20 Years SymbolicData

Hans-Gert Gräbe
Institute of Computer Science
Leipzig University
Leipzig, Germany
graebe@informatik.uni-leipzig.de

September 7, 2018

Abstract

The SymbolicData Project on testing an benchmarking Computer Algebra software grew up from the Special Session on Benchmarking at the 1998 ISSAC conference. During 20 years we collected reserach data and meta information, developed a test framework along the "cross cutting concerns" of modern software engineering and experimented with semantic technologies as a building block of a modern distributed socio-technical research infrastructure in the area of Computer Algebra. This paper presents a comprehensive survey of the most important motivations, concepts, steps, efforts and practical achievements of the SymbolicData Project to contribute to the formation of such a research infrastructure.

1 Introduction

At the OSCAR meeting[1] in Leipzig in December 2017 Bernd Sturmfels promoted a discussion "What to do with Big Old Data?"[2]. He addressed "a key problem in the development of an open source computer algebra system – the design of mechanisms and formats for dealing with big old data", i.e. "the output of a mathematical computation that is much larger than a few lines, and is intended for storage in a repository, or for further processing by a different program", and presented a list of 12 papers. In these papers authors from different areas of computational mathematics describe efforts, concepts and data stores to resolve for themselves or for a small computational mathematics subcommunity such a problem. As central goal each of these projects tries to establish from the very scratch a research infrastructure capable to present, access, inspect, exchange and maintain structured research data in both a powerful and sustainable way.

A similar situation was discussed 20 yeas ago at the *special session on Benchmarking*[3] at the 1998 ISSAC conference in Rostock. Heinz Kredel presented in his famous talk[4] also 12 examples of web resources that were set up and maintained by different people to keep track of the development concerning several of the challenges discussed at that time (Wester's CAS test suite, von zur

[1]https://www.mis.mpg.de/calendar/conferences/2017/oscar2017.html.

[2]https://www.mis.mpg.de/fileadmin/pdf/slides_oscar2017_3198.pdf.

[3]http://bwcloud-108-017.bwcloud.uni-mannheim.de/cafgbench.html.

[4]http://bwcloud-108-017.bwcloud.uni-mannheim.de/cabench/issac98/.

Gathen's polynomial factorization challenge, Montgomery's number field sieve factorizations or the SATLIB benchmark suite – to name some of them). Most of the links provided by Heinz Kredel are outdated for a long time, of course. So, didn't the world change during the past 20 years?

This paper presents a comprehensive survey of the most important motivations, concepts, steps, efforts and practical achievements of the SymbolicData Project to contribute to the formation of a reliable and sustainable research infrastructure in the area of Computer Algebra. For a more detailed explanation of different aspects of the motivations, concepts and achievements of the SymbolicData Project we refer to our publication list[5].

We start with a short discussion about the importance of research data in science in general, touch the distinction between core concerns and cross cutting concerns in modern software engineering and discuss consequences of the emerging role of semantic technologies for research infrastructures as a great difference between the situation today and 20 years ago.

A cornerstone of such concepts is the increasing importance of metadata that have to accompany research data to organize their maintainability, searchability and exchangeability. Within SymbolicData we developed the concept of *semantic-aware fingerprints* to emphasize that the definition of such metadata is a process of social coordination between the "suppliers" and the "customers" of research data to agree upon a *domain ontology* and not merely a question of just technically presenting the data.

Our experience indicates that the importance to organize the formation of a research infrastructure as a social process is underestimated. Whereas in small subcommunities of Computer Algebra research infrastructures usually grow and develop in a "natural way", it requires great efforts to organize such a social process on the level of inter-subcommunity communication within Computer Algebra. Such inter- and infradisciplinary engagement is little recognized as beneficial for the scientific career of productive young researcher. We developed the concept of a *Computer Algebra Social Network* (CASN) [4] based on modern semantic technologies and maintained a prototypical distributed implementation as a showcase over several years.

In the last two years we concentrated our resources on GitHub and terminated several unstaffed activities. In particular we moved the wiki to GitHub pages, stopped to update the CASN data base and canceled our announcement mailing list as a probably outdated way of communication. It is due to the next generation to evaluate the SymbolicData heritage and update the parts that are worth to be continued.

2 Research Data in Science

Research data plays an important role within science in at least four dimensions: 1) artifacts as problem sources (in particular digitized artifacts in the humanities), 2) benchmark examples as well established challenges for different problem classes, 3) raw output of research to be analyzed and evaluated within the scientific community and 4) publications and other consolidated scientific output as part of a common scientific *social* infrastructure.

Traditionally, the Computer Algebra community focuses on dimension 4, in particular on algorithms, implementations and software. The swMATH project [15] and several big national or EU funded projects as PoSSo [13], FRISCO [2], the SPP 1489, OpenDreamKit [10] or OSCAR [11] aimed and aim at contributing to the formation of a common Computer Algebra research

[5] `http://symbolicdata.github.io/Publications`.

infrastructure in that direction.

At the 1998 ISSAC conference in Rostock, at the end of the projects PoSSo and FRISCO, dimension 2 and 3 started to play a more important role. Gert-Martin Greuel reported in his invited lecture *Computeralgebra and Algebraic Geometry, Achievements and Perspectives* in particular about the progress in the Gröbner business as one of the core algorithmic building blocks in advanced polynomial systems solving and algebraic geometry. In 1993, the FGLM algorithm was published. But there were rumors about new algorithmic ideas of a young guy in Paris, and Greuel's Singular group in particular was interested in evaluating these ideas. Unfortunately, after the end of the PoSSo project there was no established reference for the famous *PoSSo test suite* and different printed versions with a number of misprints were in circulation. It started a disputation about the timings of examples, since it was neither clear what version of the PoSSo examples was referenced nor what "timing" really means (clock time, wall time, processor time etc.). Heinz Kredel proposed to start a *Computer Algebra Benchmarking Initiative* and invited for a *special session on Benchmarking* that took place on August 15, 1998, 18:00 in room 110 in the main building of the Rostock University. "The initiative will discuss, develop, define, collect all facets of this challenging problem. It should analyse and develop test suites but also define standard examples for the various topics of computer algebra where algorithm and systems developers can test their newly developed and improved methods. Furthermore, all kinds of test examples should be collected and consolidated. "[6] Heinz Kredel presented a *Computer Algebra Benchmarks Collection from July 1998* and shortly explained the results of the *CASBENCH Computer Algebra Benchmarks* activities, started in 1995 by the German Fachgruppe. The CASBENCH setup[7] was inspired by the Parkbench activities[8] on *Public International Benchmarks for Parallel Computers* (nowadays known as the annual HPC challenge award competition[9]) but got much less support from the community. Unfortunately, even within the ISSAC Special Session on Benchmarking the community could not agree upon a further roadmap or even a commonly accepted process or dedicated resources to advance that matter. Once more the diversity of Computer Algebra challenges prevented awareness of common interests in promoting the development of a common research infrastructure.

The 1998 special session on Benchmarking was the starting point of the SymbolicData Project. Olaf Bachmann and me developed and implemented with great support by other members of the Singular team the basic concepts and a first version of a benchmarking environment in the area of polynomial systems solving, called PolyData and later on SymbolicData. We collected a considerable number of the benchmark examples used at that time for testing polynomial systems solvers (in particular the PoSSo test suite), made it publicly and reliably available in a digital exchange format and developed a standardized environment based on GNU make and GNU time to run, time and monitor test computations on such examples using different solvers. But the main conceptual goal of SymbolicData was a nontechnical one to develop a research infrastructure that is independent of (permanent) project funding but operates based on overheads of its users. This approach was inspired by the rich experience of the Open Culture movement business models to run infrastructures. It was an early attempt to emphasize the advantage of an explicitly elaborated concept of a community-based solution to the tragedy of the commons within the Computer Algebra community and to apply such a concept to run as part of its research infrastructure.

[6] `http://bwcloud-108-017.bwcloud.uni-mannheim.de/cafgbench.html`.

[7] `http://bwcloud-108-017.bwcloud.uni-mannheim.de/cabench/casbench.html`.

[8] `http://www.netlib.org/parkbench/html/`.

[9] `http://www.hpcchallenge.org/`.

Nowadays the awareness of the importance of digital research infrastructure increased both in the scientific communities and also in research politics. The development of research infrastructures coordinated by the European Research Infrastructure Consortium (ERIC) plays a crucial role in the EU funding program HORIZON 2020. *OpenDreamKit* is such a project within Computer Algebra. Unfortunately it concentrates on (technical) interoperability rather than the research data problems discussed above. "OpenDreamKit is a project that brings together a range of projects and associate software to create and strengthen virtual research environments. The most widely used research environment is the Jupyter Notebook from which computational research and data processing can be directed. The OpenDreamKit project provides interfaces to well established research codes and tools so that they can be used seamlessly and combined from within a Jupyter Notebook."

A different approach is pursued by the *Leibniz Network MMS* on modelling and simulation[10] with great focus on research data and strong cooperation with TIB Hannover[11] and the upcoming community of Research Software Engineers[12].

Several activities concerning research data are on the way in the "big scene": *FORCE-11*[13] and the *Research Data Alliance*[14] are international interdisciplinary initiatives to promote and feature research data management and the development of digitally supported research infrastructures. Much is on the way also within mathematics – international initiatives were started to digitally organize the mathematical heritage as a whole (WDML[15], IMKT[16]) and of mathematical software in particular (swMATH [15], the INRIA Software Code Archive[17]), see [8] for a survey. SymbolicData provides 20 years of experience, in particular with semantic technologies, for these broader initiatives.

3 Testing and Benchmarking in Modern Software Engineering

Testing and benchmarking is a common task in software engineering. Modern software development concepts for enterprise middleware provide architectural and technical support (git workflows, virtualization tools) for agile approaches as continuous integration, continuous development and continuous deployment of modular software pieces in distributed environments. All this makes software development more complex and requires a good theoretical understanding of the corresponding architectural concepts. To maintain such diverging goals within a single software, modern software engineering distinguishes between core and cross cutting concerns. *Core concerns* define the main goal of the implementation that is mission critical or requires special domain specific knowledge, insight and experience. *Cross cutting concerns* (logging, testing, profiling, data management, security concerns, inter process communication etc.) are implemented using well established generic approaches. This eases both the maintenance of the software and the training of the developers. Frameworks as Spring or EJB realize the concept of *context aware programming*. The core concerns

[10] `https://www.wias-berlin.de/research/Leibniz-MMS/index.jsp?lang=en`.
[11] `https://www.tib.eu/de/`.
[12] `https://www.de-rse.org/de/index.html`.
[13] `https://www.force11.org/`.
[14] `https://www.rd-alliance.org/`.
[15] `https://www.mathunion.org/ceic/library/world-digital-mathematics-library-wdml`.
[16] `https://imkt.org`.
[17] `https://www.softwareheritage.org/`.

are implemented in a runtime that is embedded into another generic runtime – the *context* – that provides cross cutting objects to be injected as dependencies into the core runtime. Such an approach allows to concentrate combined long time efforts on the development of the commonly used context environment and to realize short term core business in a more efficient way.

Such approaches are nowadays common also within the big Computer Algebra software projects SageMath, OpenDreamKit and OSCAR. The SymbolicData benchmark activities designed by Olaf Bachmann 20 years ago anticipated several such concepts still unknown at that time. Since 2013 this framework was enhanced, advanced and consequently used in the working group of Viktor Levandovskyy at RWTH Aachen, This *SDEval*[18] *Testing and Benchmarking Environment* provides an easy way to generate executable code for benchmarks of computer algebra systems (like Singular, Magma etc.) on SymbolicData benchmark data and a framework for trustfully reproducing computation results from current research papers. SDEval is intensively used in particular in projects from TRR 195 for, e.g., finitely presented associative algebras. Its current developer team includes Karim Abou Zeid and Viktor Levandovskyy; the beginnings were laid by Albert Heinle and Benjamin Schnitzler. A list of benchmark results created using SDEval and used in published papers, can be found at [7]. There is also a video tutorial/introduction[19] for SDEval on Youtube. It covers the main functionality of the provided scripts.

Andreas Nareike implemented within a project funded by the Saxonian E-Science Initiative *SDSage*[20] – a module for the SageMath [14] generic environment to access the SymbolicData database as injected dependency object. We don't know to what extend this embedding is used by the SageMath community. The documentation refers[21] only to an earlier implementation provided by Martin Albrecht.

4 Semantic Technologies

With the consolidation of concepts as Open Access, Open Data and the emerging semantic web the general understanding of the importance of community-based efforts to develop common research infrastructures matured. This development was accompanied with conceptual, technological and architectural standardization processes that had also impact on the development of concepts and data structures within the SymbolicData Project. In 2009 we started to refactor the data along standard Semantic Web concepts based on the Resource Description Framework (RDF). With a new SymbolicData version released in September 2013 we completed the redesign of the data along RDF based semantic technologies, set up a Virtuoso based RDF triple store and an SPARQL endpoint as Open Data services along Linked Data standards[22]. The importance of the yet heavily growing Linked Open Data Cloud[23] is hardly to underestimate.

[18] https://symbolicdata.github.io/SDEval.

[19] https://www.youtube.com/watch?v=CctmrfisZso.

[20] https://symbolicdata.github.io/PolynomialSystems.Sage.

[21] http://doc.sagemath.org/html/en/reference/databases/sage/databases/symbolic_data.html.

[22] https://en.wikipedia.org/wiki/Linked_data.

[23] http://lod-cloud.net.

5 Semantic Aware Fingerprints

The main goal of SymbolicData is on data – structure, maintain and present research data in a digitally and publicly available way. 20 years ago we started with examples from polynomial systems challenges, with the PoSSo test suite and other sources. To make such a collection searchable one has to define and compile meta information about the different objects to cluster them or even identify a single one. This is challenging in particular for polynomial systems since the same example can be noted with different variable names and different term orders. Hence a pure string matching doesn't work. As a first approach we compiled invariants of polynomial normal forms and stored it together with the basis itself in a single information object. Such a strategy – to combine data and metadata in a single object – is commonly used also nowadays, e.g., within the LOM standard – the Learning Objects Metadata are tightly coupled with the Learning Objects themselves.

For our use case such a concept turned out to be suboptimal since it led to an explosion of data: polynomial systems can be interpreted in different ways, e.g., keeping a part of the variables as parameters, as homogenized ideals, bounding variables to special values etc.. Each such version had to be kept as a new data object since the metadata changed even if the basis was the same or could be easily (i.e., in polynomial time) generated from another example. In later versions we used the universal property of the ring $\mathbb{Z}[x_1, \ldots, x_n]$, decided to reduce the number of stored systems and keep track of the way how derived systems are generated from basic ones.

RDF strongly supports such a distinction between data (*resources* in the RDF terminology) and meta information (*resource descriptions*). Data is represented by URI's that can point even to remote locations. Hence RDF is well suited to describe also distributed research data even if the data is maintained by different stakeholders and only the metadata is federated in a common RDF store for search and data analytics.

SymbolicData operates such a central RDF store [17] and Andreas Nareike enhanced our metadata during his e-science project funded in 2012–2013 with metadata from two distinguished sources – the polytopes database of Andreas Paffenholz [12] and the transitive groups Database for Number Fields of Gunter Malle and Jürgen Klüners [9]. A central challenge was the definition of the metadata as a social process that requires not only sufficient domain specific insight but also resilient agreements about responsibilities to update and maintain the data and metadata. The first part of this challenge is reflected in our concept of *semantic-aware fingerprints* that focuses on the usage condition of *any* research data (awareness of the semantics) and our special use case for the meta information – search. The needle in the haystack, the fingerprint in the police file or the puzzle piece in the stack: To find it a clear domain model and a clear search strategy are required, and both are not independent from each other. For more details we refer to [5].

After a certain consolidation process on March 1, 2016, version 3.1 of the SymbolicData tools and data was released. The new release contained new resource descriptions ("fingerprints") of remotely available data on transitive groups (*Database for Number Fields* of Gunter Malle and Jürgen Klüners [9]) and polytopes (databases of Andreas Paffenholz [12] within the *polymake* project [3]), a recompiled and extended version of test sets from integer programming – work by Tim Römer (*normaliz* group [1]) –, an extended version of the *SDEval benchmarking environment* – work by Albert Heinle, Benjamin Schnitzler and Viktor Levandovskyy [6] – and a partial integration (SymbolicData People database, databases of upcoming and past conferences) of data from the CASN – the Computer Algebra Social Network subproject. Furthermore, our GitHub account[24]

[24] `https://github.com/symbolicdata`.

was transformed into an organizational account and the git repository structure was redesigned better to reflect the special life-cycle requirements of the different parts of our activities.

6 Research Infrastructure as a Social Project

So far we mainly discussed technical questions of structuring data, defining and compiling metadata and designing tools and workflows for local testing and benchmarking activities. But benchmarking – as *any* process of scientific evaluation – is primarily a *social* process. In other areas of science there are well established benchmark competitions for different algorithmic problem classes with clearly defined rules and places.

In 2012 we organized a workshop on benchmarking[25] with people from communities close to Computer Algebra. Satya Samal presented the PoCaB Project – Platform of Chemical and Biological Analysis Using Computer Algebra Methods – and explained in detail structural approaches within the PoCaB Databases and how data is generated within the PoCaB framework. PoCaB is interlinked with different communities within Computer Algebra (the polynomial systems solving and the polymake communities) and also beyond. It heavily exploits biological databases (BioModel Database, KEGG Database) that come with their own language SMBL and experiences how to express semantic aspects in a computer readable way. This example showed very clearly that communities are not interested in advice from outside how to reinvent wheels properly running for a long time within the community but acknowledge support and advice to organise intercommunity communication more smoothly in a world of evolving Linked Open Data standards.

Johannes Waldmann gave a talk about Benchmarks and Competitions in Theoretical Computer Science presenting best practices of three TCS Communities: Termination, SAT and SMT. For Termination he explained TPBD – the Termination Problems Data Base – and their way of benchmarking: They regularly organize termination competitions on previously agreed data from different problem categories in a similar way as the Formula I car race is organized: Upload tools to a single dedicated server that runs all tools on all problems and collects the results in aggregated form on a web page. Usually such a competition runs accompanying the annual large conference in the field. Similarly structured competitions take place in other areas of science, e.g., in High Performance Computing[26] or in the SAT Solver community[27].

The 1998 Special Session on Benchmarking stated that such contests with clear rules are lacking in the area of Computer Algebra. This did not change during the last 20 years. Evaluating the reason for such a longstanding deficit we observed that socially mounted benchmarking cultures live in certain Computer Algebra subcommunities but are rarely communicated beyond their scope. So what about communication between Computer Algebra subcommunities in general? RDF concepts are well suited not only to describe collections of benchmark data but also to support communication on other scientific activities and achievements between different subcommunities. Properly organized metadata generated by different stakeholders can easily be collected not only in a central store but also in a well organized distributed environment as a "scientific Facebook" – we called such a concept *Computer Algebra Social Network* (CASN) [4] – that could be implemented as a network of *CASN nodes* as part of a social research infrastructure within the Linked Open Data Cloud.

[25] `https://symbolicdata.github.io/Events.2012-12`.
[26] `http://www.hpcchallenge.org`, discontinued after 2014.
[27] `http://www.satcompetition.org/`.

Since 2012 we tried to identify problem settings of common interest, implemented building blocks of such a network, tried to get showcases socially running and promoted our CASN idea. We report shortly about three of these showcases and refer to our wiki[28] for more details.

Conferences in Computer Algebra. Reporting about upcoming conferences seems to be a common need in many Computer Algebra subcommunities and could be a first class service of a CASN. The German Fachgruppe set up such a service for a long time in printed form within their Rundbrief. Upcoming conferences are listed independently on the websites of both SIGSAM and the German Fachgruppe. We maintained for several years such information about upcoming and (archiving the entries) past conferences in a structured RDF format that can be used to extract the different web and printed views from a single commonly maintained source. Defining such an exchange format the entries can even be produced by the subcommunities and the boards have merely to collect the information. We terminated that service due to limited staff capacity. A presentation of our past conferences collection can be found at our SymbolicData demonstration site [16].

RDF is well suited to combine such conference announcements with more detailed information about the conferences (tracks and sessions, papers and authors etc.) that is compiled anyway, e.g., for the web presentation of the conference. In many cases such information is already stored in a structured way and the web site of the conference is generated from that source. In particular, Serge Autexier as the publicity chair invented such a model for the CICM conferences[29] and compiled all information of each of the 12 conferences in publicly available XML files, thus arriving at level 3 of the 5 stars scale[30] for Open Data of Tim Berners-Lee.

As a showcase we transformed four of these presentations into RDF and stored it in our CASN node[31]. This is level 4 of the 5 stars scale since the data is available as RDF but not operated within a RDF store and thus not directly accessible for SPARQL query exploration. This could be part of an upcoming conference reporting structure within an emerging CASN.

The SymbolicData People Database. Conference announcements are a first class resource of information about people actively working in Computer Algebra. We attached to our conference records information about organizers, invited speakers, program committees etc. We have more than 1000 entries in our database and joint forces[32] with Wolfram Sperber and Uwe Schöneberg (Zentralblatt) to solve the problem of identification of those people in the Zentralblatt and partly also in the MathReviews. A presentation of this database can be found at our SymbolicData demonstration site [16].

Such a People Database maintains a set of established URI's and thus is a central building block to get activities in Computer Algebra recognized within the Linked Open Data world. It allows to embed the "stories" told within Computer Algebra and its subcommunities into a bigger world, to join forces with the author disambiguation projects of "big players" (Zentralblatt, Math Reviews, ACM digital library, Springer, Elsevier, ORCID, ResearchGate, VIAF, GND) and thus actively to promote the visibility of Computer Algebra research in the emerging digital world.

[28] `https://symbolicdata.github.io/CASN`.
[29] `https://www.cicm-conference.org/cicm.php`.
[30] `https://5stardata.info/de/`.
[31] `http://symbolicdata.org/rdf`.
[32] See `https://symbolicdata.github.io/Events.2014-07` for details.

Computer Algebra Software. Another central problem within benchmarking Computer Algebra software is software disambiguation. SymbolicData started 20 years ago to maintain a consolidated list of Computer Algebra software. With the maturing swMATH project [15] we stopped in 2012 such activities and compiled together with Wolfram Sperber and Hagen Chrapary (Zentralblatt) a translation list between our URI's and those of swMATH. In the last years there was much discussion (What is a software, what a package? How to deal with libraries or different versions of the same software?) but little practical progress to prepare that collection for the Linked Open Data world of the 21st century. Being a first class reference of mathematical software swMATH achieves only 2 of the 5 stars of Tim Berners-Lee since it has no open interface to the data itself.

As a showcase we compiled in a common effort with Wolfram Sperber a consolidated RDF based version of Computer Algebra software (that is only a *part* of swMATH, since swMATH addresses mathematical software in general) combining URI's and descriptions from swMATH, the SIGSAM list of Computer Algebra software[33] and also the (very outdated) overview[34] on the website of the German Fachgruppe. We used an undocumented feature of swMATH to compile also links to Zentralblatt reviews of 10 papers related to that software. Since the data is also available from our RDF store it earns all 5 stars of Tim Berners-Lee. A presentation of this database can be found at our SymbolicData demonstration site [16].

7 SymbolicData as Non-Project

A project is usually defined by a *goal*, *attached resources* (money, web space, human resources) and a *time span* (as basis for planning, work packages, milestones etc.). 20 years ago SymbolicData grew up from the relicts remaining after the end of two such projects – PoSSo and FRISCO – and was designed from the very beginning as *non-project* – it was driven by casual volunteers, bringing in their own resources (time, web space), it was partly supported by different community structures (the Singular group, UMS Medicis, the German Fachgruppe) and it had never a defined project end but survived several "dry periods" almost without activities.

Such a situation is typical for research infrastructures and it is hard to allocate resources for such non-projects in a time of increasing importance of project-oriented research funding. The problems and workarounds are described on the pages of the OEIS Foundation as the goals of another old (since 1964) research infrastructure non-project – *The Online Encyclopedia of Integer Sequences* – in the following way: "1) own the intellectual property, 2) maintain the infrastructure as a service that is freely accessible by the general public, 3) act so as to maintain its own existence indefinitely, 4) collect and distribute funds in order to carry out the first three goals."[35]

During the last years the SymbolicData team spent efforts on goal 3 to prepare for another "dry period" since we didn't succeed with goal 4[36]. We concentrated the SymbolicData data and wiki at our GitHub account and terminated several of our ongoing activities (updating the record of upcoming conferences, advancing the alignment with swMATH or the dissertations project[37]).

The domain `http://symbolicdata.org` as a prefix of the SymbolicData ontologies is one

[33] `https://www.sigsam.org/Resources/Software.html`.
[34] `http://www.fachgruppe-computeralgebra.de/systeme/`.
[35] `http://oeisf.org/\#GOALS`.
[36] For details we refer to `https://symbolicdata.github.io/New.html`.
[37] `https://symbolicdata.github.io/Dissertations`.

of the core semantic web facilities of SymbolicData. By the RDF best practices it is of great importance to own that domain and to set up and operate an RDF store under that web address. This domain is owned and sponsored by the German Fachgruppe since 2005 and currently operated on a server at Leipzig University. Unfortunately, the current board of the German Fachgruppe doesn't understand well enough the importance to keep such an arrangement running "indefinitely" (private communication with Gregor Kemper).

8 What Else?

We acknowledge the strong support from the Board of the German Fachgruppe over many years who sponsors the domain `symbolicdata.org` since 2005 and was the power partner in our experiments towards a CASN. During the last years (2012–2017) we presented SymbolicData at several international conferences and submitted 6 papers for publication (2 accepted, 4 rejected[38]), not counting our contributions to the Rundbrief of the German Fachgruppe.

Stephen Watt asked in the discussion to my presentation in the *Work in Progress* session at CICM 2014 "How will you sustainably attract resources for your project?" In my response I shortly explained our non-project philosophy, the role of casual volunteers and ended with the famous answer of Linus Torvalds on a similar question posed by Andrew Tannenbaum: "I won't." But time certainly changed, nowadays there is a big competition between projects resting on such "casual volunteers" and one has to spent much time in advertising the own projects.

We did so and tried to align SymbolicData not only with swMATH but also with other big community projects as OpenDreamKit (Michael Kohlhase), OSCAR (Wolfram Decker), SIGSAM (Ilias Kotsireas, Matthew England) or people showing interest in "big old data" (Bernd Sturmfels). Such advertisement could only be done with very restricted resources since the single volunteer actively developing SymbolicData at the moment is a specialist on semantic technologies but far away from core Computer Algebra for many years. The results were disappointing. Even the German Fachgruppe stopped with the relaunch of their website its direct cooperation with SymbolicData and moved the pages[39] with input from the CASN node of the German Fachgruppe[40] into the background.

A great number of people (Gert-Martin Greuel, Gerhard Pfister, Winfried Neun, Wolfram Sperber, Hannes Schnemann, me) involved in one way or another with SymbolicData already retired or will retire during the next years. Other people (Olaf Bachmann, Ralf Hemmecke, Andreas Nareike, Albert Heinle) timely involved in SymbolicData left Computer Algebra or are inactive with the project at the moment.

In this paper we described the main achievements and conceptual points of the SymbolicData project so far. It is up to the next generation to take over the baton, to evaluate the SymbolicData heritage and update the parts that are worth to be continued. If any.

9 Acknowledgement

We are grateful to the German Fachgruppe to help us reach a wider audience by additionally publishing this article in their Rundbrief.

[38]For details we refer to `https://symbolicdata.github.io/Publications.html`.
[39]See the overview at `http://www.fachgruppe-computeralgebra.de/symbolicdata/`.
[40]`http://www.fachgruppe-computeralgebra.de/rdf/`.

References

[1] Winfried Bruns, Bogdan Ichim, Tim Römer, Richard Sieg, Christof Söger: Normaliz. Algorithms for Rational Cones and Affine Monoids. https://www.normaliz.uni-osnabrueck.de. [2018-09-02]

[2] FRISCO – A Framework for Integrated Symbolic/Numeric Computation. (1996–1999). https://cordis.europa.eu/project/rcn/31471_de.html. [2018-09-02]

[3] Ewgenij Gawrilow, Michael Joswig: Polymake: a Framework for Analyzing Convex Polytopes. In: Gil Kalai, Günter M. Ziegler (eds.), Polytopes – Combinatorics and Computation (Oberwolfach, 1997), pp. 43–73, DMV Sem., 29, Birkhäuser, Basel (2000).

[4] Hans-Gert Gräbe, Simon Johanning, Andreas Nareike: The SymbolicData Project – Towards a Computer Algebra Social Network. In: Workshop and Work in Progress Papers at CICM 2014, CEUR-WS.org, vol. 1186 (2014).

[5] Hans-Gert Gräbe. Semantic-aware Fingerprints of Symbolic Research Data. In Gert-Martin Greuel, Thorsten Koch, Peter Paule, Andrew Sommese (eds.). *Mathematical Software – ICMS 2016*. Volume 9725 of Lecture Notes in Computer Science, page 411–418, 2016. DOI 10.1007/978-3-319-42432-3.

[6] Albert Heinle, Viktor Levandovskyy: The SDEval Benchmarking Toolkit. ACM Communications in Computer Algebra, vol. 49.1, pp. 1–10 (2015).

[7] Albert Heinle: Benchmarks created using SDEval. https://cs.uwaterloo.ca/~aheinle/software_projects.html [2018-09-07]

[8] Albert Heinle, Wolfram Koepf, Wolfram Sperber. Some steps to improve software information. Computeralgebra-Rundbrief 60 (March 2017) *and* Communications in Computer Algebra 51.1 (March 2017), pp. 1–11.

[9] Jürgen Klüners, Gunter Malle: A Database for Number Fields. http://galoisdb.math.uni-paderborn.de/. [2018-09-02]

[10] OpenDreamKit: Open Digital Research Environment Toolkit for the Advancement of Mathematics. http://opendreamkit.org/. [2018-09-01]

[11] OSCAR: The OSCAR project. https://oscar.computeralgebra.de/. [2018-09-01]

[12] Andreas Paffenholz: Polytope Database. http://www.mathematik.tu-darmstadt.de/~paffenholz/data/. [2018-09-02]

[13] The PoSSo Project. Polynomial Systems Solving – ESPRIT III BRA 6846. (1992–1995). https://cordis.europa.eu/project/rcn/9106_en.html. [2018-09-02]

[14] The SageMath Project. http://www.sagemath.org/. [2016-03-16]

[15] swMATH – an Information Service for Mathematical Software. http://swmath.org. [2018-09-02]

[16] The SymbolicData Demonstration site. http://symbolicdata.org/info. [2018-09-02]

[17] The SymbolicData RDF Data Store. http://symbolicdata.org/Data. [2018-09-02]