

Integrated Research Infrastructure(IRI) Vision and Directions for OLCF

Mallikarjun Shankar

Section Head, Advanced Technology Section

Feiyi Wang

Group Lead, Analytics and AI Methods at Scale

ORNL is managed by UT-Battelle, LLC for the US Department of Energy



OLCF and ORNL Institutional CADES have supported Cross-Facility Workflows and Data Science for several years



E.J. Lingerfelt, A. Belianinov, E. Endeve, O. Ovchinnikov, S. Somnath, J.M. Borreguero, N. Grodowitz, B. Park, R.K. Archibald, C.T. Symons, S.V. Kalinin, O.E.B. Messer, M. Shankar, S. Jesse, BEAM: A Computational Workflow System for Managing and Modeling Material Characterization Data in HPC Environments, Procedia Computer Science, Volume 80, 2016, Pages 2276-2280,



Continued Work on Edge to Exascale Use-cases: E.g., Autonomous Microscopy (more from Rama Vasudevan)



Cross-Facility AI – SNS to OLCF: Bragg Peak Detection Workflow







User Display and Feedback

Toward an Autonomous Workflow for Single Crystal Neutron Diffraction, Junqi Yin, Guannan Zhang, Huibo Cao, Sajal Dash, Bryan C. Chakoumakos, and Feiyi Wang; https://link.springer.com/chapter/10.1007/978-3-031-23606-8 15



Using Edge-to-Exascale/Converged AI Pipelines



CAK RIDGE

5

Toward an Autonomous Workflow for Single Crystal Neutron Diffraction, Junqi Yin, Guannan Zhang, Huibo Cao, Sajal Dash, Bryan C. Chakoumakos, and Feiyi Wang; https://link.springer.com/chapter/10.1007/978-3-031-23606-8_15

Open slide master to edit

Also Moving Towards Unified Management of Data across Facilities



Let's make these more uniform: INTERSECT

AI-Workflow



Feiyi Wang; https://link.springer.com/chapter/10.1007/978-3-031-23606-8_15





INTERCONNECTED SCIENCE ECOSYSTEM

Across DOE, innovators have been taking similar concerted steps towards integration through research, partnerships, and lab-level projects

LBNL's Superfacility project ORNL's INTERSECT initiative ANL's ALCF-APS Nexus project

CAK RIDGE

These are all **separate** initiatives with **similar** integration goals. Let's now row in the same direction!





DOE Role

DOE is positioned to lead the new era of integrated science within the US Government and the world.

Linking distributed resources is becoming paramount to modern collaborative science.

The challenges of our time call upon DOE and its national laboratories to be an open innovation ecosystem:

Accelerating discovery & innovation Democratizing access Drawing new talent Advancing open science



5. Department of Emergy

National Virtual

Biotechnology Laboratory



IRI vision: A DOE/SC integrated research ecosystem that transforms science via seamless interoperability



CAK RIDGE Slide Credit: DOE SC/ASCR Dr. Benjamin Brown

The IRI Blueprint Activity Created a Framework for IRI Implementation



The IRI Framework comprises:

- IRI Science Patterns (3) represent integrated science use cases across DOE science domains.
 - > Provide the basis for organizing diverse program requirements into strategic priorities.
- IRI Practice Areas (6) represent critical topics that require close coordination to realize and sustain a thriving IRI ecosystem across DOE institutions.
 Provide the basis for organizing the program
 - governance model and cross-cutting efforts.

Convened over **150 DOE national laboratory experts** from **all 28 SC user facilities** across **13 national laboratories** to consider the **technological, policy, and sociological challenges** to implementing IRI.



The IRI Blueprint Activity Created a Framework for IRI Implementation

CENERGY Construction THE DOC OFFICE OF SCIENCE Infograted Research Infrastructure Architecture Blueprint Activity PRAN REPORT 2000

IRI Science Patterns (3)

Time-sensitive pattern has *urgency*, requiring real-time or end-to-end performance with high reliability, e.g., for timely decision-making, experiment steering, and virtual proximity.

Data integration-intensive pattern

requires combining and analyzing data from multiple sources, e.g., sites, experiments, and/or computational runs.

Long-term campaign pattern requires sustained access to resources over a long period to accomplish a well-defined objective.

Note: the patterns are not mutually exclusive.

IRI Practice Areas (6)

User experience practice will ensure relentless attention to user perspectives and needs through requirements gathering, user-centric (co)-design, continuous feedback, and other means.

Resource co-operations practice is focused on creating new modes of cooperation, collaboration, co-scheduling, and joint planning across facilities and DOE programs.

Cybersecurity and federated access practice is focused on creating novel solutions that enable seamless scientific collaboration within a secure and trusted IRI ecosystem.

Workflows, interfaces, and automation practice is focused on creating novel solutions that facilitate the dynamic assembly of components across facilities into end-to-end IRI pipelines.

Scientific data life cycle practice is focused on ensuring that users can manage their data and metadata across facilities from inception to curation, archiving, dissemination, and publication.

Portable/scalable solutions practice is focused on ensuring that transitions can be made across heterogeneous facilities (portability) and from smaller to larger resources (scalability).



Multi-Lab and Program Participants

SC Science Programs IRI Coordination Group

- BER Paul Bayer, Jay Hnilo, Resham Kulkarni
- BES Dava Keavney, Tom Russell, Misha Zhernenkov
- FES Josh King, Matt Lanctot
- HEP Jeremy Love, Eric Church
- IP Julie Ezold, Kristian Myhre
- NP Xiaofeng Guo, Paul Mantica, Jim Sowinski

ASCR Colleagues

13

ASCR Facility Directors

IRI Blueprint Activity participants!

IRI Blueprint Activity Leadership Group

HQ Executive Leadership



Ben Brown Director ASCR Facilities Division

Debbie Bard Group Lead for Data Science Engagement NERSC, LBNL

Amber Boehnlein Kjiersten Fagnan Chief Information Officer Chief Informatics Officer JLab

IRI-ABA Leadership Group



JGI, LBNL

OLCF/NCCS, ORNL



Chin Guol Group Lead for Planning and Architecture ESnet, LBNL



Fric Lancor Senior Technical Advisor Director, Scientific Data ASCR Facilities Division and Computing Center

Jini Ramprakash Deputy Division Director ALCF, ANL

Ariun Shankar Section Head, Advanced Technologies



Nicholas Group Leader, Scientific Software Eng. & Data Mgmt.. APS, ANL

ASCR Facilities IRI Task Force

BNI

Corey Adams Katie Antypas Debbie Bard Shane Canon Eli Dart Chin Guok Ezra Kissel

Eric Lancon Bronson Messer Sarp Oral Jini Ramprakash Arjun Shankar Tom Uram



ASCR is taking the first big steps

- Invest in IRI foundational infrastructure
- Bring existing IRI projects into formal coordination
- Deploy an IRI Pathfinding Testbed across the four ASCR Facilities
- Stand up an IRI Program structure (HQ and Field)

Implementing IRI: Today ASCR is taking the first big steps

1. Invest in IRI foundational infrastructure

Early focus: Report IRI Science Pattern requirements across all SC Programs and assess IRI readiness Early focus: Advance the NERSC-10, High Performance Data Facility, and OLCF-6 projects to CD-1

- 2. Bring existing IRI projects into formal coordination Early focus: Move towards a common user experience across the ASCR HPC Facilities Early focus: Create an IRI HPC allocation
- 3. Deploy an IRI Pathfinding Testbed spanning the four ASCR Facilities Early focus: Initiate projects for the Time-Sensitive and Data-Integration Intensive patterns Early focus: Engage ASCR Facilities with new IRI-oriented research projects
- 4. Stand up an IRI Program structure (HQ and Field) Early focus: Develop a DOE Authentication/Authorization standard Early focus: Devise and revise governance structure with the community

Time-Sensitive Pattern

Data-Integration Intensive Pattern

Long-Term Campaign Pattern



Office of Science

U.S. Department of Energy Selects the High Performance Data Facility Lead

OCTOBER 16, 2023

The High Performance Data Facility is envisioned as a national resource that will serve as the foundation for advancing DOE's ambitious Integrated Research Infrastructure (IRI) program. The IRI aims to provide researchers the ability to seamlessly meld DOE's unique data resources, experimental user facilities, and advanced computing resources to accelerate the pace of discovery. The mission of the HPDF will be to enable



OLCF Supporting the IRI Testbed: Enabling Facilities Integration





OLCF6 - Draft Technical Requirements Offers Workflows Context to Enable IRI



Front-end-nodes, network connectivity, storage, and software target requirements will offer greater workflow facilitation



Thank You!

