Mathematical Models as Research Data

—

why do need precise and well-written information about mathematical models and what can we do

Michael Kohlhase

Professur für Wissensrepräsentation und -verarbeitung
Informatik, FAU Erlangen-Nürnberg
http://kwarc.info

13. August 2018, Math Models and Math Software as Research Data

SOME RIGHTS RESERVED
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

1 Introduction

Mathematical Modeling and Simulation

- **Definition 1.1.** Mathematical Modeling and Simulation (MMS) as a research method
  1. fix an object and properties of interest (e.g. electron distribution in an electronic device)
  2. determine the quantities and physical laws involved (e.g. the electrostatic potential and the Poisson Equation)
  3. solve equations symbolically or numerically for given boundary conditions (complex software stacks)
  4. publish 1./2./3. in a paper and 3. in a data store (software on GitHub/GitLab)

Mathematical Modeling and Simulation

- **Definition 1.1.** Mathematical Modeling and Simulation (MMS) as a research method
  1. fix an object and properties of interest (e.g. electron distribution in an electronic device)
  2. determine the quantities and physical laws involved (e.g. the electrostatic potential and the Poisson Equation)
  3. solve equations symbolically or numerically for given boundary conditions (complex software stacks)
  4. publish 1./2./3. in a paper and 3. in a data store (software on GitHub/GitLab)

  MMS has been established as a primary scientific research method alongside the classical methods of experiment and theory.

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Mathematical Modeling and Simulation

- **Definition 1.1.** Mathematical Modeling and Simulation (MMS) as a research method
  1. fix an object and properties of interest (e.g. electron distribution in an electronic device)
  2. determine the quantities and physical laws involved (e.g. the electrostatic potential and the Poisson Equation)
  3. solve equations symbolically or numerically for given boundary conditions (complex software stacks)
  4. publish 1./2./3. in a paper and 3. in a data store (software on GitHub/GitLab)

  MMS has been established as a primary scientific research method alongside the classical methods of experiment and theory.

- Research in of MMS is characterized by mathematical models, scientific software, and numerical data from computations (input, output, parameters) (see [KT16])

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Mathematical Modeling and Simulation

- **Definition 1.1.** Mathematical Modeling and Simulation (MMS) as a research method
  1. fix an object and properties of interest (e.g. electron distribution in an electronic device)
  2. determine the quantities and physical laws involved (e.g. the electrostatic potential and the Poisson Equation)
  3. solve equations symbolically or numerically for given boundary conditions (complex software stacks)
  4. publish 1./2./3. in a paper and 3. in a data store (software on GitHub/GitLab)

  MMS has been established as a primary scientific research method alongside the classical methods of experiment and theory.

- Research in of MMS is characterized by mathematical models, scientific software, and numerical data from computations (input, output, parameters) (see [KT16])

  MMS faces a reproducibility crisis: success and proliferation puts strains on quality of models, software, and data.

Mathematical Modeling and Simulation

- **Definition 1.1.** Mathematical Modeling and Simulation (MMS) as a research method
  1. fix an object and properties of interest (e.g. electron distribution in an electronic device)
  2. determine the quantities and physical laws involved (e.g. the electrostatic potential and the Poisson Equation)
  3. solve equations symbolically or numerically for given boundary conditions (complex software stacks)
  4. publish 1./2./3. in a paper and 3. in a data store (software on GitHub/GitLab)

  MMS has been established as a primary scientific research method alongside the classical methods of experiment and theory.

- Research in of MMS is characterized by mathematical models, scientific software, and numerical data from computations (input, output, parameters) (see [KT16])

  MMS faces a reproducibility crisis: success and proliferation puts strains on quality of models, software, and data.

- Idea/Vision: Treat all three kinds of artefacts above as "Research Data", represent all aspects explicit $\rightsquigarrow$ establish machine support for

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

MMS Reproducibility Crisis

- Models (are published in mathematica/physical papers)
  - no standardization of naming, notation, constructors, …?
  - how are the formulae derived from the physical laws?
  - what are the side conditions/constraints under which the model is accurate?
- MMS Software (can only be understood wrt. the underlying models)
  - what are the underlying assumptions/constraints?
  - what are the admissible boundary conditions?
  - where does the iteration converge (well)?
- Data (needs specification to become information)
  - which software/model/discretization was used?
  - what quantity was measured in what unit?

SOME RIGHTS RESERVED

MMS Reproducibility Crisis

- Models (are published in mathematica/physical papers)
  - no standardization of naming, notation, constructors, . . . ?
  - how are the formulae derived from the physical laws?
  - what are the side conditions/constraints under which the model is accurate?
- MMS Software (can only be understood wrt. the underlying models)
  - what are the underlying assumptions/constraints?
  - what are the admissible boundary conditions?
  - where does the iteration converge (well)?
- Data (needs specification to become information)
  - which software/model/discretization was used?
  - what quantity was measured in what unit?
- Models are applied by people who did not develop them.
  - Implicit knowledge about the constraints, domains of applicability are lost.
- Models are applied by people who did not develop them.
  - Implicit knowledge about the constraints, domains of applicability are lost.

SOME RIGHTS RESERVED

State of the Art: FAIR Principles for the Data Aspect

- FAIR: data should be Findable, Accessible, Interoperable, and Reusable
  1. To be Findable:
     - F1 (meta)data are assigned a globally unique and eternally persistent identifier.
     - F2 data are described with rich metadata.
     - F3 (meta)data are registered or indexed in a searchable resource.
     - F4 metadata specify the data identifier.
  2. To be Accessible:
     - A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
     - A1.1 the protocol is open, free, and universally implementable.
     - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
     - A2 metadata are accessible, even when the data are no longer available.
  3. To be Interoperable:
     - I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
     - I2 (meta)data use vocabularies that follow FAIR principles.
     - I3 (meta)data include qualified references to other (meta)data.
  4. To be Re-usable:
     - R1 meta(data) have a plurality of accurate and relevant attributes.
     - R1.1 (meta)data are released with a clear and accessible data usage license.
     - R1.2 (meta)data are associated with their provenance.
     - R1.3 (meta)data meet domain-relevant community standards.

Ongoing. . . : how to implement these into repositories, protocols, and services?

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

State of the Art in 5 Dimensions

- Overview: Current Systems/Formats for Models, MMS Software, and Data can be characterized along five dimensions:

1: Coverage	2: Description	3: Formality	4: Computational	5 Immediacy
Domain-Independent	Continuous	Informal	Expressive	Domain Semantics
	Weak Formulations	Semi-Formal	Built-in special cases e.g. PDEs	Reformulation
Domain-Specific	Discrete	Formal	Solvable	Dedimensionalized Equations

$\leadsto$ continuous trade-off between "Specification" (hh) and "Implementation" (ll)

SOME RIGHTS RESERVED

State of the Art in 5 Dimensions

- Overview: Current Systems/Formats for Models, MMS Software, and Data can be characterized along five dimensions:

1: Coverage	2: Description	3: Formality	4: Computational	5 Immediacy
Domain-Independent	Continuous	Informal	Expressive	Domain Semantics
	Weak Formulations	Semi-Formal	Built-in special cases e.g. PDEs	Reformulation
Domain-Specific	Discrete	Formal	Solvable	Dedimensionalized Equations

$\rightsquigarrow$ continuous trade-off between "Specification" (hh) and "Implementation" (ll)

- Classifying Some Systems:

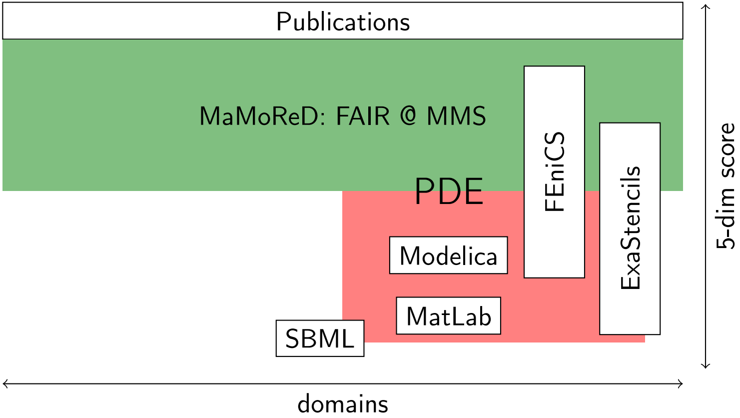
System	1	2	3	4	5
Publications	hh	hh	hh	hh	hh
Modelica	m	m	ll	ll	m
MatLab	h	ll	ll	ll	ll
FAIR @ MMS	hh-m	hh-m	hh-m	hh-m	hh-m

SOME RIGHTS RESERVED

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

FAIR Principles for Models and Simulation Software?

- Current Systems/Formats and proposed FAIR-like treatment of Models and MMS Software

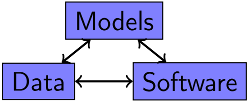


SOME RIGHTS RESERVED

2 The MaMoReD Vision (Details in later talks)

The MaMoReD Vision

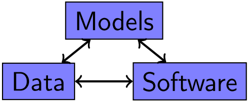
- Recap: Reproducibility of MMS requires precise information on the mathematical models, software, and data.



SOME RIGHTS RESERVED
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

The MaMoReD Vision

- Recap: Reproducibility of MMS requires precise information on the mathematical models, software, and data.

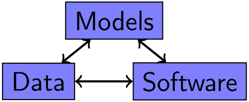


- Idea: FAIR principles for models & Software (exists for research data)
  - treat models/software as research data to make them machine-actionable
  - in particular: represent models and mathematical background knowledge explicitly/flexiformally

SOME RIGHTS RESERVED
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

The MaMoReD Vision

- Recap: Reproducibility of MMS requires precise information on the mathematical models, software, and data.

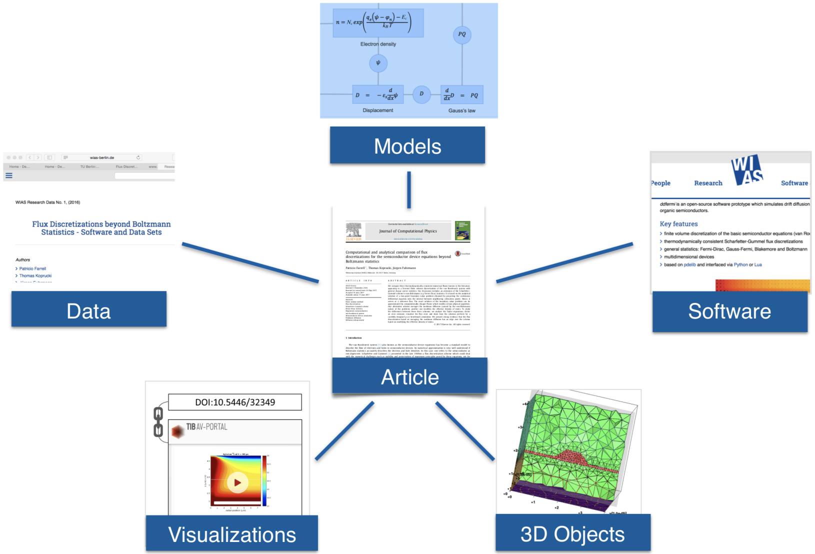


- Idea: FAIR principles for models & Software (exists for research data)
  - treat models/software as research data to make them machine-actionable
  - in particular: represent models and mathematical background knowledge explicitly/flexiformally
- Technically: Start with publications for coverage, repeat the following (conceptually)
  1. formalize, make implicit knowledge explicit
  2. organize into reusable components

  until we have enough structure to support semantic services(FAIR) do not forget to publish everything!

SOME RIGHTS RESERVED

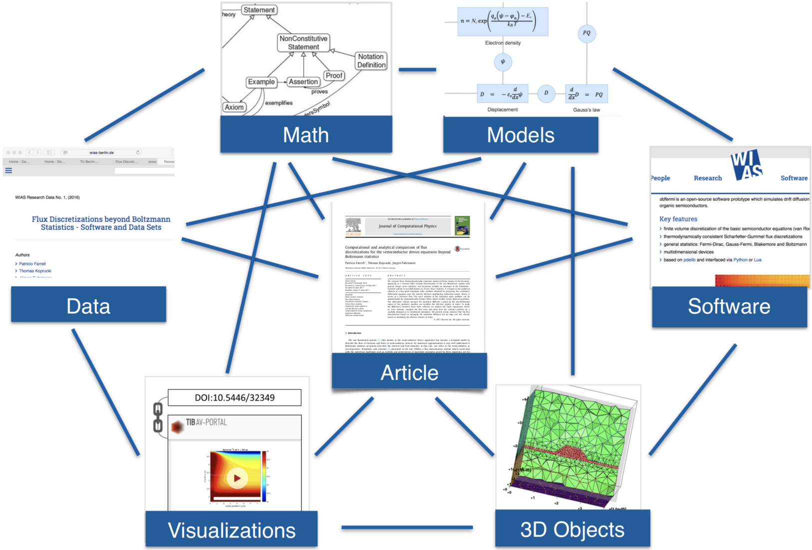
MaMoReD: Start by Publishing the Whole Story



SOME RIGHTS RESERVED

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

MaMoReD: Complex/Comprehensive Knowledge Graphs

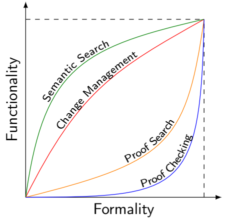


SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Content Representation and Services

- active documents adapt to audience (concise, enhanced papers)
  - e.g., "variables as functions for mathematicians",
  - in-document incremental flattening



- Flexiformal Model repositories
  - DOIs for models (MMT URIs)
  - integration with MathSearch
  - Model finder ⇝ applicable models
  - Model refactoring

- Integration of MMS software and Computer-Algebra Systems ⇝ MitM (OpenDreamKit)

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

3 MaMoRed: Modular Knowledge Representation for Model Application

Framing for Problem Solving (The FrameIT Method)

- **Example 3.1 (Problem 0.8.15).**

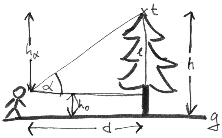
  How can you measure the height of a tree you cannot climb, when you only have a protactor and a tape measure at hand.

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Framing for Problem Solving (The FrameIT Method)

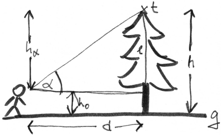
- **Example 3.1 (Problem 0.8.15).**
  How can you measure the height of a tree you cannot climb, when you only have a protactor and a tape measure at hand.



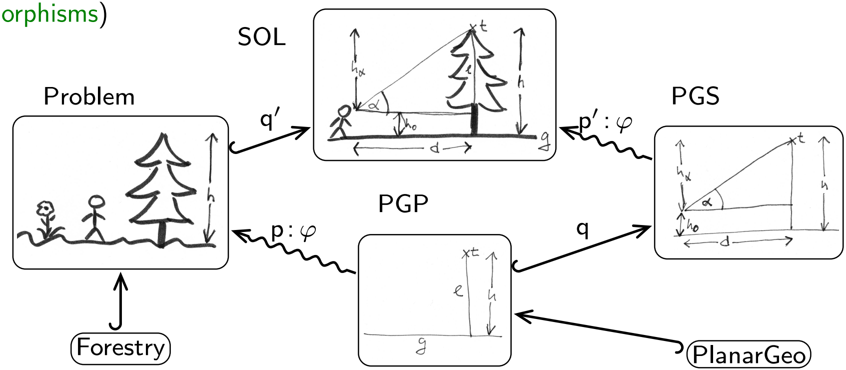
SOME RIGHTS RESERVED

Framing for Problem Solving (The FrameIT Method)

- **Example 3.1 (Problem 0.8.15).**
  How can you measure the height of a tree you cannot climb, when you only have a protactor and a tape measure at hand.



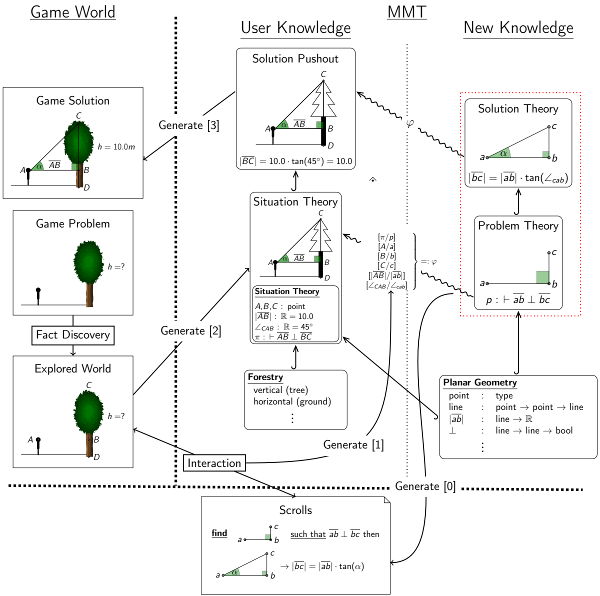
- Framing: view the problem as one that is already understood (using theory morphisms)



- squiggly (framing) morphisms guaranteed by metatheory of theories!

SOME RIGHTS RESERVED

Example Learning Object Graph



SOME RIGHTS RESERVED

4 The Math-in-the-Middle Paradigm for Interfacing Software Systems/Components — Interoperability via a Joint Meaning Space —

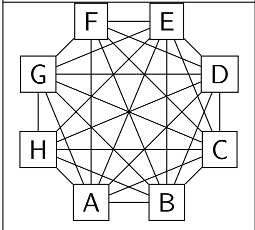
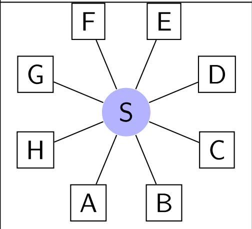
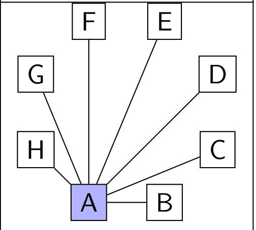
SOME RIGHTS RESERVED

Interoperability in OpenDreamKit

- OpenDreamKit (ODK): EU Project 2015-19, 16 Partners
  $\leadsto$ build a “mathematical VRE (Virtual Research Environment) toolkit”
- ODK Approach: VRE by connecting existing OSS systems. (and improve them)
- Advantages: well-known Open Source Software
  1. Let the specialists do what they do best and like (and avoid what they don't)
  2. collaboration exponentiates results
  3. competition fosters innovation (+ no vendor lock-in)
- Problem: does an elliptic curve mean the same in GAP, SageMath, LMFDB?
  - otherwise delegating computation becomes unsound
  - storing data in a central KB becomes unsafe
  - the user cannot interpret the results in an UI
- Idea: Need a common meaning space for safe distributed computation in a VRE!

Obtaining a Common Meaning Space for our VRE

- Three approaches for safe distributed computation/storage/UIs

<table>
<tr><th>peer to peer</th><th>open standard</th><th>industry standard</th></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>$n^2/2$ translations
symmetric</td><td>$2n$ translations
symmetric</td><td>$2n - 2$ translations
asymmetric</td></tr>
</table>

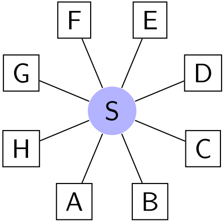
- Observation: We already have a "standard" for expressing the meaning of concepts/objects/models: mathematical vernacular! (e.g. in math. documents)
- Problem: mathematical vernacular is too
  - ambiguous: need a human to understand structure, words, and symbols
  - redundant: every paper introduces slightly different notions.
- Math-in-the-Middle Paradigm: encode math knowledge in modular flexiformal format as a frame of reference for joint meaning (OMDoc/MMT)

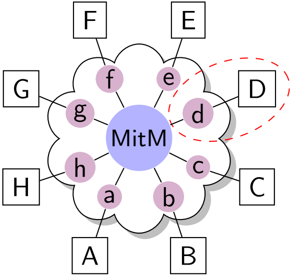
SOME RIGHTS RESERVED

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Standardization with Interfaces

- **Problem**: We are talking about knowledge-based systems (large investment)
- **Problem**: Knowledge is part of both the
  - System $\rightsquigarrow$ system-specific representation requirements and release cycle
  - Interoperability Standard $\rightsquigarrow$ stability and generality requirements.
- **Idea**: Open standard knowledge base with API theories





- **Definition 4.1.** API theories are
  - system-near (import/export facilities maintained with system)
  - declarative, in standard format (refine general theories, relation documented)

OpenMath System Dialects

- Observation: Every system has its own input language (optimized to domain)
- Idea: Abstract away from system surface languages (use internal syntax trees)

SOME RIGHTS RESERVED

OpenMath System Dialects

- Observation: Every system has its own input language (optimized to domain)
- Idea: Abstract away from system surface languages (use internal syntax trees)
- Observation: There are two kinds of symbols in syntax trees of a system S
  - constructors build primitive objects without involving computation, and
  - operations compute objects from other objects.
- **Definition 4.2.** The API theories $A(S)$ of S document them $\rightsquigarrow$ we can represent the API of S as *OpenMath* objects with constants from $A(S)$ (the $A(S)$-objects). We call the set of $A(S)$-objects the system dialect of S.

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

OpenMath System Dialects

- Observation: Every system has its own input language (optimized to domain)
- Idea: Abstract away from system surface languages (use internal syntax trees)
- Observation: There are two kinds of symbols in syntax trees of a system S
  - constructors build primitive objects without involving computation, and
  - operations compute objects from other objects.
- **Definition 4.2.** The API theories $A(S)$ of S document them $\leadsto$ we can represent the API of S as *OpenMath* objects with constants from $A(S)$ (the $A(S)$-objects). We call the set of $A(S)$-objects the system dialect of S.
- Idea: For each system S generate the API theories $A(S)$ and a serializer/deserializer into the system dialect: an OpenMath phrasebook.
- Progress: For system interoperability we only need to relate system dialects meaningfully.

Meaning-Preserving Relations between System Dialects

- **Definition 4.3.** We call a pair of identifiers (a_1, a_2) that describe the same mathematical concept an alignment.
  We call an alignment perfect, if it induces a total, truth-preserving translation. (e.g. alignment up to argument order)
- Intuition: Alignments don't need to be perfect to be useful!
  - Alignment up to Totality of Functions (e.g. division undefined on 0 and with $\frac{x}{0} = 0$)
  - Alignment for Certain Arguments (e.g. Addition on natural numbers and addition on real numbers)
  - Alignment up to Associativity (e.g. binary addition and "sequential" addition)

  They still allow for translating expressions between libraries. (under certain conditions)

MitM-Based Distributed Computation

- Observation: For interoperability between systems A and B with OpenMath phrasebooks and API theories, we only need
  1. a way of transporting *OpenMath* objects between systems A and B
  2. a system dialect mediator that translates A-objects into B-objects based on alignments.
- Idea: Mediator-based architecture

- Idea for 1.: translate A-objects to B-objects in two steps: A to ontology and ontology to B.
  Implemented in [Mül+17] based on the MMT system [Rab13; MMT], which implements the OMDoc/MMT format.
- Idea for 2.: Use the OpenMath SCSCP (Symbolic Computation Software Composability) protocol [Fre+] for that.
  Implemented SCSCP clients/server by for various OpenDreamKit systems.

5 The Flexiformalist Program: Introduction

SOME RIGHTS RESERVED

Background: Mathematical Documents

- Mathematics plays a fundamental role in Science, Technology, and Engineering (learn from Math, apply for STEM)
- Mathematical knowledge is rich in content, sophisticated in structure, and technical in presentation,
- its conservation, dissemination, and utilization constitutes a challenge for the community and an attractive line of inquiry.
- Challenge: How can/should we do mathematics in the 21st century?
- Mathematical knowledge and objects are transported by documents
- Three levels of electronic documents:
  0. printed (for archival purposes) ($\sim$90%)
  1. digitized (usually from print) ($\sim$50%)
  2. presentational: encoded text interspersed with presentation markup ($\sim$20%)
  3. semantic: encoded text with functional markup for the meaning ($\leq$0.1%)

  transforming down is simple, transforming up needs humans or AI.
- Observation: Computer support for access, aggregation, and application is (largely) restricted to the semantic level.
- This talk: How do we do maths and math documents at the semantic level?

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Hilbert's (Formalist) Program

- **Definition 5.1.** Hilbert's Program called for a foundation of mathematics with
  - A formal system that can express all of mathematics (language, models, calculus)
  - Completeness: all valid mathematical statements can be proved in the formalism.
  - Consistency: a proof that no contradiction can be obtained in the formalism of mathematics.
  - Decidability: algorithm for deciding the truth or falsity of any mathematical statement.
- Originally proposed as "metamathematics" by David Hilbert in 1920.
- Evaluation: The program was
  - successful in that FOL+ZFC is a foundation [Göd30] (there are others)
  - disappointing for completeness [Göd31], consistency [Göd31], decidability [Chu36; Tur36]
  - inspiring for computer Scientists building theorem provers
  - largely irrelevant to current mathematicians (I want to address this!)

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Formality in Logic and Artificial Intelligence

- AI, Philosophy, and Math identify formal representations with Logic
- **Definition 5.2.** A formal system $S := \langle\mathcal{L}, \mathcal{M}, \mathcal{C}\rangle$ consists of
  - a (computable) formal language $\mathcal{L} := \mathcal{L}(S)$ (grammar for words/sentences)
  - a model theory $\mathcal{M}$, (a mapping into (some) world)
  - and a sound (complete?) proof calculus $\mathcal{C}$ (a syntactic method of establishing truth)

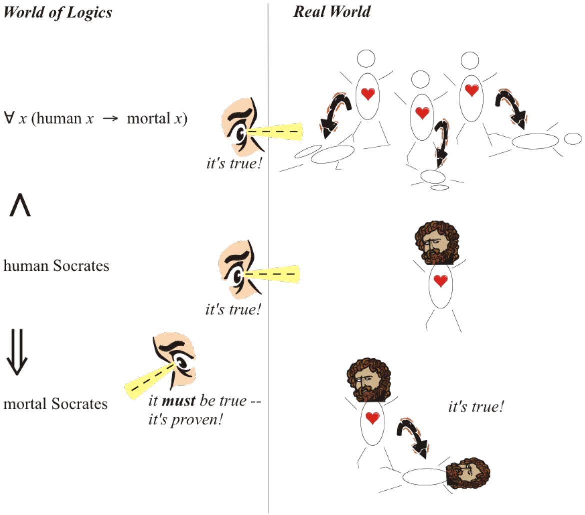
  We use $\mathfrak{F}$ for the class of all formal systems
- Reasoning in a formal system proceeds like a chess game: chaining "moves" allowed by the proof calculus via syntactic (depending only on the form) criteria.
- Observation: computers need $\mathcal{L}$ and $\mathcal{C}$ (adequacy hinges on relation to $\mathcal{M}$)
- Formality is a "all-or-nothing property".(a single "clearly" can ruin a formal proof)
- Empirically: formalization is not always achievable (too tedious for the gain!)
- Humans can draw conclusions from informal (not $\mathcal{L}$) representations by other means (not $\mathcal{C}$).

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

The miracle of logics

- Purely formal derivations are true in the real world!



SOME RIGHTS RESERVED
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Formalization in Mathematical Practice

- To formalize maths in a formal system $\mathcal{S}$, we need to choose a foundation, i.e. a foundational $\mathcal{S}$-theory, e.g. a set theory like ZFC.
- Formality is an all-or-nothing property (a single "obviously" can ruin it.)
- Almost all mathematical documents are informal in 4 ways:
  - the foundation is unspecified (they are essentially equivalent)
  - the language is informal (essentially opaque to MKM algos.)
  - even formulae are informal (presentation markup)
  - context references are underspecified
    - mathematical objects and concepts are often identified by name
    - statements (citations of definitions, theorems, and proofs) underspecified
    - theories and theory reuse not marked up at all
- The gold standard of mathematical communication is "rigor" (cf. [BC01])

SOME RIGHTS RESERVED

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Formalization in Mathematical Practice

- To formalize maths in a formal system $\mathcal{S}$, we need to choose a foundation, i.e. a foundational $\mathcal{S}$-theory, e.g. a set theory like ZFC.
- Formality is an all-or-nothing property (a single "obviously" can ruin it.)
- Almost all mathematical documents are informal in 4 ways:
- The gold standard of mathematical communication is "rigor" (cf. [BC01])
  - **Definition 5.3.** We call a mathematical document rigorous, if it could be formalized in a formal system given enough resources.
  - This possibility is almost always unconsummated
  - Why?: There are four factors that disincentivize formalization for Maths
    propaganda: *Maths is done with pen and paper*
    tedium: de Bruijn factors ~ 4 for current systems (details in [Wie12])
    inflexibility: formalization requires commitment to formal system and foundation
    proof verification useless: peer reviewing works just fine for Math
  - **Definition 5.4.** The de Bruijn factor is the quotient of the lengths of the formalization and the original text.

Formalization in Mathematical Practice

- To formalize maths in a formal system $\mathcal{S}$, we need to choose a foundation, i.e. a foundational $\mathcal{S}$-theory, e.g. a set theory like ZFC.
- Formality is an all-or-nothing property (a single "obviously" can ruin it.)
- Almost all mathematical documents are informal in 4 ways:
- The gold standard of mathematical communication is "rigor" (cf. [BC01])
  - **Definition 5.3.** We call a mathematical document rigorous, if it could be formalized in a formal system given enough resources.
  - This possibility is almost always unconsummated
  - Why?: There are four factors that disincentivize formalization for Maths
    propaganda: *Maths is done with pen and paper*
    tedium: de Bruijn factors ~ 4 for current systems (details in [Wie12])
    inflexibility: formalization requires commitment to formal system and foundation
    proof verification useless: peer reviewing works just fine for Math
  - **Definition 5.4.** The de Bruijn factor is the quotient of the lengths of the formalization and the original text.

  In Effect: Hilbert's program has been comforting but useless

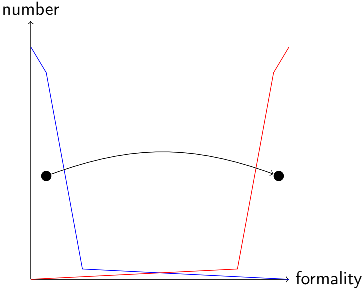
- Question: What can we do to change this?

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Migration by Stepwise Formalization

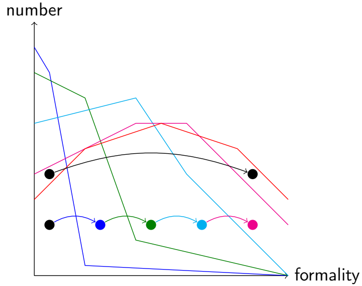
- Full Formalization is hard (we have to commit, make explicit)
- Let's look at documents and document collections.



SOME RIGHTS RESERVED
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Migration by Stepwise Formalization

- Full Formalization is hard (we have to commit, make explicit)
- Let's look at documents and document collections.
- Partial formalization allows us to
  - formalize stepwise, and
  - be flexible about the depth of formalization.



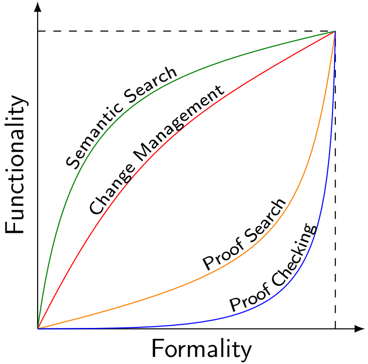
SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Functionality of Flexiformal Services

- Generally: Flexiformal services deliver according to formality level (GIGO: Garbage in $\rightsquigarrow$ Garbage out!)
- But: Services have differing functionality profiles.

  - Math Search works well on informal documents
  - Change management only needs dependency information
  - Proof search needs theorem formalized in logic
  - Proof checking needs formal proof too



SOME RIGHTS RESERVED
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

The Flexiformalist Program (Details in [Koh13])

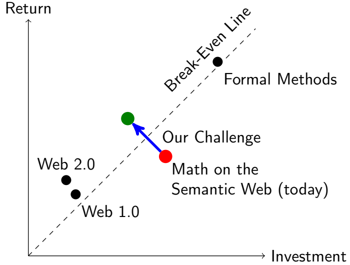
- The development of a regime of partially formalizing
  - mathematical knowledge into a modular ontology of mathematical theories (content commons), and
  - mathematical documents by semantic annotations and links into the content commons (semantic documents),
- The establishment of a software infrastructure with
  - a distributed network of archives that manage the content commons and collections of semantic documents,
  - semantic web services that perform tasks to support current and future mathematic practices
  - active document players that present semantic documents to readers and give access to respective
- the re-development of comprehensive part of mathematical knowledge and the mathematical documents that carries it into a flexiformal digital library of mathematics.

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Applications!

- A Business model for a Semantic Web for Math/Science?
- For uptake it is essential to match the return to the investment!



- Need to move the technology up (carrots) and left (easier)

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Conclusion/Take-Home Message

- Mathematical Modelling and Simulation is very successful (third pillar of science)
- MMS: Simulation software solving the equations from mathematical models produces data
- Problem: MMS has a reproducibility crisis (brought on by widespread usage)
- MaMoReD Proposal: use MKM techniques (Math Models as Research Data)
  - flexible formalization: from active articles to formalized physical laws to discrete iterations
  - modular representations for re-use and

SOME RIGHTS RESERVED

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

References I

Henk Barendregt and Arjeh M. Cohen. “Electronic communication of mathematics and the interaction of computer algebra systems and proof assistants”. In: *Journal of Symbolic Computation* 32 (2001), pp. 3–22.

Alonzo Church. “A note on the Entscheidungsproblem”. In: *Journal of Symbolic Logic* (May 1936), pp. 40–41.

Sebastian Freundt et al. *Symbolic Computation Software Composability Protocol (SCSCP)*. Version 1.3. URL: https://github.com/OpenMath/scscp/blob/master/revisions/SCSCP_1_3.pdf (visited on 08/27/2017).

Kurt Gödel. “Die Vollständigkeit der Axiome des logischen Funktionenkalküls”. In: *Monatshefte für Mathematik und Physik* 37 (1930). English Version in [**Heijenoort67**], pp. 349–360.

References II

Kurt Gödel. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". In: *Monatshefte der Mathematischen Physik* 38 (1931). English Version in [**Heijenoort67**], pp. 173–198.

Michael Kohlhase. "The Flexiformalist Manifesto". In: *14th International Workshop on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2012)*. Ed. by Andrei Voronkov et al. Timisoara, Romania: IEEE Press, 2013, pp. 30–36. ISBN: 978-1-4673-5026-6. URL: http://kwarc.info/kohlhase/papers/synasc13.pdf.

Thomas Koprucki and Karsten Tabelow. "Mathematical Models: A Research Data Category?" In: *Mathematical Software - ICMS 2016 - 5th International Congress*. Ed. by Gert-Martin Greuel et al. Vol. 9725. LNCS. Springer, 2016, pp. 423–428. DOI: 10.1007/978-3-319-42432-3. URL: http://www.wias-berlin.de/preprint/2267/wias_preprints_2267.pdf.

SOME RIGHTS RESERVED

References III

- *MMT – Language and System for the Uniform Representation of Knowledge*. project web site. URL: https://uniformal.github.io/ (visited on 08/30/2016).
- Dennis Müller et al. “Alignment-based Translations Across Formal Systems Using Interface Theories”. In: *Fifth Workshop on Proof eXchange for Theorem Proving - PxTP 2017*. 2017. URL: http://jazzpirate.com/Math/AlignmentTranslation.pdf.
- Florian Rabe. “The MMT API: A Generic MKM System”. In: *Intelligent Computer Mathematics*. Conferences on Intelligent Computer Mathematics (Bath, UK, July 8–12, 2013). Ed. by Jacques Carette et al. Lecture Notes in Computer Science 7961. Springer, 2013, pp. 339–343. ISBN: 978-3-642-39319-8. DOI: 10.1007/978-3-642-39320-4.
- Alan Turing. “On computable numbers, with an application to the Entscheidungsproblem”. In: *Proceedings of the London Mathematical Society, Series 2* 42 (June 1936), pp. 230–265.

SOME RIGHTS RESERVED

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

References IV

Freek Wiedijk. *The "de Bruijn factor"*. web page at http://www.cs.ru.nl/~freek/factor/. Mar. 1, 2012. URL: http://www.cs.ru.nl/~freek/factor/.