EGU24-12859, updated on 13 Aug 2024
https://doi.org/10.5194/egusphere-egu24-12859
EGU General Assembly 2024

The site selection data hub: a data-centric approach for integrated simulation workflow management in radioactive waste disposal site selection

Qian Chen[1], Marc S. Boxberg[2], Nino Menzel[2], Maria F. Morales Oreamuno[3], Wolfgang Nowak[3], Sergey Oladyshkin[3], Florian M. Wagner[2], and Julia Kowalski[1]
[1]Methods for Model-based Development in Computational Engineering, RWTH Aachen University, Aachen, Germany
[2]Geophysical Imaging and Monitoring, RWTH Aachen University, Aachen, Germany
[3]Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Germany

Given the importance of ensuring the safe disposal of radioactive waste, it is vital to understand the targeted subsurface systems and to build physics-based models to predict their dynamic responses to human interventions. Constructing robust predictive models, however, is very challenging due to the systems' complexity as well as the scarcity and cost of geophysical data acquisition. Optimal matching of data acquisition and predictive simulations is therefore necessary and can be achieved via integrating predictive process modeling, Bayesian parameter estimation, and optimal experimental design into a modular workflow. This allows to quantify the information content of measurement data and therefore enables optimal planning of data acquisition and monitoring strategies. Conducting such data-integrated simulation studies, however, requires a robust workflow management that ensures reproducibility, error management, and transparency.

To meet this demand, we established a data-centric approach to workflow control combining error-managed simulations with a functional data hub, providing simulations with direct access to a database of essential material properties. The latter are being made available as site specific scenario compilations along with uncertainty margins and meta information.

The data hub serves as an interface facilitating seamless data and simulation exchange to support subsequent model-driven decision-making processes and guarantees that simulations are conducted using manageable, comparable, and reproducible test cases. Furthermore, it ensures that the simulation results can be readily transferred to a designated repository allowing for real-time updates of the model. The implementation of the data hub is based on a Python-based framework for two different use cases:

1) GUI-based use case: The graphical user interface (GUI) facilitates data import, export, and visualization, featuring distinct sections for geographic data representation, structured table organization, and comprehensive visualization of physical properties in varying dimensions.

2) Module-based use case: Built on the YAML-based data-hub framework, it enables direct integration of simulation modules storing measurements and model parameters in the YAML data

format.

The data is systematically organized to furnish a versatile data selection framework that allows information to be extracted from a variety of references, including specific on-site measurements, laboratory measurements and other references, thereby enabling a comprehensive exploration of different reference-oriented scenarios.

This study showcases the data hub as a management infrastructure for executing a modular workflow. Multiple models—such as process and impact models as well as their surrogates and geophysical inverse models—are generated within this workflow utilizing scenarios provided by the data hub. Our study shows that adopting a data-centric approach to control the simulation workflow proves the feasibility of conducting different data-integrated simulations and enhances the interchangeability of information across different stages within the workflow. The paradigm of sustainable model development ensures reproducibility and transparency of our results, while also offering the possibility of synergetic exchange with other research areas.