The “NASA Effect” of a Global Digital Math Library

Stephen M. Watt

David R. Cheriton School of Computer Science

University of Waterloo

Big Proof 2019
International Centre for Mathematical Sciences
Edinburgh, 27-31 May 2019

UNIVERSITY OF WATERLOO
FACULTY OF MATHEMATICS

Outline

- Personal Background
- Viewpoints
- Digital Math Libraries
- Flexibility
- IMKT
- The “NASA Effect”

Personal Background

- **Computer Algebra**
  - As a user, creating Maple (1981), creating Axiom (1991)
  - Symbolic algorithms and software systems
    (infinite structures, symbolic exponents, approximate polynomials, …)
- **Programming Languages** for Symbolic Computation
  - Aldor (1994), dependent types λC, optimizing compiler, algebra library
- Mathematical **Knowledge Management**
  - TeX, MathML, OpenMath content conversion
  - *n*-gram analysis, document classification
- Mathematical **Handwriting Recognition**
  - Curve classification via LS, handwriting neatening
  - Computer-aided collaboration
- Academic administration
  - Waterloo has 8600 students taking math degrees

- *Not* math formalization, ATP, ITP

A Computer Algebra Viewpoint

- Programming, along with user errors.
- Keeping track of obligations, your responsibility.
- Division by zero, your fault.

- Primary job: Compute expressions for things we don’t know.
  Can verify by independent calculation, if necessary.

Computer Algebra Viewpoint

- Certain domains with solid theory treated with great care.
- Library built on polynomial algebra, linear algebra, differential algebra.

- In others, ignore theoretical impediments, e.g. simplification.
- User wants results, even if they are not universal or exhaustive.

THE JOURNAL OF SYMBOLIC LOGIC
Volume 33, Number 4, Dec. 1968

SOME UNDECIDABLE PROBLEMS INVOLVING ELEMENTARY FUNCTIONS OF A REAL VARIABLE

DANIEL RICHARDSON

Introduction. Let E be a set of expressions representing real, single valued, partially defined functions of one real variable. E^* will be the set of functions represented by expressions in E.

If A is an expression in E, $A(x)$ is the function denoted by A.

It is assumed that E^* contains the identity function and the rational numbers as constant functions and that E^* is closed under addition, subtraction, multiplication and composition. In every case it is also supposed that given A and B in E there is an effective procedure for finding expressions in E to represent

$$A(x) + B(x),$$
$$A(x) - B(x),$$
$$A(x) \cdot B(x),$$
$$A(B(x)).$$

$A(x) \equiv B(x)$ will mean that $A(x)$ and $B(x)$ are defined at the same points and equal wherever they are defined.

The identity problem for (E, E^*) is the problem of deciding, given A in E, whether $A(x) \equiv 0$.

The integration problem for (E, E^*) is the problem of deciding, given A in E, whether there is a function $f(x)$ in E^* so that $f'(x) \equiv A(x)$.

Properties in Axiom (c. 1990s)

```
Ring: Category ==  Join(AbelianGroup, Monoid) with
        Distributive(*, +)

Quaternion(R: Ring): Join(Ring, Module R) with
        quaternion: (R, R, R, R) -> %
== add
        Rep == Record(r: R, i: R, j: R, k: R)
        (a: %) + (b: %) ==  [a.r + b.r, a.i + b.i, a.j+b.j, a.k+b.k]::%
        -- Distributivity is not verified!
        (a: %) * (b: %) ==  [ a.r*b.r - a.i*b.i - a.j*b.j - a.k*b.k,     a.r*b.i + a.i*b.r + a.j*b.k - a.k*b.j,
                              a.r*b.j - a.i*b.k + a.j*b.r + a.k*b.i,   a.r*b.k + a.i*b.j - a.j*b.i + a.k*b.r]::%
        ...
```

Number of Users

Excel :: Maple :: Lean

O(10^9) :: O(10^7) :: O(10^3) ??

Document Point of View

- Metadata (Author, Title, Publication date, …)
- Subject classification, keywords
- References

- Metadata validation
- Document analysis systems
- Collection management

Formalized Mathematics Point of View

- Formal definitions and claims
- Machine-verified proofs, all cases covered
- Examples:
  Math libraries
  4 colour map theorem, Kepler conjecture, Feit-Thompson odd order

Mathematical Knowledge Management (MKM)

- Unification of these areas
- Conferences since 2001

- Advance state-of-art in each of these areas

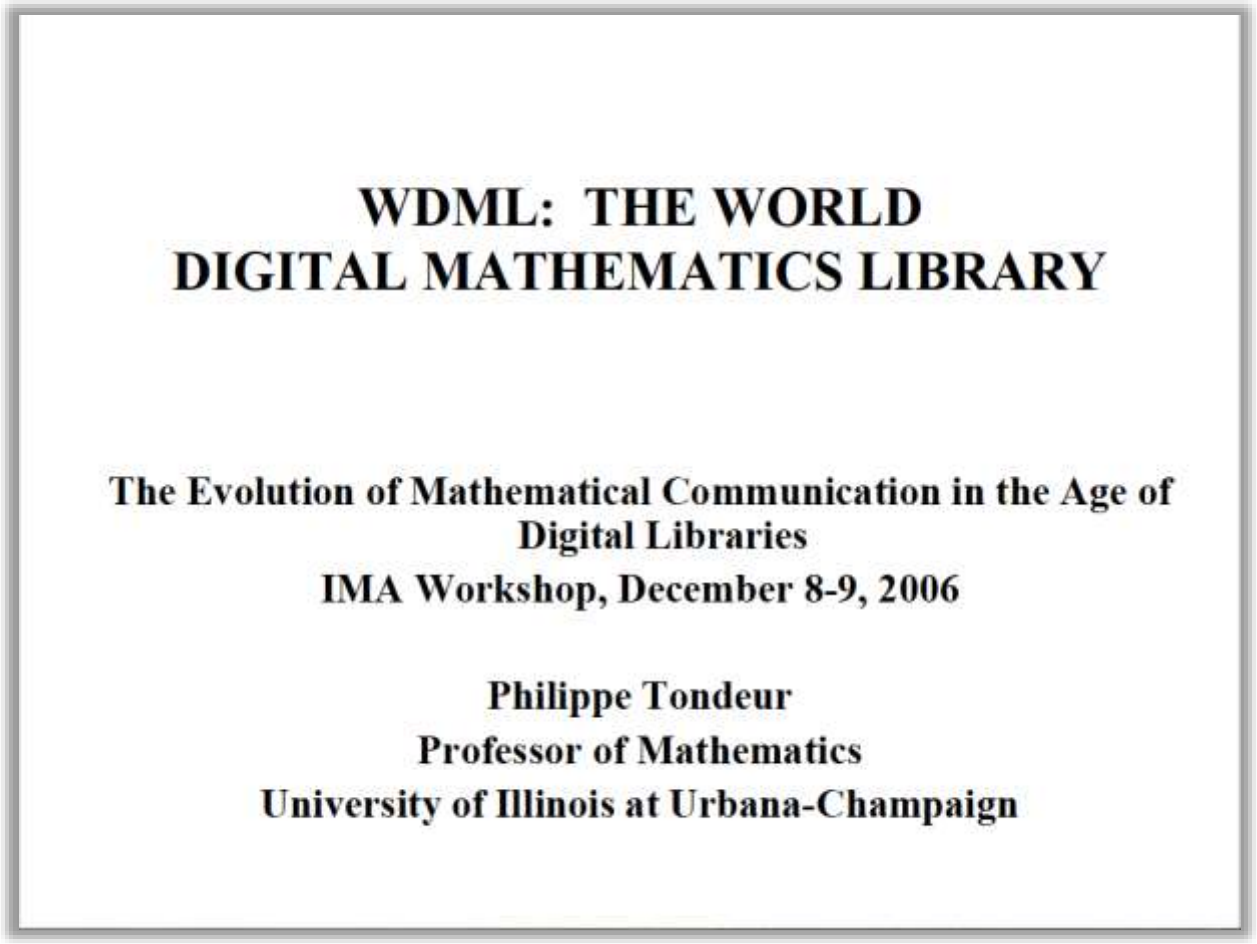
- E.g. from computer algebra point of view
  - Help systems -- what can I do here?
  - Evidence-based simplification rules

Digital Math Libraries – Early Concept

- Digitize all past mathematical literature and link it together with the present literature
- Initiators included:
  Bernd Wegner (ZentralblattMath),
  Keith Dennis (MathSciNet),
  Paul Ginsparg (arXiv), and others

15 Years Ago

Philippe Tondeur – 1999-2002 Director of Division of Mathematical Sciences
US National Science Foundation



WDML: THE WORLD DIGITAL MATHEMATICS LIBRARY

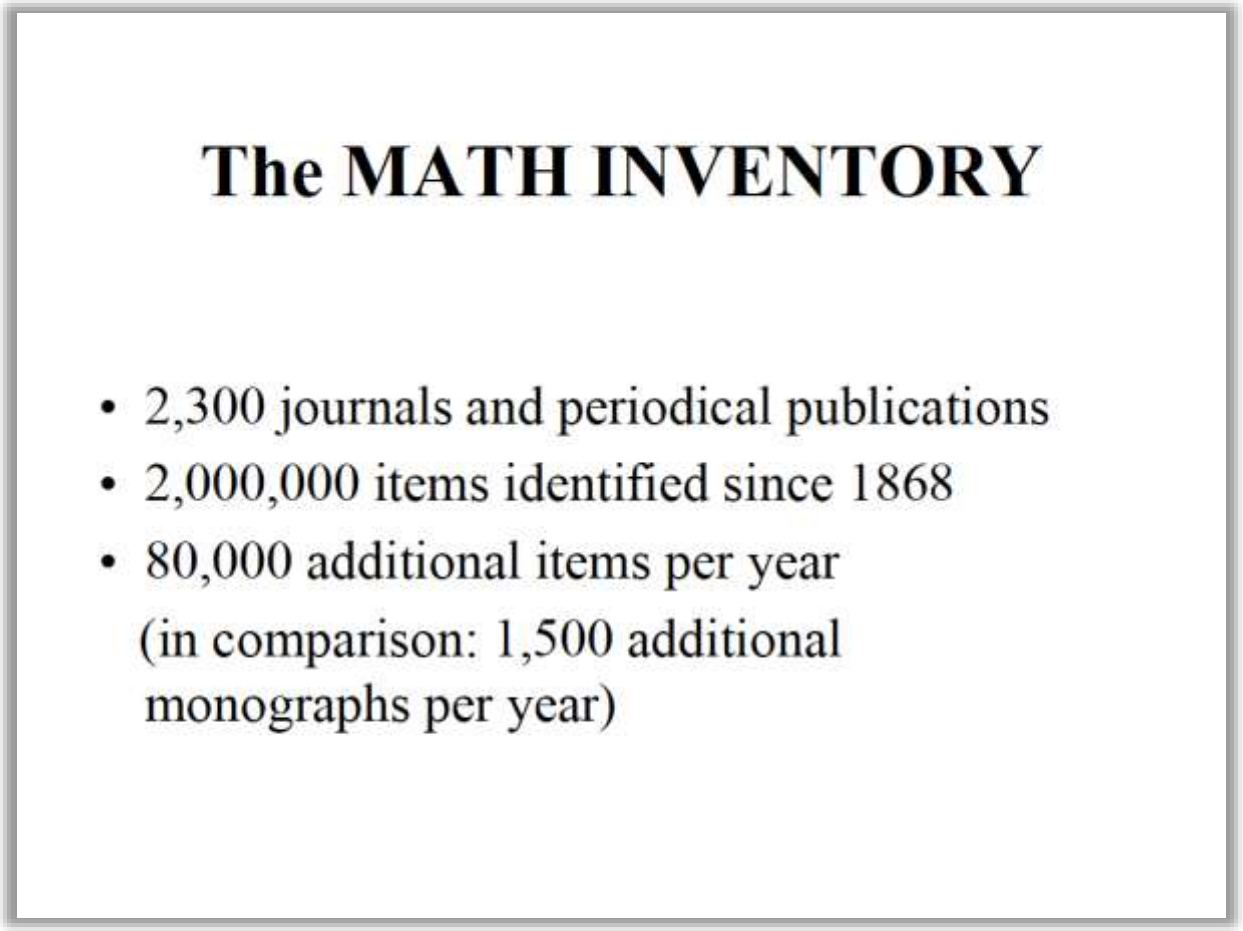
The Evolution of Mathematical Communication in the Age of Digital Libraries

IMA Workshop, December 8-9, 2006

Philippe Tondeur
Professor of Mathematics
University of Illinois at Urbana-Champaign

15 Years Ago

Philippe Tondeur – 1999-2002 Director of Division of Mathematical Sciences
US National Science Foundation



What the Future Was

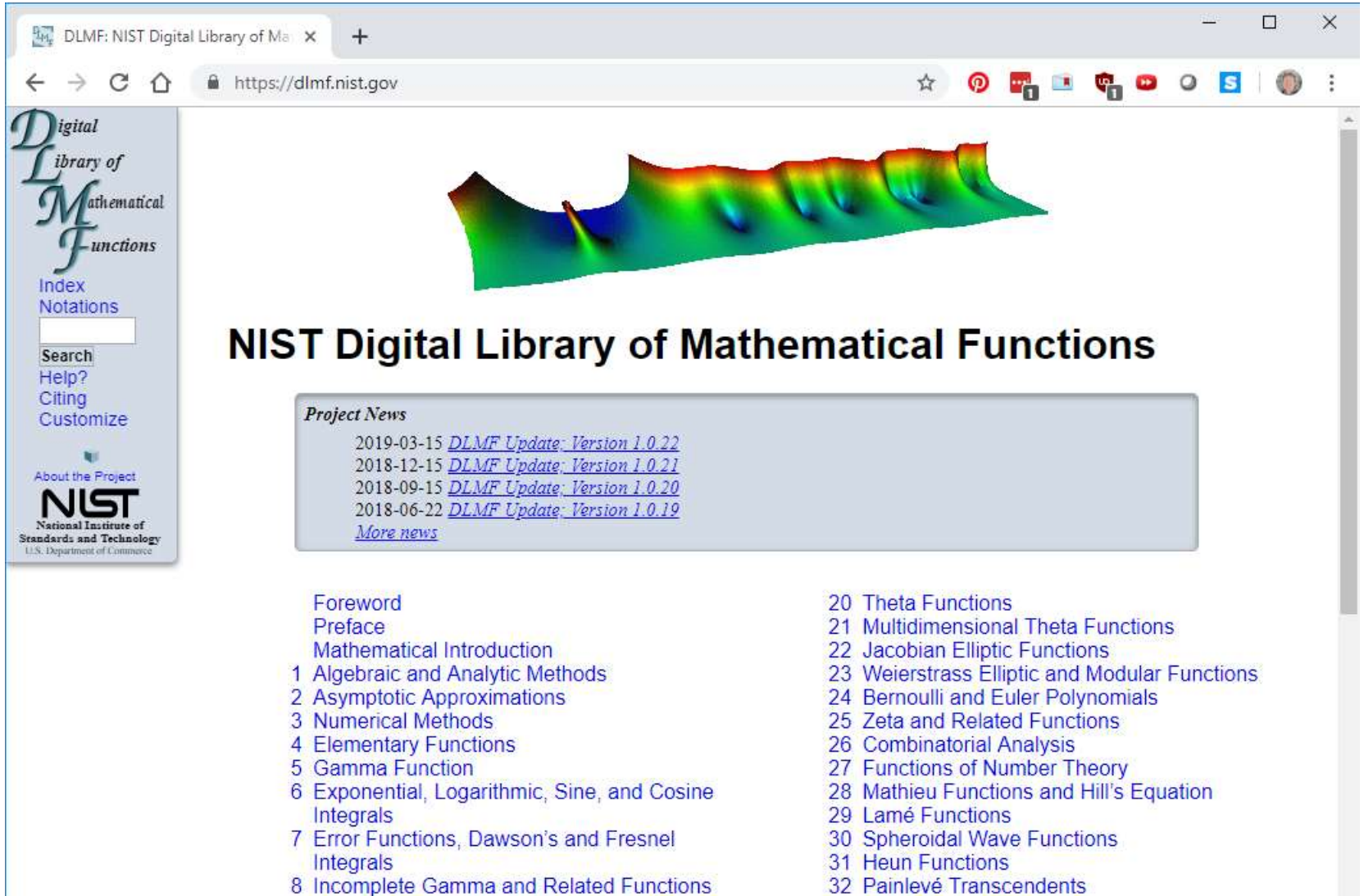
- A linked repository of articles is nice, but don't we have this already?

paleofuture.gizmodo.com

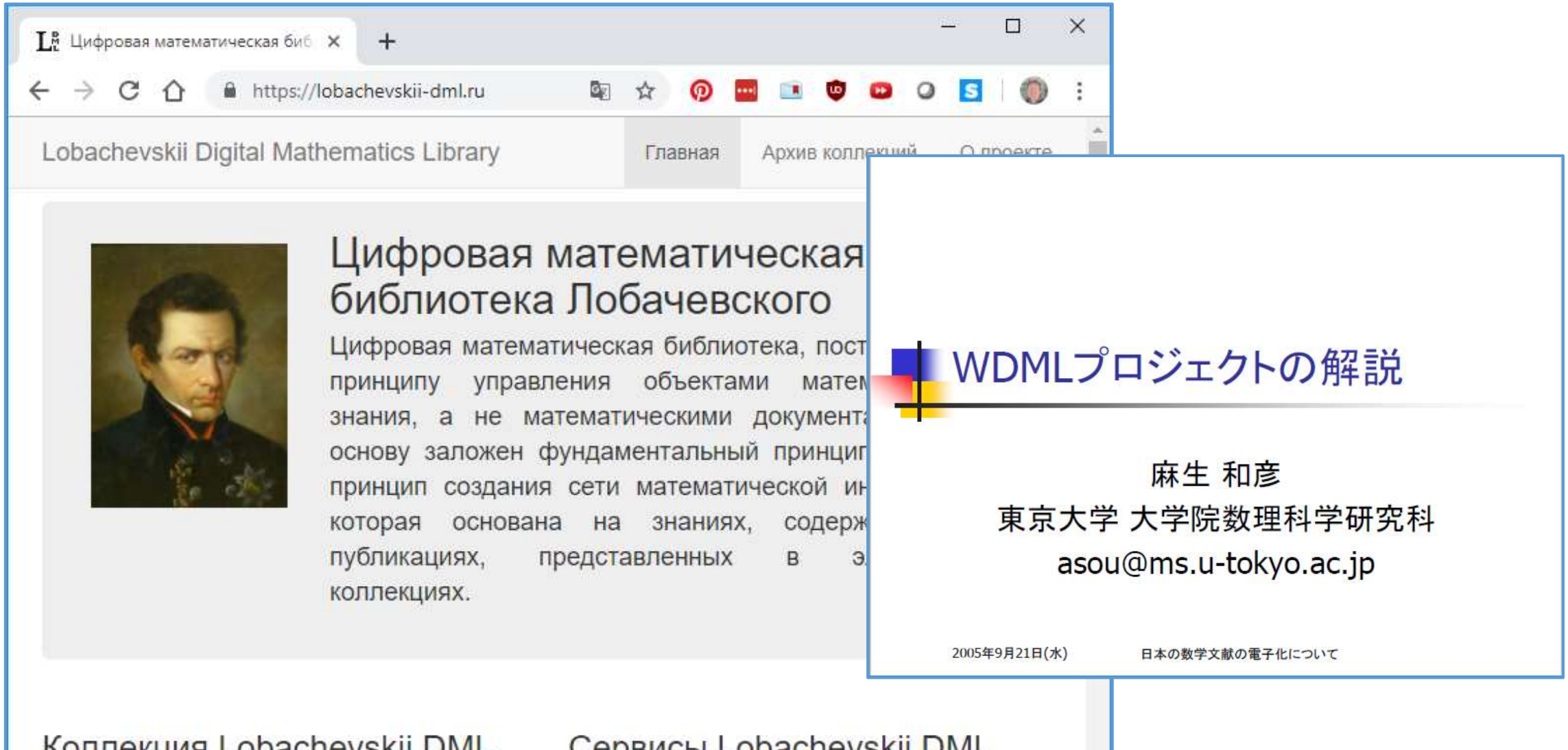
A Modern View

- Semantic capture
- Machine-accessible knowledge base
- Natural language queries
- Personalized learning
- Historical perspectives
- Conjecture generation

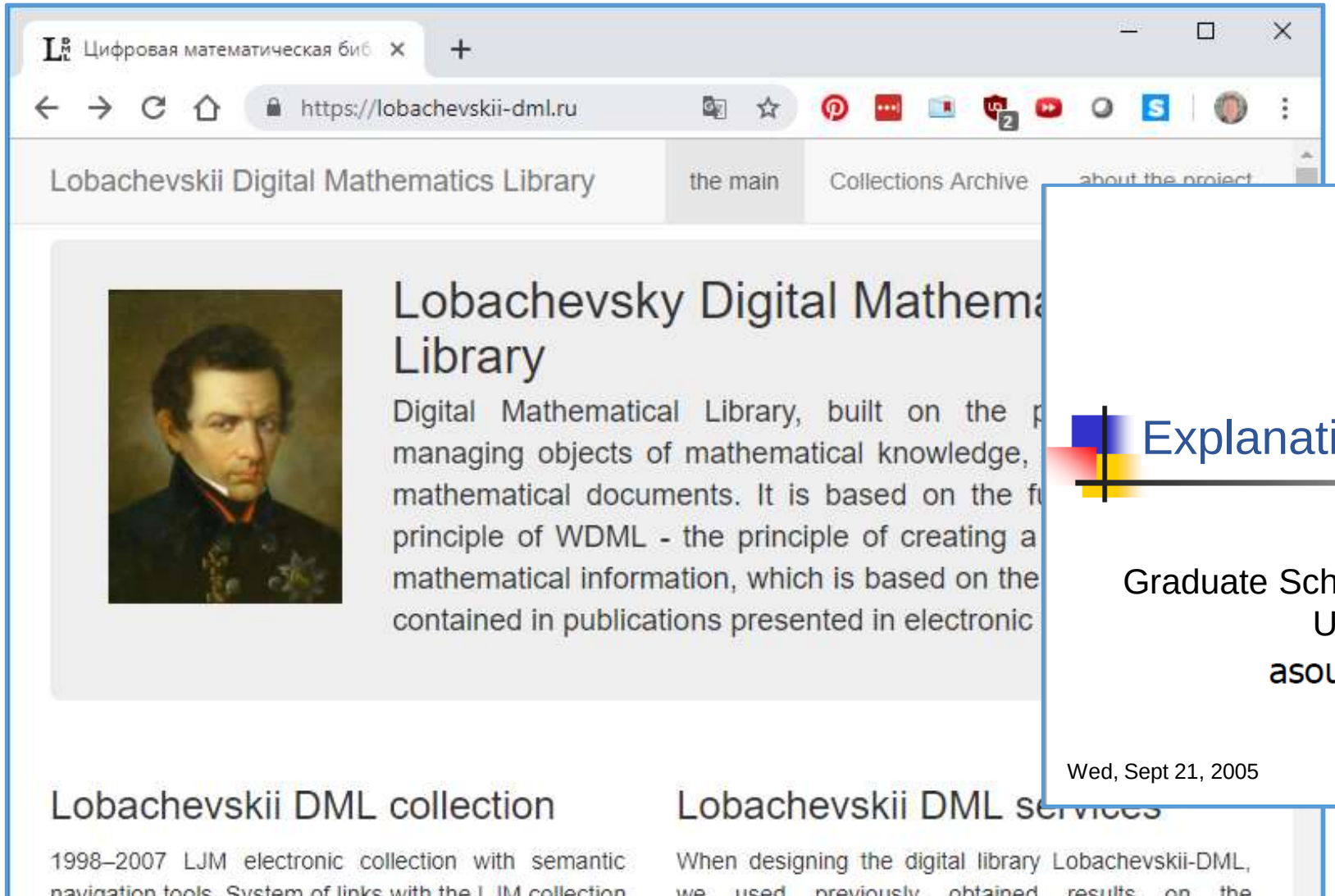
A Modern View



International



International



Explanation of the WDML Project

Kazuo Asou
Graduate School of Mathematical Sciences
University of Tokyo
asou@ms.u-tokyo.ac.jp

Wed, Sept 21, 2005

Computerization of Japanese math literature

Capturing Math

PDF page images

vs

Expression OCR

vs

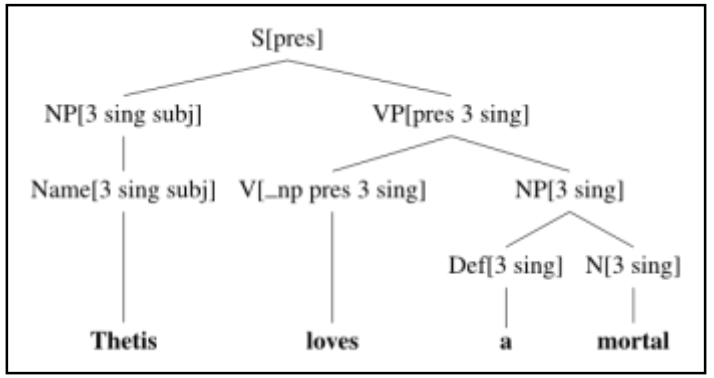
Natural language statements

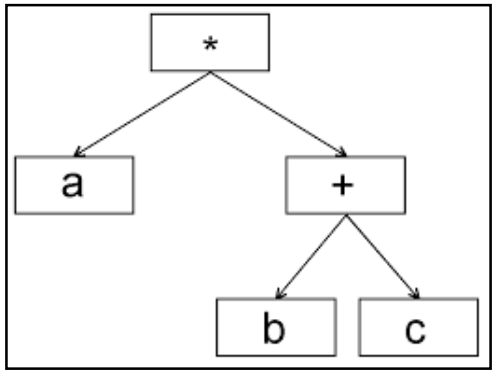
vs

Formalized mathematics

Capturing Math

- Page images *or* born-digital pages
- Structure analysis (DAS)
- Natural language and math recognition (e.g. FK, MG)
- Identification of glyphs.
- Parsing
- Semantic capture





Semantic Capture

- Extract the mathematical content of the paper.

- Can mean many things:
  - Statement of main definitions and main results.
  - Links to relevant literature.
  - Proof of results
    (at the level of formality in the paper or completely formal)
  - Discourse

Semantic Capture

- rich: allow all kind of mathematical statements
- flexible: allow human and computational annotation
- trust in the future: first instances can cost much more than subsequent
  - Increased automation, better sw + hw infrastructure
- future-proof: allow enhanced processing and annotation over time
- allow different ”levels” of annotation

Different Meanings for Basic Concepts

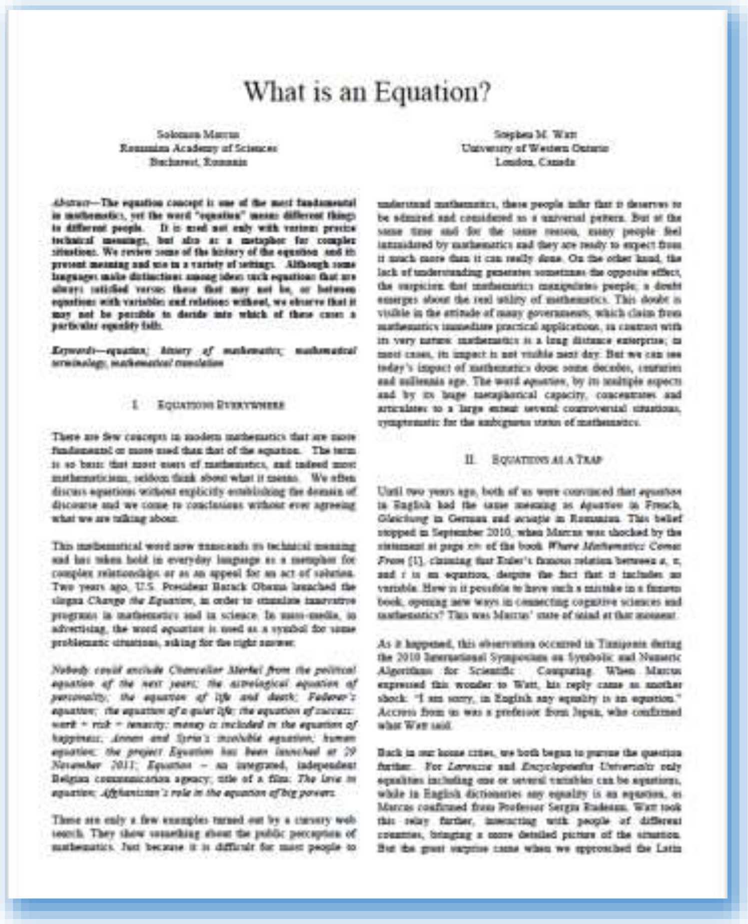
- 1391 Chaucer: ”Equation”
- 1557 Recorde: “=“
- 1637 Descartes: “x”

1. $x^4 - a = 0$
2. $xa = ax$
3. $8 + 2 = 10$

A. 1,2,3 “equation”
B. 1 “équation”, 2 “identité”, 3”rélation”

- English, Latin A
- Chinese, French, German, Russian,… B

- 1 vs 2 undecidable



What is an Equation?

Solomon Marcus
Romanian Academy of Sciences
Bucharest, Romania

Stephen M. Watt
University of Western Ontario
London, Canada

Abstract—The equation concept is one of the most fundamental in mathematics, yet the word "equation" means different things to different people. It is used not only with various precise technical meanings, but also as a metaphor for complex situations. We review some of the history of the equation and its present meaning and use in a variety of settings. Although some languages make distinctions among ideas such equations that are always satisfied versus those that may not be, or between equations with variables and relations without, we observe that it may not be possible to decide into which of these cases a particular equality falls.

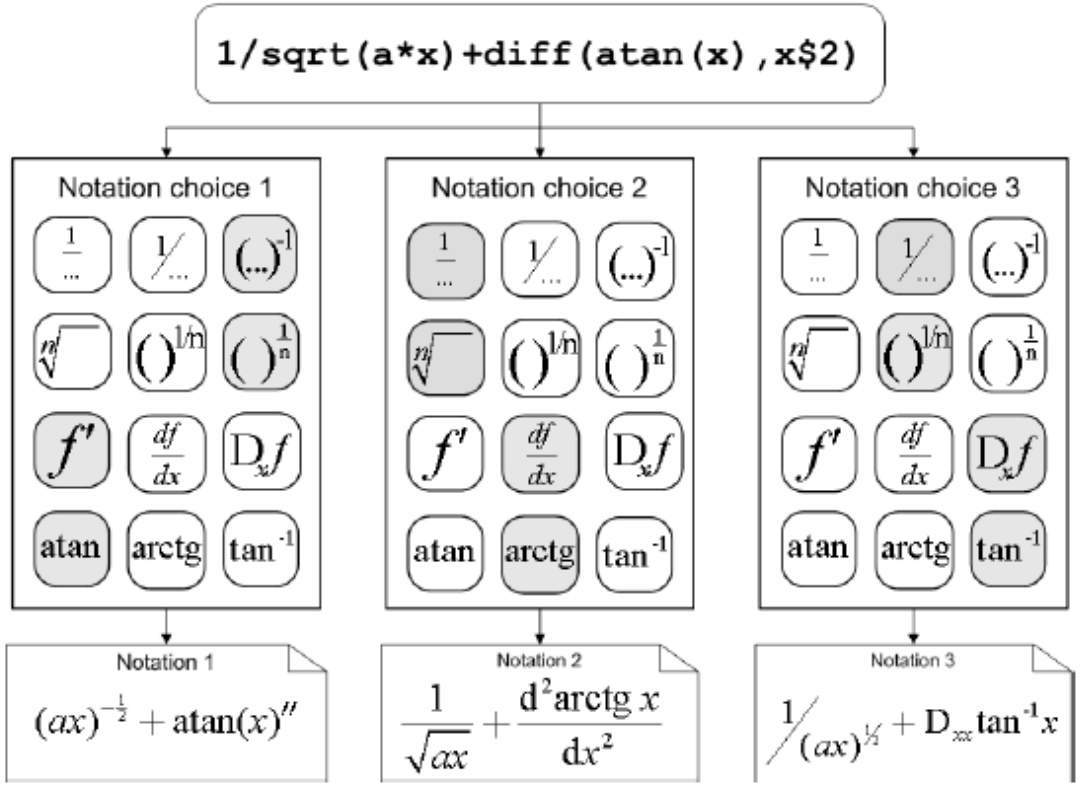
Keywords—equation; history of mathematics; mathematical terminology; mathematical translation

I. Equations Everywhere

II. Equations as a Trap

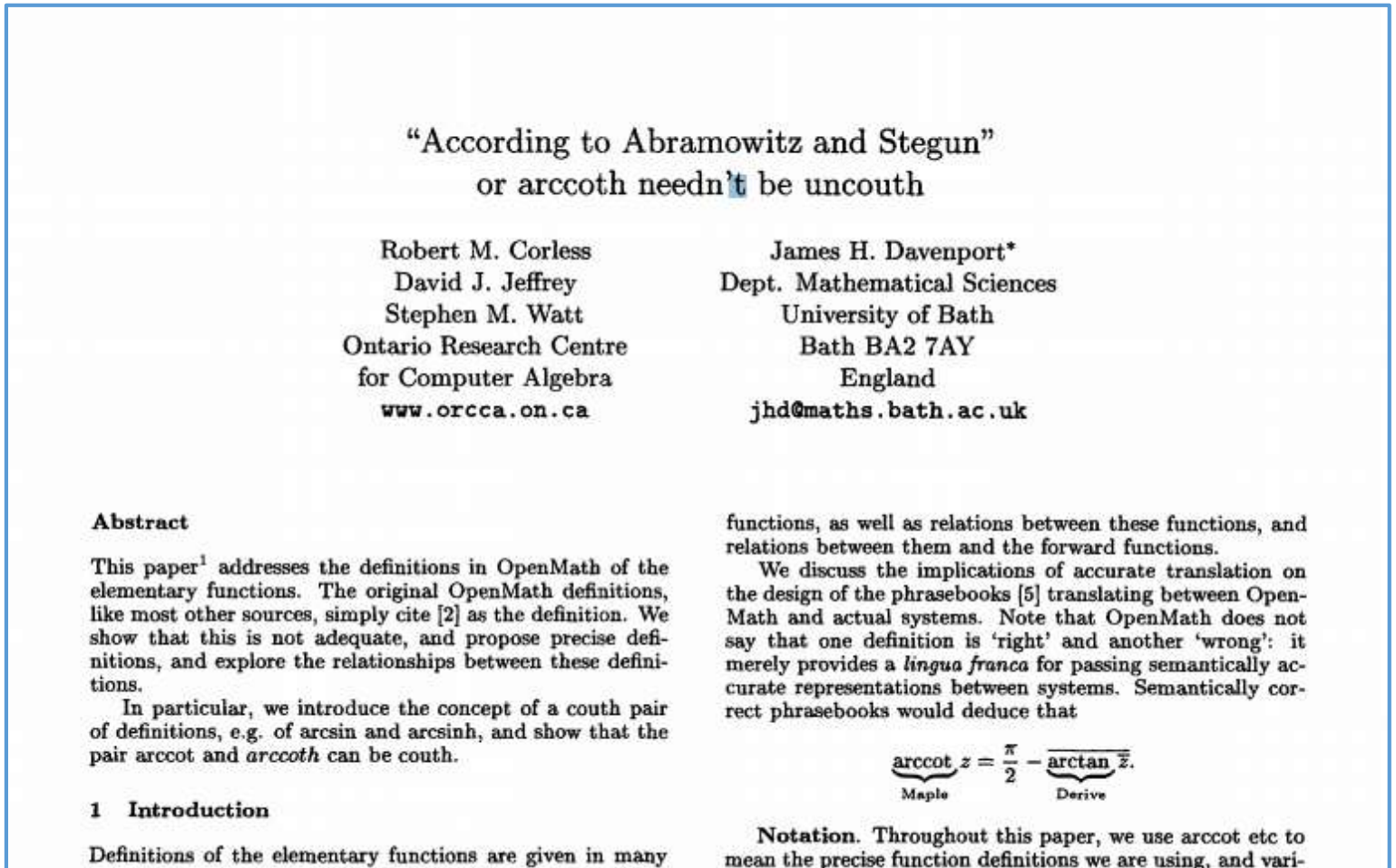
Different Notations for the Same Thing

- TeX ⇔ MathML, Presentation ⇔ Content



Different Things for the Same Notation

- J_ν
  - Bessel function
  - Angular momentum
  - 4-currrent



"According to Abramowitz and Stegun"
or arccoth needn't be uncouth

Robert M. Corless
David J. Jeffrey
Stephen M. Watt
Ontario Research Centre
for Computer Algebra
www.orcca.on.ca

James H. Davenport*
Dept. Mathematical Sciences
University of Bath
Bath BA2 7AY
England
jhd@maths.bath.ac.uk

Abstract

This paper[1] addresses the definitions in OpenMath of the elementary functions. The original OpenMath definitions, like most other sources, simply cite [2] as the definition. We show that this is not adequate, and propose precise definitions, and explore the relationships between these definitions.

In particular, we introduce the concept of a couth pair of definitions, e.g. of arcsin and arcsinh, and show that the pair arccot and *arccoth* can be couth.

1 Introduction

Definitions of the elementary functions are given in many

functions, as well as relations between these functions, and relations between them and the forward functions.

We discuss the implications of accurate translation on the design of the phrasebooks [5] translating between OpenMath and actual systems. Note that OpenMath does not say that one definition is 'right' and another 'wrong': it merely provides a *lingua franca* for passing semantically accurate representations between systems. Semantically correct phrasebooks would deduce that

$$\underbrace{\operatorname{arccot} z}_{\text{Maple}} = \frac{\pi}{2} - \underbrace{\overline{\arctan \overline{z}}}_{\text{Derive}}.$$

Notation. Throughout this paper, we use arccot etc to mean the precise function definitions we are using, and vari-

Informal vs Formal Mathematics and DMLs

Example: Derivation of the Dirac equation.

$$
\begin{aligned}
E^2 - p^2 &= m^2 \\
p^\mu p_\mu - m^2 &= (\beta^k p_k + m)(\gamma^\ell p_\ell - m) \\
&= \beta^k \gamma^\ell p_k p_\ell - m^2 + m\gamma^\ell p_\ell - m\beta^k p_k \\
&= 0 \qquad \Rightarrow \qquad \beta^k = \gamma^k, \quad \gamma^i\gamma^j + \gamma^j\gamma^i = 2g^{ij} \\
(i\gamma^\mu \partial_\mu - m)\psi &= 0
\end{aligned}
$$

Informal vs Formal Mathematics and DMLs

Example: Derivation of the Dirac equation.

$$
\begin{aligned}
E^2 - p^2 &= m^2 \\
p^\mu p_\mu - m^2 &= (\beta^k p_k + m)(\gamma^\ell p_\ell - m) \\
&= \beta^k \gamma^\ell p_k p_\ell - m^2 + m\gamma^\ell p_\ell - m\beta^k p_k \\
&= 0 \qquad \Rightarrow \qquad \beta^k = \gamma^k, \quad \gamma^i\gamma^j + \gamma^j\gamma^i = 2g^{ij} \\
(i\gamma^\mu \partial_\mu - m)\psi &= 0
\end{aligned}
$$

$$
\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ -\sigma^\mu & 0 \end{pmatrix} \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
$$

Different Levels of Formalism

- Exist in the literature
- ”Flexi-formalism” (Kohlhase)

- The de Bruijn factor (Freek): ~4

- “A little inaccuracy saves a world of explanation.”

www.loc.gov

Clovis on the Alleged Romance of Business

1924

"IT is the fashion nowadays," said Clovis, "to talk about the romance of Business. There isn't such a thing.

. . .

It was the unledgered wanderer, the careless-hearted seafarer, the aimless outcast, who opened up new trade routes, tapped new markets, brought home samples or cargoes of new edibles and unknown condiments. It was they who brought the glamour and romance to the threshold of business life, where it was promptly reduced to pounds, shillings, and pence; invoiced, double-entried, quoted, written-off, and so forth; most of those terms are probably wrong, but a little inaccuracy sometimes saves tons of explanation.

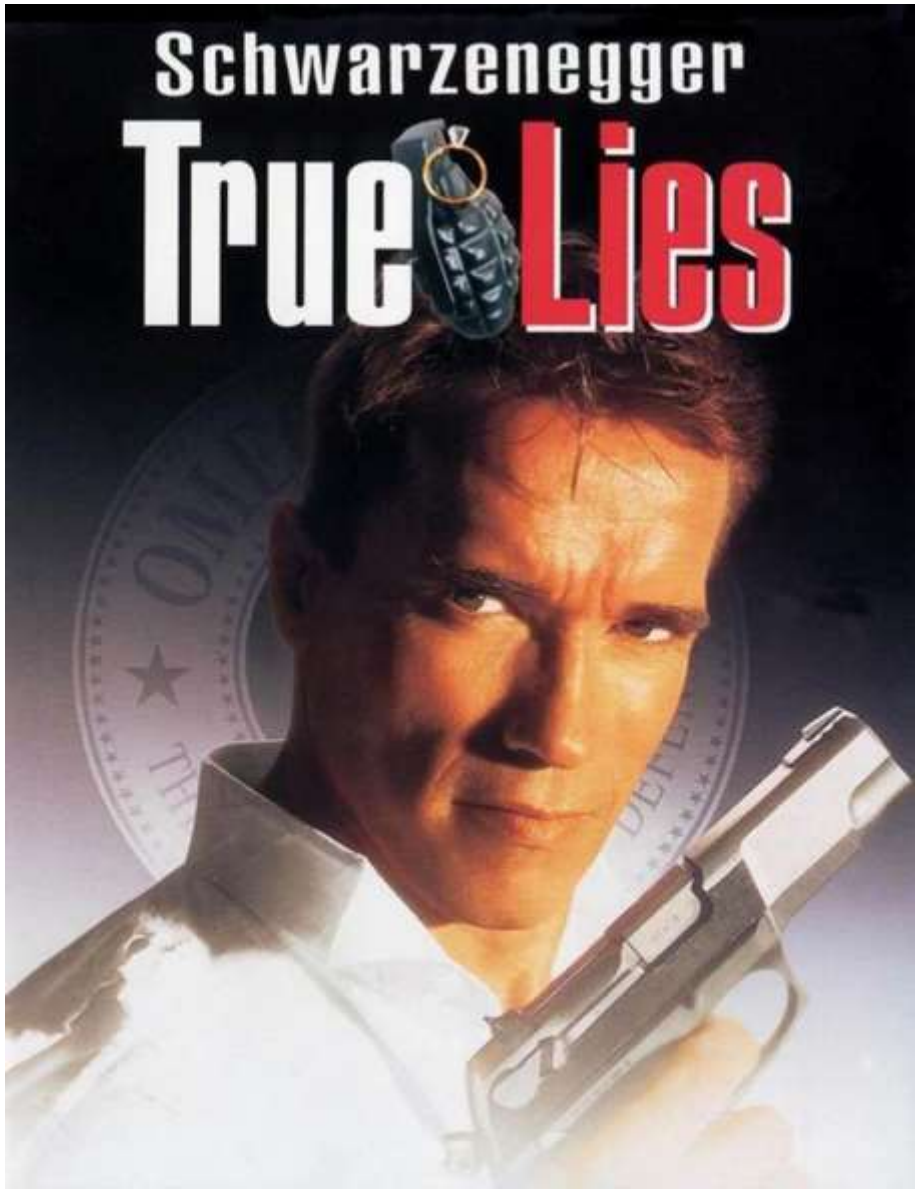
. . .

Whenever I feel in the least tempted to be business-like or methodical or even decently industrious I go to Kensal Green and look at the graves of those who died in business."

H. H. Munro ("Saki")
1870-1916
E. O Hoppé

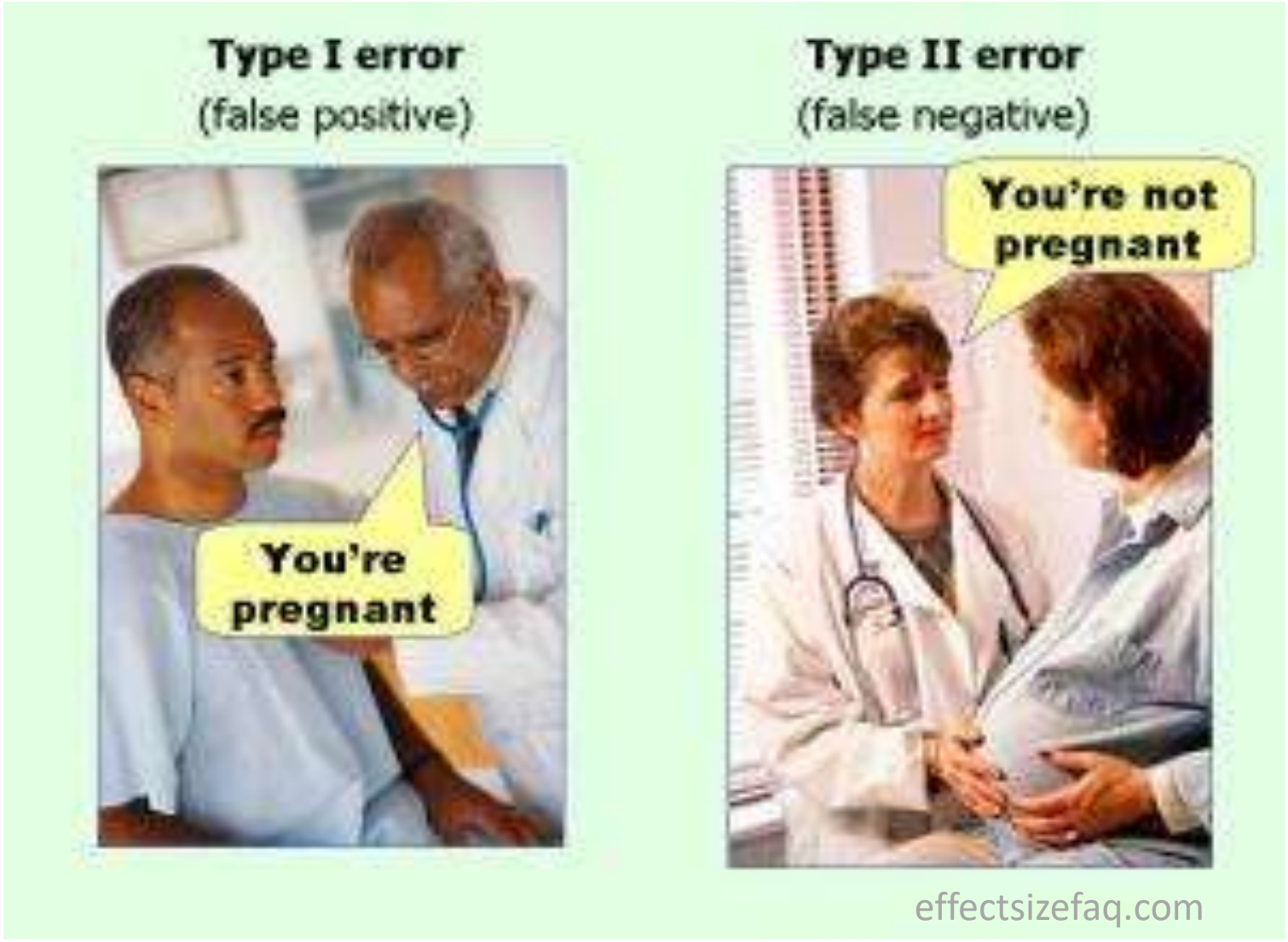
Discussion about Math

- True statements about false things.
- False statements about true things.



top2010oldiblog.com

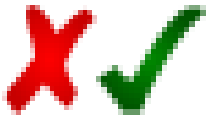
Errors


Type I error
(false positive)
You're pregnant
Type II error
(false negative)
You're not pregnant
effectsizefaq.com

Types of Errors (Math Overflow)

- ✓ Result is correct and was later rigorously proved.
- ✗✓ Result is wrong as stated, but a modified version was later rigorously proven.
- ? Status of the result is unclear
- ✗? Result is wrong as stated, but a modified version was suggested whose status is unclear.
- ✗ Result is wrong

Examples of Errors (Math Overflow)

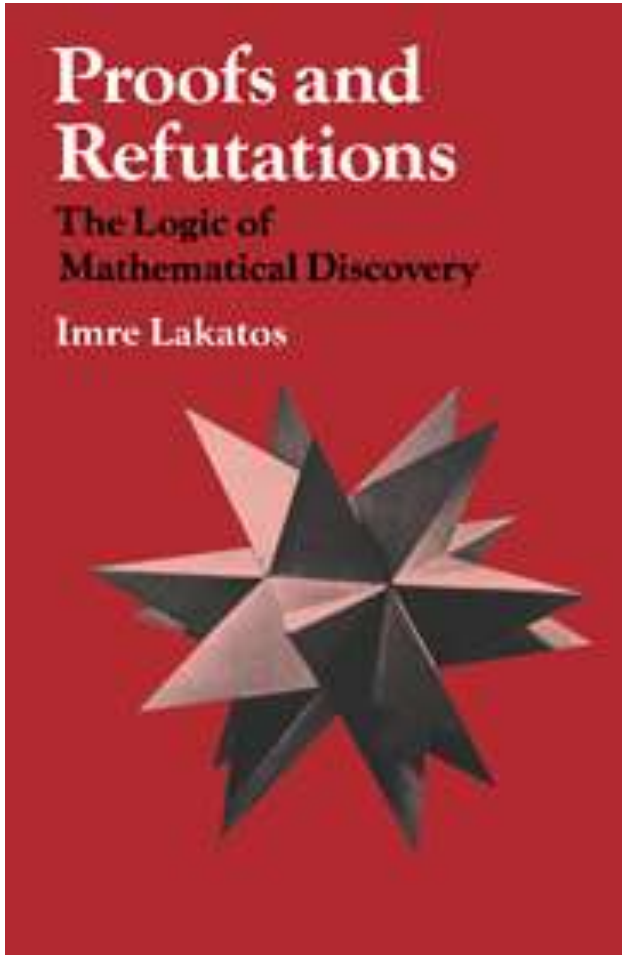


- Dirichlet's Principle
- Federov's class'n of convex polyhedra
- Mansion's Wronskians
- Plemelj & Hilbert's 21st problem
- Grunwald-Wang theorem

- Ampère differentiaility of continuous fns
- Frege Begriffsshrift consistency
- Severi finite dim space of rat equiv classes
- Vahlen's example of 3-d algeb curve

On Being Wrong

- Sometimes correct new mathematics is built on unjustified, but correct claims.
- Sometimes refutation of incorrect claims leads to important new mathematics.
- A DML must be able to represent and discuss incorrect claims.



amazon.com

Hybrid Libraries

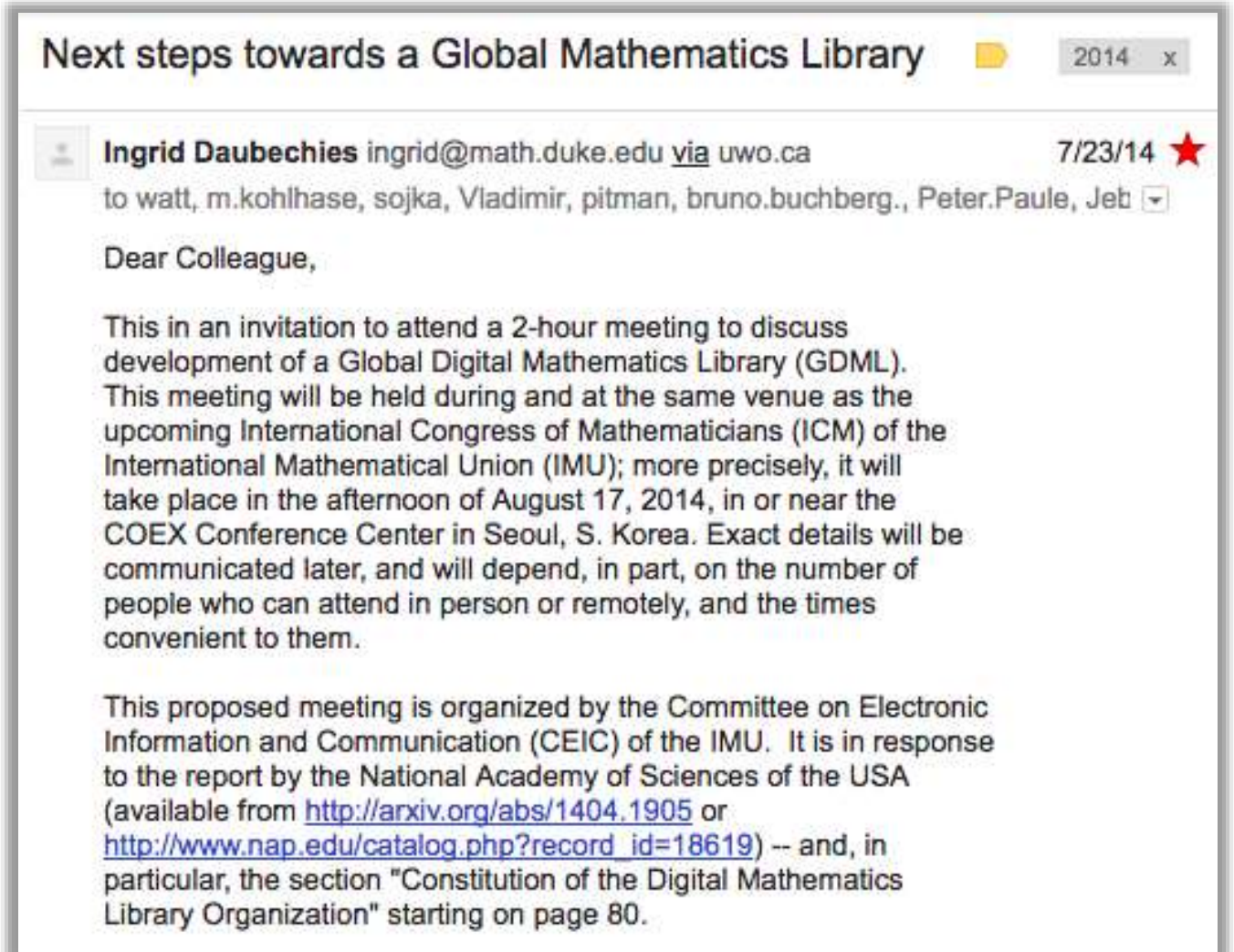
- Safe modules relying on unsafe code

Some capability is deemed unsafe, where the compiler can no longer guarantee the results will be consistent (for example, when interfacing to the C programming language). The keyword UNSAFE prefixed in front of INTERFACE or MODULE, may be used to tell the compiler to enable certain low level features of the language. For example, an unsafe operation is bypassing the type system using LOOPHOLE to copy the bits of an integer into a floating point REAL number.

An interface that imports an unsafe module must itself be unsafe. A safe interface may be exported by an unsafe implementation module. This is the typical usage when interfacing to external libraries, where two interfaces are built (one unsafe, the other safe).

5 Years Ago

Ingrid Daubechies – 2011-2014 President of the International Mathematical Union



Next steps towards a Global Mathematics Library 2014 x

Ingrid Daubechies ingrid@math.duke.edu via uwo.ca 7/23/14

to watt, m.kohlhase, sojka, Vladimir, pitman, bruno.buchberg., Peter.Paule, Jeb

Dear Colleague,

This in an invitation to attend a 2-hour meeting to discuss development of a Global Digital Mathematics Library (GDML). This meeting will be held during and at the same venue as the upcoming International Congress of Mathematicians (ICM) of the International Mathematical Union (IMU); more precisely, it will take place in the afternoon of August 17, 2014, in or near the COEX Conference Center in Seoul, S. Korea. Exact details will be communicated later, and will depend, in part, on the number of people who can attend in person or remotely, and the times convenient to them.

This proposed meeting is organized by the Committee on Electronic Information and Communication (CEIC) of the IMU. It is in response to the report by the National Academy of Sciences of the USA (available from http://arxiv.org/abs/1404.1905 or http://www.nap.edu/catalog.php?record_id=18619) -- and, in particular, the section "Constitution of the Digital Mathematics Library Organization" starting on page 80.

5 Years Ago

Ingrid Daubechies – 2011-2014 President of the International Mathematical Union

The meeting's goal is to assess interest of organizations and individuals in contributing to this effort and to discuss setting up a small organization that would

* coordinate efforts of several existing DML efforts, some started after the IMU's call for a World Digital Mathematics Library (WDML) several years ago;
* encourage the development of tools and a platform that would add functionality to these DMLs beyond the simple availability of documents, and that would be useful to mathematical researchers worldwide;
* help build a consortium of institutions to support the GDML.

5 Years Ago

- IMU Working Group established to
  - Develop a concrete roadmap
  - Incremental budget
  - Organize proposals to funders

- Members:

  Thierry Bouche (EuDML)
  Bruno Buchberger (Risc LINZ)
  Patrick Ion (Math Reviews (ret.), MathML)
  Michel Kohlhase (Jacobs, OpenMath)
  Jim Pitman (Berkeley)
  Olaf Teschke (ZbMath)
  Stephen Watt (ORCCA, MathML, doc analysis)
  Eric Weisstein (MathWorld, Wolfram Research)

WG Activities

- Workshops, sessions and panels
- Specific presentations and conference presentations
- Seed funding grant
- Founding not-for-profit
- Seed projects

International Mathematical Knowledge Trust

The International Mathematical Knowledge Trust (IMKT) has as its long-term goal 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.

Governance

The IMKT was first set up as a step on the path to a global digital mathematics library, as that may be understood in the present day, with a grant from the Alfred P. Sloan Foundation to principal investigators associated with the Global Digital Mathematics Library Working Group of the International Mathematical Union's Committee on Electronic Information and Communication under its 2015 remit.

MORE

Sponsors

UNIVERSITY OF WATERLOO

International Mathematical Union

Alfred P. Sloan FOUNDATION

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

Events

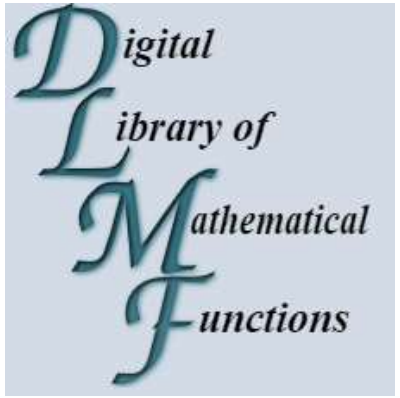
- Mathematical Information in the Digital Age of Science, JMM 2016
- Semantic Representation of Mathematical Knowledge, Fields Institute, 2016
- 10th Conference on Intelligent Computer Mathematics, 2017
- Applications of Computer Algebra, 2017
- Mathematical Information in the Digital Age of Science, JMM 2018
- Computer Aided Mathematical Proof, INI, 2017
- ICM, 2018
- Big Proof 2019

IMKT Initial Objectives

- F Abstracts
- Formal Harmony
- Special Function Concordance
- Mathematical Document Classification
  (N-Grams and Machine Learning)

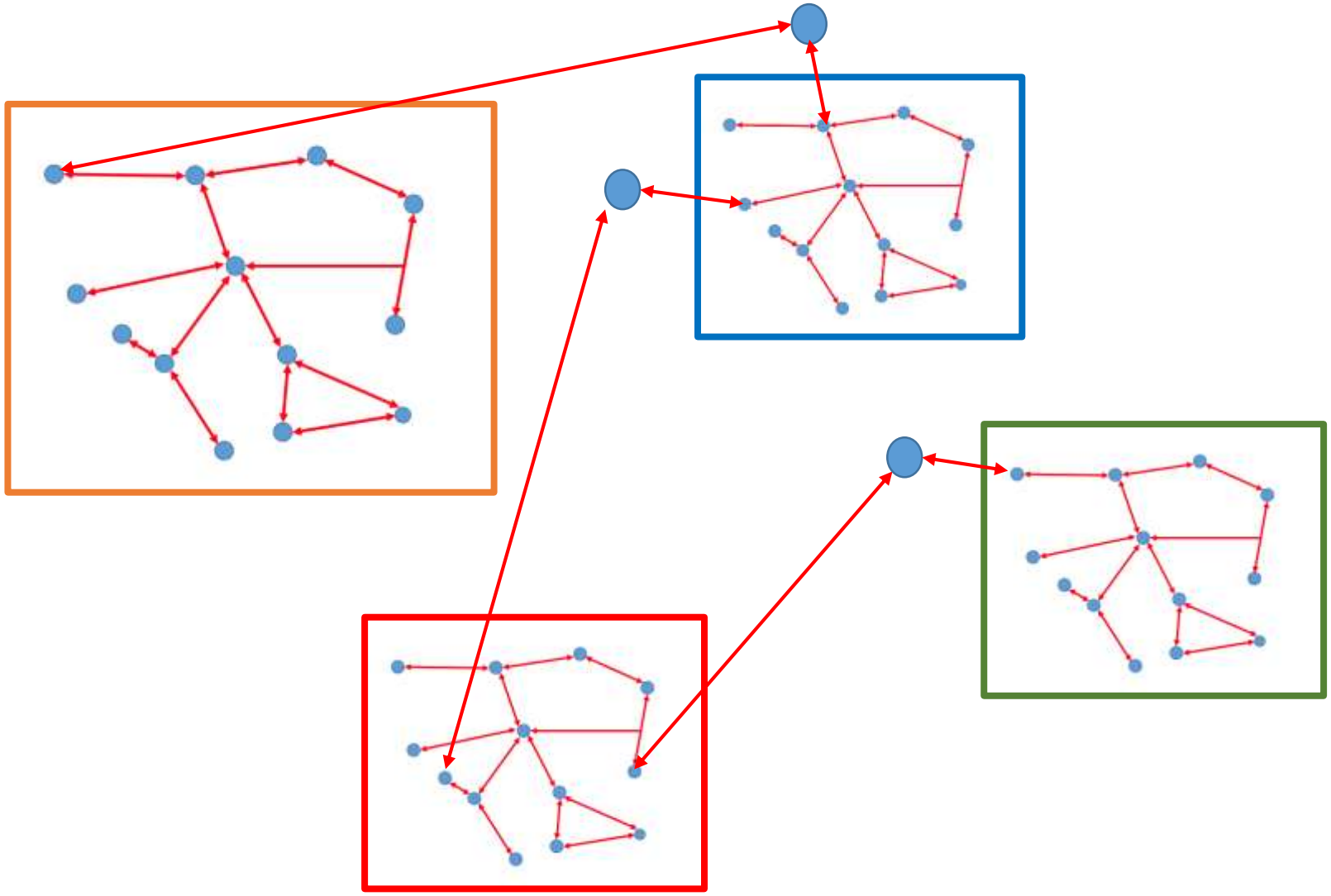
Tom Hales

Gilles Dowek



NIST

Several Groups' Work



Slightly Less Obvious Statements about Getting Something Done

- Can afford to hand-annotate/capture a small subset
- Rely on improvements of technology/AI
  - By Hand
  - Supervised automation
  - Automation + checking
  - Automation ☺?
- Approximation goes a long way
- Future-proof the data

The “NASA Effect”

- 1957 Sputnik
- 1958 NASA
- 1962 “We choose to go to the Moon”
- Pushback: “Costs too much”, “Money better spent”, …
- 1969 Apollo 11 Moon landing

- 1979: Majority of Americans felt Apollo program did not justify costs
- 1999: Majority felt it did

  https://news.gallup.com/poll/3712/landing-man-moon-publics-view.aspx

The “NASA Effect”

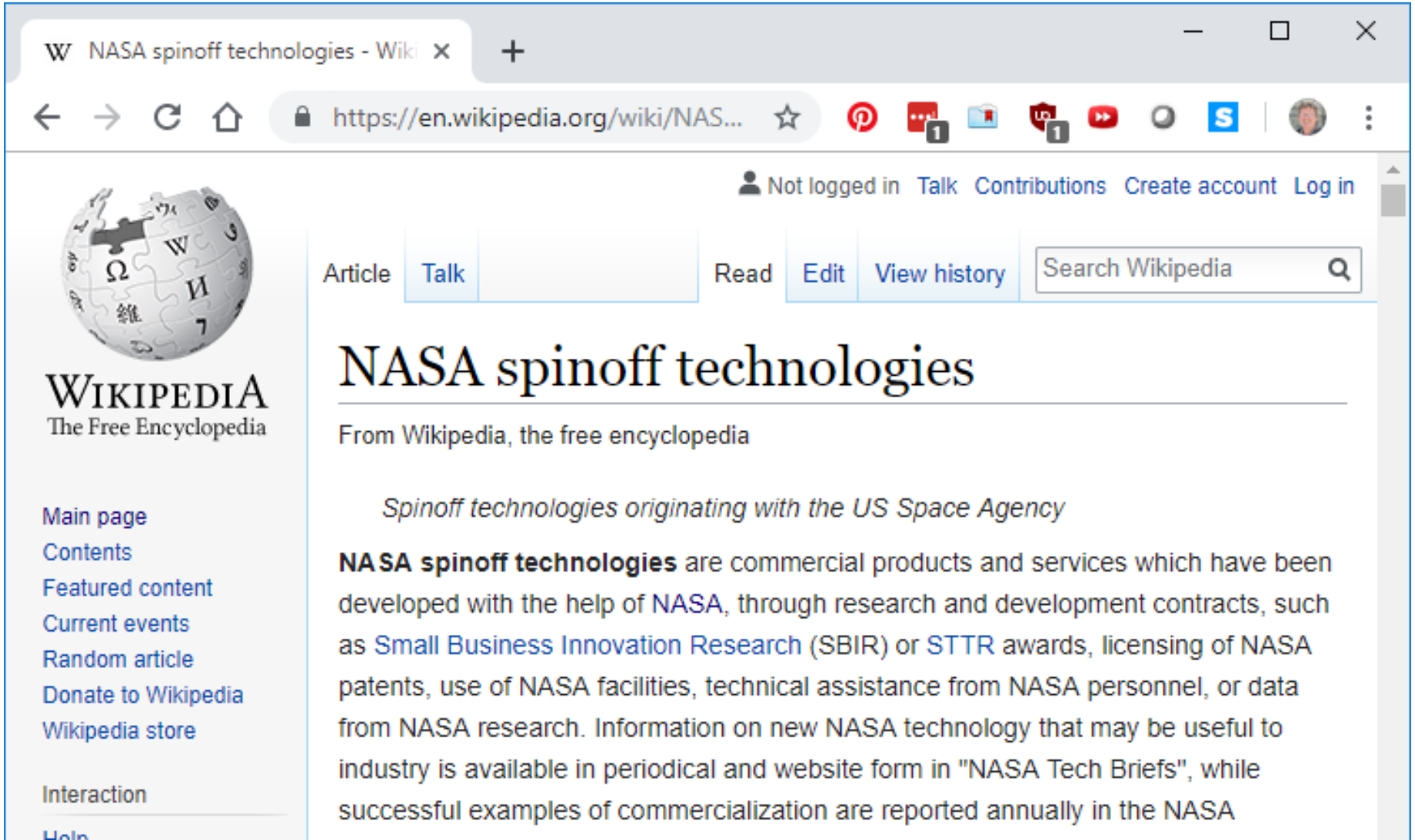
Claimed benefits:

- Increased public engagement in science
- Technological advance (remote sensing, sea floor maping, electronics)
- Spin-offs
- Political benefit
  https://www.nasa.gov/centers/goddard/news/series/moon/first_lunar_program.html
    - Sakharov’s open letter about democracy

The “NASA Effect”

- Making America cleverer: Intake of physics PhDs 3x
- In-flight computers, integrated circuits, Fairchild, Intel
- Surrey Satellite Technology Limited (Sweeting)
- Students for Exploration and Development of Space, Bezos president 1 yr
- Jobs

The “NASA Effect”

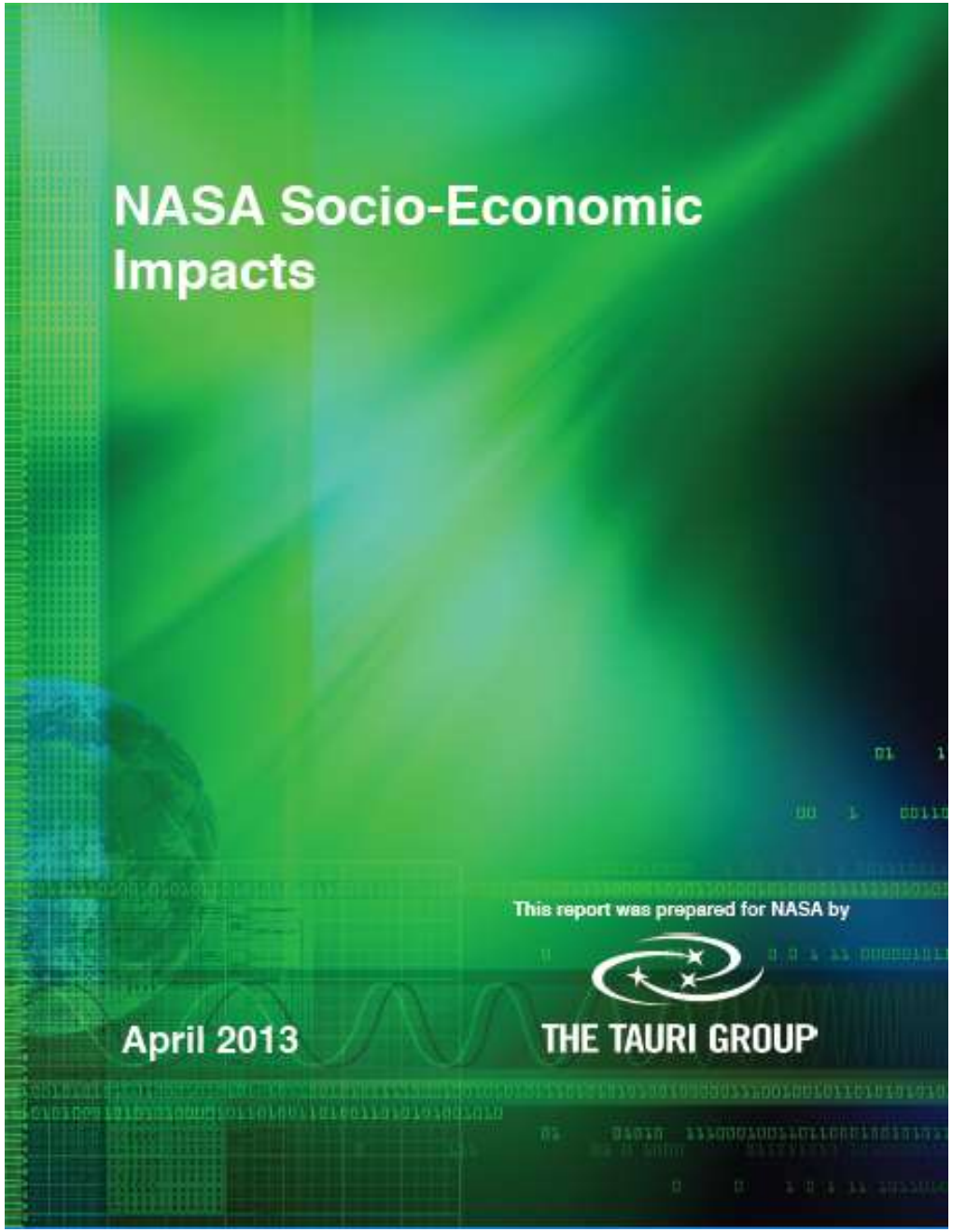


The “NASA Effect”

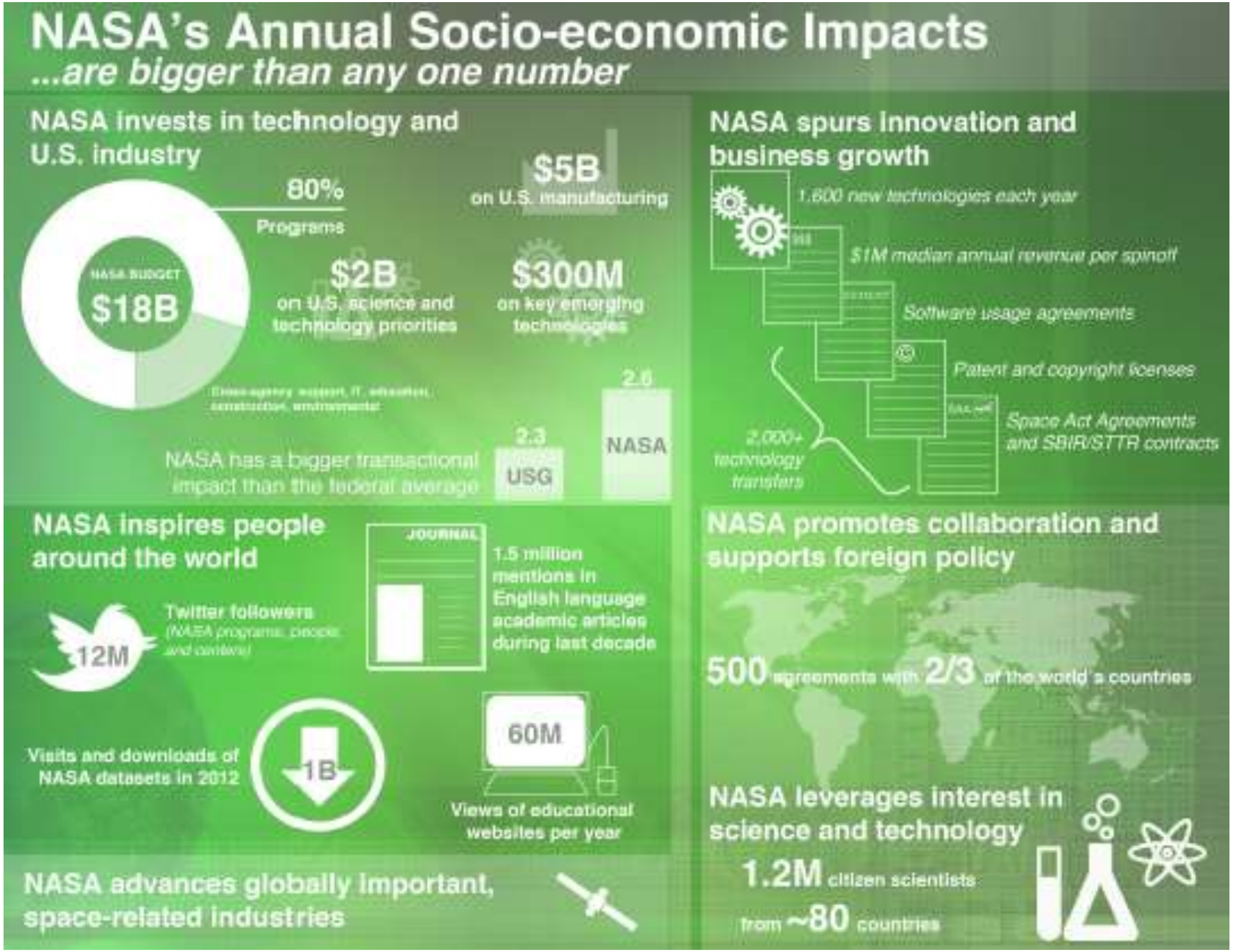
https://en.wikipedia.org/wiki/NASA_spinoff_technologies

Not Tang, Velcro or Teflon

The “NASA Effect”



The "NASA Effect"



Perspective

Ensuring Mutual Intelligibility

Ptolemaic vs Copernican Architecture

en.wikiquote.org

gocomics.com

Esperanto vs Francophonie