Contribution ID: **224**   Type: **Presentation**

Infrastructure and Practices for Sharing and Disseminating In-Silico Medicine Research Data

Friday 21 March 2025 12:30 (15 minutes)

A significant amount of research data remains underutilized due to being unpublished or poorly described, leading to a loss of funding and scientific potential. To recover lost data and prevent further waste, researchers must be encouraged to use FAIR data sharing repositories and adopt good publishing practices, such as providing descriptive metadata and utilizing datasets from the community. Since this involves additional complications, the data retrieval and publication process should be simplified, and extra motivation should be provided.

Our solution comprises integration between the Model Execution Environment platform for in-silico model simulations on HPC resources, and open-source data sharing repositories, along with the required infrastructure, including an instance of the Dataverse repository, and a set of practices for its effective use, to enable convenient collaboration within the community. It aims to facilitate data management for execution of medical simulations, provide scientists with tools for cooperation, and engage the scientific community in further advancing research through data contributions. Additionally, the Model Execution Environment leverages HPC resources for the scientists, providing a straightforward interface and structuring complex computational workflows.

The aforementioned practices include rule-based data sharing based on an incentive-driven mechanism, fueling research even after the data becomes public. This approach is embodied by the Sano Dataverse instance, part of the RODBUK Krakow Open Research Data Repository, through the publication of a dataset from the DPValid case study, conducted by UNIBO within the InSilicoWorld project. The publication preparation process, which includes data processing on HPC using the MEE platform, uploading to Sano Dataverse or Zenodo, configuring rule-based data sharing, and ensuring ongoing curation and support, is presented on the basis of the requirements of the ISW scientific community which the authors are part of.

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

Co-authors: Ms BOTTIN, Francesca (Alma Mater Studiorum - University of Bologna); Dr DAVICO, Giorgio (Alma Mater Studiorum - University of Bologna); Mr STANIC, Goran (Alma Mater Studiorum - University of Bologna); MEIZNER, Jan (Sano Centre for Computational Medicine); MALAWSKI, Maciej (AGH University of Science and Technology); Mr KASZTELNIK, Marek (ACC Cyfronet AGH); BUBAK, Marian (AGH Krakow); NOWAKOWSKI, Piotr (ACC Cyfronet AGH); Mr POŁEĆ, Piotr (ACC Cyfronet AGH)

Presenter: ZHYHULIN, Taras

Session Classification: Data sharing infrastructures

Track Classification: Main sessions: CS3 federations and synergies with eResearch infrastructures.