2nd MaRDI Workshop on Scientific Computing

Wednesday 16 October 2024 - Friday 18 October 2024

Max Planck Institute for Dynamics of Complex Technical Systems

Book of Abstracts

Book of abstracts

Contents

14

Welcome and Opening Remarks

Corresponding Author: benner@mpi-magdeburg.mpg.de

Invited / **20**

preCICE: a FAIR coupling tool generating FAIR simulation results

Corresponding Author: miriam.schulte@ipvs.uni-stuttgart.de

preCICE is a numerical coupling library for partitioned multi-physics simulations with more than 200 users. It is fully open source and used for applicatiom as divers as
coupling reinforment learning with training data generating, various types of surface coupling such as fluid-structure interacions, and multi-scale simulations. It is not only
a library, but equipped with a whole ecosystem of additional tools to increase its usability. These tools serve, e.g., as adpaters allowing to couple specific single-physucs, single-scale solvers to other applications via preCICE or as a wrapper for a large number of micro-simulations into a single coupling participant. Some of these
tools are provided and maintained by users.

In the presentation, we present the main features of preCICE with some application examples and the whole software ecosystems and try to answer the following questions: What is FAIR research software? Is software simply a different type of data? And how can we generate FAIR data with such a software?

Invited / **7**

UM-Bridge: Enabling complex scientific applications

Author: Anne Reinarz[1]

Co-author: Linus Seelinger [2]

[1] *Durham University*

[2] *Karlsruhe Institute of Technology*

Corresponding Author: anne.k.reinarz@durham.ac.uk

Treating uncertainties in models is essential in many fields of science and engineering. When dealing with complex and computationally costly numerical models this necessitates a combination of efficient model solvers, advanced UQ methods and HPC-scale resources. The resulting technical complexities and software engineering challenges are holding back adoption of UQ methods in many fields. I will discuss the ways in which high-level abstractions can be used to break down technical complexity and enable a separation of concerns between experts. I will introduce UM-Bridge (the UQ and Modeling Bridge), a software protocol that facilitates universal interoperability of UQ software with simulation codes. Language-specific integrations make UM-Bridge models appear as native entities (classes, function calls etc.) in the respective programming language or even as native model classes in specific UQ packages.

Finally I will present a library of UQ benchmark problems, which we make available as an open-source software repository. UM-Bridge support makes the benchmarks available to virtually any UQ software and ensures portability and reproducibility via containerization.

Contributed / 9

A concrete CSE workflow framework for FAIR numerical experiments: A multi-layered approach

Author: Pavan L. Veluvali[1]

Co-authors: Jan Heiland ; Peter Benner

[1] *Max Planck Institute for Dynamics of Complex Technical Systems*

Corresponding Authors: veluvali@mpi-magdeburg.mpg.de, heiland@mpi-magdeburg.mpg.de, benner@mpi-magdeburg.mpg.de

Numerical algorithms and computational tools are essential for managing and analyzing complex data processing tasks. As meta-data and parameter-driven simulations have become more prevalent, the need for automated workflows to reproduce computational experiments across platforms has significantly increased. In general, a computational workflow is defined as a step-by-step description for accomplishing a scientific objective. Characterized through their input-output relation, computational workflows are designed such that the associated meta-data can be used interchangeably and redundantly. In the present work, we develop a computational framework, namely, MaRDIFlow, that focuses on the abstraction of meta-data while negating the underlying dependencies through multi-layered descriptions. Notably, by allowing the complete range between abstract descriptions and concrete numerical realizations (or even plain input-output data) of the tasks to serve equivalently and possibly redundantly in the definition of the workflows, we provide the lowest possible barrier for findable, accessible, interoperable and reusable (i.e. FAIR) workflow definitions. We showcase minimum working examples and how they are systematically incorporated into our workflow framework.

Contributed / 17

Improving Interoperability in Scientific Computing via MaRDI Open Interfaces

Author: Dmitry Kabanov[1]

Co-authors: Mario Ohlberger [1]; Stephan Rave

[1] *University of Münster*

Corresponding Authors: mario.ohlberger@uni-muenster.de, stephan.rave@uni-muenster.de, dmitry.kabanov@uni-muenster.de

In scientific computing, researchers often encounter two challenges while conducting computational experiments. First, many problems necessitate the use of multiple numerical solvers implemented in different programming languages. Second, numerical solvers for the same problem type typically have discrepancies in their programming interfaces. These factors hinder interoperability: connecting solvers together requires writing bindings between programming languages, while replacing one solver with another (for the same problem type) requires adapting the calling code to different function signatures.

To diminish these challenges, we work on the Measure 2.2 "Open Interfaces"of the MaRDI project. We are developing a library that automatically handles data marshalling and function calls across programming languages eliminating the need for manual bindings. Additionally, we are developing a set of interfaces for common numerical problems. Thus, via MaRDI Open Interfaces, interoperability of numerical solvers can be improved.

In the talk, we describe the software architecture and implementation details of the library. We also demonstrate its application using the developed open interface for solving initial-value problems for ordinary differential equations, as well as provide performance study comparing the library's overhead with direct usage of numerical solvers.

Invited / **5**

Reusability and Reproducibility Challenges in Commercial Software Development at Scale

Author: Manuel Baumann[1]

[1] *Dassault Systèmes Deutschland GmbH*

Corresponding Author: manuel.baumann@3ds.com

CST Studio Suite is a high-performance 3D EM analysis software package for designing, analyzing and optimizing electromagnetic (EM) components and systems. Its mathematical foundation dates back to the 1970s when the so-called Finite Integration Technique (FIT) was developed that is still the basis of CST's time-domain solver.

Nowadays, CST is a part of the multi-physics simulator SIMULIA and it is developed further by more than 100 software engineers that commit code changes on a daily basis. In my talk, I will present tools and workflows such as our "Testsuite" that are used at CST in order to guarantee high-standard industry code development.

Contributed / **4**

The MaRDI Portal - Overview & Outlook

Author: Tim Conrad[1]

Co-authors: Daniel Mietchen [2]; Eloi Ferrer [1]; Larissa Pusch [1]; Madhurima Deb [2]; Moritz Schubotz [2]; Rim Moussa [1]

[1] *Zuse Institut Berlin*

[2] *FIZ Karlsruhe*

Corresponding Author: conrad@zib.de

The MaRDI portal is a specialized platform developed as part of the MaRDI project under the National Research Data Infrastructure (NFDI) initiative, aimed at advancing the field of mathematics. The portal provides access to a comprehensive repository of metadata within and beyond the mathematical sciences, facilitating efficient discovery, management, and reuse of mathematical research data. This metadata is generated and structured through the underlying MaRDI knowledge graph, which allows for advanced interlinking and ensures that the metadata adheres to the principles of being Findable, Accessible, Interoperable, and Reusable (FAIR).

In this talk, we will present the current status of the MaRDI portal, along with an outlook on future developments and enhancements.

Invited / **3**

RDM within CRC/TR 287 BULK-REACTION

Corresponding Author: thevenin@ovgu.de

Research in engineering science is characterised by inter- and multi-disciplinarity. This leads to a strong heterogeneity of resulting research data, in particular when experimental as well as numerical investigations are performed collaboratively. The systematic and documented storage and archiving of this data are of fundamental importance. It provides the basis for the objective verification

and validation required during the development of new methods (measurement as well as numerical simulation techniques) and physical models. This can take place at the time of data generation but often also later and by independent research groups.
One of the primary objectives of CRC/TR287 BULK-REACTION is the generation of advanced, fundamental knowledge on the transport and reaction phenomena in high-temperature moving and reacting granular assemblies. The establishment of a reliable and long-lasting data management infrastructure is therefore paramount for the fulfilment of its scientific goals. It is also a necessary foundation for the creation of a community-specific scientific database that provides over a long period of time reference data for the considered processes. The central objective of the embedded INF project is to generate a long-term Research Data Management process and infrastructure for BULK-REACTION. This includes providing guidance, training, solutions and support regarding procedures and software such that an effective and appropriate handling of research data will be possible.

Invited / 8

Benchmarking for the UK Exascale programme

Author: Garth Wells[1]

[1] *University of Cambridge*

Corresponding Author: gnw20@cam.ac.uk

Procurements of high-performance computers usually involve benchmarking. Benchmarking is the element in a procurement where the fitness of a computer for its intended research uses is assessed by actual tests. Designing and defining these tests to be representative, fair and robust is not trivial and requires specific skills and is helped by experience. The 'return on investment' of good benchmarking can be very substantial. There is also a growing in interest in benchmarking for high-performance computing outside of procurements, particularly from cloud vendors, for assessing and demonstrating the (relative) performance of available systems.

I will talk about the developing benchmarking effort in the UK Pathway to Exascale programme and some similar nascent efforts. A key consideration is how we can make the development and maintenance of benchmarks a continuous and 'living' project and maintain skills in the community, rather following the historic pattern of discrete efforts. Inclusiveness, transparency and openness is central to our efforts. This promotes community building and supports our aim of a living benchmark suite.

Contributed / 12

Benchmarking-on-demand/Benchmarking-as-a-service: New concepts for numerical benchmarking in the age of Chatgpt and other AI tools

Author: Stefan Turek[1]

[1] *TU Dortmund*

Corresponding Author: stefan.turek@mathematik.tu-dortmund.de

Nowadays, it is more or less standard that newly proposed numerical algorithms and software tools are validated and evaluated by known and community-accepted benchmark results. This typically requires presenting corresponding numerical results for (at least) three different "grid sizes"(in terms of mesh widths in space and time) so that comparisons can be made with the corresponding reference results found in the literature. However, in the age of Chatgpt and similar AI tools, it seems increasingly possible to automatically provide corresponding numerical results that mimic the expected (asymptotic) behavior of the underlying methods in a way that makes it difficult even for specialists to adequately assess the quality of the newly proposed methods.

As an alternative, we want to discuss the concept of "benchmarking-on-demand"(resp. "benchmarking-as-a-service"), i.e. fully automated benchmark results for specific applications that are not known before publication, so that a more rigorous (and reliable) evaluation of new approaches becomes possible. However, this concept requires a network of participating "trusted"partners that can be certified to act as appropriate "benchmark centers" for various specific benchmarking cases. We illustrate the underlying concepts in detail with some CFD benchmarks that are commonly used and might be candidates for such specific and new benchmarking scenarios, among other cases.

Contributed / 18

Computer Algebra Data

Author: Michael Joswig[1]

[1] *TU Berlin*

Corresponding Author: joswig@math.tu-berlin.de

We will give a data-centric introduction to the new open source computer algebra system OSCAR: https://www.oscar-system.org

The OSCAR project develops a comprehensive Open Source Computer Algebra Research system for computations in algebra, geometry, and number theory, written in Julia. In particular, the emphasis is on supporting complex computations which require a high level of integration of tools from different mathematical areas.

Invited / 11

Cloud-Optimized Data and Dynamic Analysis Tools for Earth System Science

Author: Markus Reichstein[1]

Co-author: Fabian Gans

[1] *Max-Planck-Institute for Biogeochemistry*

Corresponding Author: reichstein-office@bgc-jena.mpg.de

Earth system science examines the complex interactions between the atmosphere, biosphere, hydrosphere, and other spheres, with the goal of understanding the planet's dynamic processes. A crucial aspect of this field is the collection and analysis of vast amounts of data, which provide insights into these interactions, both from simulations and observations. A prevalent data type in this field is gridded spatiotemporal data, typically represented as multidimensional arrays, also known as data cubes. In recent years storing and processing these data in cloud environments has gained adoption which resulted in development of new cloud-optimised data formats. In addition, the shift to cloud-based data analysis necessitates the evolution of analysis tools capable of handling gridded data efficiently. A significant challenge in this context is the joint analysis of data from different sources with varying spatial and temporal resolutions. Emerging solutions like on-the-fly data cube generation (kerchunk) and dynamic regridding facilitate seamless integration and analysis of diverse datasets, paving the way for more robust and flexible earth system science research.

Contributed / 19

Systems biology data management in de.NBI

Author: Wolfgang Mueller[1]

[1] *Heidelberg*

Corresponding Author: wolfgang.mueller@h-its.org

The de.NBI is the German (de) Network for Bioinformatics Infrascructure. It was initially funded by BMBF and is currently indirectly funded by the BMBF and the state of North Rhine-Westphalia via the FZ Jülich. Currently, it is organized into centers. One of these centers has systems biology as a topic, combining modelling tools COPASI and CellNetAnalyzer with data management tools. Systems biology is an interdisciplinary approach in which specialists from experiment, clinics and modelling are building an understanding of biological systems. Within the talk de.NBI, the center de.NBI-SysBio, the tools, standards, and platforms for data management within the center are described, seeking commonalities across subjects.

Invited / 6

Towards a Knowledge Graph for Models and Algorithms in Simulation Science – A Use Case in Amputation Surgery

Author: Dominik Göddeke[1]

[1] *University of Stuttgart*

Corresponding Author: dominik.goeddeke@mathematik.uni-stuttgart.de

Designing an ontology and creating a knowledge graph encompassing scientific computing, simulation science, mathematical modeling and a variety of application domains is a truly challenging task. In this talk, we will first sketch the MaRDI endeavor towards abstracting such an ontology and populating such a knowledge graph. Contributed talks by, e.g., F. Wübbeling, will dive into the details. We then sketch a multi-X forward-inverse simulate-optimise-invert use case, namely the novel limb amputation technique AMI (agonist-antagonist myoneural interface). Here, the quantity of interest is the prestretch of the artificial muscle-tendon link, applied during surgery. This renders the entire setting into a parameter identification problem, and ultimately, into an optimisation problem. We use this challenging application to showcase the power of the modeled knowledge, in particular with respect to automatically/iteratively tailoring numerical solution schemes to a given model hierarchy with variable target accuracies, precisions and resources. Based on hypothesized "what-if" questions, we demonstrate tradeoffs and benefits between precision (uncertainty), accuracy (certainty) and simulation budgets (runtime, hardware provisioning).

This is joint work with, in alphabetical order, C. Biedinger, M. Burger, C. Homs Poms, J. Fiedler, F. Huber, D. Iglezakis, H. Kleikamp, T. Klotz, T. Koprucki, R. Lautenschlager, B. Maier, M. Reidelbach, O. Röhrle, A. Shehu, B. Schembera, B. Schmidt, A. Schöbel, M. Schulte, T. Kabelow, F. Wübbeling.

Contributed / 2

Ontology for R language packages

Authors: Laszlo Nemeth[1]; Karsten Tabelow[1]; Thomas Koprucki[1]

[1] *Weierstrass Institute for Applied Analysis and Stochastics*

Corresponding Authors: nemeth@wias-berlin.de, karsten.tabelow@wias-berlin.de

Our aim is an ontology for R language packages published on The Comprehensive R Archive Network. First, we try to enhance the current metadata information available on CRAN. In addition, with a software extracting the enriched metadata information on package uploads that can fill the ontology with instances, connection to the Mathematical Research Data Initiative (MaRDI) portal and knowledge graph are established. Based on this ontology, the user experience can be further enhanced as well, e.g., automatic recommendations for CRAN task views or packages linked to publications, mathematical methods, algorithms, and models.

Contributed / **16**

MathAlgoDB –A Knowledge Graph for Scientific Computing

Author: Frank Wübbeling[1]

Co-authors: Hendrik Kleikamp ; Mario Ohlberger [1]; Stephan Rave

[1] *University of Münster*

Corresponding Authors: stephan.rave@uni-muenster.de, hendrik.kleikamp@uni-muenster.de, frank.wuebbeling@uni-muenster.de, mario.ohlberger@uni-muenster.de

In the MathAlgoDB (formerly AlgoData) project, we focus on the Findability part of the FAIR principles in scientific computing.

Currently, there is no way of directly discovering matching algorithms for a problem with specific properties, corresponding publications, benchmarks, and implementations. We fix this problem by creating a structured, curated database that adheres to a strict ontology and provide interfaces for easy access to its contents.

In a first step, an ontology for algorithms in scientific computing has been developed. A knowledge graph based on that ontology has been created in sample fields. A user interface that allows interaction with the graph has been implemented. Recently, an editor interface that allows appointed editors to easily add data and users to propose contents was established.

In the talk, we will focus on use cases, further development of the ontology, and the connection to the model ontology MathModDB.

Contributed / **13**

Imperative Knowledge Representation with PyIRK

Author: Carsten Knoll[1]

Co-author: Julius Fiedler [1]

[1] *TU Dresden*

Corresponding Authors: julius.fiedler@tu-dresden.de, carsten.knoll@tu-dresden.de

The semantic web as a concept was introduced more then 20 years ago. Nevertheless, compared to numerical AI (neural networks) the usage of semantic technologies (aka symbolic AI) in the mathematics-based sciences, including engineering, is still not widespread. In this talk we briefly evaluate existing technologies like RDF, OWL, Wikidata and ORKG for modeling mathematical knowledge such as definitions and theorems. We then present PyIRK –a Python based framework

for imperative representation of knowledge. The basic idea is to leverage the flexibility of a full-fledged programming language to formalize complex knowledge structures such as setting, premise and assertion of theorems with moderate effort while at the same time allowing for SPARQL queries and rule based reasoning. We further present the application of this approach to implement the Ontology of Control Systems Engineering (OCSE) by discussing the modeling of concepts and theorems from Lyapunov theory as examples. Finally, we also give an insight in our current attempt to leverage large language models to extract mathematical knowledge from LaTeX source code and transform it into PyIRK.

15

Closing Remarks

Corresponding Author: benner@mpi-magdeburg.mpg.de

Contributed / 21

Research data management in the curriculum

Corresponding Author: christiane.goergen@mis.mpg.de

I will provide a brief overview about successful and not-so-successful implementations of rdm topics in maths curricula, and I will discuss the aims of such implementations. I then provide a platform for discussion, together with the workshop audience, about how lecturers and researchers in scientific computing might adapt existing solutions and anchor them in their communities.

Contributed / 1

LiveDocs: Crafting Interactive Development Environments From Research Findings

Authors: Christoph Lehrenfeld[1]; Markus Osterhoff[1]; Pedro Costa Klein[1]

[1] *Univeristy of Göttingen*

Corresponding Authors: p.klein@math.uni-goettingen.de, mosterh1@gwdg.de, lehrenfeld@math.uni-goettingen.de

Open Science is a recurrent topic in scientific discussion, with a current effort to make research more accessible to a broader audience being observed. In this sense, a focus on delivering research findings that are reproducible, or even re-usable has been proposed as one way of achieving such accessibility goals.

In this talk, we will present LiveDocs: an initiative of the "Collaborative Research Center 1456 - Mathematics of Experiment" on tackling common issues of reproducibility and re-usability in scientific publications. LiveDocs is proposed as a concept alongside a collection of methods that enable scientists to provide research findings under an interactive development environment.

This allows users from a broader audience to easily reproduce research findings by re-running scripts, for instance, those that generate figures, tables, and other elements from scientific publications. Moreover, LiveDocs enable the audience to interact with code and data in such environments, thus allowing users to explore algorithms, datasets and software interfaces. This approach lowers the barriers to access and comprehend research methods and findings, which facilitates more scientific exchange and fosters knowledge advancement.