SymbolicData, Computer Algebra and Web 2.0

H.-G. Gräbe[1], A. Heinle[2], S. Johanning[3]

[1] *Leipzig University, Germany, graebe@informatik.uni-leipzig.de*
[2] *University of Waterloo, Canada, aheinle@uwaterloo.ca*
[3] *Leipzig University, Germany, simonjohanning@googlemail.com*

1   Introduction

What is Computer Algebra (CA)? Twenty years ago more than 200 leading edge computer algebraists in a worldwide joint effort compiled a description of the CA landscape [4] and defined the target of CA in the following way:

> "Computer Algebra is a subject of science devoted to methods for solving mathematically formulated problems by symbolic algorithms, and to implementation of these algorithms in software and hardware. It is based on the exact finite representation of finite or infinite mathematical objects and structures, and allows for symbolic and abstract manipulation by a computer. Structural mathematical knowledge is used during the design as well as for verification and complexity analysis of the respective algorithms. Therefore computer algebra can be effectively employed for answering questions from various areas of computer science and mathematics, as well as natural sciences and engineering, provided they can be expressed in a mathematical model." [4, p. 2]

Johannes Grabmeier, at that time head of the German CA Fachgruppe, developed an even broader view of a subject *Computer Mathematics* as a symbiosis of computer technology and mathematics at large as the true core of "Scientific Computing" [5]. Such a *technology*[1] *of quantitative methods* is a corner stone of any science, that "can be considered as developed"[2]. Figure 1 displays these concepts, the difference between deductive and numerical mathematics and the position of such a *Computer Mathematics* between mathematics and computer science.

Twenty years later a practical incarnation of such a vision is any of the mature General Purpose Computer Algebra Systems, in particular the one that claims to be "the world's definitive system for modern technical computing" [9].

---

[1]Technology considered in the broad sense of *applicable processual knowledge* in the ancient greek meaning, see e.g., [16], not in the narrow meaning of *tool, tool making, tool using* as, e.g., in the German VDI-3780 norm [17].

[2]A claim attributed to Karl Marx by Paul Lafargue. David Hilbert supports such a view: "Everything that can be an object of scientific thinking as soon as it matures to formation of theories is in the bondage of the axiomatic method and thus indirectly of mathematics." [8]

What remained from such an integrated view – a *CA Tower of Babel* – twenty years later? What is the relation between the sections of ACA 2015, that address specialized and over the years even more and more specialized topics? What are the consequences of a division of CA into more and more sub- and subsubcommunities? Has THE LORD confused their language[3] once more?

2 The SYMBOLICDATA Project

2.1 The Roots

SYMBOLICDATA hab been part of CA infrastructural efforts for more than 15 years. It grew up from the Special Session on Benchmarking at the 1998 ISSAC conference, and started to build a reliable and sustainably available reference of Polynomial Systems data, to extend and update it, to collect meta information about the records, and also to develop tools to manage the data and to set up and run testing and benchmarking computations on the data. The main design decisions and implementations of the first prototype were realized by Olaf Bachmann and Hans-Gert Gräbe in 1999 and 2000. We collected data from *Polynomial Systems Solving* and *Geometry Theorem Proving*, set up a CVS repository, and started test computations with the main focus on Polynomial Systems Solving.

In 2005 the Web site `http://www.symbolicdata.org` sponsored by the German CA Fachgruppe went online and a second phase of active development started. Data was supplied by the CoCoA group (F. Cioffi), the Singular group (M. Dengel, M. Brickenstein, S. Steidel, M. Wenk), V. Levandovskyy (non commutative polynomial systems, G-Algebras) and Raymond Hemmecke (Test sets from Integer Programming). During the Special Semester on Groebner Bases in March 2006 we discussed aspects to extend the project, in particular with Bruno Buchberger, Alexander Zapletal, and Viktor Levandovskyy.

2.2 Version 3 – SYMBOLICDATA Goes Semantic

A third phase started in 2009 when the project joined forces with the Agile Knowledge Engineering and Semantic Web (AKSW) Group at Leipzig University [1] in order to strongly refactor the data along standard Semantic Web concepts based on the Resource Description Framework (RDF) [15]. In 2012–2014 these efforts were

---

[3]But the Lord came down to see the city and the tower which the sons of men had built. And the Lord said, "Indeed the people are one and they all have one language, and this is what they begin to do; now nothing that they propose to do will be withheld from them. Come, let Us go down and there confuse their language, that they may not understand one another's speech." (Genesis 11)

supported by a 12 month grant for Andreas Nareike and another 5 month grant for Simon Johanning within the *Saxonian E-Science Initiative* [3].

Within this scope we completed a redesign of the data along the rules of Linked Data and semantic, RDF-based technology, distinguishing more consequently between data (*resources* in the RDF terminology) and meta data (*knowledge bases* in the RDF terminology). The new SYMBOLICDATA data and tools were released as version 3.0 in September 2013. Version 3.1 is to be released in the mid of 2015.

Resources (examples for testing, profiling and benchmarking software and algorithms from different CA areas) are publicly available (mainly) in XML markup, meta data in RDF notation both from a public git repo, hosted at `github.org`, and from an OntoWiki based RDF data store at `http://symbolicdata.org/Data`. Moreover, we offer a SPARQL endpoint to explore the data by standard Linked Data methods.

The website operates on a standard installation using an Apache web server to deliver the data, a Virtuoso RDF data store as data backend, a SPARQL endpoint and (optionally) OntoWiki to explore, display and edit the data. This standard installation can easily be rolled out on a local site (tested under Linux Debian and recent Ubuntu versions; a more detailed description can be found in our wiki [14]) to support local testing, profiling and benchmarking.

The distribution offers also tools for integration with a local computational environment as, e.g., provided by Sagemath [13] – the Python based *SDEval package* [7] by Albert Heinle offers a JUnit-like framework to set up, run, log, monitor and interrupt testing and benchmarking computations, and the *sdsage package* [10] by Andreas Nareike provides a showcase for SYMBOLICDATA integration with the Sagemath computational environment.

Currently the SYMBOLICDATA data collection contains resources from the areas of Polynomial Systems Solving (390 records, 633 configurations), Free Algebras (83 records), G-Algebras (8 records), GeoProofSchemes (297 records) and Test Sets from Integer Programming (49 records).

Note that such a concept is not restricted to resources centrally managed at `symbolicdata.org`, but can easily be extended to other data stores on the web that are operated by different CA subcommunities and offer a minimum of Linked Data connectivity. There are draft versions of resource descriptions about Fano Polytopes (8630 records) and Birkhoff Polytopes (5399 records) from the `polymake` project hosted by Andreas Paffenholz and about Transitive Groups (3605 records) from the Database for Number Fields of Jürgen Klüners and Gunter Malle that point to external resources. For Test Sets we joined forces with the Normaliz group — Tim Römer provided some new benchmarks and refactored the old examples to fit with the `normaliz` syntax.

2.3 The Vision: Towards a Computer Algebra Social Network

I come back to the *CA Tower of Babel* image unfolded in the introduction. If we shift the focus of any project – and we did so for SYMBOLICDATA with version 3.0 – from the *data* to the *people and their intentions* perspective – *why* they collect and manipulate such data – new organizational perspectives for common efforts are revealed. The business of any CA project is a techno-social one and we think in the spirit of sociomateriality [12], it is time to use the *technical means* of a semantically enriched Web 2.0 to also strengthen the *social* part of our cooperation and to contribute to the efforts to build up an interconnected *E-Science World*.

During the last years such efforts matured within the *Science at Large*. Services such as MathSciNet, arXiv.org, or EasyChair.org have been established and their usefulness is widely acknowledged. There are plenty of new activities, in particular by the *national libraries and organizations* [18], by the *Zentralblatt Mathematik* [11], or by the IMU, that advances the vision of a *21ˢᵗ Century Global Library for Mathematics Research* (GDML) [2].

It is a great challenge to smaller scientific communities to adopt such developments for their own scientific communication processes and to join forces with other scientific communities to get own requirements publicly recognised. A first step in such a direction could be a more detailed description of ongoing scientific processes using standard RDF terminology.

With version 3 SYMBOLICDATA started to address the technical aspects of such cooperational needs in more detail, developed a vision of a *Computer Algebra Social Network* (CASN) [6], and started to realize it.

3 About our Talk

In this abstract we tried to explain the background of our project and in particular the data services available to the public. We operate these services on a stable basis, but the current focus of the project is on the CASN concepts and the difficult (social) task to get them running. In our presentation at ACA 2015 we will concentrate on a report about the current state of our efforts towards such a CASN. To prepare for our talk we invite interested people to look at our project wiki [14] and in particular at the survey paper [6].

We expect fruitful discussions and hope to convince more people that such a project is not only a "nice to have" but also deserves own *practical* efforts. The train is ready for departure. Don't miss it!

References

[1] The Agile Knowledge Engineering and Semantic Web Group at Leipzig University, `http://aksw.org/About.html` [2014-02-19].

[2] *Developing a 21st Century Global Library for Mathematics Research. Report of the Committee on Planning a Global Library of the Mathematical Sciences*, The National Academies Press (2014).

[3] The eScience Research Network Saxony, `http://www.escience-sachsen.de` [2014-02-19].

[4] J. Grabmeier, E. Kaltofen, V. Weispfenning (eds.), *Computer Algebra Handbook. Foundations – Applications – Systems*, Springer Verlag, Berlin (2003).

[5] J. Grabmeier, *Computeralgebra – eine Säule des Wissenschaftlichen Rechnens*, it + ti, **6:5**, p. 20 (1995).

[6] H.-G. Gräbe, S. Johanning, A. Nareike, *The SYMBOLICDATA Project – from Data Store to Computer Algebra Social Network*, Computeralgebra Rundbrief, vol. **55**, pp. 22–26 (2014), see also `http://symbolicdata.org/Papers/car-55.pdf` [2015-04-12].

[7] A. Heinle, V. Levandovskyy, *The SDEval Benchmarking Toolkit*, Communications in Computer Algebra, vol. **49.1**, pp. 1–10 (2015), see also `http://wiki.symbolicdata.org/SDEval` [2015-04-12].

[8] Citation from `http://de.wikipedia.org/wiki/Wissen` [2015-04-11], attributed to Hermann Weyl, *Philosophie der Mathematik und Naturwissenschaft*, R. Oldenbourg, Munich (1976).

[9] `http://www.wolfram.com/mathematica/` [2015-04-11]

[10] A. Nareike, *The SYMBOLICDATA sdsage package*, `http://wiki.symbolicdata.org/PolynomialSystems.Sage` [2015-04-12].

[11] U. Schöneberg, W. Sperber, *POS Tagging and its Applications for Mathematics*, in *Intelligent Computer Mathematics*, LNCS vol. 8543, pp. 213–223 (2014).

[12] Sociomateriality, see `http://en.wikipedia.org/wiki/Wanda_Orlikowski` [2015-04-12].

[13] Sage – a free open-source mathematics software system, `http://www.sagemath.org` [2014-02-19].

[14] The SYMBOLICDATA Project Wiki, `http://wiki.symbolicdata.org` [2015-04-12].

[15] J. Tauberer, *Quick Intro to RDF*, `http://www.rdfabout.com/quickintro.xpd` [2014-02-20].

[16] Technology, `http://en.wikipedia.org/wiki/Technology` [2015-04-11].

[17] DIN Deutsches Institut für Normung e.V., *VDI 3780 – Technology Assessment Concepts and Foundations*, VDI-Richtlinie, ed. 2000-09.

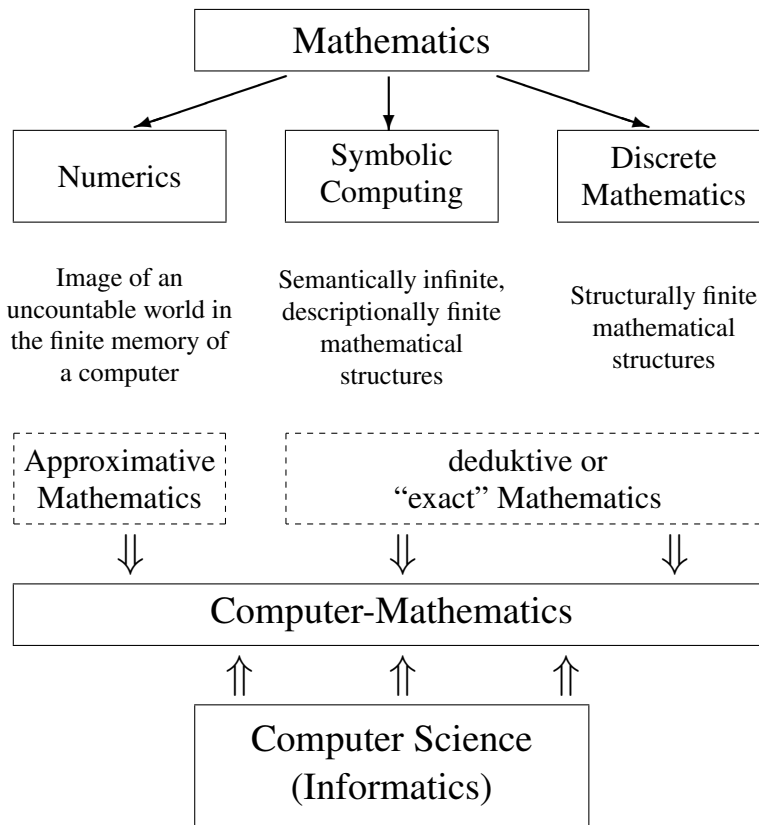[18] VIAF: Virtual International Authority File. `https://viaf.org/` [2015-04-12].

Figure 1: The Genesis of Computermathematics