompson Theorem, the Kepler Conjecture, aka the Hales-Ferguson Theorem, and so on. This was discussed in the previous panel at the Rio ICM. But there's an enormous amount of knowledge in the [Mizar Mathematical Library](#), the [Archive of Formal Proofs](#), and the [MetaMath Proof Explorer](#).

Finally there are even more miscellaneous mathematical projects and activities online such as the [Encyclopedia of Continued Fractions](#), the [Maple Function Advisor](#), [Maple implementations of Exact Solutions of the Einstein Field Equations](#), the [GAP Package liealgdb](#), [Matrix Market](#), [EqWorld](#), [Formulasheet](#), and [SymbolicData](#) site.

With all this material actively online it is clear what the current direction is. We are passing from digitized page images to the results of parsing pages, then on to semantic markup of pages, and finally to corpus analysis, once material has passed through the earlier stages or been created with a higher degree of preparation.

A full use of digital technology for mathematics will have to provide support tools for dealing with the meaning of the content of mathematical works. Then we will be able to have a real Mathematical Digital Library, and/or a Mathematical Knowledge Base with Mathematical Knowledge Tools to explore and develop it. The fundamental point is that point is mathematics can be likened to Pando, the Trembling Giant. This is a group of quaking aspen trees estimated to have 47,000 trunks on about 433,000 square meters, weighing a total of 6,000,000 Kg. It is now realized to be a single organism extensively connected by its roots below the ground, which have been growing together and expanding their cover for 80,000 years. Mathematics is perhaps not that old but it is remarkably interconnected as a subject. It too shows "the fundamental interconnectedness of all things".

To return to the more mundane Watt then mentioned the four basic Initiatives that IMKT has been trying to get started. More description of the intent of these initiatives and their progress will be found [on the IMKT site](#).

**Special Function Concordance** The special functions (SF), are traditionally those of mathematical physics, but nowadays are widely applied as well, for instance, in computer science, biology, statistics and finance. They are commonly considered well-known basic material. SF can be calculated with any of the common computer algebra systems or looked up in tables or online, to get both values and properties. However, the various systems and resources that are called upon for reference vary in details of conventions in their definitions. The DLMF from NIST (the thorough-going revision of the famous Abramowitz and Stegun handbook), the DDMF from France, the implementations of special functions by the hundreds in Wolframs Mathematica and Alpha, or Maple, or in Sage (Pari, Gap, et al.) and even the libraries of definitions in automatic theorem provers and logic assistants such as Mizar, HOL Light, Coq and LEAN, all show variations in their handling of common functions, such as the inverse trigonometric, Bessel or elliptic functions. This leads to simple non-interoperability of results that are published, and a tendency for a particular research team to stick with one vendor so as to be at least internally consistent. Only occasionally do massive difficulties result from this Babel, but theres a good deal of time and effort lost in checking consistency when users employ more

than one resource. The SFC initiative is getting representatives of groups with a stake in special functions, as mentioned above, together to find common ways to express their accords and their differences, publicly and in a machinable form. Already Bruce Miller of DLMF has developed OpenMath Content Dictionaries that may act as a springboard for further development.

**FABstracts** This project, suggested by Tom Hales and now running under a Sloan grant awarded to Hales and Jeremy Avigad, aims to provide formal abstracts for mathematical items from an initial sample of mathematics. It is using the language of the LEAN3 theorem prover in its initial efforts, and is international in that the initial teams are in Pittsburg, USA, and Hanoi, Vietnam, with other contributors from all over the world and a GitHub presence. Public results may be expected soon.

**Formal Harmony** There's need for harmonization similar to that for SFC above comes where formalization of mathematics is concerned. In fact, this is a sort of thing that practitioners of mathematical formalization can see as a useful project. There are by now many overlapping pieces of mathematics, such as basic pieces of Euclidean geometry and some fundamental lemmas and theorems of analysis that have had to be formalized in more than one system. The main systems where there is real overlap are HOL Light, Coq and Mizar. What has not been done is to make careful comparisons of the formulations of, say, the Jordan Curve Theorem, or the Fundamental Theorem of Calculus in the various contexts, and to see in what senses they really are the same and really may differ. This is not just a matter of the way the proofs proceed but may be about subtleties in the semantics resulting from different foundational bases.

Clarity should be of value to appliers of mathematics and also to educational efforts. What is required for a theorem's truth is often somewhat obscure. Researchers have expressed interest in exploring the ways systems are the same and differ. As a practical result the search for applicable results may be extended to a federated search over different library collections of already formalized material with increased clarity over what's returned.

**Mathematical Document Analysis and Classification** There are several sorts of document analysis that are possible and have not been applied to the mathematical corpus, which has its own special features. Mathematics is arguably the natural language of a globally distributed people (those with enough training in mathematics to do science and engineering) but not a typical one. For a start, its written form is not truly linear, involves a large sign vocabulary, and sometimes different grammar. We do see a need for power tools for document digestion as the GDML effort ramps up, so Natural Language Processing (NLP) methods tried and tested elsewhere will need to be adapted to handling the mathematical corpus.

$N$ -gram analysis is a relatively easy place to begin. NLP tools here are readily available for ordinary languages. Sample corpora, e.g., those of zbMATH and arXiv are readily available and their curators interested in such ideas. Some early experiments have been tried at zbMATH for keyword generation, and another small project was completed at ORCCA under Watt's direction that concentrated the  $N$ -grams on the mathematical formulas and found characteristics of different subjects could be detected.

This is a much more limited approach than deep learning and needs much fewer resources, but it is applicable in context finding and clustering tools that can also be deployed much more.

In addition, there are other projects underway using machine learning for the examination of the extant mathematical literature that hope to enhance the access to the semantics implicit in the publication data already collected through new search mechanisms and to facilitate input of mathematics through hand-writing and retrieval with math-aware search. The respective examples are being worked on by teams led by Charles S. Lafferty and David M. Blei, and by Richard Zanibbi and Lee Giles.

All these all ultimately must involve bringing together several groups' work.

Watt finished with some 'slightly less obvious statements about getting something done'. First we can afford to hand-annotate or capture the semantics and special characteristics of a small subset of mathematics, but we must rely on improvements in technology and artificial intelligence to get further. The stages would seem to be 'by hand', closely supervised automation, automation allowed to run but then checked by humans, and finally full automation. It is true that approximation can go a long way in this process. However, as projects follow this path they must consciously future-proof their data, not, for instance, throwing away too much information collected at one stage that could be useful for later automation.

Finally, he remarked that mathematics has to solve the problems of mobilizing its community to more open participation and to do this there have to be provided new inducements as returns on effort. There are the examples ranging from the now almost classic [Wikipedia](#) to the newer [Nupedia](#) to encourage us.

**Gadadhar Misra [GM]** is a [mathematics professor at the Indian Institute of Science](#), Bangalore. He is a Fellow of the Indian National Academy of Sciences and its Vice-President of the Publications and Informatics. His main field of mathematical research has been analysis and, in particular, operator theory. He has worked on the Indian efforts to provide an Open Access Repository of publications of Fellows of the Indian Academy of Sciences and to make research open and actionable in a timely manner. India is the country with the second largest population and an example to developing nations.

Misra began his view from India by remarking that in a way the situation there was rather good. The infrastructure of the internet and access to web resources was good in a land with a successful information technology business sector and IT education. Concentrating on progress in the field of mathematics, but applying more generally to much of science, preprint archives provide a platform for permanently storing soft copies of scientific manuscripts, often before peer review, and allow open access to any interested person able to get on the internet. Indeed the NIH has published the statement "Scientists issue preprints to speed dissemination, establish priority, obtain feedback, and offset publication bias", and this is true in our subject as well.

To take a quick look at recent history we can not that one of the first, and popular, online preprint archives was [arXiv](#), which in over 25 years of its existence, strongly influenced many publishers and impacted how science is disseminated.

The arXiv originated at Los Alamos (for physics) and at Duke University (for mathematics). These sites were then combined at the Los Alamos (LANL) centre and have since moved to Cornell University. After the service was widely used mirror sites were introduced.

The IIMSc mirror at <https://in.arxiv.org> is a mirror site for arXiv that has been functional since the late nineties. Currently, that mirror is connected to the internet with a sufficiently high-speed link that it is up-to-date within about 20–30 minutes from the time that the primary site is updated. There is enough local processing power and disk-space to last for quite a while.

This sort of clear success has meant that a new initiative, Open Access India, is in the process of forming a steering committee to manage a Preprints Repository for India, **indiaRxiv**. This is an Open Access Repository of publications of Fellows of the Indian Academy of Sciences. It attempts to collect, preserve and disseminate the intellectual output of the Academy available as publications by its Fellows in peer-reviewed journals.

For the Indian Institute of Science (IISc), which is a country-wide distributed system, an **ePrints@IISc** repository collects, preserves and disseminates in digital format the research output created by the IISc research community. This enables the Institute community to deposit their preprints, postprints and other scholarly publications using a web interface, and organizes these publications for easy retrieval.

Going beyond the individual postings of preprints, or indeed of reprint and alternative material, that allow open and free access there is a new approach that can be taken to publishing that could be of value to mathematics.

[Episciences.org](http://Episciences.org) is an innovative combination of the two routes of free access:

- the gold route provided by hosting journals in open access (overlay journals)
- the green route whereby articles are submitted to journals by depositing them in an open archive

The Hardy-Ramanujan Journal, a respected Indian publication since its inception in 1978 moved to publication on EpiSciences in 2014. (A diagram showing the revised workflow of publication resulting from this change is on a slide in Misra’s presentation).

The growing interest, which is significant for policy making at a national level, in facilitating wide-spread access to mathematical results can be summarized by two recent public declarations by major bodies.

On 3 February 2018 the “Declaration By Open Access Communities & Attendees of OpenCon 2018, New Delhi” stated

We will strive to publish our interim research outputs as preprints or postprints (e.g. Institutional Repositories) and encourage our peers and supervisors to do the same to make our research open and actionable in a timely manner. [See also <http://openaccessindia.org/>.]

On 2 June 2018 the “Recommendation By The Indian National Science Academy (INSA)” declared



Various agencies/organizations in India that fund research should take cognizance of articles that have been deposited in established free open access Pre-Print Archives as a proof of prior-data.

Summarizing, Gadadhar Misra said that in many respects the practical details of access in India to the new mathematical literature through preprints, or indeed to the old insofar as it generally available elsewhere, are largely solved, and there is a growing awareness of how this might be capitalized upon. However, he emphasized that there are important sociological aspects of society, and of the practice of mathematics and science in education, that still obstruct making use of what is becoming technically possible. Concomitant changes in the way mathematical work is judged for promotion and appointments are slow in coming while it is easy to use simpler older methods of counting pages or using other simple-minded statistics.

**Alf A. Onshuus [AO]** is a mathematics professor at Universidad de los Andes, Bogotá, Colombia. He is a member of the CEIC of the IMU, and also of the IMU's Committee on Developing Countries (CDC). As a mathematician his interest are logic and the foundations of mathematics, number theory and algebra. He has been involved in initiatives in Colombia to enhance access mathematical literature and concerned with the details of open access as they impact developing countries.

Onshuus titled his remarks "Some ideas from developing countries for GDML". As an initial overview he offered, and then discussed, four main points:

- Most universities in developing countries, even research universities, don't subscribe to all the main mathematical journals.
- Article access is achieved through:
  1. Personal contacts,
  2. Library access of researchers through an Alma Mater,
  3. Library agreements.
- arXiv has become a very mainstream tool.
- Internet access and speed have both become better in most places. Speed of downloading does not seem a significant obstacle, contrary to what was happening some years ago.

He then moved on to point out some initiatives in Colombia to help access to the mathematical literature; these he felt were a good start and can be examples to developing nations generally.

- Library agreements for sharing journal articles.
- Joint access to MathScinet from many small universities under a consortium agreement.

But there's a challenge we are all talking about. The "Gold Open Standard" initiative has been discussed. Changing the model from a "pay-to-read" to a "pay-to-publish" model can have complicated consequences for scientists in Colombia and the developing world who do not, in general, belong to institutions with large numbers of subscriptions to journals. This is a problem that has to be faced.

Onshuus then offered comments that should serve to emphasize that there are some details with serious consequences to be kept in mind in trying to move towards global digital access to mathematics. These resulted from an informal poll of colleagues he did.

A couple of quotations:

“It would be great to have access to information at more basic levels than what is found in Arxiv. For example, to support a graduate student in Cambodia or Honduras.”

“I have not explored what is available but what I have seen is at the level of research that a student (and most teachers) in many developing countries could not read. Something that would be of special interest at the level of developing countries would be to have documents edited by experts on how one enters some areas of research.”

A couple more replies:

“This is a very interesting initiative. It should be open access. One thing to explore is to map mathematical knowledge, understanding for example how many written articles are in the interface of two subfields.”

“It would be important to discuss first the balance between novelty and soundness, between the number of publications and their quality, between open source and funding for this purpose.”

**Liu Zheng [LZ]** is Deputy Director of the Collection & Knowledge Organization Center of the National Science Library at the Chinese Academy of Sciences, Beijing. She is by training a librarian and information specialist with, nowadays, a special interest in and responsibility for effecting the change from document retrieval to domain knowledge structure discovery using digital technologies. In particular, the mathematics collections of China, the world’s most populous country, are within her purview.

Liu Zheng’s views from China are conditioned by the very size of mathematics in China. She immediately offered a table of the official 2017 statistics estimates for mathematics degrees.

**Table 1.** Mathematics in China

Type	Graduates	University and Research Institutions
Bachelor	66,100–71,500	649
Master	6000	246
Doctor	1000	75

Further examination of mathematics research shows the number of papers published has increased year by year since 2009. Funding for mathematics research is at 500 million yuan

from the National Natural Science Foundation of China (ca. 63,200,000 Eu or 73,500,000 US \$). However, it is an official concern that perhaps the quality of mathematics research is still lower than it should be on average. Obviously there are some centers of excellence but the country is very large.

On the Digital Mathematics Library front in China, unfortunately there is no digital mathematics library as such. Much material is available through electronic database subscriptions to commercially published material. Efforts have gone into constructing mathematics library home pages and subject portals to aid in mathematics electronic journal navigation. There are some platform starts but very few mathematical papers are submitted to reprint platforms or an institutional repository. Zheng showed screen shots of some relevant examples (viewable on her slides).

In China this is a time for planning the future. We hope and suggest, Liu Zheng said, that introducing Chinese research results to international peers, and understanding the progress of similar research on a topic, will encourage mathematics research in China. We need to decrease the time required for retrieval and discovery of mathematics and hope that a GDML will be a helpful tool in this regard.

This a time when we should be passing from from document retrieval to discovery of domain knowledge structure using new digital capabilities.

It can be hoped that integrated and sharing of warehoused collections of, say, journals, proceedings of conferences, monographs, textbooks, tutorial papers etc., will lead to better networks of relationships for mathematical researchers, whether at a university, college or institution and regardless of whether a student, teacher, researcher, or just a user of mathematics.

We will need intelligent tools and assistants that are suited to particular subject domains. In early form they are the formularies, collections of mathematical modeling available today, but it seems they will only be fully developed when more is understood of the foundations and philosophy of mathematics.

In summary, the CAS library is devoted to developing the digital libraries to provide better services to Chinese researchers. It will be happy to join in the GDML effort and work for mathematics.

**Thierry Bouche [TB]** is a mathematics professor at the Université Grenoble Alpes, in Grenoble, France, and also a member of the Institut Fourier there. He is Director of [Cellule Mathdoc](#) and responsible for the significant projects [NUMDAM](#) (math serials retrodigitisation), [Mini-DML](#) (OAI-PMH integrator for digital math articles), [CEDRAM](#) (publishing platform and portal for academic math journals), Gallica-Math (article-level index of volumes digitised at [Gallica](#)) and [Mersenne](#) (extension of Cedram goes beyond math and its core services). He was scientific coordinator for the successful [EuDML](#) (European Digital Mathematics Library) project and is Chair of the succeeding [EuDML Initiative](#). He has been a member of the IMU's CEIC and of the GDML WG. His mathematical field is global analysis, but his interests and involvements have long also extended digitisation, e-publishing and DML.

Bouche discussed European DML and Open Access developments. He started by looking at the basic notion of a Mathematics Library. The librarians main functions are:

- Selection of collections by subject, document type
  - Not necessarily formally published: theses, reports, preprints...
  - Some grey literature needs to be archived or made accessible to patrons (e.g. Nash, Perelman...)
- Acquisition of actual documents (books, journals...)
- Cataloguing for collection maintenance and easy retrieval
- Archiving and Preservation of the curated collections
- Access provision maintained for patrons

EuDML tries to recreate this in the digital era with

- A curated digital archive
- Physically hosted at public (at least non-profit, science-driven) institutions
- Free to use (possibly after some moving wall)

EuDML is conceived as a regional integrator ready for worldwide integration, but is presently missing international partners.

- Worldwide content
- Enhanced with intelligent math-aware agents
- More accessible to humans with disabilities
- Free to mine (more accessible to machines)

The vision for The Digital Mathematics Library is really

- A global (distributed) facility dedicated to archive historical (digitised) material as well as newly published (born digital)
- An up-to-date registry of all available resources
- Mechanisms for interlinking the holdings with existing and future infrastructures
- Seamless navigation across the whole corpus
- Instant and perpetual access

Just let us note the size and special features of our mathematical literature. The reference mathematical corpus is

- About 4,500,000 items ( $135 \times 10^6$  pages perhaps)
- Heavily multilingual
- Growing by about 140,000 new items a year (mostly English)

But Bouche also offered some warnings. Beware confusion of the following:

**Digital Mathematics Library vs. repositories mixing flavours of eprints**

→ We need reliable references

**Trusted archive (original sources) vs. Encyclopedia**

→ we need access to the original sources!

**Mathematical corpus vs. math. knowledge as big data**

—→ True vs. plausible theorems?

**However our main current problem is that we lack support to significantly enlarge the available content.**

This leads us to the matter of Open Access. Mathematicians have an overall favorable attitude towards Open Access (OA), but not at any cost. There are

**Good options:**

- Eventual perpetual OA to all (through GDML/IMKT)
- Diamond OA, with some regulation (e.g. funded and monitored by a scientific organization, selective support by library groups. . . )
- Reasonably priced subscription journals are better than crap OA journals!

and **Concerns:**

- An obvious dangerous cost: APCs (Author Paid Publication)
- Putting societies' publications at risk
- Research integrity might be endangered by open access mandates that could encourage bad practices

On the subject of open access to scientific data and literature and the assessment of research by metrics, there is an important report developed by a Sub-Group of the International Council for Science (ICSU) Executive Board, with input from individuals attending an expert workshop, ICSU Members and the ICSU Secretariat.

The International Council for Science advocates the following goals for open access. The scientific record should be:

- free of financial barriers for any researcher to contribute to;
- free of financial barriers for any user to access immediately on publication;
- made available without restriction on reuse for any purpose, subject to proper attribution;
- quality-assured and published in a timely manner; and
- archived and made available in perpetuity.

These goals apply both to peer-reviewed research publications, the data on which the results and conclusions of this research are based, and any software or code used in the course of the research.

Generally there is the question as to whether the mathematical literature is headed towards a not-for-profit system? A plot of the numbers of journal articles listed in Mathematical Reviews from commercial and non-commercial publishers, offered by D. E. McClure of the AMS in 2010, shows a change from 56.4% commercial and 43.6% non-commercial in 2000 to the reversed position of 55.8% non-commercial and 44.2% commercial in 2009 with a crossover point in 2006.

Another effect on mathematics, which does in fact provide open access, is that of the arXiv. Florian Müller and Olaf Teschke of zbMATH have already raised the question of

“Everything on the arXiv?”, in a article in the European Mathematical Society Newsletter **99** (2016). They plotted the percentage of arXiv coverage for the main MSC subject classes. They found a reasonably complete coverage of recent research in many classes but many of the articles generally will usually not be indexed in zbMATH. On the other hand, the scopes of arXiv and zbMATH differ (for instance, certain math-ph submissions may very well be beyond the scope of zbMATH), and there is also a possible contribution due to the necessary delay in indexing by zbMATH. These two effects can be quite significant. It is possible too that the arXiv seem to be approaching a level of saturation for some subjects. Another impression to be gained from the plots is that behaviours in specific subject area can also change quickly. The study was based on a sample that included 60,000 arXiv submissions with DOIs that matched 45,000 items in zbMATH.

Bouche then passed on to mentioning several different initiatives for Open Access. In France there has been the [Jussieu Call for Open science and bibliodiversity](#) [Points 7 and 8]:

the scientific communities must be able to access national and international infrastructures which guarantee the preservation and circulation of knowledge against any privatization of contents. Business models should be found which preserve their long-term continuity;

priority should be given to business models that do not involve any payments, neither for authors to have their texts published nor for readers to access them. Many fair funding models exist and only require to be further developed and extended: institutional support, library contributions or subsidies, premium services, participatory funding or creation of open archives, etc.

There is also an effort called [MathOA](#) as part of a wider series of OA journal sites for different subjects that promulgate practical [Fair OA principles](#) and have begun ‘flipping’ mathematics journals to their form of OA.

There is a [Free Journal Network](#) with about 25 mathematical titles from all over the world as members. The Electronic Library of Mathematics, [eLibM](#) from FIZ Karlsruhe contains open access journals and proceedings volumes in all fields of mathematics, and lists 11 titles as of writing.

Bouche then passed to the brand-new [Centre Mersenne](#), which he directs. The Centre Mersenne for Open Scientific Publishing aims at supporting and fostering open access scientific publishing. It offers tools and services for scholars and editorial teams of open access journals formatted with LaTeX. Recently it took on the well-known journals “Algebraic Combinatorics” and “Annales de l’Institut Fourier” to add tot he other 7 in its list.

An alternative route to gaining access to the mathematical literature has been the German [DEAL Project](#) which has been negotiating country-wide licenses for the offerings of major publishers, notably Elsevier.

Then for mathematics there may be much to be learned from the initiatives oming out of other subjects, such as [SciPost](#) from physics:

- SciPost is a complete scientific publication portal.

- It is purely online-based, and offers freely, openly, globally and perpetually accessible science.
- Being managed by professional scientists, and making use of editor-solicited and contributed reviews, its Journals aim at the highest achievable standards of refereeing.
- SciPost Commentaries allow Contributors to seamlessly comment on all existing literature.

There is also the approach of [Peer Community in . . .](#) “A free recommendation process of scientific preprints (and published articles) based on peer-reviews”. The humanities have the [Open Library of Humanities](#), and there’s the universal [Knowledge Unlatched](#) that “offers free access to scholarly content for every reader across the world. Our online platform provides libraries worldwide with a central place to support Open Access models from leading publishing houses and new OA initiatives.”

In summary, there are many ways to approach Open Access still. What will be best for mathematics remains to be seen. It is to be hoped that the GDML/IMKT efforts can contribute to and benefit from the currently changing publishing context.

**Patrick D. F. Ion [PI]** chaired the panel. He is a member of the IMU’s CEIC and chair of the GDML WG. He spent a 30-year career as an editor at [Mathematical Reviews](#) (MR) run by the [American Mathematical Society](#), but maintains an [affiliation with the University of Michigan](#) at Ann Arbor. Before that he had positions at the University of London (UK), Rijksuniversiteit Groningen (Netherlands), RIMS Kyoto (Japan) and Universität Heidelberg (Germany); on sabbaticals he spent years at Université Strasbourg (France) and IHES, Bures-sur-Yvette (France) and months at the University of Auckland (NZ). In addition to work as a mathematician editor he was closely involved in the change at MR to the use of [T<sub>E</sub>X](#) and other digital tools, and later to its transition to a database-driven operation and ultimately to [MathSciNet](#). He took a leading role in the revisions of the Mathematics Subject Scheme (MSC) in 2000 and to the current [MSC2010](#), and its modernization to a SKOS form of Linked Open Data. His involvement in mathematical knowledge management (MKM) is of long standing, and he is on the Executive Committee of the [OpenMath Society](#). He was co-chair of the [World Wide Web Consortium Math Working Group](#) for 12 years, and an editor and author of the W3C specification MathML, now an ISO standard as [MathML3](#) and part of the HTML suite, and of [XML Entity Definitions for Characters](#). His mathematical interests have been in quantum field theory and statistical mechanics, quantum stochastic processes, noncommutative geometry,  $q$ -analogues of special functions and combinatorics, and the discrete Fourier transform in connection with elementary geometry.

As moderator, Ion tried to summarize the state of play as follows: Different panel members were able to address some of the different aspects of what’s involved in setting up a global digital mathematics library. It is obvious the GDML/IMKT effort needs to learn more about community views and to serve them. We did have on the panel geographic diversity and representatives of the world’s largest populations, but the hope has to be we are all willing to try to think globally as well.

What has happened since ICM 2014 is that, supported an Alfred P. Sloan Foundation grant, we have a legally established International Mathematical Knowledge Trust. The IMKT

web presence is expected to be able soon to provide services to the mathematical community. The four initiatives IMKT set out are starting to be realized, with FABstracts well underway. The SFC, which could have been imagined to be the easiest front on which to make open-source progress, has seen a company, Wolfram Research, make the most headway with additions to the Wolfram Language (aka Mathematica) for mathematical entity handling. With the very recent progress at DLMF this should mean SFC moves forward further. There's work afoot on the other two initiatives mentioned, and on significant projects using machine learning supposed to build tools for mathematicians to use.

One obvious GDML subgoal still to be realized is direct enhancement of access to mathematical literature. We hope to be able to build from what EuDML has achieved, but that is still in the, hopefully near, future.

If the GDML WG from 2014 has clearly achieved something, it is that it has not let the effort to develop a GDML drop, which seems to have occurred more than once before in our subject's long history. This long-term commitment will continue with the IMKT.

### 3 Questions and answers—comments from the floor

**Q—Marie Farge, ENS France** There's a lot of digitally closed material out there: will you re-digitise?

**A—SW** Though there is a good deal already in public digital form, the “copyright aggregators” are indeed an issue. Springer, for instance, does claim to be in favour of Open Access but has a commercial basis. In practice many institutions too seem wedded to the current system. Mathematics has the slight advantage that it's not as profitable as some of the other subjects, e.g., Medicine. But the area of relationship with commercial publishing is very much “work in progress”.

**Q—Gerhard Paseman, USA** There's a lot of work being done in the area of Open Access, e.g. Gowers' work. This wasn't really mentioned. Also, what about the really useful mathematical work that is posted in forums.

**A—PI** We're certainly not unaware of these issues, but we need to prioritise our work. A full discussion of both matters touched on would take a long time. But these are matters that the GDML WG has discussed, and the IMKT has to take into account.

**Q—Marie Farge, ENS France** For one article, we were ready to pay the APC (Article Processing Charge), but the publisher still insisted on retaining copyright. This is a problem for many of us.

**A—SW** IMKT is focused on technical issues, though there are legal ones. We have to deal with the current copyright aggregators, and being a copyright advocate is not IMKT's business.

**A—AO** There are ongoing discussions at IMU about this.

**A—PI** Yes, it's the IMU's CEIC that is trying to revise an earlier report on copyright issues for mathematicians, in the light of changing conditions.

**Q—Unknown** There are a lot of good lawyers working on copyright issues. But I note that the ICM 2018 copyright agreement, which speakers had to sign, is pretty draconian.



**A–PI** Clearly this should be noted, since the IMU and its CEIC will have to reach consistency.

**Q–James Davenport, UK** AO’s presentation had mentioned the importance of arXiv, especially in developing countries. But a student and I recently noted a case where the paper on arXiv was very different (and silently so) from the published paper. Isn’t this a problem?

**A–SW** But we should note that paper corrigenda are not perfect either.

**James Davenport, UK** But that’s a fault of technology: this is human weakness.

**Marie Farge, France** We need a good practice statement from the IMU.