Interoperability in the OpenDreamKit Project: The Math-in-the-Middle Approach

Paul-Olivier Dehaye[3] Michael Kohlhase[2] Alexander Konovalov[4] Samuel Lelièvre[1] Markus Pfeiffer[4] Nicolas M. Thiéry[1]

[1] Université Paris-Sud
[2] Jacobs University
[3] University of Zürich
[4] University of St Andrews

1 Introduction

From their earliest days, computers have been used in pure mathematics, either to make tables, to prove theorems (famously the four colour theorem) or, as with the astronomer's telescope, to explore new theories. Computer-aided experiments, and the use of databases relying on computer calculations such as the Small Groups Library in GAP, the Modular Atlas in group and representation theory, or the L-functions and Modular Forms Database (LMFDB, see later), are part of the standard toolbox of the pure mathematician, and certain areas of mathematics completely depend on it. Computers are also increasingly used to support collaborative work and education.

The last decades witnessed the emergence of a wide ecosystem of open-source tools to support research in pure mathematics. This ranges from specialized to general purpose computational tools such as GAP, PARI/GP, LinBox, MPIR, Sage, or Singular, via online databases like the LMFDB and does not count online services like Wikipedia, arXiv, or MathOverflow. A great opportunity is the rapid emergence of key technologies, and in particular the Jupyter (previously IPython) platform for interactive and exploratory computing which targets all areas of science.

This has proven the viability and power of collaborative open-source development models, by users and for users, even for delivering general purpose systems targeting a large public (researchers, teachers, engineers, amateurs, ...). Yet we are still missing *Virtual Research Environments* (VRE), that combine these tools and present a uniform work environment to the user, and enable groups of researchers, typically widely dispersed, to work collaboratively on a per project basis. This is exactly where the OpenDreamKit project kicks in.

2 Integrating Mathematical Software Systems via the Math-in-the-Middle Approach

To achieve the goal of assembling the ecosystem of mathematical software systems in the OpenDreamKit project into a coherent mathematical VRE, we have to make the systems interoperable at a mathematical level. In particular, we have to establish a common meaning space that allows to share computation, visualization of the mathematical concepts, objects, and models (COMs) between the respective systems. Building on this we can build a VRE with classical techniques for integrated development environments (IDE).

2.1 A Common Meaning Space for Interoperability

Concretely, the problem is that the software systems in OpenDreamKit have different coverage, and where these overlap representations of and functionalities for the COMs involved differ. This starts with simple naming issues (*e.g.* elliptic curves are named ec in the LMFDB, and as EllipticCurve in Sage), persists through the underlying data structures (permutations are represented as products of cycles in GAP, in list form in Sage, and in differing representations in the various tables of the LMFDB), and becomes virulent at the level of algorithms, their parameters, and domains of applicability.

To obtain a common meaning space for a VRE, we have the three well-known approaches in Figure 1.

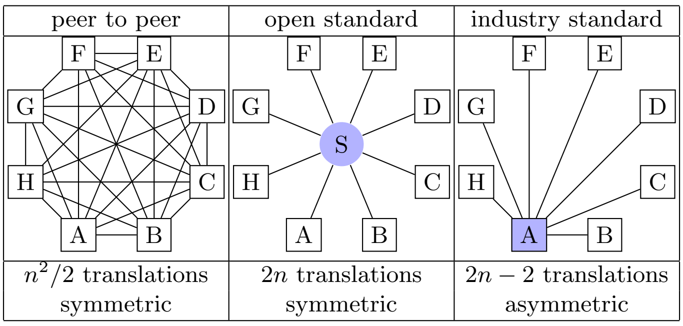

Fig. 1. Approaches for many-systems interoperability

The first does not scale to a project with about a dozen systems, for the third there is no obvious contender in the OpenDreamKit ecosystem. Fortunately, we already have a "standard" for expressing the meaning of COMs – **mathematical vernacular**: the language of mathematical communication, and in fact all the COMs supported in the OpenDreamKit VRE are documented in mathematical vernacular in journal articles, manuals, etc.

The obvious problem is that mathematical vernacular is too *i*) *ambiguous*: we need a human to understand structure, words, and symbols *ii*) *redundant*: every paper introduces slightly different notions.

Therefore we explore an approach, where we **flexiformalize** (i.e. partially formalize; see [Koh13]) mathematical vernacular to obtain a flexiformal ontology of mathematics that can serve as an open communication vocabulary. We call the approach the **Math-in-the-Middle** (MitM) Strategy for integration and the ontology the **MitM ontology**.

Before we go into any detail about how this ontology looks and how it induces a uniform meaning space, we have to address another problem: the descriptions in the MitM ontology must at the same time be system-near to make interfacing easy for systems, and serve as an interoperability standard – *i.e.* be general and stable. If we have an ontology system that allows modular/structured ontologies, we can solve this apparent dilemma by introducing **interface theories** [KRSC11], *i.e.* ontology modules (the light purple circles in Figure 2) that are at the same time system-specific in their description of COMs – near the actual representation of the system and part of the greater MitM ontology (depicted by the cloud in Figure 2) as they are connected to the core MitM ontology (the blue circle) by views we call **interviews** (see below). The MitM approach stipulates that interface theories and interviews are maintained and released together with the respective systems, whereas the core MitM ontology represents the mathematical scope of the VRE and is maintained with it. In fact in many ways, the core MitM ontology is the conceptual essence of the mathematical VRE.

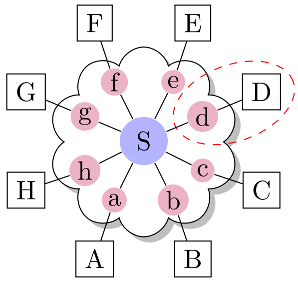

Fig. 2. Interface theories

2.2 Realizing and Utilizing a MitM Ontology

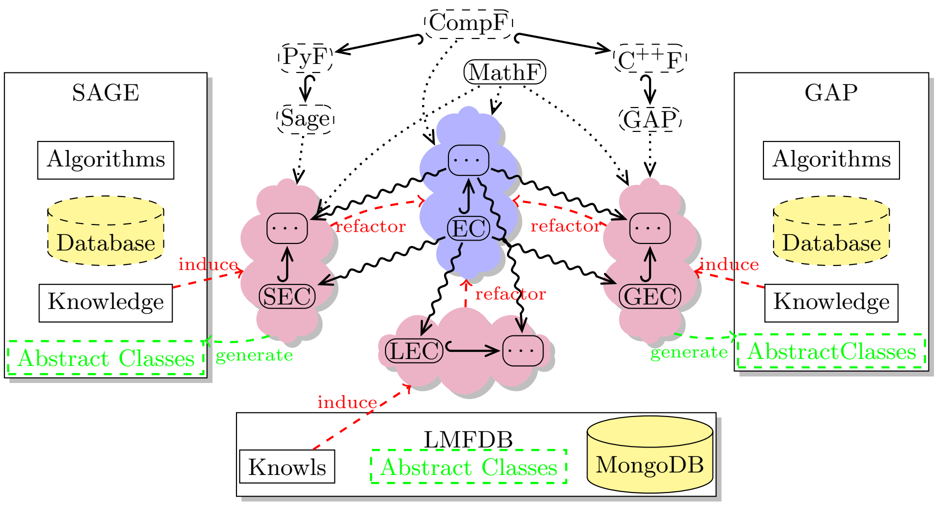

Fig. 3. The MitM Paradigm in Detail

For the MitM Ontology, we arrive at the situation in Figure 3, where we drill into the MitM information architecture from Figure 2, but restrict at this stage to three systems from the OpenDreamKit project. In the middle we see the core MitM ontology (the blue cloud) as an OMDoc/MMT theory graph connected to the interface theories (the purple clouds) via MitM interviews. Conceptually, the systems in OpenDreamKit consist of three main components:

i) a *Knowledge Representation component* that provides data structures for the COMs and their properties.
ii) a *DataBase component* that provides mass storage for objects, and
iii) a *library of algorithms* that operate on these.

To connect a system to an MitM-based VRE, the knowledge representation component is either refactored so that it can generate interface theories, or a schema-like description of the underlying data structures is created manually from which abstract data structures for the system can be generated automatically – in this version the interface theories act as an Interface Description Language.

In this situation there are two ways to arrive at a greater MitM ontology: the OpenDreamKit project aims to explore both: either *i*) standardizing a core MitM by refactoring the various interface theories where they overlap, or *ii*) flexiformalizing the available literature for a core MitM ontology. For *i*), the MitM interviews emerge as refinements that add system-specific details to the general mathematical concepts For *ii*), we have to give the interviews directly.

To see that this architecture indeed gives us a uniform meaning space, we observe that the core MitM ontology uses a mathematical foundation (presumably some form of set theory), whereas the interface theories also use system-specific foundations that describe aspects of the computational primitives of the respective systems. We have good formalizations of the mathematical foundations already; first steps towards a computational ones have been taken as well.

Our efforts also fit neatly alongside similar efforts underway across the sciences to standardize metadata formats, except that the typing taking place here tends to have much higher complexity since our objects of study are sometimes seen as types and sometimes as instances (think of groups for instance).

3 Conclusion

In this paper we have presented the OpenDreamKit project and the "Math-in-the-Middle" approach it explores for mitigating the system integration problems inherent in combining an ecosystem of open source software systems into a coherent mathematical virtual research environment. The MitM approach relies on a central, curated, flexiformal ontology of the mathematical domains to be covered by the VRE together with system-near interface theories and interviews to the core ontology that liaise with the respective systems. We have reported on two case studies that were used to evaluate the approach: an interface for the LMFDB, and a more semantic handle interface between GAP and Sage.

Even though the development of the MitM is still at a formative stage, these case studies show the potential of the approach. We hope that the nontrivial cost of curating an ontology of mathematical knowledge and interviews to the interface theories will be offset by its utility as a resource, which we are currently exploring; the unification of the knowledge representation components

- enables VRE-wide domain-centered (rather than system-centered) documentation;
- can be leveraged for distributed computation via uniform protocols like the SCSCP [HR09] and MONET-style service matching [CDT04] (the absence of content dictionaries – MitM theories – was the main hurdle that kept these from gaining more traction);
- will lead to the wider adoption of best practices in mathematical knowledge management in the systems involved – in fact, this is already happening.

Whether in the end the investment into the MitM will pay off also depends on the quality and usability of the tools for mathematical knowledge management. Therefore we invite the CICM community to interact with and contribute to the OpenDreamKit project, on this work package and the others.

Acknowledgements The authors gratefully acknowledge the other participants of the St Andrews workshop, in particular John Cremona, Luca de Feo, Mihnea Iancu, Steve Linton, Dennis Müller, Viviane Pons, Florian Rabe, and Tom Wiesing, for discussions and experimentation which clarified the ideas behind the math-in-the-middle approach.

We acknowledge financial support from the OpenDreamKit Horizon 2020 European Research Infrastructures project (#676541), from the EPSRC Collaborative Computational Project CoDiMa (EP/M022641/1) and from the Swiss National Science Foundation grant PP00P2_138906.

References

[CDT04] Olga Caprotti, Mike Dewar, and Daniele Turi. *Mathematical Service Matching Using Description Logic and OWL*. Tech. rep. The MONET Consortium, 2004. URL: `http://monet.nag.co.uk/monet/publicdocs/monet_onts.pdf` (visited on 08/11/2010).

[HR09] Peter Horn and Dan Roozemond. “OpenMath in SCIEnce: SCSCP and POPCORN”. In: *MKM/Calculemus Proceedings*. Ed. by Jacques Carette et al. LNAI 5625. Springer Verlag, July 2009, pp. 474–479. ISBN: 978-3-642-02613-3.

[Koh13] 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`.

[KRSC11] Michael Kohlhase, Florian Rabe, and Claudio Sacerdoti Coen. "A Foundational View on Integration Problems". In: *Intelligent Computer Mathematics*. Ed. by James Davenport et al. LNAI 6824. Springer Verlag, 2011, pp. 107–122. ISBN: 978-3-642-22672-4. URL: `http://kwarc.info/kohlhase/papers/cicm11-integration.pdf`.