WWU MÜNSTER

A Knowledge Graph of Numerical Algorithms

Peter Benner, Jan Heiland, Christian Himpe, Mario Ohlberger, Stephan Rave, Jens Saak

DMV-ÖMG Jahrestagung

Passau, September 27, 2021

Outline

1. Scientific Computing within MaRDI
   - Measure 2: Open Interfaces for Scientific Computing
   - Measure 3: Benchmark Framework
   - Measure 4: Description and Design of FAIR CSE Workflows
2. Measure 1: Knowledge Graph of Numerical Algorithms

Scientific Computing within MaRDI

MaRDI Task Area 2: Scientific Computing

Objects / Data

problem
input
model
benchmark
algorithm
output
solution
visualization
.....

M1 Knowledge Graph of Numerical Algorithms

M2 Open Interfaces for Scientific Computing

M3 Benchmark Framework

M4 Description and Design of FAIR CSE workflows

TA2 Objectives

- Verified research data in scientific computing and its fields of application
- FAIR principles for computer-based experiments and the entailing data
- Ontology of mathematical objects
- Confirmable workflows for trustworthy computations
- Dissemination of numerical methods and algorithms

Figure: Measures and major objectives

M3 – Benchmark Framework

A common theme in scientific computing

The race for the

- most efficient,
- most accurate,
- most elegant,
- most universal

algorithm for a class of problems.

This requires infrastructure for

- exchange of methods/algorithms and examples,
- comparison of competing implementations on (sets of) examples,
- tracking of progress in the field.

M3 Benchmark Framework

A Benchmark Framework

Create a generic toolkit to **fairly**

- compare and validate existing methods for new applications,
- compare new methods to existing ones,

in well-defined reference environments.

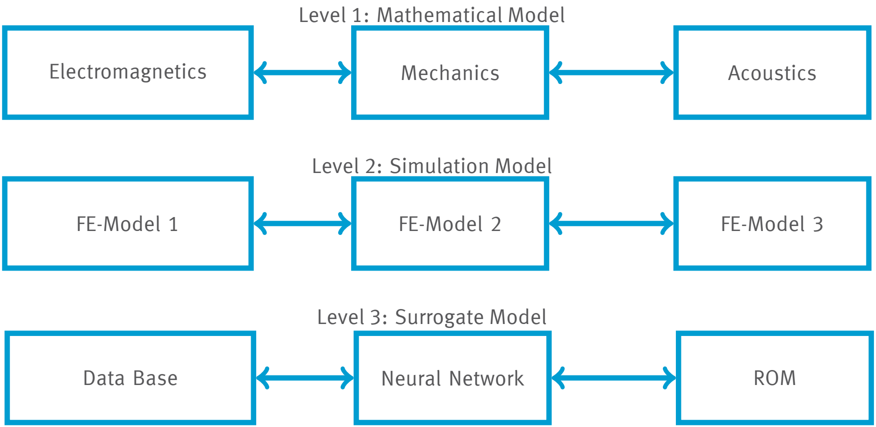
Tasks

- Assembly of domain-independent specifications
- Database of curated benchmarks

Connections in MaRDI

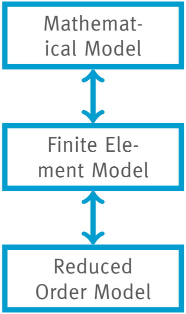
- Knowledge graph (M2.1)
- Open interfaces (M2.2)
- MaRDI Portal

M4 – Application Example: Simulation of Transformer Noise



M4 – An Electronic Lab Notebook for CSE based on Meta-Descriptions

- Every building block can be described differently.
- Only the interfaces and the meta data matter.



The Project:

1. Describe CSE workflow building block by meta data and interfaces
2. Realize the description in an *Electronic Lab Notebook*
3. so that the workflows components can be defined redundantly and interchangeable.

M2 – Software Matters

- Software is a key tool for scientific discovery.
  - We need numerical experiments to explore the world.
  - We need numerical experiments to evaluate algorithms.
  - We also need to check large problems.

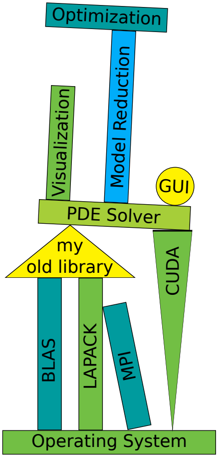
M2 – Software Matters

- Software is a key tool for scientific discovery.
  - We need numerical experiments to explore the world.
  - We need numerical experiments to evaluate algorithms.
  - We also need to check large problems.
- Software is a device for collaboration.
  - Basis for complex interdisciplinary simulation workflows.
  - What is an optimal method combining adaptive FEM, MOR and optimization?
  - Helps establish unified view on related methods.

M2 – Software Matters

- Software is a key tool for scientific discovery.
  - We need numerical experiments to explore the world.
  - We need numerical experiments to evaluate algorithms.
  - We also need to check large problems.
- Software is a device for collaboration.
  - Basis for complex interdisciplinary simulation workflows.
  - What is an optimal method combining adaptive FEM, MOR and optimization?
  - Helps establish unified view on related methods.
- Software is costly.
  - Complexity has greatly increased.
  - Getting an algorithm ‘right’ takes effort.
  - Development time costs money/grad students.
  - Domain experts required.
  - We need more sustainable software development.

A Tower of Doom



M2 – Open Interfaces to the Rescue!

- Common interfaces for scientific computing, e.g.:
  - problem description interface for ODEs / PDEs and control problems
  - high-level ODE / PDE solver interface
  - solver solution interface
  - internal solver algorithm and data structure interface
- Tools for bridging the language barrier. Easy interoperability between C++, Python, Matlab, Julia, Fortran, R
- Specification freely available and published under open licenses.
- Community driven development process.

Measure 1 – Knowledge Graph of Numerical Algorithms

Two User Stories

Alice

Two User Stories

Alice

- Alice is a **computational biologist**.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Bob

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Bob

- Bob is an **expert** in model reduction.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Bob

- Bob is an **expert** in model reduction. He wants to **follow current research in his field** and **compare his algorithms** to the state of the art of **competing methods**.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Bob

- Bob is an **expert** in model reduction. He wants to **follow current research in his field** and **compare his algorithms** to the state of the art of **competing methods**.
- He's clearly too **busy** to follow `math.NA`, and not everything is there.
- There is **no useable AMS classification**. Keywords? But which?

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Bob

- Bob is an **expert** in model reduction. He wants to **follow current research in his field** and **compare his algorithms** to the state of the art of **competing methods**.
- He's clearly too **busy** to follow `math.NA`, and not everything is there.
- There is **no useable AMS classification**. Keywords? But which?
- Luckily, he can follow **all relevant researchers** on ResearchGate and meet them at conferences.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Bob

- Bob is an **expert** in model reduction. He wants to **follow current research in his field** and **compare his algorithms** to the state of the art of **competing methods**.
- He's clearly too **busy** to follow `math.NA`, and not everything is there.
- There is **no useable AMS classification**. Keywords? But which?
- Luckily, he can follow **all relevant researchers** on ResearchGate and meet them at conferences.
- **Implementing a competitor's method** and comparing it to his own is **too much work**, so he does not do it.

Two User Stories

Alice

- Alice is a **computational biologist**. In her work, she encounters a **non-smooth optimization** problem she needs to solve. She has **no expericence** with this.
- She wants to find the **best algorithm** for her **specific problem** and a software library she can use.
- She does not really know how to get an overview, of what her options are.
- A google scholar search is **daunting**.
- Luckily, she has a **colleague**, who has had a similar problem before, she can ask for **advice**.

Bob

- Bob is an **expert** in model reduction. He wants to **follow current research in his field** and **compare his algorithms** to the state of the art of **competing methods**.
- He's clearly too **busy** to follow `math.NA`, and not everything is there.
- There is **no useable AMS classification**. Keywords? But which?
- Luckily, he can follow **all relevant researchers** on ResearchGate and meet them at conferences.
- **Implementing a competitor's method** and comparing it to his own is **too much work**, so he does not do it. On the upside, his competitors **don't prove** his method to be **inferior** either.

Let’s help Alice and Bob become better researchers!

- Numerical algorithms are main research artifacts produced by scientific computing.
- They are foundational for a large part of science.

Let’s help Alice and Bob become better researchers!

- Numerical algorithms are main research artifacts produced by scientific computing.
- They are foundational for a large part of science.
- No way to directly search/get suggestions for:
  - journal articles discussing a specific algorithm
  - (benchmark) experiments comparing algorithms for a specific problem
  - implementing software
  - algorithms that solve a certain mathematical problem
  - algorithms similar to a given algorithm

Let’s help Alice and Bob become better researchers!

- Numerical algorithms are main research artifacts produced by scientific computing.
- They are foundational for a large part of science.
- No way to directly search/get suggestions for:
  - journal articles discussing a specific algorithm
  - (benchmark) experiments comparing algorithms for a specific problem
  - implementing software
  - algorithms that solve a certain mathematical problem
  - algorithms similar to a given algorithm

Our goal

Build and maintain a knowledge graph of numerical algorithms, which interlinks those algorithms with the addressed mathematical problems and associated research data such as journal papers, benchmarks or implementing software packages.

What is a knowledge graph?

One possible definition:[1] A knowledge graph represents a collection of interlinked descriptions of entities – real-world objects and events, or abstract concepts (e.g., documents) – where:

- Descriptions have formal semantics that allow both people and computers to process them in an efficient and unambiguous manner;
- Entity descriptions contribute to one another, forming a network, where each entity represents part of the description of the entities, related to it, and provides context for their interpretation.

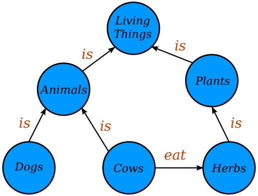


Figure: A knowledge graph[2]

[1] Source: ontotext.com

[2] Jayarathina, CC BY-SA 4.0 ‹https://creativecommons.org/licenses/by-sa/4.0›, via Wikimedia Commons

Example: Wikidata

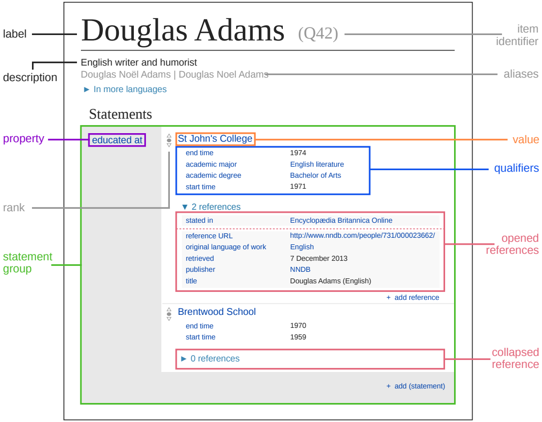
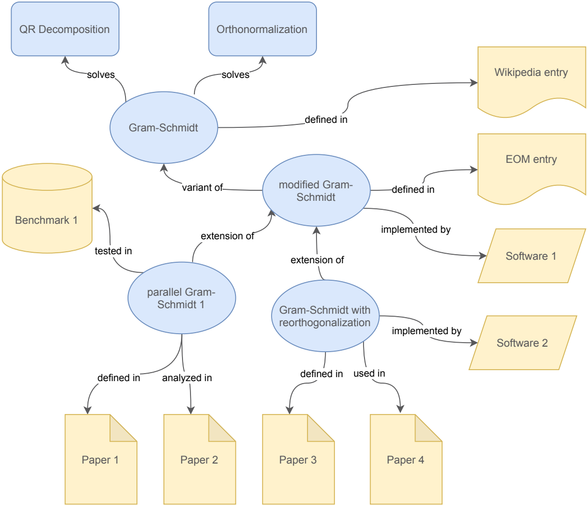


Figure: Wikidata datamodel[3]

[3] Charlie Kritschmar (WMDE), CC0, via Wikimedia Commons

A Knowledge Graph of Numerical Algorithms



Building the knowledge graph

- Establish editorial board of domain experts
  - define and update the nodes (algorithms) in the graph
  - decide which algorithms are different/essentially the same
  - choose short descriptions
  - link defining articles/software/benchmarks and related algorithms

Building the knowledge graph

- Establish editorial board of domain experts
  - define and update the nodes (algorithms) in the graph
  - decide which algorithms are different/essentially the same
  - choose short descriptions
  - link defining articles/software/benchmarks and related algorithms
- Workflow
  - dashboard which list suggestions for extending/modifying the graph
  - mostly simple yes/no decisions
  - data sources: AI-powered text mining, manuscript meta-data, user suggestions

Building the knowledge graph

- Establish editorial board of domain experts
  - define and update the nodes (algorithms) in the graph
  - decide which algorithms are different/essentially the same
  - choose short descriptions
  - link defining articles/software/benchmarks and related algorithms
- Workflow
  - dashboard which list suggestions for extending/modifying the graph
  - mostly simple yes/no decisions
  - data sources: AI-powered text mining, manuscript meta-data, user suggestions
- Costs/benefits
  - workload: < 30min per week

Building the knowledge graph

- Establish editorial board of domain experts
  - define and update the nodes (algorithms) in the graph
  - decide which algorithms are different/essentially the same
  - choose short descriptions
  - link defining articles/software/benchmarks and related algorithms
- Workflow
  - dashboard which list suggestions for extending/modifying the graph
  - mostly simple yes/no decisions
  - data sources: AI-powered text mining, manuscript meta-data, user suggestions
- Costs/benefits
  - workload: $<$ 30min per week
  - help your community and science in general!
  - honor and prestige!
  - power!

Integration with other services

- The knowledge graph will have it's own searchable/browsable web frontend.

- It will also be integrated with other MaRDI or external services:
  - public API
  - when viewing a paper, get suggestions for papers discussing the same algorithm
  - when looking at benchmarks, find links to implementing software
  - make suggestions for linking to algorithms directly from the arXiv/zbMATH/publisher's homepage

Thank you for your attention!

- We have the money, but we need your input!
- Soon: Community-building workshop
- Please subscribe for updates:
  https://www.listserv.dfn.de/sympa/subscribe/mardi-scicomp