

Hewlett Packard Enterprise



Enrichment and Acceleration of Edge to Exascale Computational Steering STEM Workflow using Common Metadata Framework

Enabling efficiency and reproducibility in AI-Driven Research

Gayathri Saranathan^{*}, Martin Foltin^{*}, Aalap Tripathy^{*}, Annmary Justine^{*}, Ayana Ghosh⁺, Maxim Ziatdinov[^], Kevin Roccapriore⁺, Suparna Bhattacharya^{*}

May 8th, 2024

- * Hewlett Packard Labs HPE
- + Oak Ridge National Laboratory
- ^ Pacific Northwest National Laboratory



Evolving Landscape of Scientific Research



Challenges in Federated Research Environments

Agenda



Introducing FAIR Data Principles

Federated Common Metadata Tracking Framework

Deep Kernel Learning & Meta-learning in Microscopy



Case Study: From Experiment to Simulation in Microscopy





Questions?

Cray User Group

Edge-to-Exascale Autonomous Instrumentation Workflow – Vision









Data silos

Current Challenges in Federated Research Environments



Time consuming data transfers



Expensive workflow iterations



Difficulty in tracking metadata and reproducibility

FAIR Data Principles





Common Metadata Framework

Open-source Metadata and Lineage Tracking Tool



End-to-End Tracking with CMF from Edge-to-HPC





Scientific Workflow with CMF

Case Study : From Experiment to Simulation in Microscopy



Deep Kernel Learning for Microscopy



Federated CMF for Microscopy Workflow: From Experiment to Simulations







CMF Logged Lineage for Experimental Data

Data Processing and Meta-model Training





CMF Logged Lineage for Experimental Data – Meta model Training (3 Iterations)



CMF Logged Lineage for Experimental Data – Active Learning







Ongoing Work: Identifying Stable Structures with MD Simulations



Mask S traj 0







Simulated Dataset – MoS₂



Ground TruthActive Learning Points

Experiment Guided Active Learning Exploration













> cmf metadata pull> cmf artifact pull





Increased research efficiency: Spend less time on data management and more time on sci entific discovery.



Improved collaboration: Seamlessly share data and models with colleagues across differen t institutions and computing environments.

Benefits of CMF for Scientific Workflows



Enhanced reproducibility: Ensure the reliability and validity of research findings.



Simplified compliance: Adhere to FAIR principles and other data management standards more easily.



Cost optimization: Make efficient use of resources across Edge to HPC platforms.



Questions ?

CALL FOR ACTION



CMF

GIT HUB https://github.com/HewlettPackard/cmf/

Slack commonmetadata.slack.com <u>https://join.slack.com/t/commonmetadata/shared_invite/zt-2cwfar9cl-55cY6Ugh_p5GJRuRRkcQBA</u>

