Contribution ID: **41** Type: **Poster**

Unified metadata handling for reproducible simulation workflows

Tuesday 10 October 2023 14:30 (15 minutes)

Computer simulations are an essential pillar of knowledge generation in science. Understanding, reproducing, and exploring the results of simulations relies on tracking and organizing metadata describing numerical experiments. However, the models used to understand real-world systems, and the computational machinery required to simulate them, are typically complex, and produce large amounts of heterogeneous metadata. Capturing and structuring these metadata along the processing chain is a vital requirement, for example, to make numerical experiments reproducible, to enable systematic benchmarking and validation of simulation software and models, to assess the reliability of simulations, and to foster data exploration and comparison [1,2]. Providing the ability to search, share, and evaluate metadata from heterogeneous simulations and environments is however a major challenge. The availability of a common metadata management framework, which can be adopted by scientists from different scientific domains, would therefore be highly desirable and foster the meta-analysis of HPC simulation workflows [3].
Here, we present a general concept for acquiring and handling metadata that is agnostic to software and hardware, and highly flexible for the user. It consists of two steps: 1) recording and storing raw metadata, and 2) selecting and structuring metadata in a configurable manner. We implement this concept in tools that can be attached to existing simulation workflows, and demonstrate it by applying our tools to distinct high-performance computing use cases from hydrology and neuroscience.

1. Guilyardi, E., et. al. (2013) doi: 10.1175/BAMS-D-11-00035.1
2. Manninen, T., et. al. (2018) doi: 10.3389/fninf.2018.00020
3. Ivie, P., & Thain, D. (2018). doi: 10.1145/3186266

Acknowledgments
The authors would like to thank Jan Bumberger, Helen Kollai, Michael Denker, Dennis Terhorst, Rainer Stotzka, Guido Trensch, and Stefan Sandfeld for ongoing Aug 21, 2023, 12:00 PM fruitful discussion. This project was funded by Helmholtz Metadata Collaboration (HMC) ZT-I-PF-3-026, EU Grant 945539 (HBP), Helmholtz IVF Grant SO-092 (ACA), and Joint lab SMHB; compute time was granted by VSR computation grant JINB33, Jülich. The work was carried out in part within the HMC Hub Information at the Forschungszentrum Jülich.

Please assign your contribution to one of the following topics

Infrastructure and common practices for consolidating (meta)data

Please specify ”other” (stakeholder)

In addition please add keywords.

Simulation workflow; metadata management framework.

Please assign yourself (presenting author) to one of the stakeholders.

Researchers

Primary authors: VILLAMAR, Jose (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research

Centre, Jülich, Germany; RWTH Aachen University, Aachen, Germany); KELBLING, Matthias (Dept. Computational Hydrosystems, Helmholtz-Centre for Environmental Research, Leipzig, Germany); MORE, Heather (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; Institute for Advanced Simulation (IAS-9), Jülich Research Centre, Jülich, Germany); TETZLAFF, Tom (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany); SENK, Johanna (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany); THOBER, Stephan (Dept. Computational Hydrosystems, Helmholtz-Centre for Environmental Research, Leipzig, Germany)

Presenter: VILLAMAR, Jose (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; RWTH Aachen University, Aachen, Germany)

Session Classification: Poster session

Track Classification: Poster session