

# Building a Semantically Linked Data Web for Symbolic Computation Communities

Hans-Gert Gräbe – Andreas Nareike

University of Leipzig, Germany

eScience Network Conference

June 11, 2013

Dresden

# What is Symbolic Computation?

- area of mathematics concerned with exact (as opposed to numerical) solutions, often in form of mathematical expressions
- study and classification of different mathematical structures (semigroups, polytopes, transitive groups, number fields)
- number theory and cryptography
- development of algorithms to solve various algebraic problems, e.g.
  - calculation of Gröbner bases
  - automatic differentiation
- evaluating the performance of algorithms (benchmarking)

# What is Benchmarking?

**Benchmarking** is the evaluation of algorithms on a given problem with respect to

- memory consumption
- running time

Problems with benchmarking in Symbolic Calculation:

- no common repository
- problems often vary unintentionally due to transcribing mistakes
- results are unreliable for the community

There are *a lot* Computer Algebra Systems, each of which with their own file format and syntax.

Produced data is stored in databases (more or less):

- some are just resource files in a ZIP archive
- some are only available as a package for a specific Computer Algebra System
- some (few) are available as a relational database

# Why Change it?

- There are 'fundamental' databases that are interesting to many people working in Symbolic Computation.
- There are often links between different fields of Computer Algebra and thus between databases.
- Existing databases often only allow questions that are interesting from one point of view.

# SymbolicData: Where we Started

- **main incentive:** revise an unstructured collection of polynomial systems from different areas
- **tools:** Semantic Web technologies
- **goals:** inclusion of more Computer Algebra related data

- Which format should be used for **resources**, e.g. data of polynomial systems?
- Concerning **metadata**, i.e. information *about* the resources, questions are:
  - how to store
  - which information to store

- Resources are stored in XML as an exchange format, or in any format convenient.
- Metadata is stored in a RDF/OWL with a SPARQL endpoint.



- With RDF, OWL and SPARQL it is possible to advance beyond single databases towards a **Semantically Linked Data Web** of Symbolic Computation.
- SPARQL allows to pose very detailed questions while providing a much broader scope of possible question. This enables mathematicians to use a database in new ways and discover new results.

# SymbolicData as a Base for Further Development

- SDEval is a software system that transforms SymbolicData resources to various Computer Algebra Systems and runs multiple benchmarks (possibly on a remote server).
- There is a SymbolicData package for Sage, which allows to load and transform resources from the SymbolicData repository.

Not only resources and metadata about them is interesting, but also

- in which paper are certain resources are studied
- which papers are related to a given problem (where 'related' will vary)
- which researchers are working on a given problem

To further help researchers

- collection algorithm and resource management
- collection of best practices in Computer Science to help developers

# The Bigger Picture (Regarding Symbolic Computation)

- There are already attempts to unify researches in Symbolic Computation, mostly on the software side (Sage, swMath).
- We are going to forge **stronger connections**, i.e. we will use a Sage server behind Symbolic Data to perform calculations and transformations as needed.

# The Bigger Picture (Regarding Semantic Web)

- There are already techniques for the transformation of 'classical' SQL data into RDF, developed by the AKSW (Agile Knowledge Engineering and Semantic Web, [aksw.org](http://aksw.org)) in Leipzig, e.g.:
  - SPARQLify: a SPARQL  $\rightarrow$  SQL translator
  - Triplify: providing plugins for database-backed Web applications in order to expose semantics as RDF
- These tools are also interesting for people in Symbolic Computation who do not want to fully convert to RDF but want some of the benefits regardless.

# Ways of Using and Contributing to SymbolicData

- actively generating RDF metadata, official repository:  
`https://github.com/symbolicdata/symbolicdata`
- learning (only) SPARQL to present metadata
- including metadata from other sources (e.g. a MySQL database) into SymbolicData

**PoCab** ([pocab.cg.cs.uni-bonn.de](http://pocab.cg.cs.uni-bonn.de)) is a collection of models coming from the field of biology and chemistry. They examine different algebraic entities given in those models and apply algebraic methods. Their data of interest is coming from two renowned and publicly available databases, namely KEGG and the BIOMODELS DATABASE.

## The **DFG Priority Program SPP 1489**

([www.computeralgebra.de](http://www.computeralgebra.de)) aims to considerably further the algorithmic and experimental methods in Symbolic Computation and to combine the different methods where needed.

There are interesting databases of different accessibility (from Postgres databases to ZIP archives), some of which are already accessible via Symbolic Data, some that we will work on in the next weeks.



Thanks for your attention.