



Contribution ID: 151

Type: **Presentation**

Seamless Integration of Data Sharing Repositories with High-Performance Computing Simulation Platform

Tuesday 12 March 2024 14:30 (15 minutes)

Scientific advancements increasingly rely on complex computational models fueled by diverse datasets. However, the collaborative sharing of these datasets poses significant challenges, hindering progress of research within scientific communities. This paper addresses the pivotal issue of efficient data sharing among scientists engaged in advanced simulations and computational modeling.

The proposed solution integrates the Model Execution Environment (MEE) with widely used data repositories such as Dataverse or Zenodo for seamless connection to data. This integration empowers scientists to access external data securely and in accordance with predefined rules directly within their workflow templates on the simulation platform. Notably, this approach eliminates the need for users to allocate local disk space, as the High-Performance Computing (HPC) system fetches the required data directly to the executing job directory.

A compelling use case scenario illustrates the efficiency of this solution: a research team collaboratively working on a computational model relies on external data stored in repositories. Seamless integration features allow team members to execute simulations without utilizing storage on their devices, streamlining the research process. Furthermore, the valuable research data generated during the simulation is being positioned for storage on the data sharing platform. The research team can control access to this data, ensuring that it remains within the consortium and preventing public dissemination until establishing a publication agreement.

This publication is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement Sano No 857533. This publication is supported by Sano project carried out within the International Research Agendas programme of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund. This publication is (partly) supported by the European Union's Horizon 2020 research and innovation programme under grant agreement ISW No 101016503. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2023/016227.

Primary authors: Mr ZAJĄC, Karol (Sano Centre for Computational Medicine); Mr ZHYHULIN, Taras (Sano Centre for Computational Medicine)

Co-authors: MEIZNER, Jan (Sano Centre for Computational Medicine); MALAWSKI, Maciej (AGH University of Krakow (PL)); Mr KASZTELNIK, Marek (ACC Cyfronet AGH); BUBAK, Marian (AGH Krakow); NOWAKOWSKI, Piotr (ACC Cyfronet AGH); Mr POŁEĆ, Piotr (ACC Cyfronet AGH)

Presenter: Mr ZHYHULIN, Taras (Sano Centre for Computational Medicine)

Session Classification: FAIR Data Management

Track Classification: Main sessions: User Voice: Innovative Applications, Data Science Environments & Open Data