

### Neues vom SYMBOLICDATA-Projekt

**Hans-Gert Gräbe**  
(Leipzig)

graebe@informatik.uni-leipzig.de



---

### Zur Geschichte des Projekts

---

Die Idee zum SYMBOLICDATA-Projekt entstand 1998 am Rande der ISSAC in Rostock, als zunehmend sichtbar wurde, dass die Datensammlungen, welche im Rahmen der Projekte PoSSo<sup>1</sup> und FRISCO<sup>2</sup> entstanden waren, ohne aktives Zutun der Community nicht dauerhaft digital verfügbar sein würden.

In den folgenden Jahren wurde dieses Projekt vor allem von Olaf Bachmann und dem Autor dieses Beitrags mit weiterer Unterstützung aus dem Umfeld der Singular-Gruppe vorangetrieben. Im CAR 26 (März 2000) und CAR 28 (März 2001) wurde das SYMBOLICDATA-Projekt im damaligen Zuschnitt ausführlich vorgestellt. Der Schwerpunkt der Datensammlung lag im Bereich polynomialer Gleichungssysteme mit ganzzahligen Koeffizienten, die zu jener Zeit insbesondere im Zusammenhang mit Varianten des Buchberger-Algorithmus intensiv studiert wurden, wobei die verschiedenen Benchmark-Beispiele aus der Literatur und den genannten EU-Projekten systematisiert und digital aufbereitet sowie entsprechende Werkzeuge zur Verwaltung und zum Aufsetzen von Benchmarks entwickelt wurden.

Nach dem Weggang von O. Bachmann aus Kaiserslautern Ende 2000 wurde mit dem geometrischen Theorembeweisen ein zweites größeres Sammelgebiet erschlossen, das über die Koordinatenmethode eng mit polynomialen Gleichungssystemen verbunden ist. Einfache Beispiele lassen sich auf geometrische Sätze *vom konstruktiven Typ* zurückführen, deren Beweis auf eine „einfache“ Normalformberechnung rationaler Ausdrücke zurückgeführt werden kann. Obwohl solche Normalformberechnungen im Prinzip verstanden sind, laufen sie aber oft praktisch nicht durch, während alternative Koordinatisierungen *vom Gleichungstyp* und ge-

schickte Ansätze des Lösens der entsprechenden polynomialen Gleichungssysteme zum Ziel führen. M. Witte, damals Student in Leipzig, untersuchte und digitalisierte insbesondere eine größere Anzahl von Beispielen aus dem Buch von S.-C. Chou<sup>3</sup>.

Weiteres Beispielmaterial wurde von der CoCoA-Gruppe (F. Cioffi), der Singular-Gruppe (M. Dengel, M. Brickenstein, S. Steidel, M. Wenk), J.-C. Faugère (weitere Polynomsysteme), V. Levandovskyy (nicht-kommutative Polynomsysteme, G-Algebren) sowie von Raymond Hemmecke (Testmengen aus der ganzzahligen Programmierung) beige-steuert. Die Projektdaten wurden über ein CVS-Repository verwaltet. Die Domäne `symbolicdata.org` wurde und wird durch die CA-Fachgruppe finanziert und seit 2005 von der GI-Geschäftsstelle gehostet.

Leider standen ab 2003 keine personellen Ressourcen zur systematischen weiteren Arbeit am SYMBOLICDATA-Projekt zur Verfügung, so dass das Datenmaterial seither allein auf einer „as is“ Basis zur Verfügung stand und auch nur gelegentliche Erweiterung erfahren hat. Insbesondere stellte sich heraus, dass die ambitionierten Vorstellungen, eine *einheitliche Werkbank für Benchmarking im symbolischen Rechnen* mit entsprechenden Werkzeugen und Konzepten zu entwickeln, mit den verfügbaren Ressourcen nicht zu stemmen war und – mit Blick auf die Vielfalt der Rechnerarchitekturen, Problemstellungen und verwendeten Programmiersprachen – wohl auch am realen Bedarf vorbei ging. Statt dessen konzentriert sich dieser Teil des Projekts nunmehr darauf, den Code von Best Practice Beispielen zu sammeln und zur Nachnutzung zur Verfügung zu stellen.

Ein weiterer Versuch, die Arbeiten am SYMBOLICDATA-Projekt mit vereinten Anstrengungen zu intensivieren und mit angrenzenden Bemühungen in dieser Richtung, etwa der Datenbank *Gröbner Bases Imple-*

<sup>1</sup><http://posso.dm.unipi.it>

<sup>2</sup><http://www.nag.co.uk/projects/frisco.html>

<sup>3</sup>S.-C. Chou. *Mechanical Geometry Theorem Proving*. Reidel, Dordrecht, 1988.

<sup>4</sup><http://www.risc.jku.at/Groebner-Bases-Implementations>

mentations. *Functionality Check and Comparison*<sup>4</sup> oder dem *Gröbner Bases Bibliography Project*<sup>5</sup> zu koordinieren, wurde im März 2006 während des *Special Semester on Groebner Bases* am RISC Linz gestartet. Jedoch fand auch dieses Projekt im Nachgang keine nachhaltige Resonanz in der Community.

---

## Entwicklungen nach 2003 – SYMBOLICDATA meets RDF

---

In den letzten 10 Jahren nahm das (semantische) Web der Daten eine besonders stürmische Entwicklung, was die konzeptionellen Entwicklungen zur Datenhaltung aus den Anfangszeiten des SYMBOLICDATA-Projekts entwertet hat, über die seinerzeit in den entsprechenden Aufsätzen im CAR berichtet wurde. Mit dem *Semantic Stack* haben sich inzwischen Standards durchgesetzt und stehen Werkzeuge zur Verfügung, mit denen sich Verwaltungsaufgaben auf verschiedenen Ebenen nach einheitlichen Prinzipien organisieren lassen. Dies sind

- XML und XSchema als Standards zur Darstellung strukturierter Daten auf Ressourcen-Ebene sowie
- RDF und OWL als Standards zur Beschreibung semantischer Aspekte der Daten.

Während ersteres prinzipiell geeignet ist, dezentral betreute Datenbestände in einem SYMBOLICDATA-Netz zusammenzuführen und durch geeignete URL's zu adressieren (die Frage der *nachhaltigen* Verfügbarkeit derartiger dezentral betreuter Datenbestände sei hier ausgeklammert), ist die Aggregation der Beschreibungen semantischer Aspekte unbedingt erforderlich, um sinnvolle Suchprozesse nach Informationen auf diesen vernetzten Beschreibungen zu ermöglichen. Während wir in der Anfangszeit davon ausgingen, dass dazu ein wesentlicher Teil der Informationen auch an einem Ort zusammengetragen sein muss – etwa als Datenbestand in unserem CVS-Repository – zeigen modernere Entwicklungen der Linked Open Data Cloud, in welchem Umfang auch diese Informationen dezentral verwaltet werden können, wenn entsprechende Protokolle des Datenaustauschs vereinbart werden. Genauer geht es nicht um die Protokolle selbst, die als *Metasprache* mit dem RDF-Framework und dessen Erweiterungen weitgehend standardisiert sind, sondern um die gemeinsam zu vereinbarende Ontologie als *domänenspezifische Sprache*, mit der RDF-Tripel der Wissensbasis auf einheitliche Weise „verstanden“ werden.

Die wenigen verfügbaren Ressourcen der letzten Jahre wurden darauf verwendet, einen Migrationsprozess der Datenbasis hin zu diesen Standards in die Wege zu leiten, also

- die Benchmark-Daten aus den verschiedenen Bereichen auf je standardisierte Weise in XSchema-beschriebene XML-Ressourcen zu transformieren,
- die im ursprünglichen SYMBOLICDATA-Konzept eng mit den Ressourcen gekoppelten Beschreibungsdaten zu separieren und über RDF-Standards (insbesondere einen SPARQL-Endpunkt<sup>6</sup>) für Suchanfragen verfügbar zu machen sowie
- eine SYMBOLICDATA-Ontologie als domänenspezifische vom Computer lesbare Sprache für diese Beschreibungsdaten zu entwickeln.

Die Erfahrungen aus vergleichbaren Projekten der AKSW-Gruppe<sup>7</sup> an unserem Lehrstuhl legten nahe, dass dieser Migrationsprozess am besten als agiler Prozess anzulegen ist, also das Prinzip des „a little bit thinking, a little bit coding“ aus der Anfangszeit des SYMBOLICDATA-Projekts auch hierbei anwendbar ist, um mit kleinen Schritten die sehr begrenzten Ressourcen einzusetzen. Insbesondere lehren diese Beispiele, dass es – im deutlichen Gegensatz zu den Erfahrungen der ER-Modellierung – nicht sinnvoll ist, bereits am Anfang zu viel Mühe auf die genaue Fassung der SYMBOLICDATA-Ontologie zu verwenden. Hierbei wird der Vorteil von RDF gegenüber herkömmlichen Datenbankansätzen deutlich, denn Prädikate lassen sich auf einfache Weise in der Wissensbasis selbst hinzufügen oder redefinieren, was bei Datenbanken ein aufwändiges Redesign der Tabellenstruktur erfordert. Die Konsolidierung der Ontologie selbst kann so auf einen späteren Zeitpunkt verschoben werden, wenn eine ausreichende Zahl von „Sätzen“ der zu definierenden Sprache bereits vorliegen.

Im Repository sind größere Teile der Daten (insbesondere ein wesentlicher Teil der Polynomsysteme) inzwischen in XSchema-beschriebene XML-Ressourcen transformiert. Etwas schwieriger gestaltet sich der Transformationsprozess bei den Geometrie-Beispielen, da dort zusätzlich die Ähnlichkeiten zu den Beschreibungssprachen verschiedener DGS, also neue semantische Aspekte, berücksichtigt werden sollen.

Anfang 2011 wurde die Projektverwaltung<sup>8</sup> auf Mercurial umgestellt, da so Entwicklungszweige deutlich flexibler angelegt und wieder mit dem Hauptstamm vereinigt werden können. In einem solchen Zweig<sup>9</sup> wurden weitere Beispiele zu freien Algebren zusammengetragen.

---

<sup>5</sup><http://www.risc.jku.at/Groebner-Bases-Bibliography>

<sup>6</sup>Provisorisch unter <http://hgg.ontowiki.net> verfügbar.

<sup>7</sup><http://wiki.aksw.org/About>

<sup>8</sup>Siehe <https://bitbucket.org/hgg/symbolicdata>

<sup>9</sup>Siehe [https://bitbucket.org/levandov/symbolicdata\\_freealgebra](https://bitbucket.org/levandov/symbolicdata_freealgebra)

---

## **Das Benchmarking-Projekt im Rahmen der Sächsischen E-Science-Initiative**

---

Im Rahmen der Sächsischen E-Science-Initiative<sup>10</sup> wurde ein Projekt „Benchmarking im symbolischen Rechnen“ bewilligt, über das ab September 2012 für 12 Monate eine halbe Mitarbeiterstelle finanziert und mit Andreas Nareike besetzt ist. Damit besteht die Möglichkeit, die bisherigen Migrationen zu konsolidieren und dem SYMBOLICDATA-Projekt, wenigstens für eine gewisse Zeit, neuen Schwung zu verleihen.

Wir möchten dabei weitere interessierte Partner und Datenbestände einbeziehen und insbesondere die Beschreibungstechniken auf RDF-Ebene weiter qualifizie-

ren. Zur Intensivierung der Kommunikation haben wir dazu neben der schon länger existierenden Mailingliste `symbolicdata` bei Google Groups ein Mediawiki<sup>11</sup> an der Uni Leipzig aufgesetzt, in das die bisherigen Inhalte der Webseiten sowie weitere Informationen migriert werden sollen.

Während eines

### **Arbeitstreffens und Workshop am 13.–15. Dezember 2012 in Leipzig**

sollen die bis dahin erreichten Ergebnisse vorgestellt und das weitere Vorgehen abgestimmt werden. Dazu laden wir interessierte Mitsreiter ein. Der Workshop ist als „Zero Budget Workshop“ geplant und wird voraussichtlich an der HTWK Leipzig stattfinden. Mehr dazu im SYMBOLICDATA-Wiki.

---

<sup>10</sup><http://www.escience-sachsen.de>

<sup>11</sup><http://symbolicdata.uni-leipzig.de>