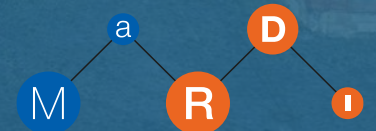
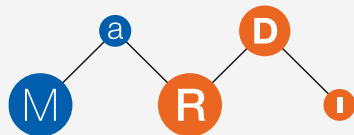


Mathematics Research Data Initiative - MaRDI4NFDI for PUNCH4NFDI

Frank Wübbeling (University of Münster)





Who we are



UNIVERSITÄT
LEIPZIG



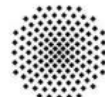
Mathematisches
Forschungsinstitut
Oberwolfach



TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN



Universität
Münster



Universität
Stuttgart

Technical
University
of Munich



IMAGINARY
open mathematics

Mathematics
Cross-Disciplinary
Infrastructure
Providers

Clusters of Excellence

- SimTech
- Mathematics
Münster
- MATH+
- STRUCTURES

Scientific Associations and Societies

- DMV
- GAMM
- GOR
- EMS
- FhG IUK

IMAGINARY
open mathematics

M

a

D

I

R

■ **WIAS**

- ZIB
- MPI DCTS Magdeburg
- FIZ Karlsruhe
- MPI MIS Leipzig
- Fraunhofer ITWM
- MFO

- TU Berlin
- FU Berlin
- U Stuttgart
- LMU München
- TU München
- WWU Münster
- TU Kaiserslautern

Leibniz Network
Mathematical
Modeling and
Simulation

M
M | **S**

Computer
Algebra



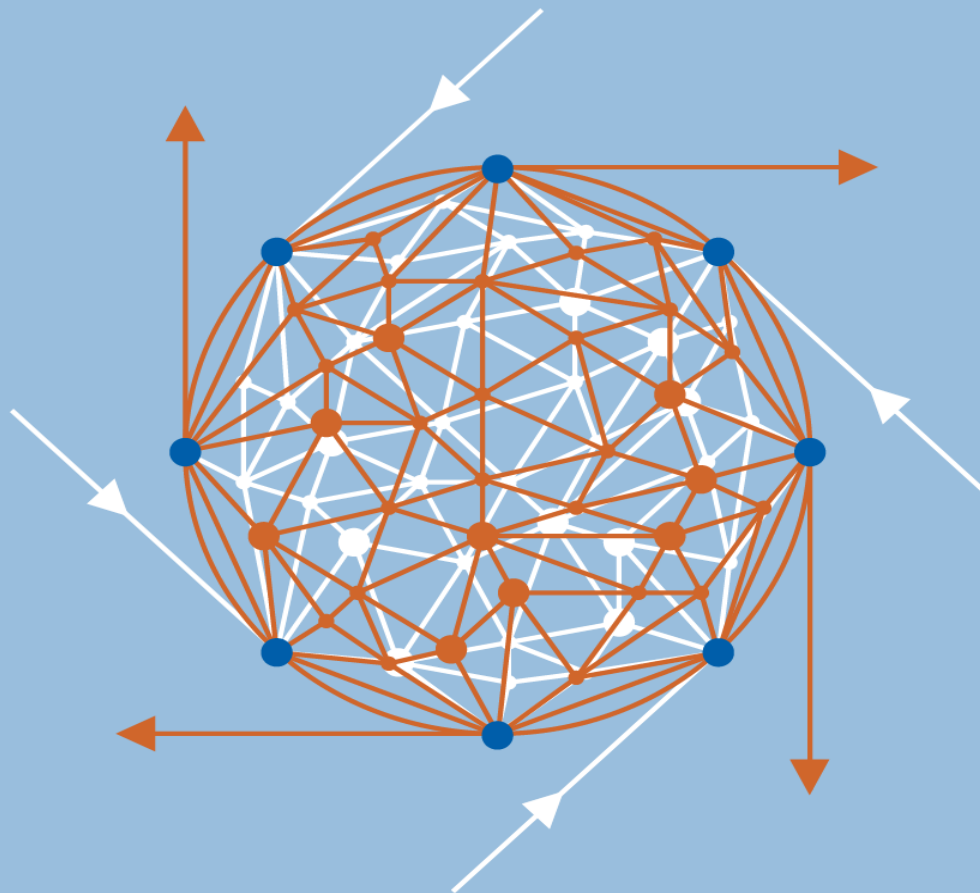
Scientific
Computing

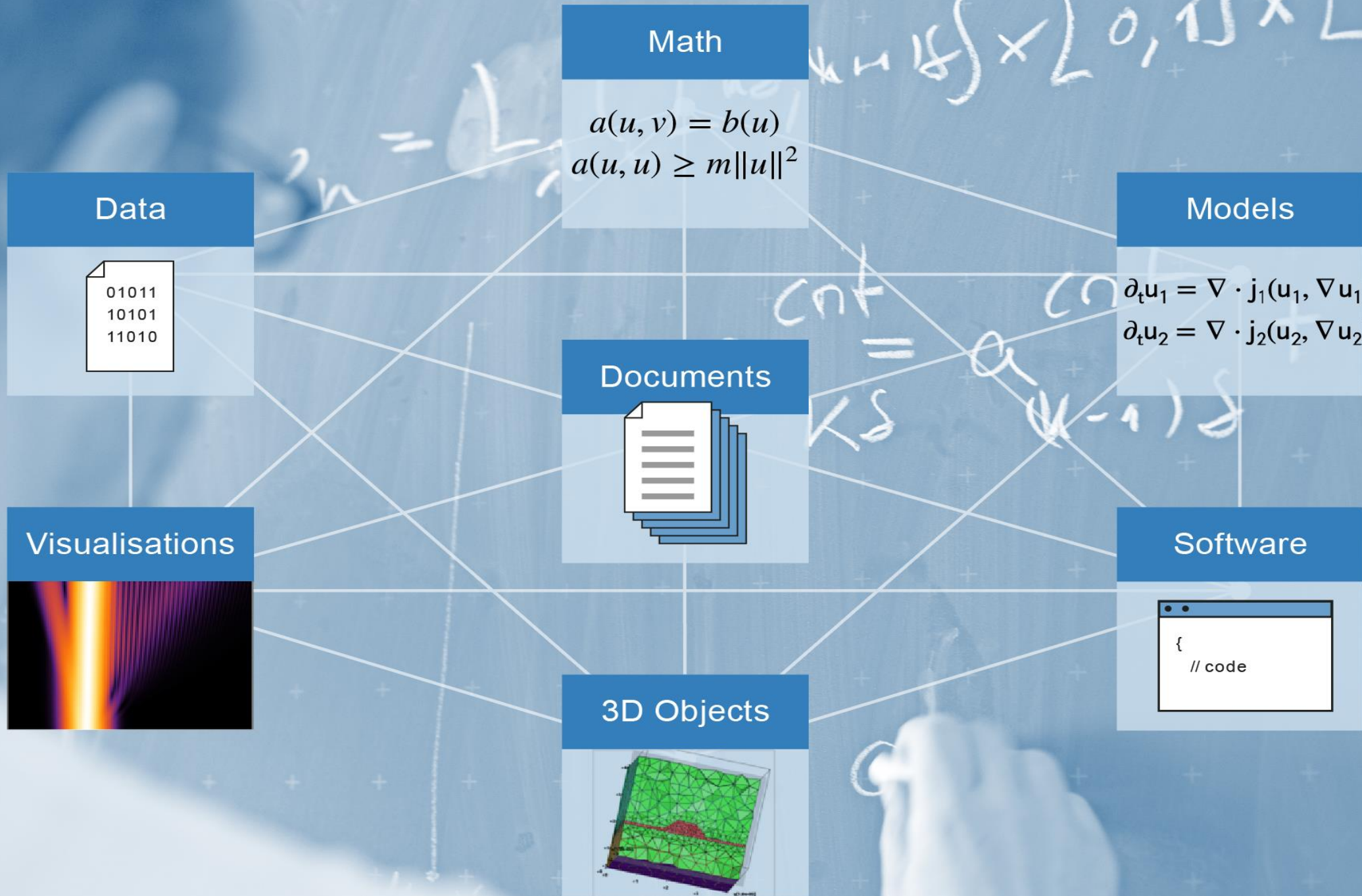


Inter-
disciplinary



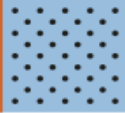
Statistics
and Machine
Learning





TA

6



Data Culture and Community Integration

TA

5



MaRDI Portal

Create and provide Services

Synergies
within NFDI

TA

1

Computer
Algebra

TA

2

Scientific
Computing

TA

3

Statistics and
Mach. Learning

TA

4

Inter-
disciplinary

Algorithms

Mathematical Models

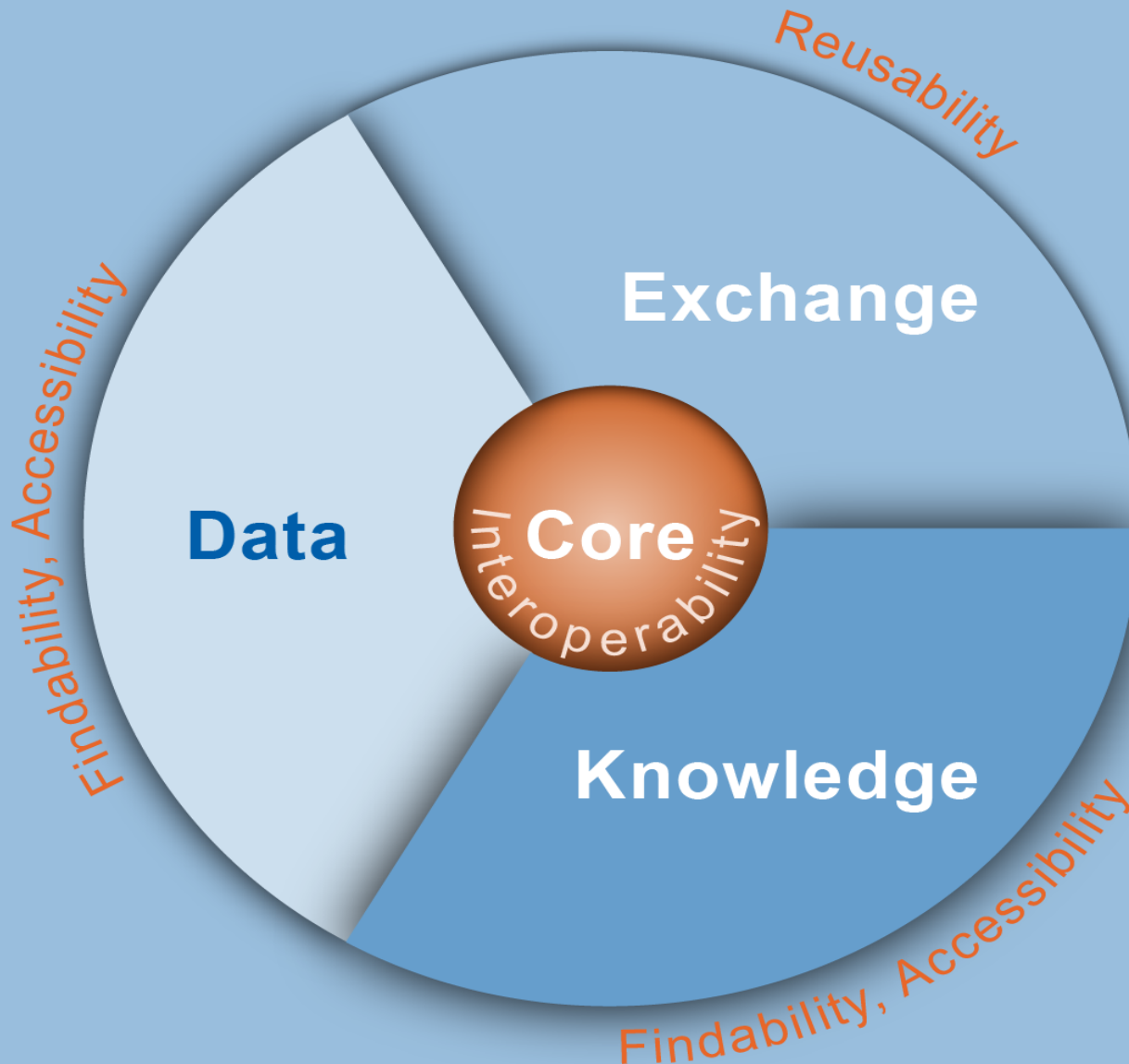
Workflows

TA

7



Governance



Example: Linear Solver $Ax=b$

X1: Core

- Data structures for matrices and vectors
- Representation formats
- Exchange formats
- Application programming interfaces (APIs)

X2: Data

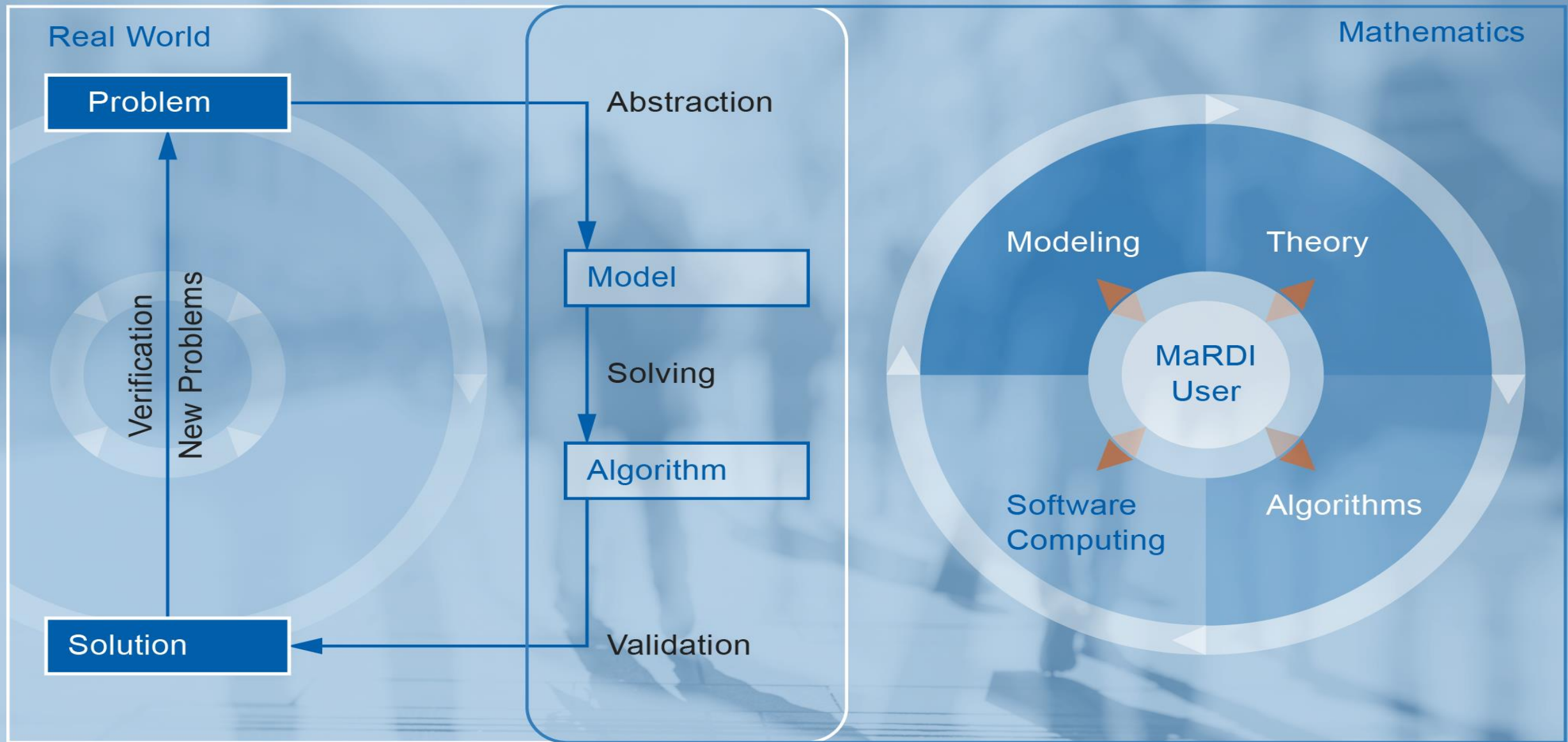
- Test cases (matrices, solutions)
- Matrix properties (meta-data)
- Persistent identifiers (PIDs)

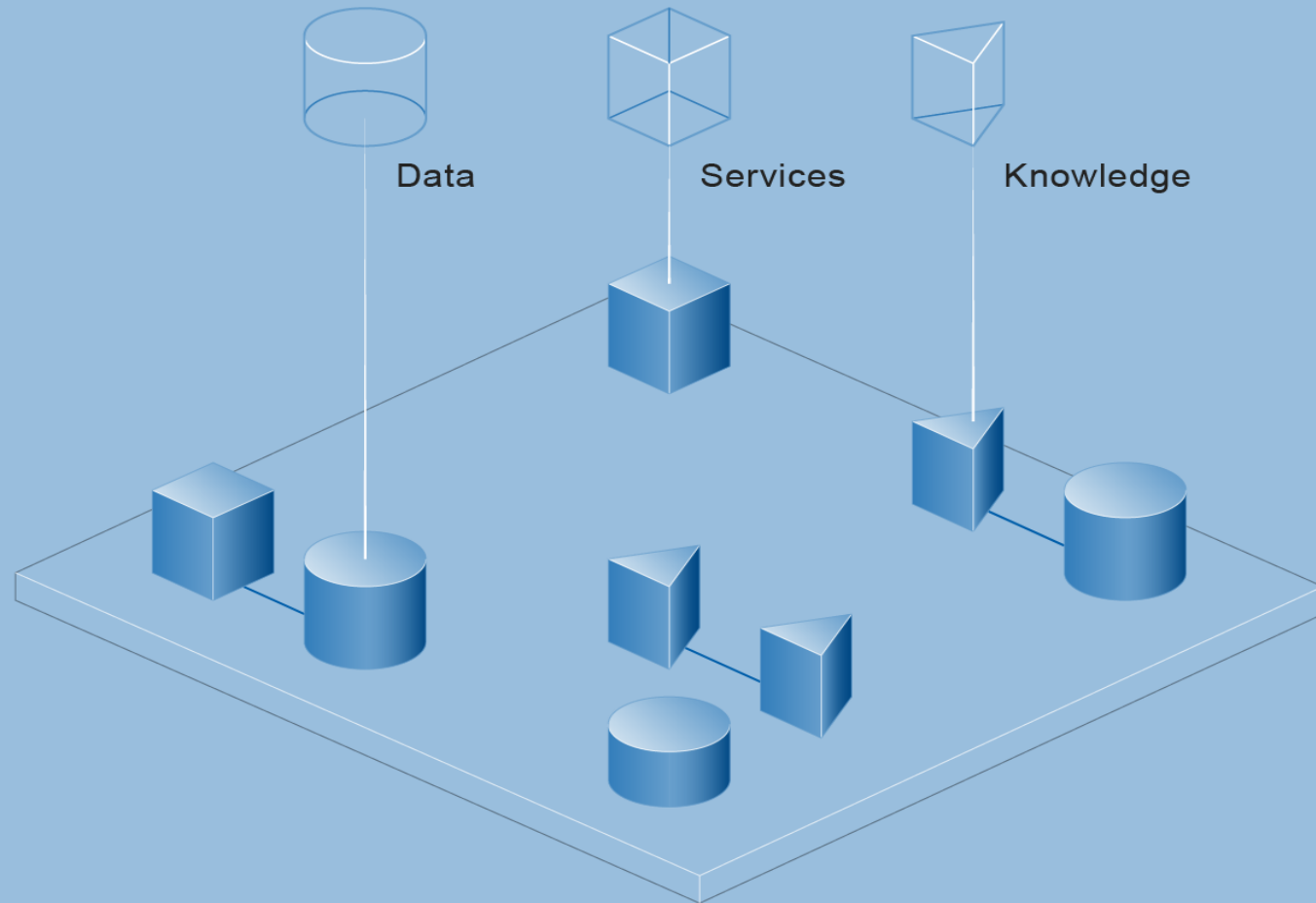
X3: Exchange

- Benchmark framework
- Pre-defined software environments
- Workflows, continuous-benchmarking

X4: Knowledge

- Ontology of linear problems and solvers
- Link to algorithms, publications software and test-data





Examples for Services

- OpenML
- MORwiki
- polyDB & SmallGrp

Examples for Infrastructures

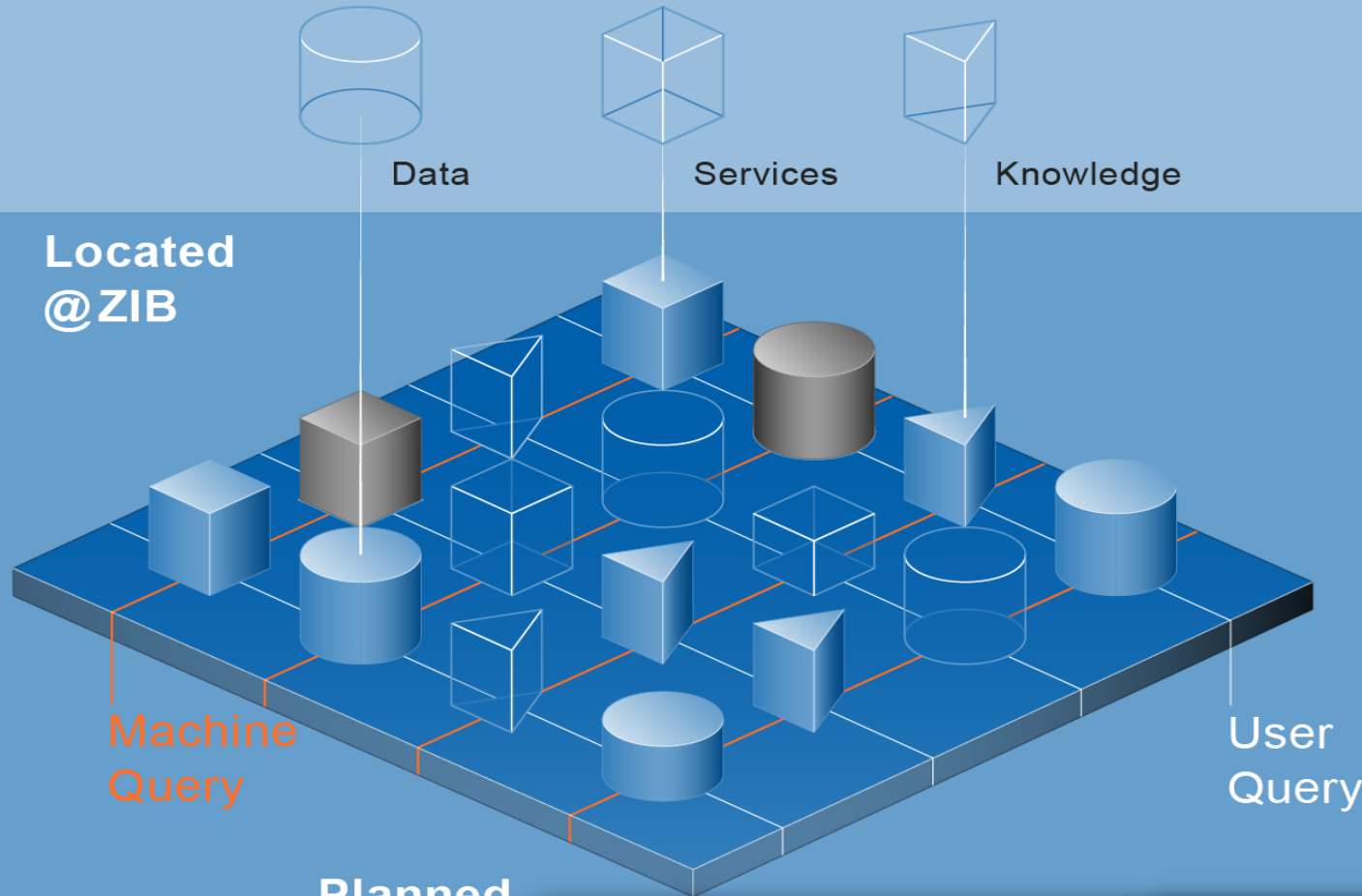
- zbMATH & swMATH
- Encyclopedia of Mathematics
- RADAR

Located @ Provider-Site

External

- OpenAIRE
- Zenodo
- DLMF

Located @ ZIB



Examples for Services

- OpenML
- MORwiki
- polyDB & SmallGrp

Examples for Infrastructures

- zbMATH & swMATH
- Encyclopedia of Mathematics
- RADAR

Planned Services:

Data

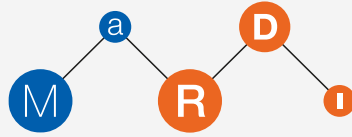
- Repositories for Computer Algebra
- Library of Benchmark Data
- Persistent Identifier Registry

Research

- Benchmark Framework
- Library of Statistical Analysis
- Workflow and Data Certification Service

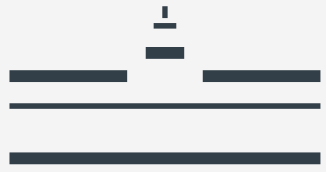
Knowledge

- Knowledge Graph of Numerical Algorithms
- Knowledge Graph for Modeling and Simulation
- Terminology Service



Existing New Services

- Computer Algebra (OSCAR):
 - **FAIR file formats; MaRDI data packaging system; FAIR Computer Algebra Workflow guidelines**
- Scientific Computing:
 - **Algorithm, Mathematical Model Knowledge Graph; OpenInterfaces; Benchmarking and Workflow**
- Statistics/Machine Learning
 - **OpenML datasets and FAIR algorithms for data-driven approaches**
- Central
 - **The MaRDI portal (portal.mardi4nfdi.de)**
 - **Teaching Lecture Notes and Slides; Research Guidelines; Help Desk**

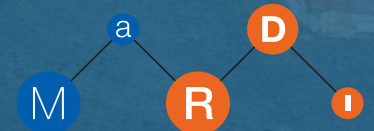


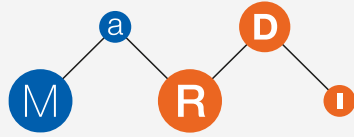
Universität
Münster

Establishing a Knowledge Graph for CSE: From Prototype to Community Project

Frank Wübbeling
René Fritze (former)
Christian Himpe (former)
Hendrik Kleikamp
Mario Ohlberger
Stephan Rave
(all University of Münster)

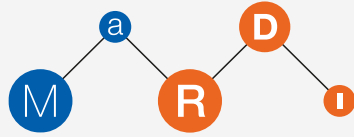
living.knowledge





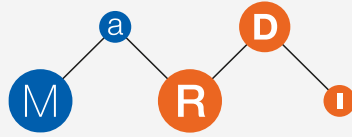
Why?

- Scientific computing focusses on algorithms. But:
 - **There is currently no way of systematically searching for data associated with an algorithm**
 - **Where ``data`` is**
 - **Articles inventing, discussing, testing, analyzing, applying ... it**
 - **Software implementing it**
 - **Benchmarks testing it**
 - **Problems they solve (the reverse is more interesting)**
 - **Required/exploited prerequisites (spd, sparse), properties (order, complexity)**



What?

- Knowledge graphs are a well-established technique for representing knowledge in a graph-structured data model. In a nutshell, they represent **relations** between **objects** (triples).
- Application to scientific computing
 - **Algorithms** solve **problems**
 - **Possess** mathematical **properties**
 - **Implemented** in **software**
 - **Tested** by **benchmarks**
 - **Documented** in **publications**



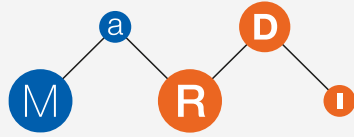
What's the benefit? Get answers to...

Starting as a practitioner from a (practical) problem:

- What are the available algorithm options for my problem? (Hint: Need link to Engineering graph)
- Are the prerequisites for an algorithm satisfied by my problem? Does it exploit its properties?
- Which implementations are available? Where do I find a common analysis?
- How can I document which algorithms I used for my problem?
- Need references for all.

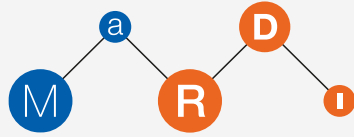
Starting as an algorithm developer:

- Which methods am I competing against? (keeping up with developments in vivid fields ain't easy)
- Which properties distinguish my algorithm from others?
- Which benchmarks should I run? Which implementations should I use for competing algorithms?
- Where do I find problems that used competing algorithms?
- Need references for all.



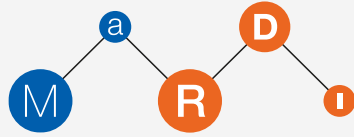
Which structure?

- The ontology formalizes this structure using predefined **relations** between **objects**.
- Which questions do you want answered?
 - **Which algorithms solve emission tomography problems?**
 - **Which publication invented Maximum Likelihood Expectation Maximization?**
 - **Which software implements filtered backprojection?**
 - **Which benchmark tests Ordered Subset Expectation Maximization?**

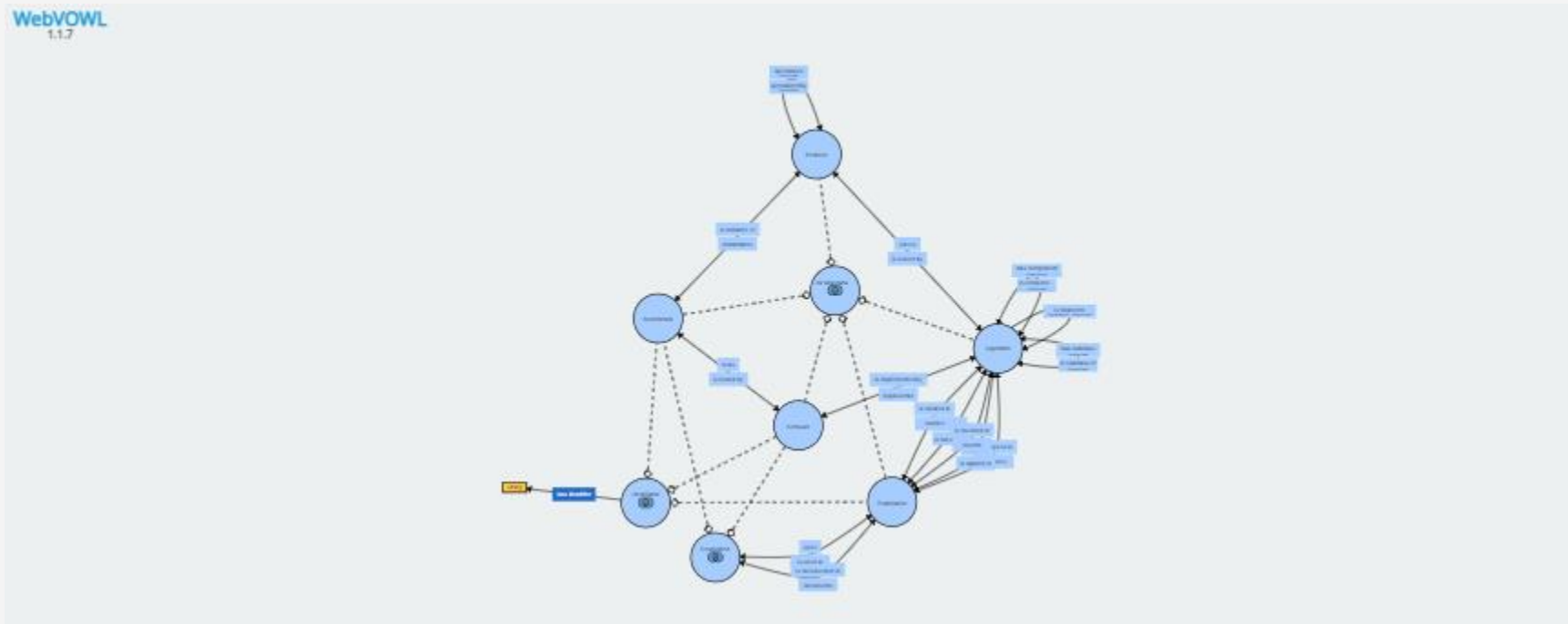


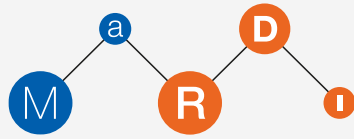
What does it finally consist of?

- Set of objects and relations using ontology web language (OWL) standards
 - **Objects: Problem, Algorithms, Software, Benchmark, Publication, Mathematical Property**
 - **Hierarchy: Problems, Algorithms, Software**
 - **Relations: implements, implementedBy, solves, solvedBy, analyzedIn...**
(total 35 relations, but most of these with inverse, 20 actual relations)
- **Our Goal: Create a **curated** knowledge graph for scientific computing in this ontology.**

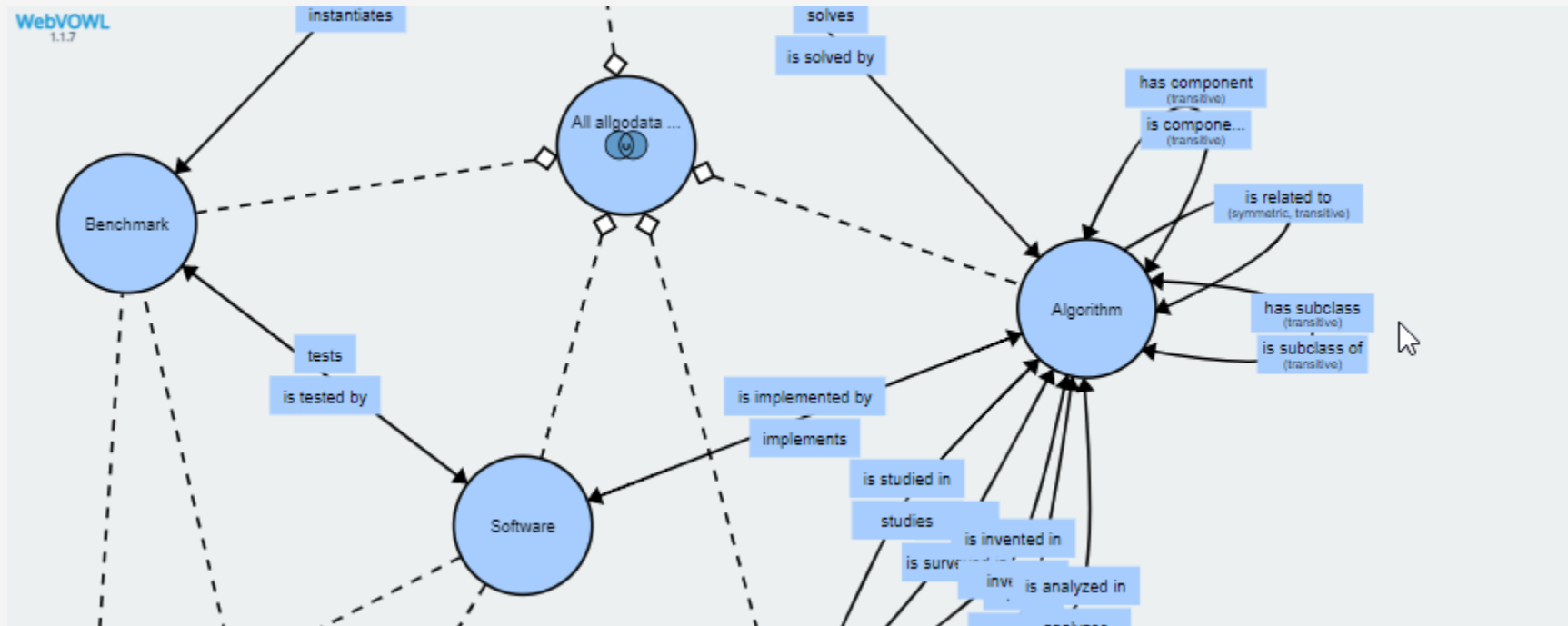


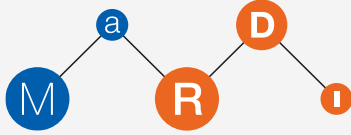
OWL Graph



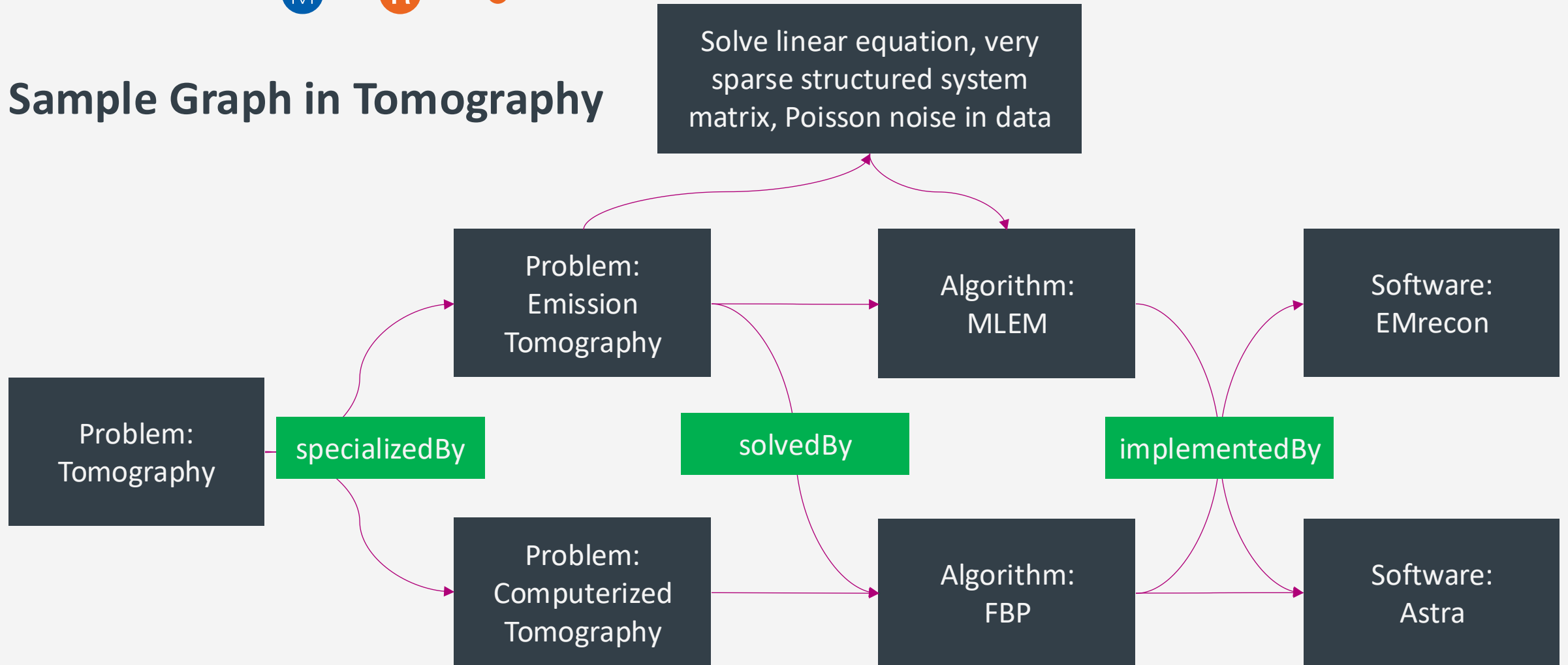


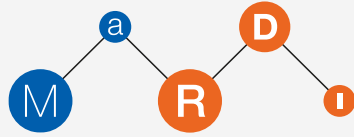
OWL Graph





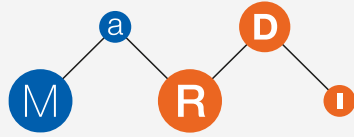
Sample Graph in Tomography





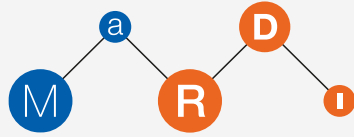
What is available?

- Sample knowledge graphs in two fields:
 - **Numerical Linear Algebra (basic)**
 - **Model order reduction:**
22 problems/subproblems, 90 algorithms, 17 implementations,
38 benchmark problems, 604 publications
- Sample open source implementation of query infrastructure and GUI
 - **Django on top of Jena-Fuseki with OWL reasoner in a container setup**
 - **Public available**



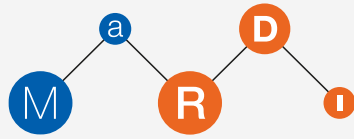
How can I use it?

- Raw graph in standard data formats (turtle, RDF, triples, ...)
- SPARQL database access
- Standardized REST/API access
 - ex: return all publications that analyze MLEM (as JSON)
 - ex: convenience functions for resolution of DOIs etc.
- Send natural language requests through an LLM
- Sample GUI
 - relies on the API (and has minimal functionality in itself)
- **Public on <https://mathalgotdb.mardi4nfdi.de>**
- **Beta of unified graphs with TA4 on <https://mtsr2024.m1.mardi.ovh/>**



Processing natural language requests with LLMs

<https://www.uni-muenster.de/FB10srvi/service/openai/query.php>



GUI

[Home](#) [Graph Query](#) [Documentation](#) [Feedback](#) [Admin](#)

Graph Query

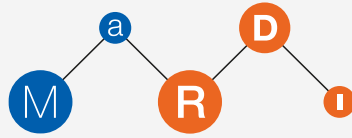
First, select to query the whole graph (all fields) or select a specific field of applied mathematics.

1. Field:

Second, form a question by selecting from the given options, for example: What ? [Answer](#)

2. Query: What ?

AlgoData - Algorithm Knowledge Graph — A [MaRDI](#) Measure — [Share](#)



GUI: Guided Queries

[Home](#) [Graph Query](#) [Documentation](#) [Feedback](#) [Admin](#)

Graph Query

First, select to query the whole graph (all fields) or select a specific field of applied mathematics.

1. Field:

Second, form a question by selecting from the given options, for example: What ? [Answer](#)

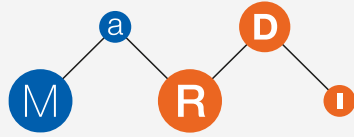
2. Query: What ?

What Software implements Filtered Backprojection / FBP?

→ ASTRA tomographic toolbox / ASTRA

→ TOMOPY - tomographic reconstruction in Python / TOMOPY

AlgoData - Algorithm Knowledge Graph — A [MaRDI](#) Measure — [Share](#)



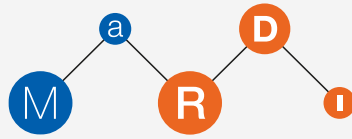
GUI: keyword search

[Home](#) [Graph Query](#) [Documentation](#) [Feedback](#) [Admin](#) [Search](#)

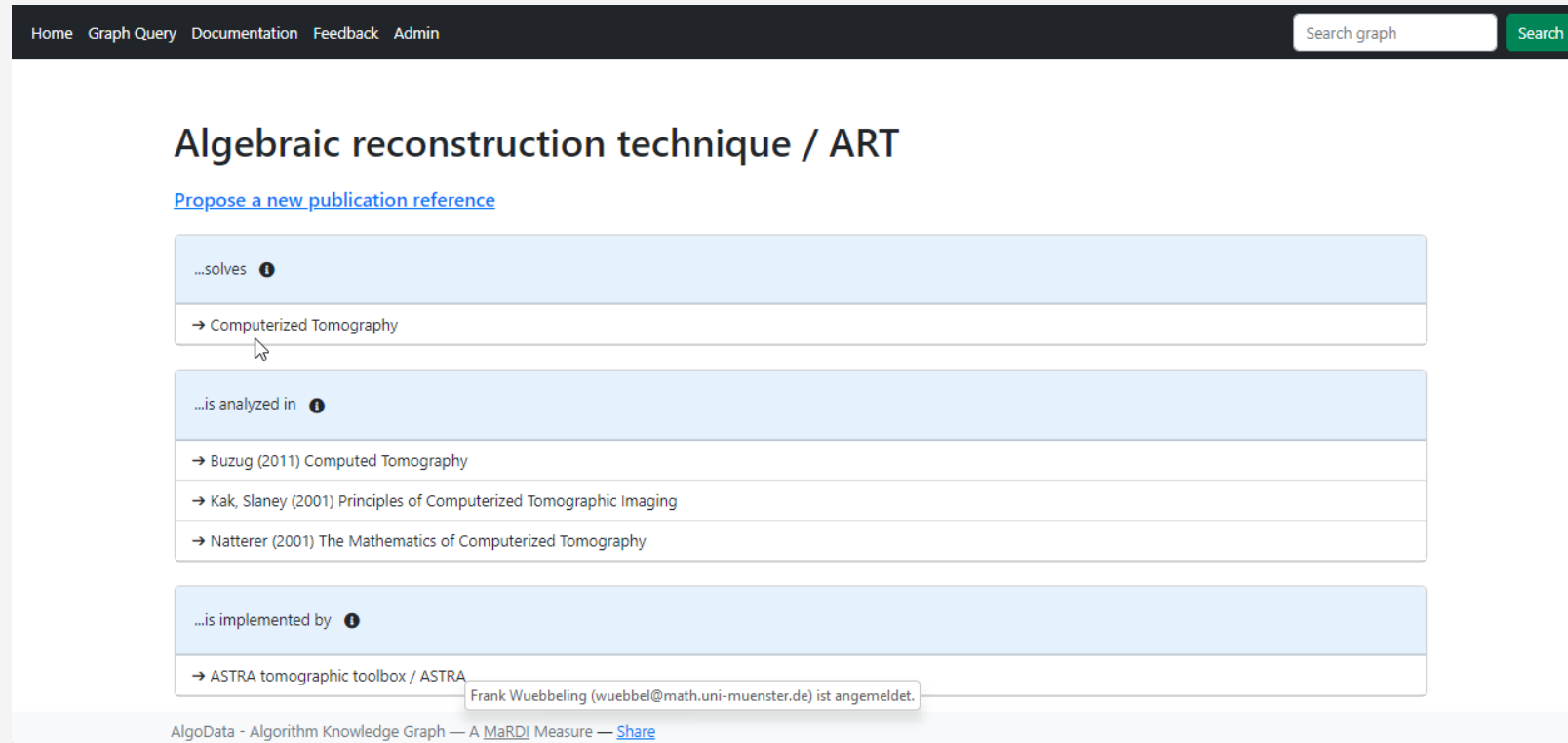
Search results for request "tomography" (9 results)

Problem:
→ Computerized Tomography
→ Emission Tomography
→ Tomography

Publication:
→ Buzug (2011) Computed Tomography
→ Faridani (2006) Fan-beam tomography and sampling theory
→ Kusters, Schafers, Wubbeling (2011) EMRECON: An expectation maximization based image reconstruction framework for emission tomography data
→ Natterer (1986) The Mathematics of Computerized Tomography
→ Natterer (2001) The Mathematics of Computerized Tomography
→ van Aarle, Palenstijn, De Beenhouwer, Altantzis, Bals, Batenburg, Sijbers (2015) The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography



GUI result display



Home Graph Query Documentation Feedback Admin

Algebraic reconstruction technique / ART

[Propose a new publication reference](#)

...solves ⓘ

- Computerized Tomography

...is analyzed in ⓘ

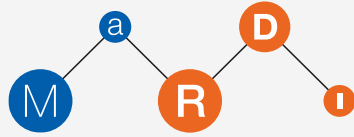
- Buzug (2011) Computed Tomography
- Kak, Slaney (2001) Principles of Computerized Tomographic Imaging
- Natterer (2001) The Mathematics of Computerized Tomography

...is implemented by ⓘ

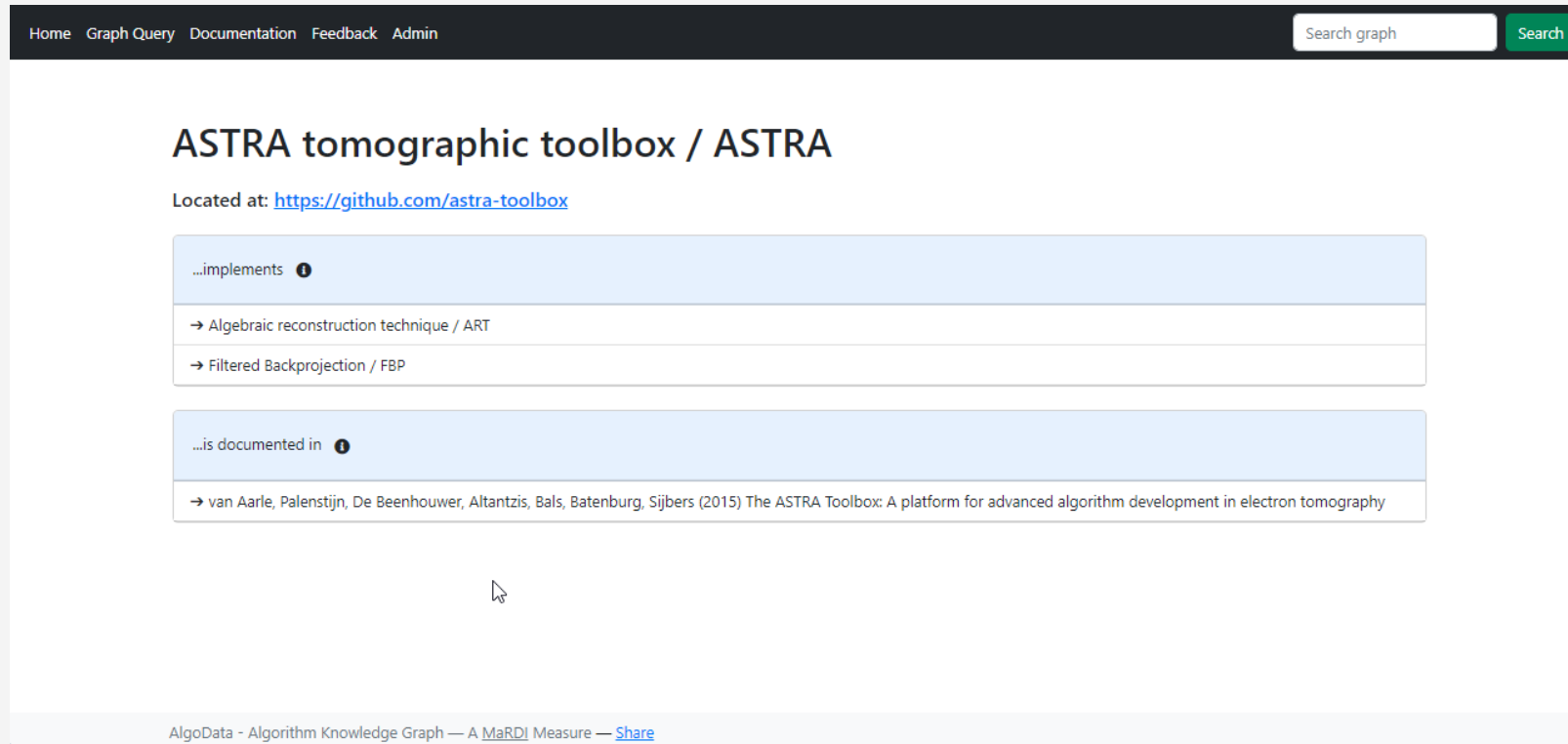
- ASTRA tomographic toolbox / ASTRA

Frank Wuebbeling (wuebbel@math.uni-muenster.de) ist angemeldet.

AlgoData - Algorithm Knowledge Graph — A [MaRDI](#) Measure — [Share](#)



GUI result display



Home Graph Query Documentation Feedback Admin

ASTRA tomographic toolbox / ASTRA

Located at: <https://github.com/astra-toolbox>

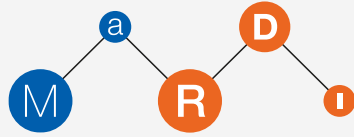
...implements ⓘ

- Algebraic reconstruction technique / ART
- Filtered Backprojection / FBP

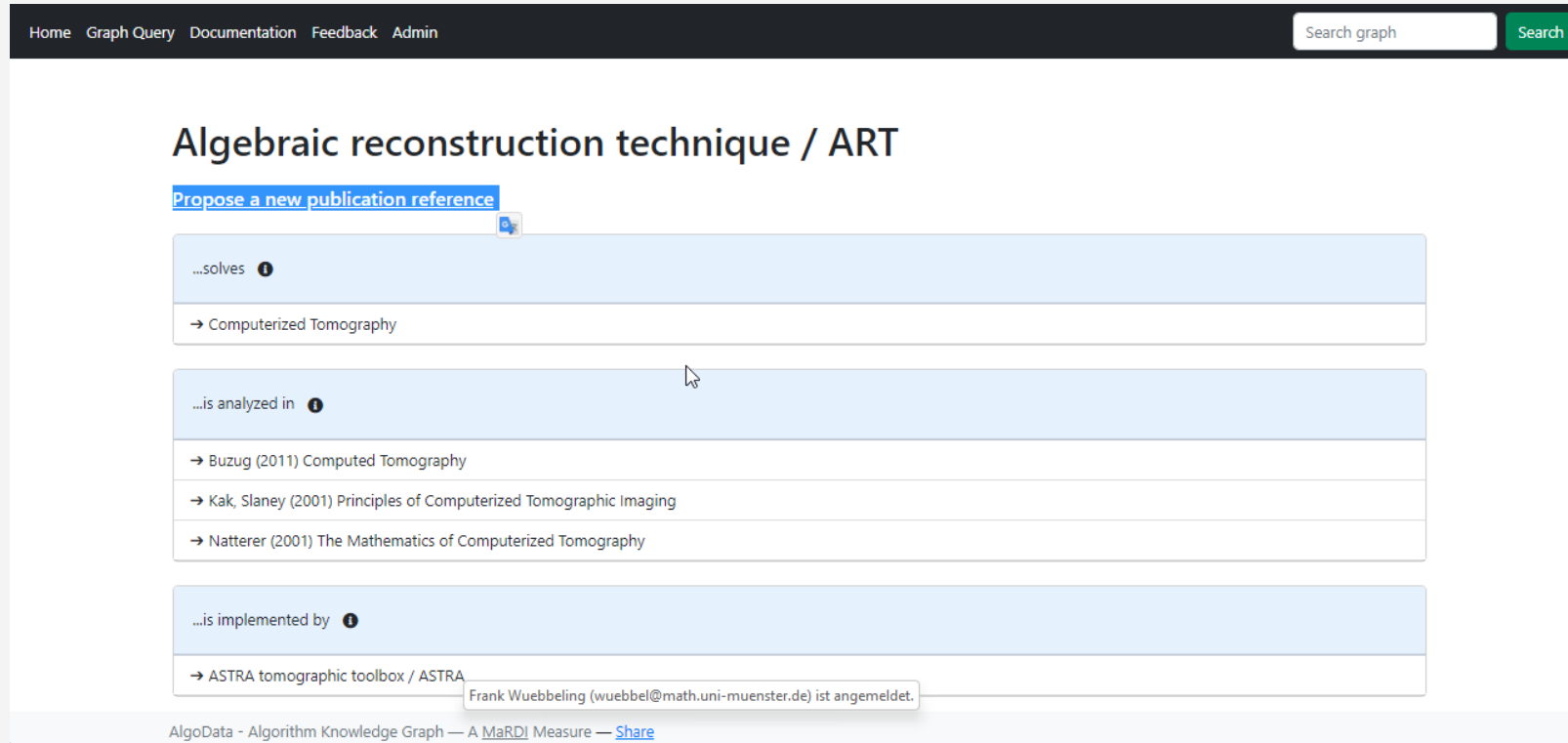
...is documented in ⓘ

- van Aarle, Palenstijn, De Beenhouwer, Altantzis, Bals, Batenburg, Sijbers (2015) The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography

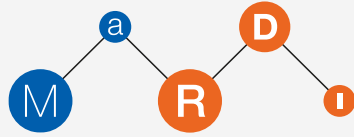
AlgoData - Algorithm Knowledge Graph — A [MaRDI](#) Measure — [Share](#)



GUI: Invites to play around, discover, explore

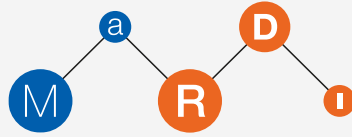


The screenshot shows a web application interface for a knowledge graph. At the top, there is a navigation bar with links: Home, Graph Query, Documentation, Feedback, and Admin. A search bar with the placeholder text "Search graph" and a green "Search" button is located on the right. The main content area is titled "Algebraic reconstruction technique / ART". Below the title, there is a link "Propose a new publication reference" with a small icon. The interface displays three expandable sections, each with a light blue header and a white content area. The first section is labeled "...solves" and contains the text "→ Computerized Tomography". The second section is labeled "...is analyzed in" and contains three entries: "→ Buzug (2011) Computed Tomography", "→ Kak, Slaney (2001) Principles of Computerized Tomographic Imaging", and "→ Natterer (2001) The Mathematics of Computerized Tomography". The third section is labeled "...is implemented by" and contains the text "→ ASTRA tomographic toolbox / ASTRA". A small tooltip is visible over the third section, displaying the text "Frank Wuebbeling (wuebbel@math.uni-muenster.de) ist angemeldet.". At the bottom of the page, there is a footer with the text "AlgoData - Algorithm Knowledge Graph — A MaRDI Measure — Share".

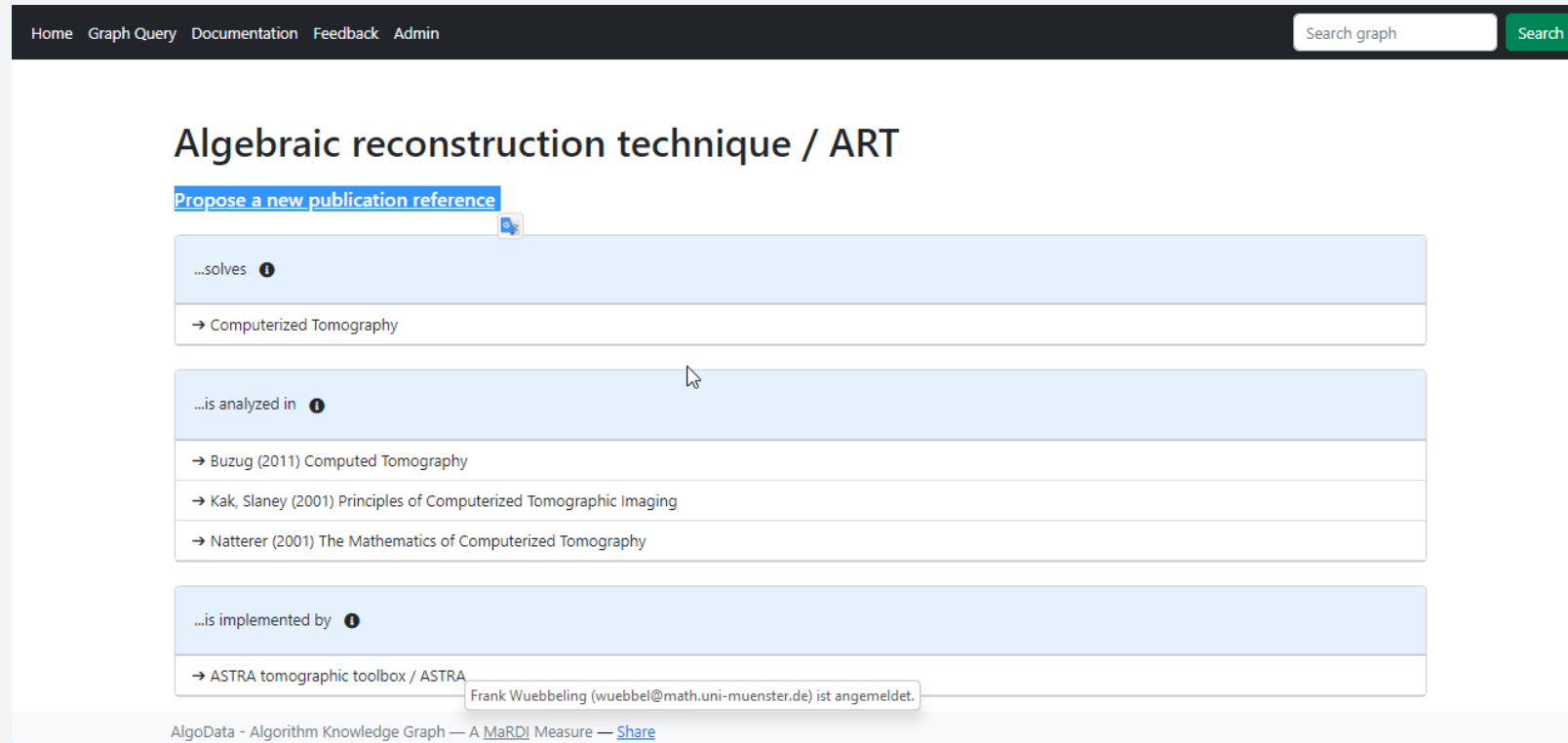


How can we overcome the hen and egg problem?

- The project will succeed only with support from the communities.
- It is vital that the work of adding pieces to the graph is as simple as possible
- Workflow:
 - **An initial graph for a field, formalizing the general structure (basic problems, algorithms, ...) is created by editors from the community.**
 - **The graph is continually updated using proposals from the community, accepted or denied by the editors.**



Contributing to the knowledge graph



Home Graph Query Documentation Feedback Admin

Algebraic reconstruction technique / ART

[Propose a new publication reference](#)

...solves ⓘ

- Computerized Tomography

...is analyzed in ⓘ

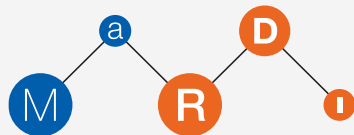
- Buzug (2011) Computed Tomography
- Kak, Slaney (2001) Principles of Computerized Tomographic Imaging
- Natterer (2001) The Mathematics of Computerized Tomography

...is implemented by ⓘ

- ASTRA tomographic toolbox / ASTRA

Frank Wuebbeling (wuebbel@math.uni-muenster.de) ist angemeldet.

AlgoData - Algorithm Knowledge Graph — A [MaRDI](#) Measure — [Share](#)



Contributing to the knowledge graph

[Home](#) [Graph Query](#) [Documentation](#) [Feedback](#) [Admin](#)

Add publication to database

Algorithm: Algebraic reconstruction technique

Choose relation for the publication:*

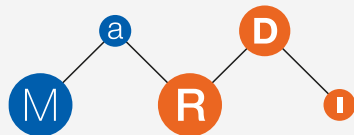
Enter:
DOI for the publication or
search by ?author:title (provided by crossref, may be slow) or
enter !authors (year) title
Press Tab after entry for completion.*

Label for the publication (not editable):*

Unique ID for the publication (not editable):*

Enter your Email address (you will be notified on acceptance)

AlgoData - Algorithm Knowledge Graph — A [MaRDI](#) Measure — [Share](#)



Checked by editorial board

[Home](#) [Graph Query](#) [Documentation](#) [Feedback](#) [Admin](#)

Admin interface

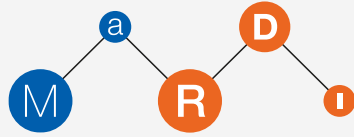
The following suggestions have been made in your category.

Submitted at 2022-10-23 11:32:29.272429 from 10.68.194.214 by wuebbel@wwu.de
Proposed relation: [Algebraic reconstruction technique](#) is invented in
Underlying DOI is [doi:10.1093/comjnl/4.3.265](https://doi.org/10.1093/comjnl/4.3.265). Category: tomo
Identifier:*

Label:*

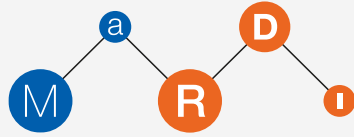
☐ Accept

AlgoData - Algorithm Knowledge Graph — A [MaRDI](#) Measure — [Share](#)



Summary

- We defined a knowledge graph for Scientific Computing
- The knowledge graph supports scientists exploring fields
- Structure and data access are open, sample implementations are based on open source standards.
- Available access levels are raw data, database access, REST API, GUI
- Sample datasets have been constructed (with only MOR being fairly wide)
- Adding data to the graph and creating a new field is easy
- **Public on <https://mathalgodb.mardi4nfdi.de/>**
- **Public beta of unified graph with MathModDB on <https://mtsr2024.m1.mardi.ovh/>**



Outlook

- Integration into larger ontology for mathematics / engineering
- More fields
- Permanent production deployment
- Community building
- Board of experts in relevant fields