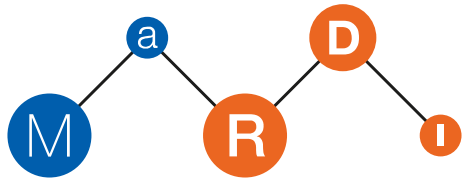


Scientific Computing

The Benchmark Framework

One Path to Standardize the Comparison of Numerical Methods

Peter Benner, Kathryn Lund, Jens Saak

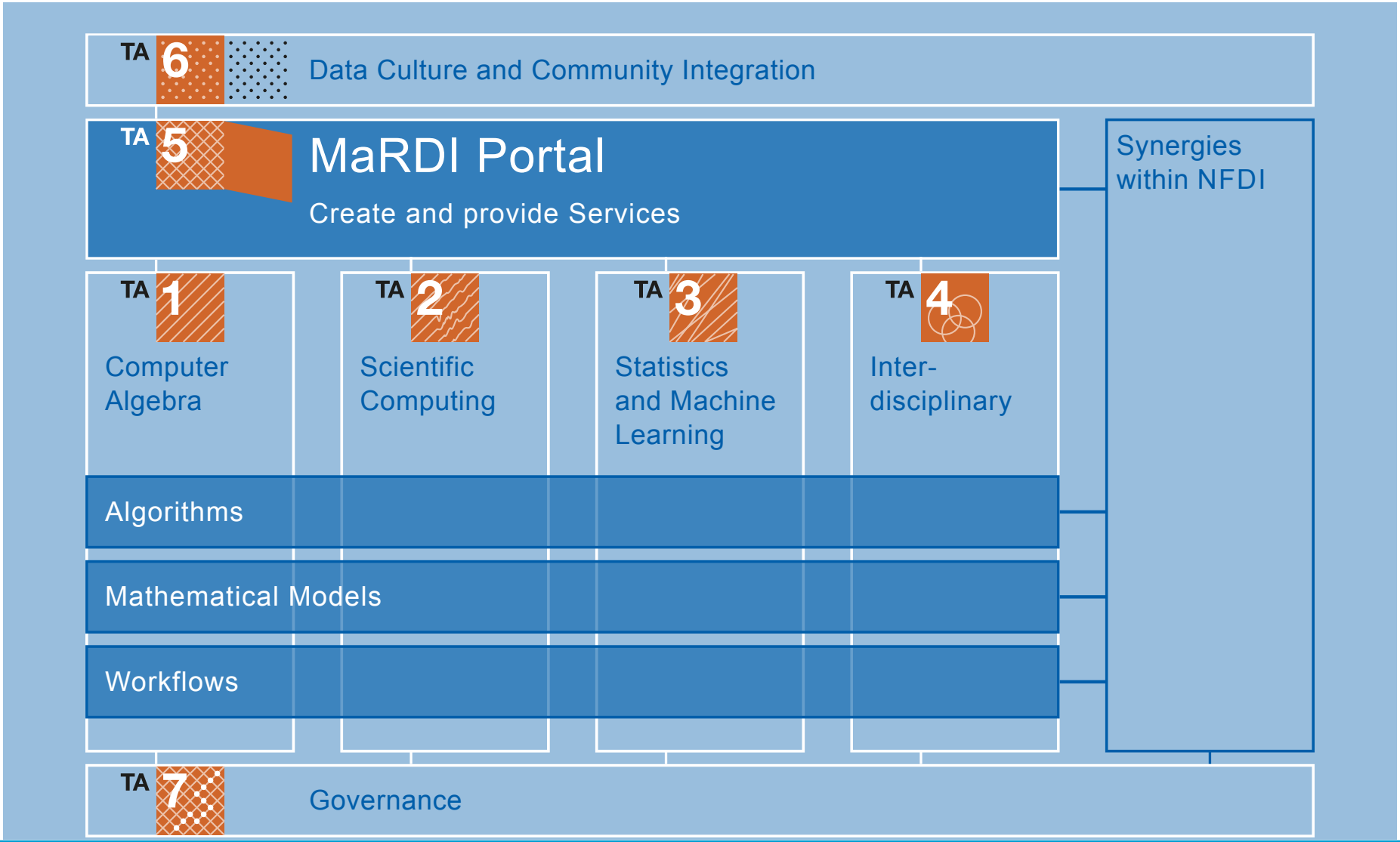
September 14, 2022

MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

Outline

Outline

TA2 within MaRDI



MaRDI Task Area 2: Measures and major objectives

Objects / Data

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

M1 Knowledge Graph of Numerical Algorithms

M2 Open Interfaces for Scientific Computing

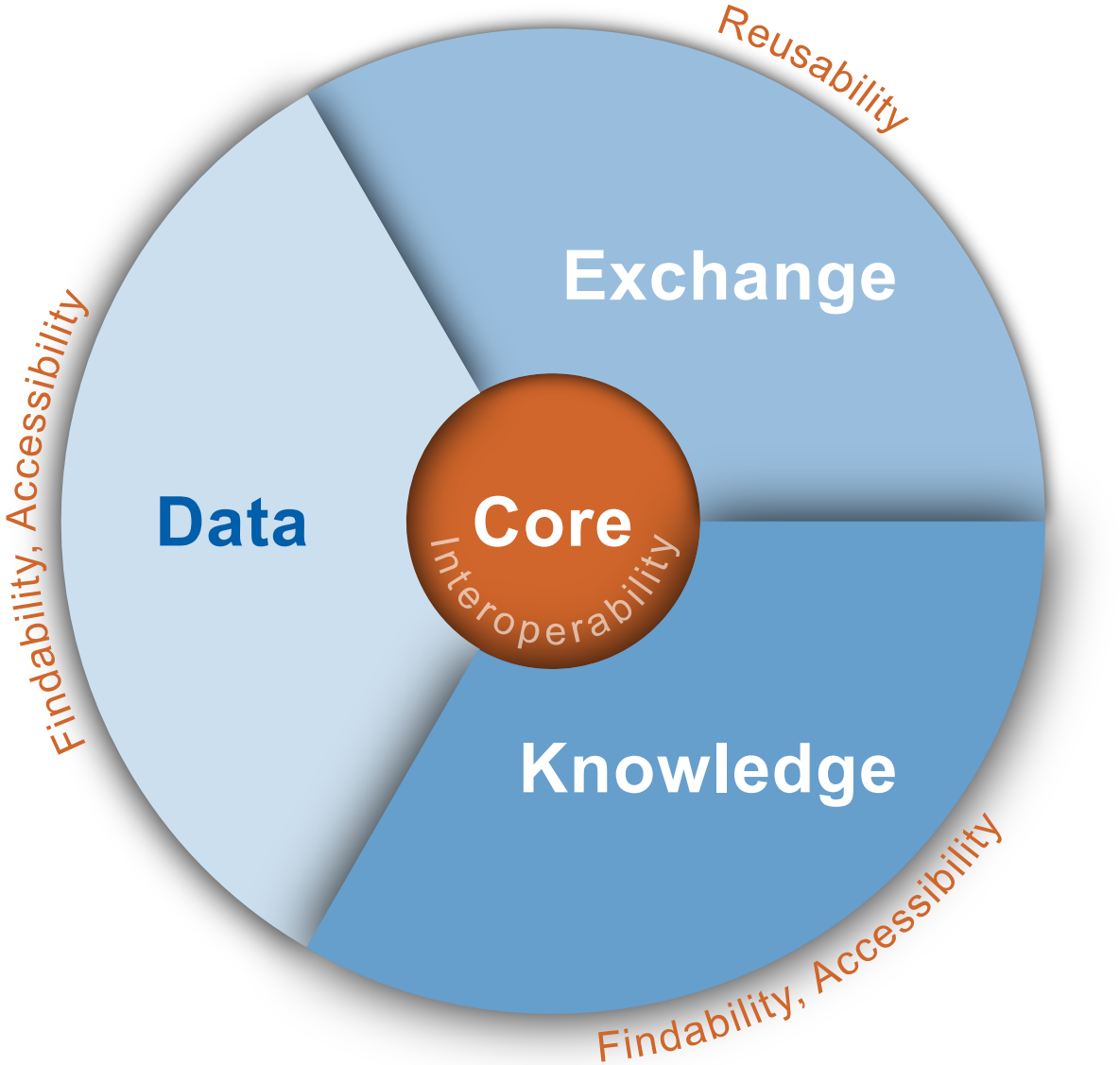
M3 Benchmark Framework

M4 Description and Design of FAIR CSE workflows

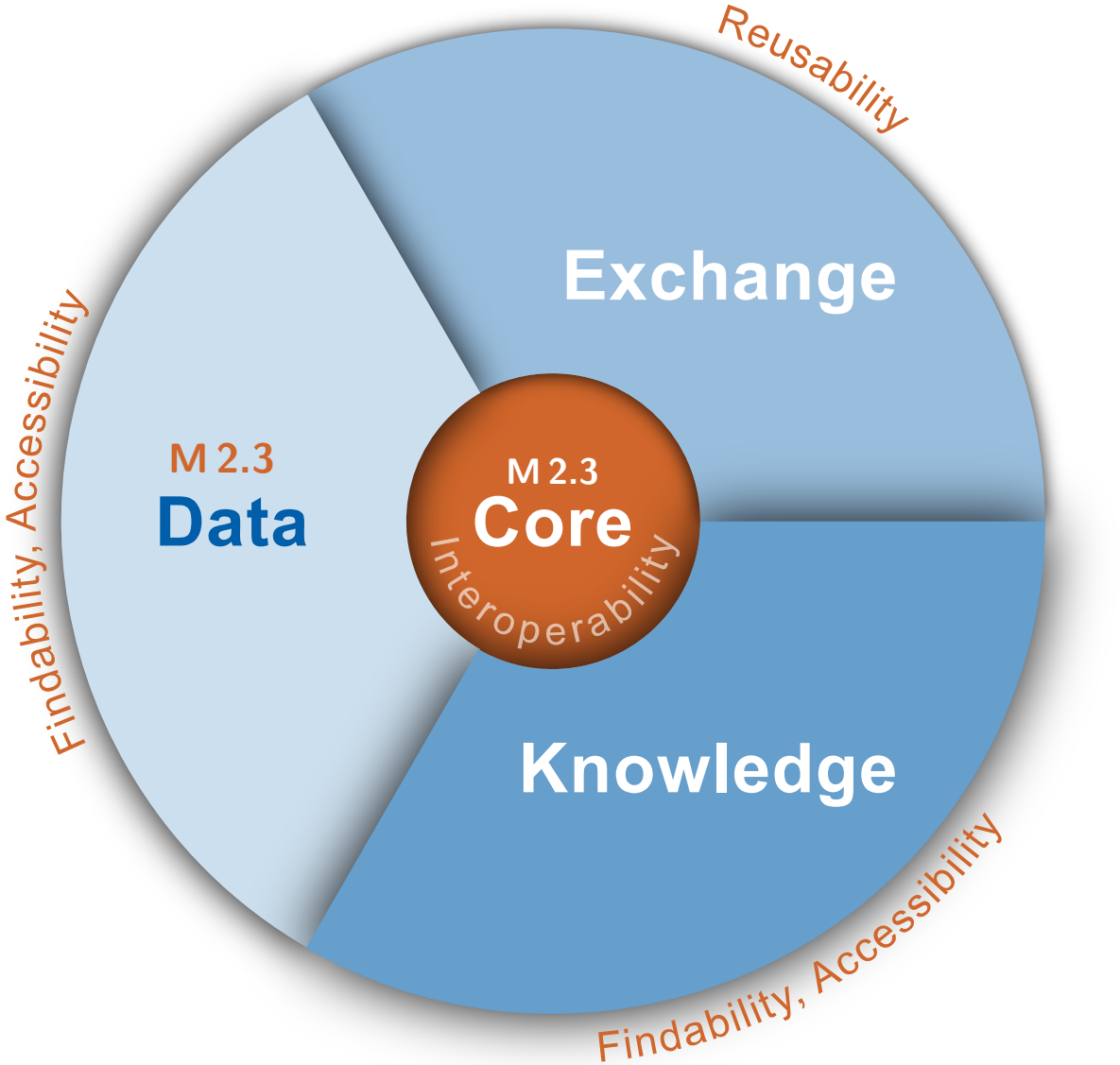
TA2 Objectives

- 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

MaRDI Layer structure



MaRDI Layer structure



Outline

Some Motivating Remarks

compare DOI:bsb2

Numerical experiments have become increasingly more important

1971 J. Nitsche. **Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind**. *Abh. Math. Semin. Univ. Hambg.*, 36(1):9–15.
No numerical example/experiment at all

Some Motivating Remarks

compare DOI:bsb2

Numerical experiments have become increasingly more important

1971 J. Nitsche. **Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind**. *Abh. Math. Semin. Univ. Hambg.*, 36(1):9–15.
No numerical example/experiment at all

1986 Y. Saad and M. H. Schultz. **GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems**. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869
2 out of 14 pages devoted to numerical experiments.

Some Motivating Remarks

compare DOI:bsb2

Numerical experiments have become increasingly more important

1971 J. Nitsche. **Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind**. *Abh. Math. Semin. Univ. Hambg.*, 36(1):9–15.
No numerical example/experiment at all

1986 Y. Saad and M. H. Schultz. **GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems**. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869
2 out of 14 pages devoted to numerical experiments.

2010 S. Chaturantabut and D. C. Sorensen. **Nonlinear model reduction via discrete empirical interpolation**. *SIAM J. Sci. Comput.*, 32(5):2737–2764.
consists of more than 30% of numerical examples or reasoning based on numerical experiments

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**

Some Motivating Remarks

The situation we are facing

Common Quotes

- **“The PhD student who conducted the experiments is no longer working in our group.”**
- **“The code is in no shape to share.”**

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."**

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."** arxiv:2203.17174v2

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."** arxiv:2203.17174v2

Quality of Codes & Data

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."** arxiv:2203.17174v2

Quality of Codes & Data

- **Code is incomplete**

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."** arxiv:2203.17174v2

Quality of Codes & Data

- **Code is incomplete**
- **Input data are unavailable or even unknown**

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."** arxiv:2203.17174v2

Quality of Codes & Data

- **Code is incomplete**
- **Input data are unavailable or even unknown**
- **"Accidental obfuscation": existing code and data are useless**

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."** arxiv:2203.17174v2

Quality of Codes & Data

- **Code is incomplete**
- **Input data are unavailable or even unknown**
- **"Accidental obfuscation": existing code and data are useless**
- **Code & data are available ☺,**

Some Motivating Remarks

The situation we are facing

Common Quotes

- **"The PhD student who conducted the experiments is no longer working in our group."**
- **"The code is in no shape to share."**
- **"All the other datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request."** arxiv:2203.17174v2

Quality of Codes & Data

- **Code is incomplete**
- **Input data are unavailable or even unknown**
- **"Accidental obfuscation": existing code and data are useless**
- **Code & data are available ☺, but all publications use different computers, running different software on different examples ☹**

Outline

M 2.3 — MaRDIMark

An general-purpose benchmarking framework for comparing implementations of algorithms using problems native to a community

Aims:

- **Generic, extensible toolkit**

M 2.3 — MaRDIMark

An general-purpose benchmarking framework for comparing implementations of algorithms using problems native to a community

Aims:

- **Generic, extensible toolkit**
- **Language-agnostic interoperability**

M 2.3 — MaRDIMark

An general-purpose benchmarking framework for comparing implementations of algorithms using problems native to a community

Aims:

- **Generic, extensible toolkit**
- **Language-agnostic interoperability**
- **Fair comparison among different implementations of algorithms**

M 2.3 — MaRDIMark

An general-purpose benchmarking framework for comparing implementations of algorithms using problems native to a community

Aims:

- **Generic, extensible toolkit**
- **Language-agnostic interoperability**
- **Fair comparison among different implementations of algorithms**
- **Flexible (community-driven) performance measures**

M 2.3 — MaRDIMark

Features:

- Compare existing implementations on new problems

M 2.3 — MaRDIMark

Features:

- Compare existing implementations on new problems
- Compare new algorithms/implementations to existing ones

M 2.3 — MaRDIMark

Features:

- Compare existing implementations on new problems
- Compare new algorithms/implementations to existing ones
- Use benchmark examples native to the community

M 2.3 — MaRDIMark

Features:

- Compare existing implementations on new problems
- Compare new algorithms/implementations to existing ones
- Use benchmark examples native to the community
- Measure performance according to community's standards

M 2.3 — MaRDIMark

Main Elements

Problems	Methods	Driver	Analysis	Explorer
data, metadata	code, executable, metadata	interfaces, parameters	performance measures	distillation, result browser
X2: Data	X2: Data	X1: Core	X2: Data	X4: Knowledge

Tasks

- Assembly of domain-independent specifications
- Database of curated benchmarks
- Result data (schemes, amounts, formats, raw or analyzed?)
- Classification, visualization?

Connections

- Uses knowledge graph (M 2.1)
- Uses open interfaces (M 2.2)
- Uses confirmable workflows (M 2.4)
- Has high synergetic potential (TA3)
- Integrates into MaRDI Portal (TA5)

M 2.3 — MaRDIMark

Main Elements

Problems	Methods	Driver	Analysis	Explorer
data, metadata	code, executable, metadata	interfaces, parameters	performance measures	distillation, result browser
X2: Data	X2: Data	X1: Core	X2: Data	X4: Knowledge

Tasks

- Assembly of domain-independent specifications
- Database of curated benchmarks
- Result data (schemes, amounts, formats, raw or analyzed?)
- Classification, visualization?

Connections

- Uses knowledge graph (M 2.1)
- Uses open interfaces (M 2.2)
- Uses confirmable workflows (M 2.4)
- Has high synergetic potential (TA3)
- Integrates into MaRDI Portal (TA5)

M 2.3 — MaRDIMark

Main Elements

Problems	Methods	Driver	Analysis	Explorer
data, metadata	code, executable, metadata	interfaces, parameters	performance measures	distillation, result browser
X2: Data	X2: Data	X1: Core	X2: Data	X4: Knowledge

Tasks

- Assembly of domain-independent specifications
- Database of curated benchmarks
- Result data (schemes, amounts, formats, raw or analyzed?)
- Classification, visualization?

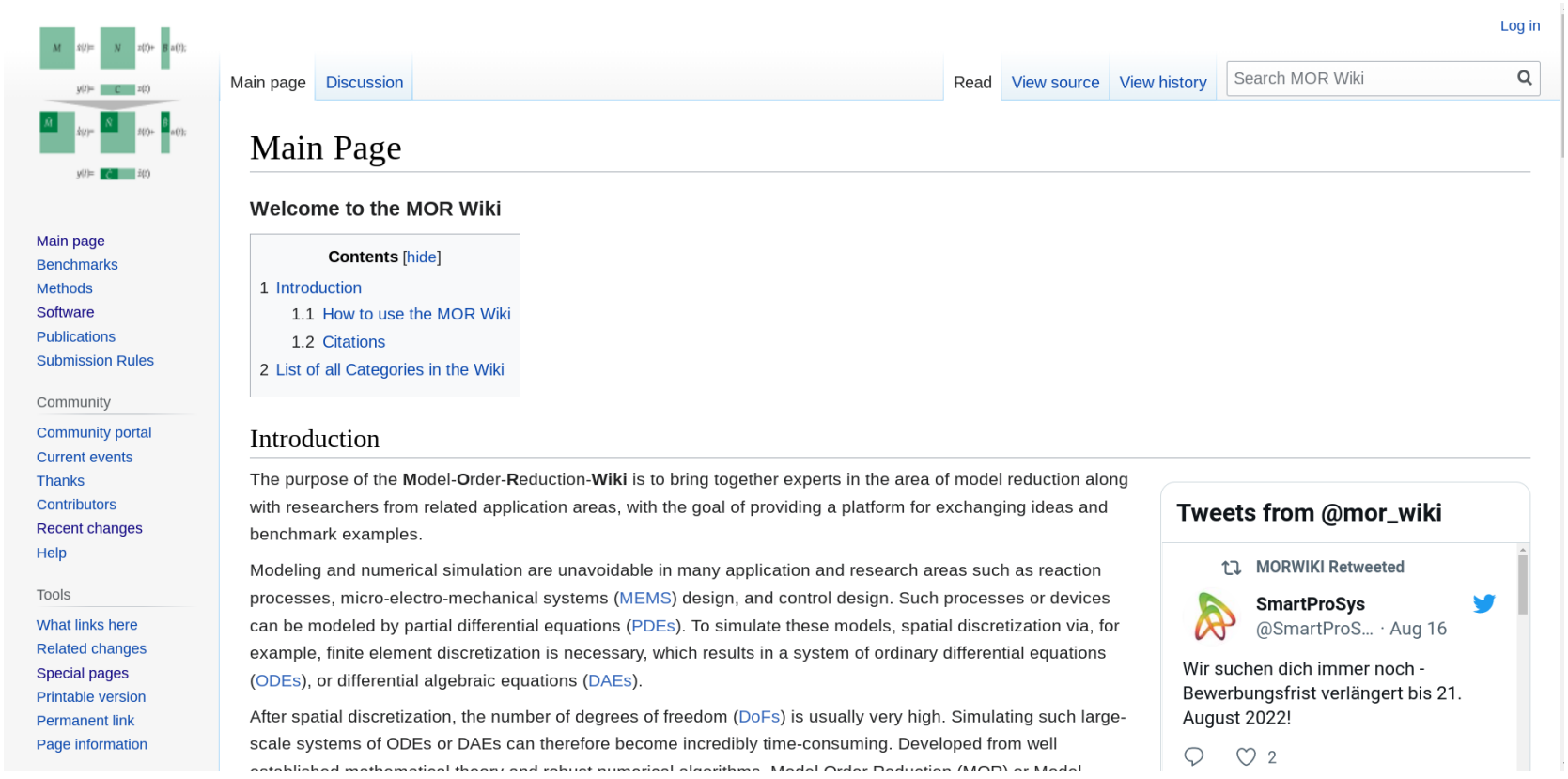
Connections

- Uses knowledge graph (M 2.1)
- Uses open interfaces (M 2.2)
- Uses confirmable workflows (M 2.4)
- Has high synergetic potential (TA3)
- Integrates into MaRDI Portal (TA5)

Outline

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection



http://modelreduction.org

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods

Lessons learned

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods
- **Collection of curated benchmark examples**

Lessons learned

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods
- **Collection of curated benchmark examples**
- **Description and comparison of available MOR software**

Lessons learned

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods
- **Collection of curated benchmark examples**
- **Description and comparison of available MOR software**
- MOR literature aggregation and BibTeX data

Lessons learned

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods
- **Collection of curated benchmark examples**
- **Description and comparison of available MOR software**
- MOR literature aggregation and BibTeX data
- Compilation of community events

Lessons learned

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods
- **Collection of curated benchmark examples**
- **Description and comparison of available MOR software**
- MOR literature aggregation and BibTeX data
- Compilation of community events

Lessons learned

- Encourage community engagement via low contribution barrier:
  - Easy access
  - Simple formats
  - Small rule-sets

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods
- **Collection of curated benchmark examples**
- **Description and comparison of available MOR software**
- MOR literature aggregation and BibTeX data
- Compilation of community events

Lessons learned

- Encourage community engagement via low contribution barrier:
  - Easy access
  - Simple formats
  - Small rule-sets
- Ensure content licensing and create proper citation culture

Model Order Reduction Wiki (MORWiki)

A community platform as a prototype for a curated benchmark collection

Services provided

- Descriptions of basic MOR methods
- **Collection of curated benchmark examples**
- **Description and comparison of available MOR software**
- MOR literature aggregation and BibTeX data
- Compilation of community events

Lessons learned

- Encourage community engagement via low contribution barrier:
  - Easy access
  - Simple formats
  - Small rule-sets
- Ensure content licensing and create proper citation culture
- Incentivize users for their contributions

Model Order Reduction Wiki (MORWiki)

Tasks and challenges

Licensing

- Old benchmark models without licenses
  - SLICOT Collection collected 2002–2006
  - Oberwolfach Collection circa 2005
- Benchmark descriptions and illustrations in the wiki

Model Order Reduction Wiki (MORWiki)

Tasks and challenges

Licensing

- Old benchmark models without licenses
  - SLICOT Collection collected 2002–2006
  - Oberwolfach Collection circa 2005
- Benchmark descriptions and illustrations in the wiki

Assembling metadata

- Classic data properties (authors, etc.)
- Mathematical properties of
  - the data implementing the benchmarks
  - the mathematical objects represented by the benchmarks
  - the mathematical properties of the objects of interest (theoretical / numerical)

Outline

Model Order Reduction Benchmarker (MORB)

The general specifications of the M 2.3-milestone

- Due after 36-month
- MORB will compute assessments of all benchmarks in the MORWiki
  - Fixed matrices
  - Data-only (trajectory snapshots, frequency samples, . . .)
  - Procedural generated
  - Linear and nonlinear
  - Dynamical systems, stationary and nonstationary PDEs

Model Order Reduction Benchmarker (MORB)

MORB 0.1 — benchmarking LTI Systems in MATLAB

Linear Time-Invariant (LTI) System

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t). \end{aligned} \quad \Leftrightarrow \quad H(s) = C(sE - A)^{-1}B$$

Model Order Reduction Benchmarker (MORB)

MORB 0.1 — benchmarking LTI Systems in MATLAB

Linear Time-Invariant (LTI) System

$$E\dot{x}(t) = Ax(t) + Bu(t), \qquad y(t) = Cx(t) + Du(t). \quad \Leftrightarrow \quad H(s) = C(sE - A)^{-1}B$$

Why so restrictive at the moment?

- KISS design principle
- Majority of MORWiki benchmarks is LTI or parametric LTI
- Most MOR software for LTI systems is written in MATLAB
- Systems theory knows a zoo of properties $\rightsquigarrow$ community-driven metadata!

Model Order Reduction Benchmarker (MORB)

MORB 0.1 — benchmarking LTI Systems in MATLAB

Challenges

- Ensuring all benchmark data is encoded uniformly
- Calling external software fairly, without unnecessary overhead
- Determining what counts as a unique implementation of an algorithm (algorithm isotope)
- Choosing subroutines that compute measures efficiently and accurately

Model Order Reduction Benchmarker (MORB)

MORB 0.1 — benchmarking LTI Systems in MATLAB

Challenges

- Ensuring all benchmark data is encoded uniformly
- Calling external software fairly, without unnecessary overhead
- Determining what counts as a unique implementation of an algorithm (algorithm isotope)
- Choosing subroutines that compute measures efficiently and accurately

Tasks

- Upload uniform benchmark data to Zenodo
- Automate computation of mathematical metadata
- establish database of benchmark-metadata
- Write download tool and wrappers for algorithm isotopes
- Establish unit tests

MaRDI TA2 Team


Jens Saak


Christian Himpe


Kathryn Lund


Mario Ohlberger


Stephan Rave


Peter Benner


René Fritze


Pavan Veluvali


Jan Heiland

MaRDI TA2 Team