Computational Science

The Role of FAIR in Machine Learning and Simulation Workflows

Dr. Polina Shpilker, Postdoctoral Research Associate
Workflow and Ecosystem Services
Oak Ridge National Laboratory, TN

Findable, Accessible, Interoperable, Re-usable (FAIR) principles have been explored, applied, and implemented for over a decade, following the coalescing of ideas and the development of tools around data, software, workflows, and now machine learning. Scientific workflows, an emerging execution mode for complex scientific experiments in many domains, including molecular biology and high-energy physics, are a preferred way for integrating machine learning and traditional simulations. Workflows enable the reproducibility of computational experiments by formally separating a workflow specification from its instantiation and the execution of a particular workflow. Reproducibility, a fundamental tenet of the scientific method, benefits from the implementation of FAIR principles in experiments, when data and software are published with the goal of reproducing results. To improve reproducibility, one must measure and characterize it at scale. In HPC workflows, providing the ability to reproduce results may entail extracting, curating and reconciling metadata, comparing execution patterns, and measuring result variability. Experiments that include machine learning models add requirements to make data and models FAIR; adhering to FAIR is necessary but not sufficient to ensure trust in ML results. This talk will discuss FAIR and its limitations for reproducibility in high performance and data-intensive computing. We will present some solutions and research directions developed in RECUP, a US Department of Energy Advanced Scientific Computing Research program-sponsored project, for scalable, reproducible workflows.

Monday, April 27, 2026, 4:00 PM

Please pre-register for the Zoom meeting at
https://us06web.zoom.us/meeting/register/oQUgmZkwR2SXv_mHGVLMqA

We acknowledge the Sustainable Horizons Institute and DOE national laboratories for organizing this Computational Research Leadership Council Seminar Series.
For more information, please contact us at computationalscience@utep.edu.