Characterization of Data Provenance in Computational Engineering by an Ontological Representation of Simulation Workflows

Martin Thomas Horsch,[1] Christoph Niethammer,[2] Silvia Chiacchiera,[1] Peter Schiffels,[3] Michael A. Seaton,[1] Ilian T. Todorov,[1] Jadran Vrabec,[4] Philipp Neumann,[5] and Welchy Leite Cavalcanti[3]

[1] UK Research and Innovation, STFC Daresbury Laboratory, Keckwick Ln, Daresbury, Cheshire WA4 4AD, United Kingdom
[2] High Performance Computing Center Stuttgart, Nobelstr. 19, 70569 Stuttgart, Germany
[3] Fraunhofer Institute for Manufacturing Technology and Advanced Materials, Wiener Str. 12, 28359 Bremen, Germany
[4] Technische Universität Berlin, Thermodynamics and Process Engineering, Ernst-Reuter-Platz 1, 10587 Berlin, Germany
[5] Universität Hamburg, Scientific Computing, Bundesstr. 45A, 20146 Hamburg, Germany

martin.horsch@stfc.ac.uk

Abstract. This demo presents the Ontology for Simulation, Modelling, and Optimization (OSMO), i.e., a semantic asset that can be employed to formally represent simulation workflows in computational molecular engineering. It discusses technical aspects of the ontology and its OWL DL implementation as well as two related diagram notations for workflows: Reduced workflow graphs, i.e., MODA notation, and extended workflow graphs where MODA elements are connected to nodes that represent logical resources.

Keywords: Ontology, simulation workflow, materials modelling

1 Demo: Ontology for Simulation, Modelling, and Optimization

Where databases and platforms using different data structures and file formats interoperate, or where data from various sources are combined, agreement on semantics becomes a necessity. By an ontology, rules are formulated for entities (i.e., objects) and relations (i.e., properties of objects), which can guide the data ingest into databases and extend the possibilities of data extraction and evaluation by automated logical reasoning.

For data technology in computational molecular engineering, the characterization of workflows is relevant in two major ways. First, workflows are designed and communicated to simulation environments where materials models are evaluated to generate data by simulation. Second, in order to integrate data obtained in different ways (e.g., from simulation and experiment, or from simulations with different models or solvers), simulation results need to be stored together with metadata that describe their provenance, i.e., the process by which they have been produced. Hence, simulation workflows need to be described in a machine-readable way. There are a great variety of environments and languages for workflows (e.g., file formats and graph languages). Many of these, including AiiDA [1], MODA [2], Salome/YACS [3], and the TaLPas workflow and

performance modelling environment [4], are applicable to simulation workflows in materials modelling.

OSMO [5] is compatible with RoMM [6], and it is directly based on MODA; by providing a common semantic basis for workflows that were designed with different tools, OSMO can be employed to consistently integrate data provenance descriptions for materials modelling data from diverse sources [5]. The demonstration at DACOMSIN will illustrate how OSMO may be used in practice to provide a common framework for previously non-interoperable workflow environments.

Acknowledgements

The co-authors M. T. H., S. C., P. S., M. A. S., I. T. T., and W. L. C. acknowledge funding within the Virtual Materials Marketplace (VIMMP) project; this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 760907. The co-authors C. N., J. V., and P. N. acknowledge funding from the German Federal Ministry for Education and Research (BMBF) under grant agreement No. 01IH16008 (TaLPas).

References

1. Pizzia, G., Cepellotti, A., Sabatini, R., Marzari, N., and Kozinsky, B.: AiiDA – automated interactive infrastructure and database for computational science. Computational Materials Science **111**, 218–230 (2016).
2. CEN-CENELEC Management Centre: CEN Workshop Agreement 17284. CEN, Brussels (2018).
3. Ribes, A. and Caremoli, C.: Salomé platform component model for numerical simulation. In: Proceedings of COMPSAC 2007, pp. 553–563. IEEE Computer Society, Los Alamitos (2007).
4. Shudler, S., Vrabec, J., and Wolf, F.: Understanding the scalability of molecular simulation using empirical performance modelling. In: Programming and Performance Visualization Tools, pp. 125–143. Springer: Heidelberg (2018).
5. Horsch, M.T., Niethammer, C., Boccardo, G., Carbone, P., Chiacchiera, S., Chiricotto, M., Elliott, J.D., Lobaskin, V., Neumann, P., Schiffels, P., Seaton, M.A., Todorov, I.T., Vrabec, J., and Cavalcanti, W.L.: Semantic interoperability and characterization of data provenance in computational molecular engineering. To appear (2019).
6. De Baas, A.F.: What Makes a Material Function? EU Publications Office, Luxembourg (2017).