TEX and Global Mathematics

Patrick D. F. Ion

GDML WG, IMKT & University of Michigan, MI USA, pion@umich.edu
(Mathematical Reviews & MathSciNet, AMS ret'd)

26 July 2020 / TUG 2020 – Zoom 15:00 EST

Abstract

TEX was developed as a way of communicating mathematics; it has been very successful for that and much more. But TEX did not completely dominate publishing, though it much expanded the community able to write mathematics directly. MathML (Mathematics Markup Language) was specified as a markup for mathematics in the W3C (World Wide Web Consortium) context; it is both officially part of the web's basic HTML and an ISO standard. The idea that there should be a Global Digital Mathematics Library (GDML) is an obvious one. There's an International Mathematical Knowledge Trust (IMKT) devoted to eventually realizing a GDML, growing out of efforts by the International Mathematical Union. Some of how the present situation came to be and what's evolving now will be examined.

Outline

- Mathematics
- TEX
- MathML
- GDML / IMKT

- Math is a natural language
- spoken by a globally distributed tribe (science is influential)
- jargon: the technical terminology ... of a special activity or group
- an artificial language to discuss natural patterns
- pasigraphy (ICM 1898)
- now 2-4K + years since inception (or 30-40K yr)

Math
Ishango closeup

- Developed by Don Knuth (with many followers)
- for communicating mathematics (writing formulas)
- became pervasive in scientific and multilingual publishing
- much expanded the community directly writing mathematics
- now 40 + years since inception

[Illustrated here from my experience.]

- developed by a working group within
- World Wide Web Consortium (W3C) from 1997
- W3C standard as 1.0 in 1998 (one of the first)
- part of official HTML standard since 2015
- MathML 3 an ISO standard since 2016
- its use is spreading; in part over accessibility issues
- fits XML publishing
- now 20 + years on

- Global Digital Mathematics Library
- International Mathematical Union (IMU) — 2006
- GDML WG — Seoul ICM 2014
- International Mathematical Knowledge Trust (IMKT) — Waterloo 2016
- 5 + years on
- what's happened

My experience

- Associate Editor, Mathematical Reviews, Ann Arbor, MI; 1980 from
  - University of Heidelberg, Germany, 1974–1980
  - RIMS, Kyoto, Japan, 1972–1974
  - Rijksuniversiteit Groningen, Netherlands, 1971–1972
  - Bedford College, University of London, 1970-71
- quantum stochastic processes (QSP) with coauthors in Nottingham and India
- QSP equations used lots of tensor product signs

My experience

- Associate Editor, Mathematical Reviews, Ann Arbor, MI; 1980 from
  - University of Heidelberg, Germany, 1974–1980
  - RIMS, Kyoto, Japan, 1972–1974
  - Rijksuniversiteit Groningen, Netherlands, 1971–1972
  - Bedford College, University of London, 1970-71
- quantum stochastic processes (QSP) with coauthors in Nottingham and India
- QSP equations used lots of tensor product signs

From (6.1) it is clear that, if $C = C_1 \oplus \ldots \oplus C_n$ is a direct sum, then

$$\Gamma(C) = \Gamma(C_1) \otimes \ldots \otimes \Gamma(C_n). \tag{6.5}$$

Let J be the inclusion injection from $\mathcal{h}$ into a larger space $\mathcal{h}'$. Then, for $A \in \mathbf{B}(\Gamma(\mathcal{h}'))$, $\Gamma(J)^* A \Gamma(J)$ is the vacuum conditional expectation of A given $\mathbf{B}(\Gamma(\mathcal{h}))$, that is, the unique operator in $\mathbf{B}(\Gamma(\mathcal{h}))$ such that, for arbitrary $\psi, \chi \in \Gamma(\mathcal{h})$,

$$\langle \psi, \Gamma(J)^* A \Gamma(J) \chi \rangle = \langle \psi \otimes \Omega(\mathcal{h}^\perp), A\chi \otimes \Omega(\mathcal{h}^\perp) \rangle,$$

where $\mathcal{h}^\perp$ is the orthogonal complement of $\mathcal{h}$ in $\mathcal{h}'$ and we have used the identifications

$$\mathcal{h}' = \mathcal{h} \oplus \mathcal{h}^\perp, \qquad \Gamma(\mathcal{h}') = \Gamma(\mathcal{h}) \otimes \Gamma(\mathcal{h}^\perp).$$

The Fock representation of the canonical commutation relations (RCCR) over $\mathcal{h}$ is the family of unitary Weyl operators $(W(f) : f \in \mathcal{h})$ in $\Gamma(\mathcal{h})$, whose actions on exponential vectors are

$$W(f)e(g) = \exp(-\tfrac{1}{4}\|f\|^2 + 2^{-1/2} i \langle f, g \rangle) e(g + 2^{-1/2} i f).$$

The map $f \mapsto W(f)$ is continuous in the norm topology of $\mathcal{h}$ and the strong operator topology of $\mathbf{B}(\Gamma(\mathcal{h}))$, the Weyl relation

$$W(f)W(g) = \exp\{-\tfrac{1}{2} i \operatorname{Im} \langle f, g \rangle\} W(f+g) \quad (f, g \in \mathcal{h}) \tag{6.6}$$

holds and the vacuum expectation functional is

$$F(f) = \langle \Omega, W(f)\Omega \rangle = \exp(-\tfrac{1}{4}\|f\|^2)$$

My experience

- Associate Editor, Mathematical Reviews, Ann Arbor, MI; 1980 from
  - University of Heidelberg, Germany, 1974–1980
  - RIMS, Kyoto, Japan, 1972–1974
  - Rijksuniversiteit Groningen, Netherlands, 1971–1972
  - Bedford College, University of London, 1970-71
- quantum stochastic processes (QSP) with coauthors in Nottingham and India
- QSP equations used lots of tensor product signs
- TEX at JMM, Jan 1981, presented by Don Knuth and Mike Spivak
- TEX could 'do' special symbols

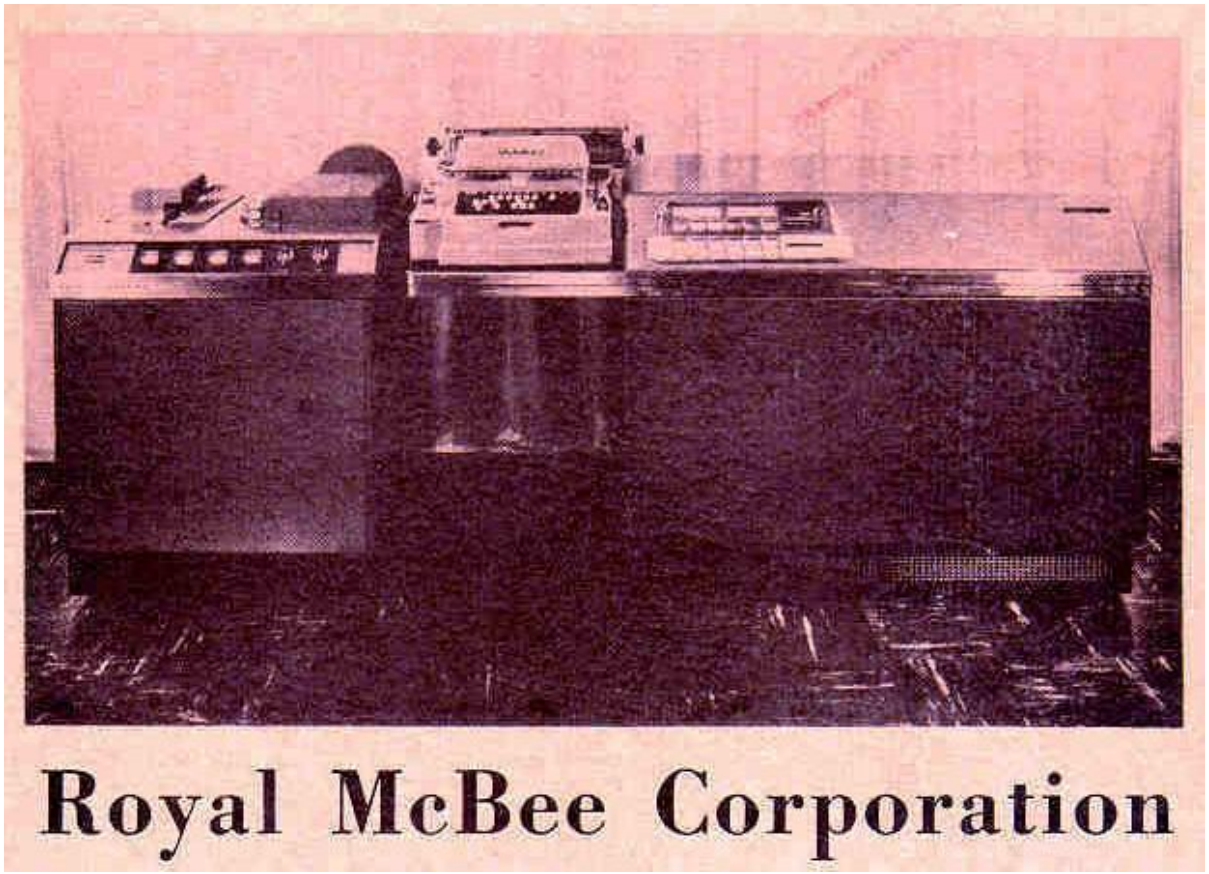
- AMS had TEX!
  - DEC 2020-60 (Tops 20) in Providence RI with 300 Bd phone access
  - proof pages off a Florida Data 24-pin printer
  - driver Monolithic box with 2 $\times$ Z80 boards; assembler
  - worked for me for 3 weeks and stalled
- Was tooling up in Heidelberg for numerics of Lorenz attractor (after 15 yr non-computing)
- Previous experience
  - Royal McBee LGP-30 (1960)
  - Burroughs Datatron 200 (1960)
  - Univac , IBM, ICL Manuals (ca. 1962)
  - Algol 60 and Atlas (1964)

My experience

- AMS had TEX!
  - DEC 2020-60 (Tops 20) in Providence RI with 300 Bd phone access
  - proof pages off a Florida Data 24-pin printer
  - driver Monolithic box with 2 $\times$ Z80 boards; assembler
  - worked for me for 3 weeks and stalled
- Was tooling up in Heidelberg for numerics of Lorenz attractor (after 15 yr non-computing)
- Previous experience
  - Royal McBee LGP-30 (1960)
  - Burroughs Datatron 200 (1960)
  - Univac , IBM, ICL Manuals (ca. 1962)
  - Algol 60 and Atlas (1964)

- AMS had TEX!
  - DEC 2020-60 (Tops 20) in Providence RI with 300 Bd phone access
  - proof pages off a Florida Data 24-pin printer
  - driver Monolithic box with 2 $\times$ Z80 boards; assembler
  - worked for me for 3 weeks and stalled
- Was tooling up in Heidelberg for numerics of Lorenz attractor (after 15 yr non-computing)
- Previous experience
  - Royal McBee LGP-30 (1960)
  - Burroughs Datatron 200 (1960)
  - Univac , IBM, ICL Manuals (ca. 1962)
  - Algol 60 and Atlas (1964)

Royal McBee LGP-30



My experience

- I found out
  - TEX was written in SAIL (Stanford Artificial Intelligence Language)
  - Stanford Artificial Intelligence Lab provided reports to U Michigan
  - SAIL Manual was one of them
  - SAIL was an extension of Algol 60
  - I was hooked on getting stuff working
- AMS was supporting TEX's development and encouraging internal use
- was learning to use computers; sent people to workshops at Stanford etc.
- Richard Palais' Dream

- I found out
  - TEX was written in SAIL (Stanford Artificial Intelligence Language)
  - Stanford Artificial Intelligence Lab provided reports to U Michigan
  - SAIL Manual was one of them
  - SAIL was an extension of Algol 60
  - I was hooked on getting stuff working
- AMS was supporting TEX's development and encouraging internal use
- was learning to use computers; sent people to workshops at Stanford etc.
- Richard Palais' Dream

- MR production was moving to TEX-based
  - many involved (Beeton, other TUG members)
  - did header for MR issue 1984-5 (hundreds of pages; special formatting rules)
  - Jan 1985 ‘hit newstands’ on time
  - AlphaType (a 5,000 dpi photo printer) in RI

TEX Timeline

- Don Knuth wants good corrections to "*Art of Computer Programming*" vol. 2
- 1978 May: First prototype; June 100 users; July 1,000 users
- 1983 public release
  interim: change from SAIL to extended Pascal; MetaFont
- Web literate programming; ethernet;
- 1985 Math Reviews and AMS Publishing are using TEX
- personal computing; PCTEX, ArborText;
- Paul Ginsparg's arXiv for preprints with much TEX
- 1994 at MSRI Electronic Communications of Mathematics
- . . . Jim Gosling announced Oak (Java)

MathML

- developed by a working group within
- World Wide Web Consortium (W3C) from 1997
- W3C standard as 1.0 in 1998 (one of the first)
- part of official HTML standard since 2015
- MathML 3 an ISO standard since 2016
- its use is spreading; in part over accessibility issues
- fits XML publishing
- now 20 + years on

MathML

- World Wide Web Consortium (W3C) [Berners-Lee]
- ca. 1995: a burgeoning web using HTTP and HTML
- Working groups emphasizing consensus
- Producing Recommendations
- No standards for markup of mathematical formulas
- Math WG 1997
- those who had ways of math markup (or were concerned)
  - IBM (Scratchpad — Axiom)
  - Mathematica [operator precedence parsing]
  - Maple
  - TEX
  - Elsevier
  - Microsoft
  - AMS
  - ...

MathML

- World Wide Web Consortium (W3C) [Berners-Lee]
- ca. 1995: a burgeoning web using HTTP and HTML
- Working groups emphasizing consensus
- Producing Recommendations
- No standards for markup of mathematical formulas
- Math WG 1997
- those who had ways of math markup (or were concerned)
  - IBM (Scratchpad — Axiom)
  - Mathematica [operator precedence parsing]
  - Maple
  - TEX
  - Elsevier
  - Microsoft
  - AMS
  - ...

MathML

- W3C standard as 1.0 in 1998 (one of the first)
- "*MathML: A Key to Math on the Web* 1999 TUGBoat, vol. 20, no. 3
- MathML part of official HTML standard since 2015
- Presentation MathML and Content MathML
- MathML 3: ISO standard ISO/IEC DIS 40314 since 2015
- WG continued to 2018: Co-chairs. Robert Miner + PI, Angel Diaz + PI, David Carlisle + PI
- Unicode (Murray Sargent) especially' v. 6
- XML Character Entity Names (David Carlisle) — goes back to TEX names

MathML

- MathML use is spreading; in part driven by accessibility issues
- fits XML publishing — was XHTML-based
- was and is being re-written to harmonize with newer web technology developed since early days (HTML5, CSS, SVG, ARIA, ECMAscript = Javascript) and to deprecate ones which didn't persist well (namespaces, XSLT, XML, . . .)
- browser manufacturer attention was a problem (e.g., 2014)
- now 20 + years on
- MathML Refresh Community Group, Chair: Neil Soiffer from 2019
- splitting off a MathML Core from MathML4 and considering additional markup options to carry semantics

GDML — Global Digital Mathematics Library

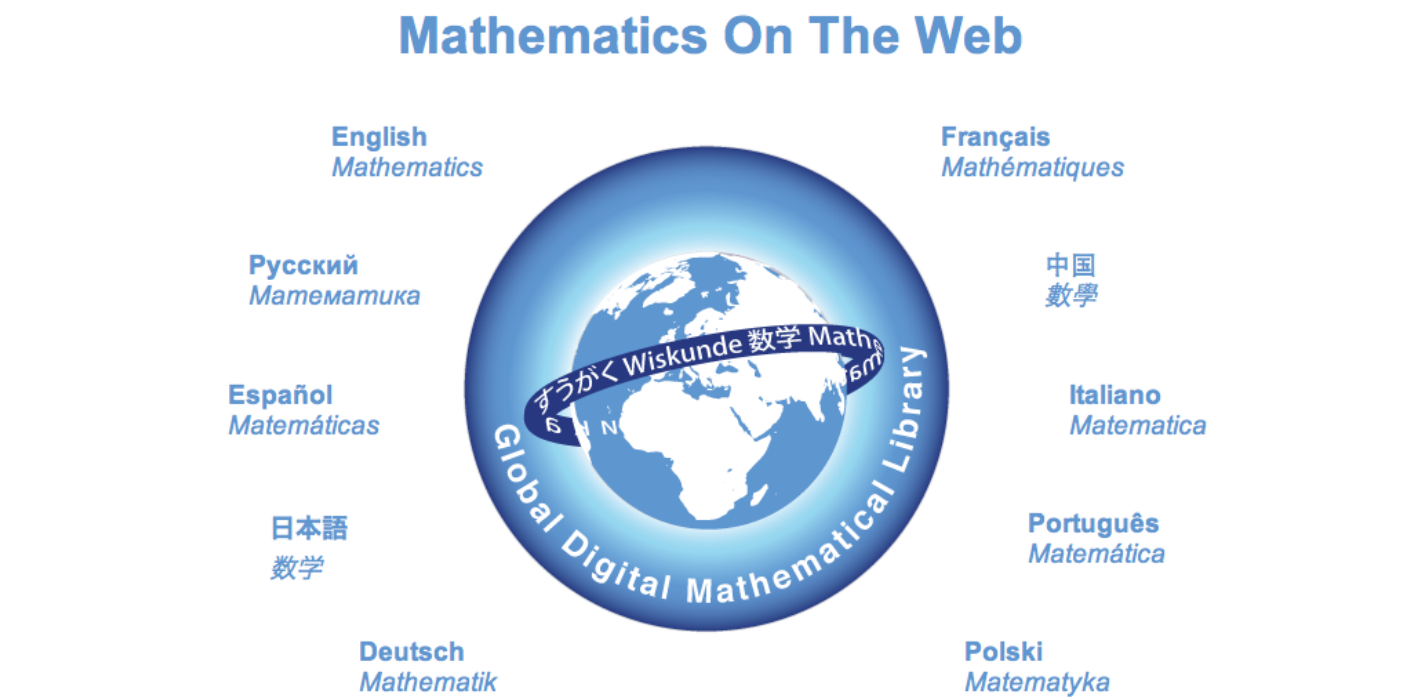
What is it?

- Global — for all the World, drawn from all the World
- Digital — using current technology
- Mathematics — our subject, especially research
- Library — a knowledge repository

[sometimes World Digital Mathematics Library (WDML)]

Worldwide Information System
for
Digitally Organized Mathematics
[WISDOM]

a web service maybe:

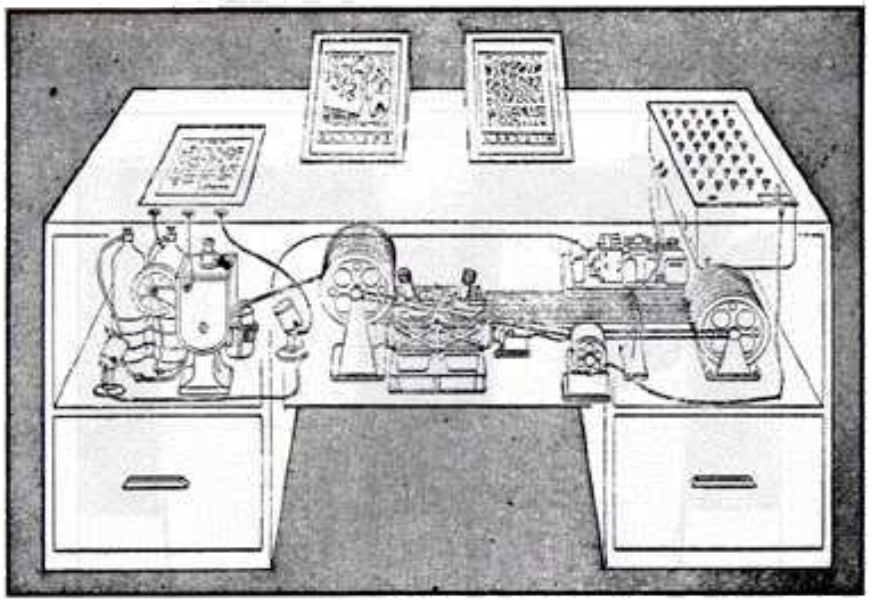

Mathematics On The Web
English
Mathematics
Français
Mathématiques
Русский
Математика
中国
數學
Español
Matemáticas
Italiano
Matematica
日本語
数学
Português
Matemática
Deutsch
Mathematik
Polski
Matematyka
Global Digital Mathematical Library
Mission Initiatives Projects News Portal
Comments to: gdml@mathontheweb.org
© Copyright 2014, Associated Ions
Privacy Statement

History
very potted

- Great Library of Alexandria in the Mouseion founded ca. 323 BCE by Ptolemy.
- Archimedes (287–212 BCE); Eratosthenes (276–195 BCE); Apollonius (262–190 BCE); Aristarchus of Samos (310–230 BCE); Hero (ca. 10 CE–70 CE); Hypatia, last director of the Mouseion lynched by a rabble in 415 CE
- Ca. 1200 years
- Leibniz and *Calculus Ratiocinator*
- Pasigraphy: E. Schröder, G. Peano at ICM 1897
- Georg Valentin's comprehensive bibliography to 1928
- Paul Otlet and Henri La Fontaine, about 1895: Mundaneum to ca. 1941
- Vannevar Bush imagined Memex in 1945 (Shannon)

Mundaneum
Telegraph room

Mundaneum
post WW II



Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p. 123).

- Late 1990's: initial vision
- 1998: WDML endorsed by the International Mathematical Union (IMU)
- 2001: IMU issues "Call to All Mathematicians to Make Publications Electronically Available"
- 2000's: large digitization projects [Google Books, Hathi Trust, national]
- 2006: IMU Report "Digital Mathematics Library: A Vision for the Future"
- 2006: IMU GA Resolution

- 2010: European Digital Mathematics Library (EuDML)
- 2010: Digital Public Library of America launches with support of Sloan Foundation
- 2011: Alfred P. Sloan Foundation funds WDML workshop at NAS November, 2012
- 2013: US NAS Digital Math Library Committee Report, [Daubechies, Lynch]
- 2013: “The Mathematical Sciences in 2025” US NAS

GDML

- 2014: Seoul ICM Meeting
- 2014: Creation of GDML WG
- at inception
  WG: Austria, Canada, France, Germany 2, USA 3
- 2015: WG of IMU Committee on Electronic Communication and Information

To construct, as a global public good, an open knowledge base encompassing the results of the world's mathematics through collaborations deploying both present and new technology, and to foster a supporting community.

- To enhance openness and accessibility of all mathematical knowledge world-wide, present, past and future.
- To serve research mathematics, education and the scientific and technological use of mathematics.
- To be a resource for developing tools to promote use and development of mathematics.
- To facilitate creation, dissemination and archiving of semantically annotated mathematical material.
- To encourage the collaborative development of services based on semantic annotation.

The GDML tries to achieve its goals by building collaborations. The effort involves the creation of standards and indications of best practices, encouraging the instantiation of such standards with content, and making such content openly available.

- Organization, Governance & Community
- Corpus & Collection
- Tools & Services
- Knowledge Management

Issues
Organization, Governance & Community

- *Chicken and egg*: International Mathematical Union, WG
- International — legal, communication: examples
  - HathiTrust, DPLA, JSTOR, COS, ...
- Mathematics as a Universal Language: math community

- Boundaries
  - Advanced Research Mathematics (mostly)
  - Applied Mathematics (the theoretical)
  - Any natural language (mostly English presently)
  - MR $\cup$ ZM $\cup$ swMATH $\cup$ MathDataHub ?
  - Legacy material vs. broad present

- Ownership
  - Publishing is a business
  - Mathematics is a branch of knowledge, that is a fact collection
  - Mathematical facts are not patentable
  - Much publication metadata is public
  - Collections of such are not intrinsically held to be public
- Competition to be replaced by collaboration

- Materials:
  - EuDML, arXiv, IMU proceedings, open legacy material
  - Euclid?, HathiTrust?, JSTOR?, CALIS?
- Cataloging: metadata standards; EuDML
  - zbMATH, MathSciNet?
  - EuDML, Beebe, DRM, other public aggregations
- Authority, Trust, Provenance: current standards?
- Crowd sourcing: current groupings; current technology, Mendeley, Bibsonomy

- Multilingual: Unicode
- Formulas: MathML (W3C renewal), OpenMath, TEX / LATEX, OverLeaf,. . .
- Multiform: XML for description, or whatever's needed
- Listings; Annotation: lack of full support; W3C Annotation?
- Data-mining: LDA; NLP+: MathWordNet
- Corpus structure: graph analysis & visualization;
  — simplicial complex homology; persistent homology

- Classification: MSC in SKOS (Linked Open Data)
  - MSC 2020 revision [Master is a TEX file still]
- Ontology
- Issues of proof
  - Computer Assisted
  - Four Color, Kepler-Hales, Odd Order
  - JVM and chip verification
  - coloring Pythagorean triples; Sudoku

- Semantic Intermediate Abstraction Language
  - between basic markup and formalization
  - flexiformality
  - Part of Math tagging
  - semantic search, . . .
  - constrained natural language
- Previous attempts
  - Automath (de Bruijn 1960), . . .
  - Maple, Mathematica

- JMM Special Session on Mathematical Information in the Digital Age of Science, Seattle Jan 9-11 2016
- Semantic Representation of Mathematical Knowledge Workshop, Fields Institute February 3–5 2016, with Wolfram Research as Sloan grant recipient
- Applied for and received Sloan grant to found an International Mathematical Knowledge Trust (IMKT)

- Legal Foundation of IMKT based in Waterloo ON, Canada
  - Boards: Governing and Scientific Advisory
  - Work groups
- Short term: Outreach, seed projects, coordination
- Long term: Make available the “totality” of mathematical knowledge in digital form employing human- and machine-usable knowledge tools
- Initiatives
  - Special Function Concordance [Semantic TEX macros — Bruce Miller]
  - FABstracts [TEX is almost a basic Controlled Natural Language]
  - FHarmony
  - Document analysis: *n*-gram studies

- JMM Special Session on Mathematical Information in the Digital Age of Science, San Diego, Jan 9–11 2018
- ICMS 2018, 25–29 July 2018, Tom Hales presents FABstracts
- ICM 2018, 1–9 August 2018, Panel on Digital Libraries: Canada, China, Colombia, France, India, US represented

- Mathematical Research Data Initiative (MaRDI)
  `https://wias-berlin.de/mardi/`
- FAIR Math

- Joint Mathematics Meetings 2020 @ Denver, CO
  Special Session 78 - Mathematical Information in the Digital Age of Science
  12 events
- Coronavirus pandemic

- zbMATH Open !!!!
- EMS Newsletter Issue June 2020, pp. 44Ð47; DOI: 10.4171/NEWS/116/12; Online: 2020-06-08
  The Transition of zbMATH Towards an Open Information Platform for Mathematics, Klaus Hulek and Olaf Teschke, `https://www.ems-ph.org/journals/show_pdf.php?issn=1027-488X&vol=6&iss=116&rank=12`
- Short term: Outreach, seed projects, coordination

International Mathematical Knowledge Trust

International Mathematical Knowledge Trust

The long-term goal of the International Mathematical Knowledge Trust (IMKT) is the creation of a comprehensive mathematical knowledge base, to be used by people and software systems world-wide.

The IMKT supports a variety of digitization and mechanization projects for mathematical data and knowledge. These efforts are coordinated with a number of commercial and not-for-profit partners, consistent with a commitment to Open Data.

Board Membership

Jennifer Chayes
Boston, MA, USA

Ingrid Daubechies
Durham, NC, USA

Patrick Ion
Ann Arbor, MI, USA

Ursula Martin
Oxford, UK

Scott Pagan
Toronto, ON, Canada

Bernard Saint-Donat
New York , NY, USA

Stephen M. Watt
Waterloo, ON, Canada

Glen Whitney
Cambridge, MA, USA

Friday January 17, 2020, 8:00 a.m.-11:00 a.m.
AMS Special Session on Mathematical Information in the Digital Age of Science, I

08:00 Ingrid Daubechies 1154-00-957 Towards a Global Digital Mathematics Library

08:30 Katya Bercic 1154-00-1073 Research data in mathematics: taking the high road

09:00 Mila Rünnwerth 1154-00-953 The Neverending Story of a Holistic Research Infrastructure for Mathematics

09:30 Bruce Miller 1154-33-1352 Writing Mathematics in the Digital Age.

10:00 Mitch Keller 1154-01-1017 The Mathematics Genealogy Project as a Dataset.

10:30 Public Discussion of GDML Issues

Saturday January 18, 2020, 8:00 a.m.-12:00 p.m.
AMS Special Session on Mathematical Information in the Digital Age of Science, II

08:00 John Harrison 1154-03-1280 Automated Reasoning: retrospective and current progress

09:00 Tom Hales 1154-00-1029 The Formalization of Mathematics and Controlled Natural Language.

09:30 Gilles Dowek 1154-03-825 Logipedia: towards a Wikipedia of formal proofs.

10:00 Richard Zanibbi & Anurag Agarwal1154-00-1116 Progress Report from the MathSeer project.

10:30 Stephen Watt 1154-68-1086 Progress in Mathematical Information and Knowledge Bases.

11:00–12:00 Panel Discussion with Jean-Pierre Bourguignon and others

- Machine Learning on math corpus – Lafferty-Blei; Zanibbi-Giles
- Classification – Mathematica
- Visualization – Mathematica (Whitney at Brown)
- MathML 4 – Core and Full
- MGP - API
- Mathematical Data - Bercic - WDS?
- Mathematical Software - swMATH?

Future

- Organization, Governance & Community
  - Community building, Asian, US and European Trust entities,
  - Web presence and Wiki on the initiatives
- Collection Development
  - Collaboration with EuDML, arXiv and Euclid
  - Collaboration with Wikipedia, WikiData
  - Contact with potential Asian partners
- Tools & Services
  - Mathematical Object Identifiers (MOI)
  - Proposal toward open access book identification
- Knowledge Management
  - Initiatives
  - Portal
  - Stacks Project? Lurie?
  - Machine Learning results: Lafferty & Blei; Zanibbi & Giles
  - Learning from WRI; blockchains
  - Wikipedia & WikiData

Future

- Outreach
  - ICM 2022
- Open zbMATH !?

Our Mission

International Mathematical Knowledge Trust

To construct, as a global public good, an open knowledge base encompassing the results of the world's mathematics through collaborations deploying both present and new technology, and to foster a supporting community.

- Resources
  - Funding very conventional: cf 2 $\times$ \$ 900K for ML on math
  - Business model is charity or academic: AI startups promise more
  - Foundations and UNO
- Awareness
- Getting involvement
  - Patreon gives artistic works
  - Kickstarter promises rewards
  - Academia provides reputation
- See Pitman's 2014 Papers on ODML

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.