



The Next Generation Earth System Grid Federation (ESGF2)

Forrest M. Hoffman (ORNL), Ian Foster (ANL), and Sasha Ames (LLNL)

DOE Urban-IFL Kickoff Meeting
November 29, 2022



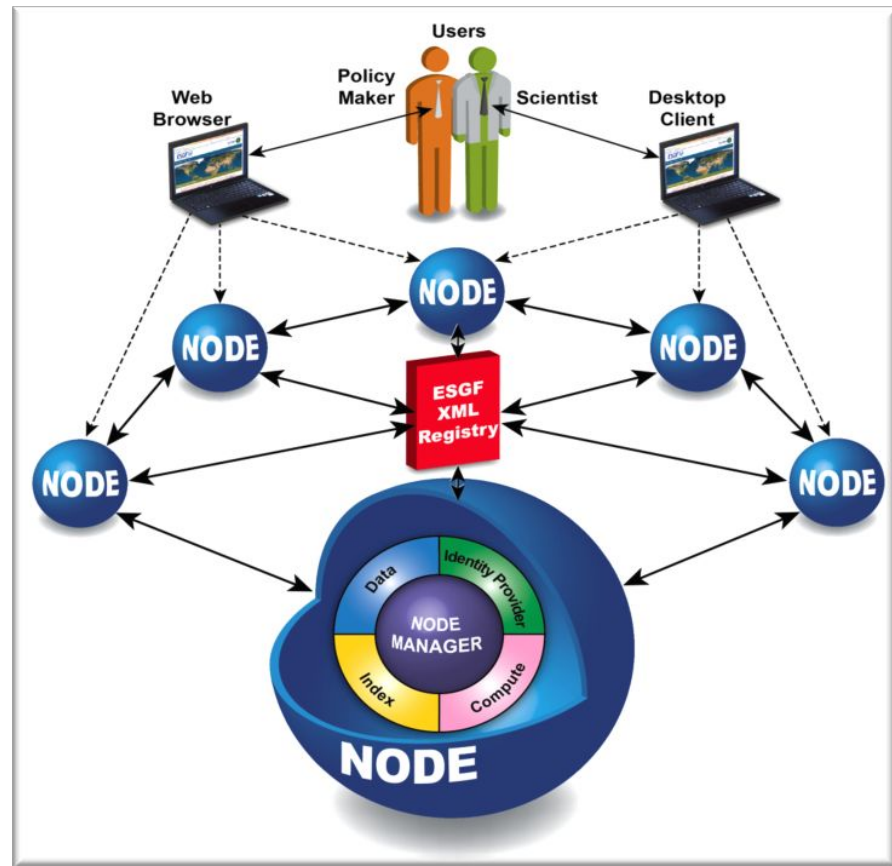
U.S. DEPARTMENT OF
ENERGY

Office of
Science



What is the Earth System Grid Federation?

- The **Earth System Grid Federation (ESGF)** is a globally distributed peer-to-peer network of data servers using a common set of protocols and interfaces to archive and distribute Earth system model (ESM) output
- ESM output data are used by scientists all over the world to investigate consequences of possible climate change scenarios and the resulting Earth system feedbacks



ESGF2 IPCC AR6 Released

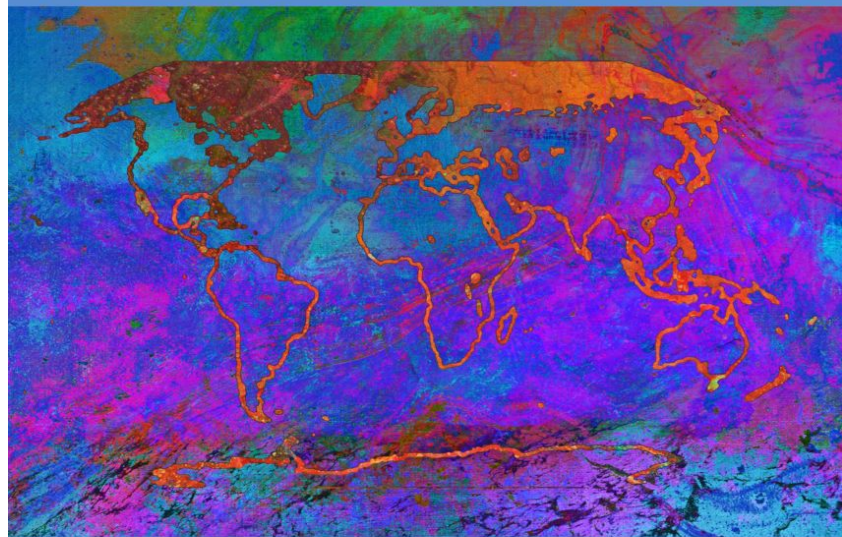
- The United Nations' Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report from Working Group I was released on Monday, August 9, 2021
- Most of the climate and Earth system model simulation output underpinning this report was produced by modeling centers participating in the World Climate Research Programme's (WCRP's) sixth phase of the Coupled Model Intercomparison Project (CMIP6)
- Nearly all of that model output was stored in and distributed to researchers via ESGF
- **Data are about the future of life on Earth!**

ipcc

INTERGOVERNMENTAL PANEL ON Climate change

Climate Change 2021

The Physical Science Basis



WGI

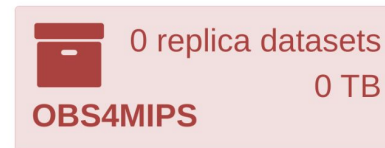
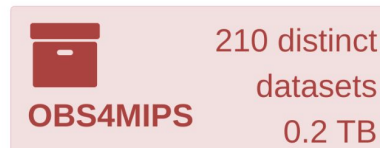
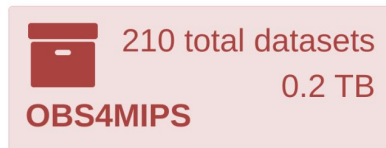
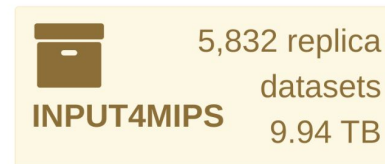
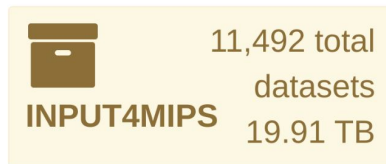
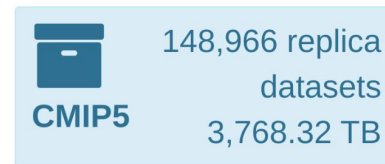
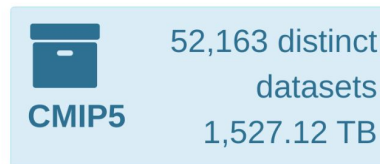
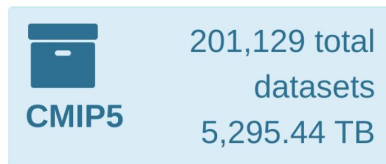
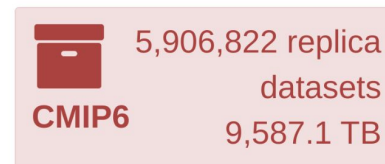
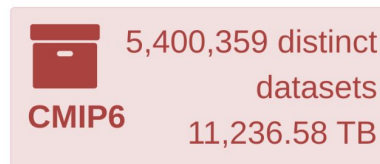
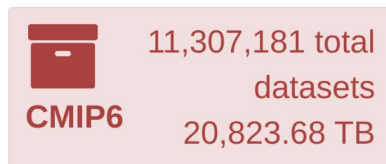
Working Group I contribution to the
Sixth Assessment Report of the
Intergovernmental Panel on Climate Change





ESGF Holdings are Large and Growing

- CMIP5 totals >5 PB
- CMIP6 totals >20 PB
- We expect CMIP7 output, including high resolutions simulations and more ensembles, to total >100 PB
- We plan to expand Federation holdings by adding other Earth science data projects





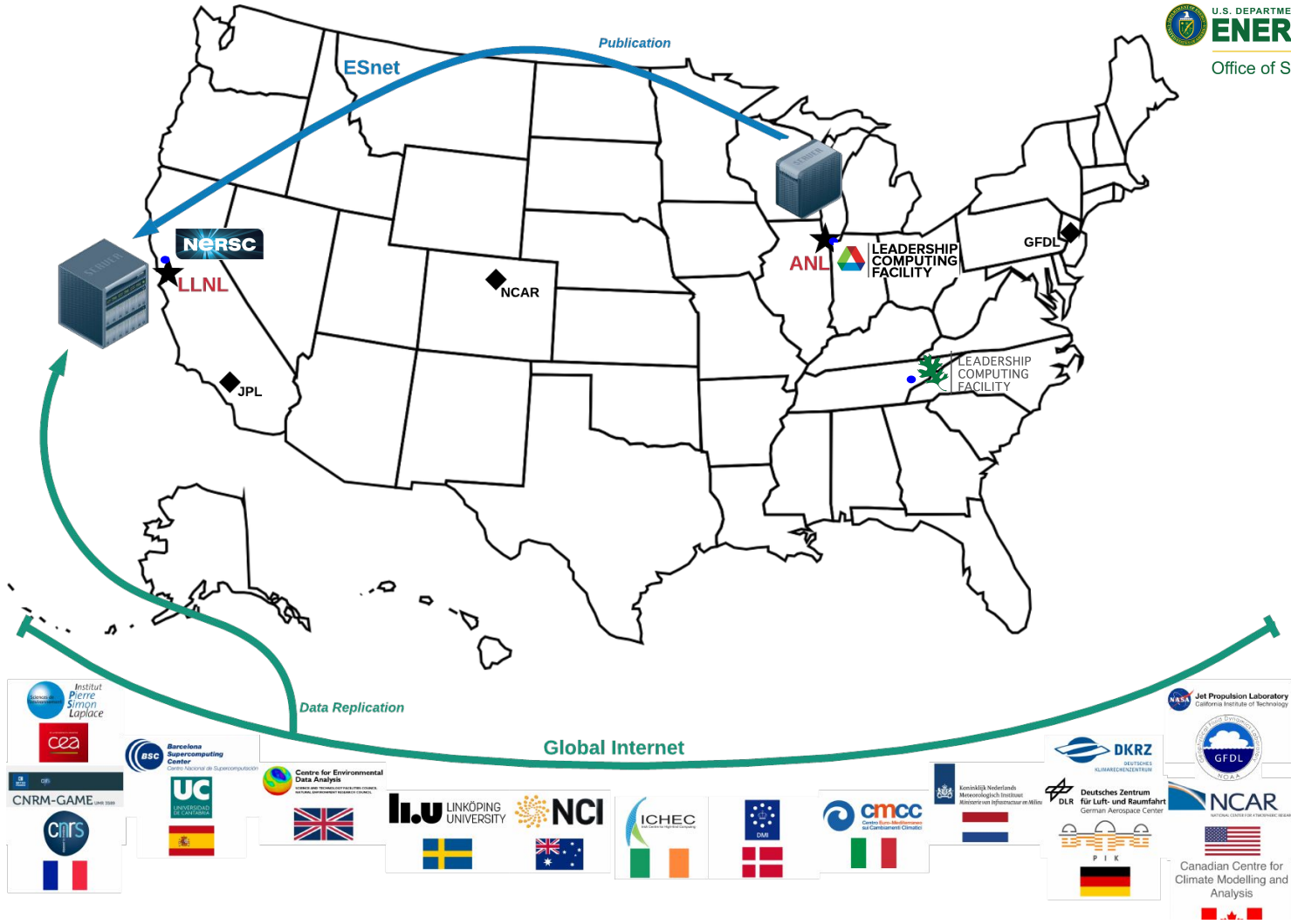
ESGF2 A New Consortium Project in the USA

- New team from **Oak Ridge National Laboratory**, **Argonne National Laboratory**, and **Lawrence Livermore National Laboratory** proposed to modernize the data backplane and deliver new computational capabilities
- In collaboration with the **ESGF Executive Committee**, we will develop and deploy a new architecture based on the *Future Architecture Roadmap*
- In addition, we will develop new **data discovery tools and data access interfaces**, **server-side computing** (subsetting & summarizing), and **user computing** (Kubernetes & JupyterHub) with improved **user & system metrics**
- We will add a **Resource & Project Liaison** group and a **Science, User & Facility Advisory Board**; hold outreach activities; and offer a help desk/user support
- **ESGF2-US will add new data projects**, going beyond CMIP-related data, to support needs of Earth and environmental science projects, especially for DOE



DOE's Current Earth System Grid Federation

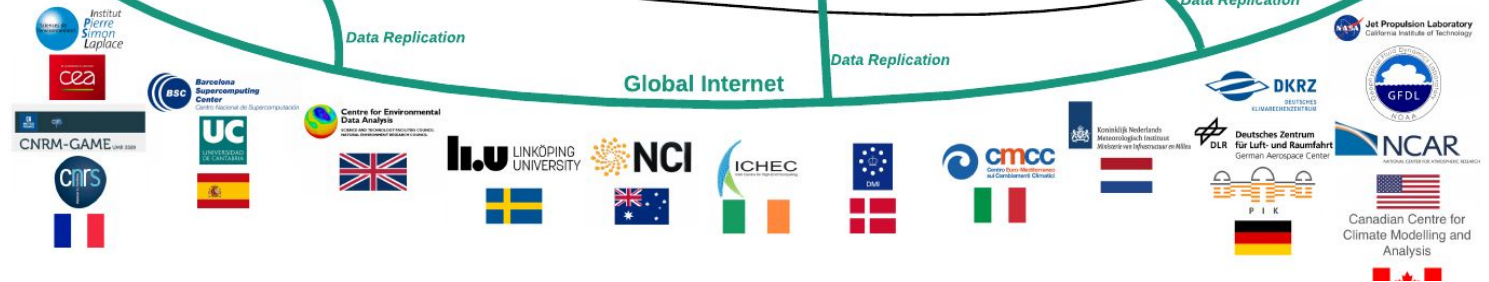
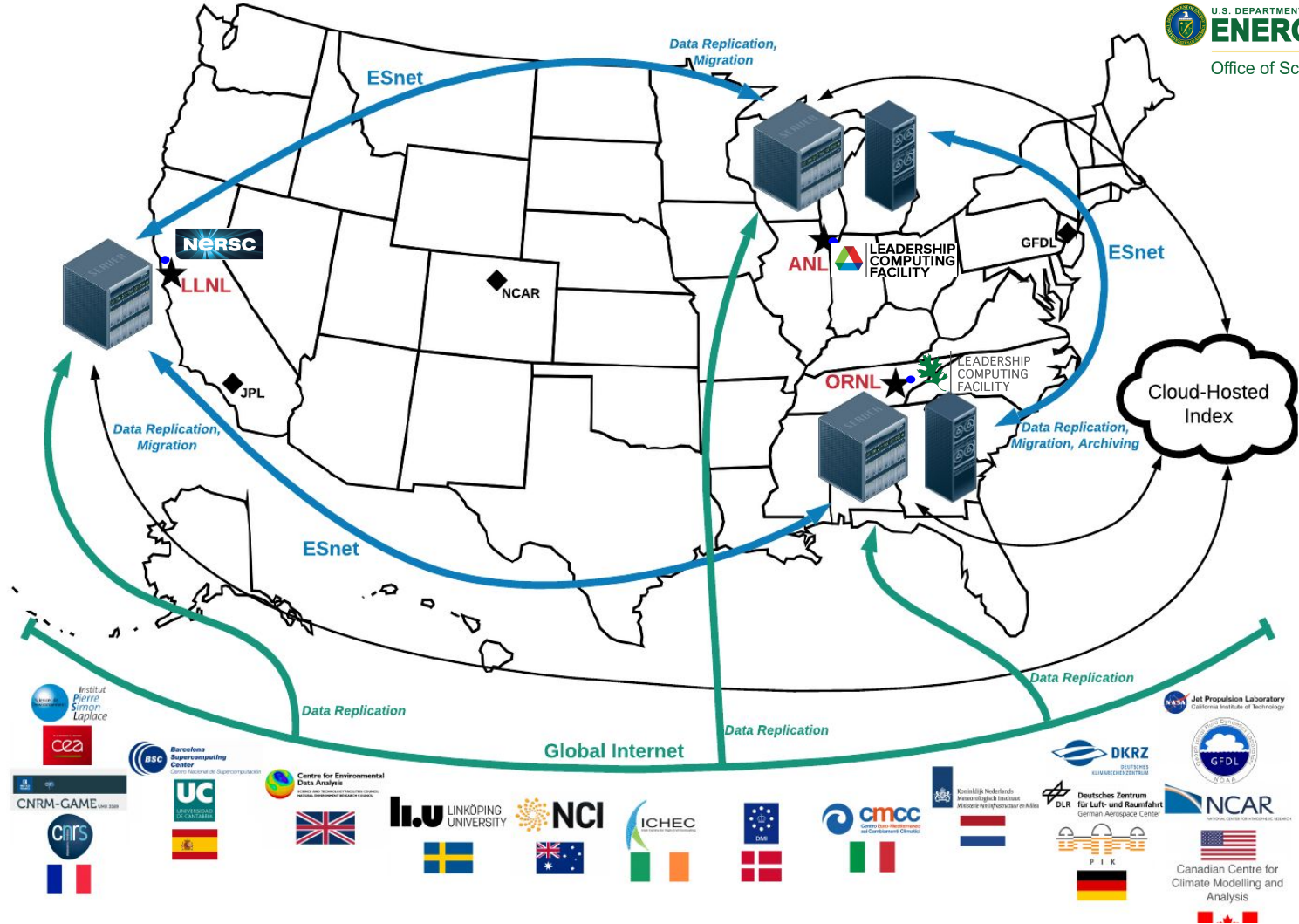
- Primary server at LLNL
- Replicating data from the global Federation
- Independent data node at ANL



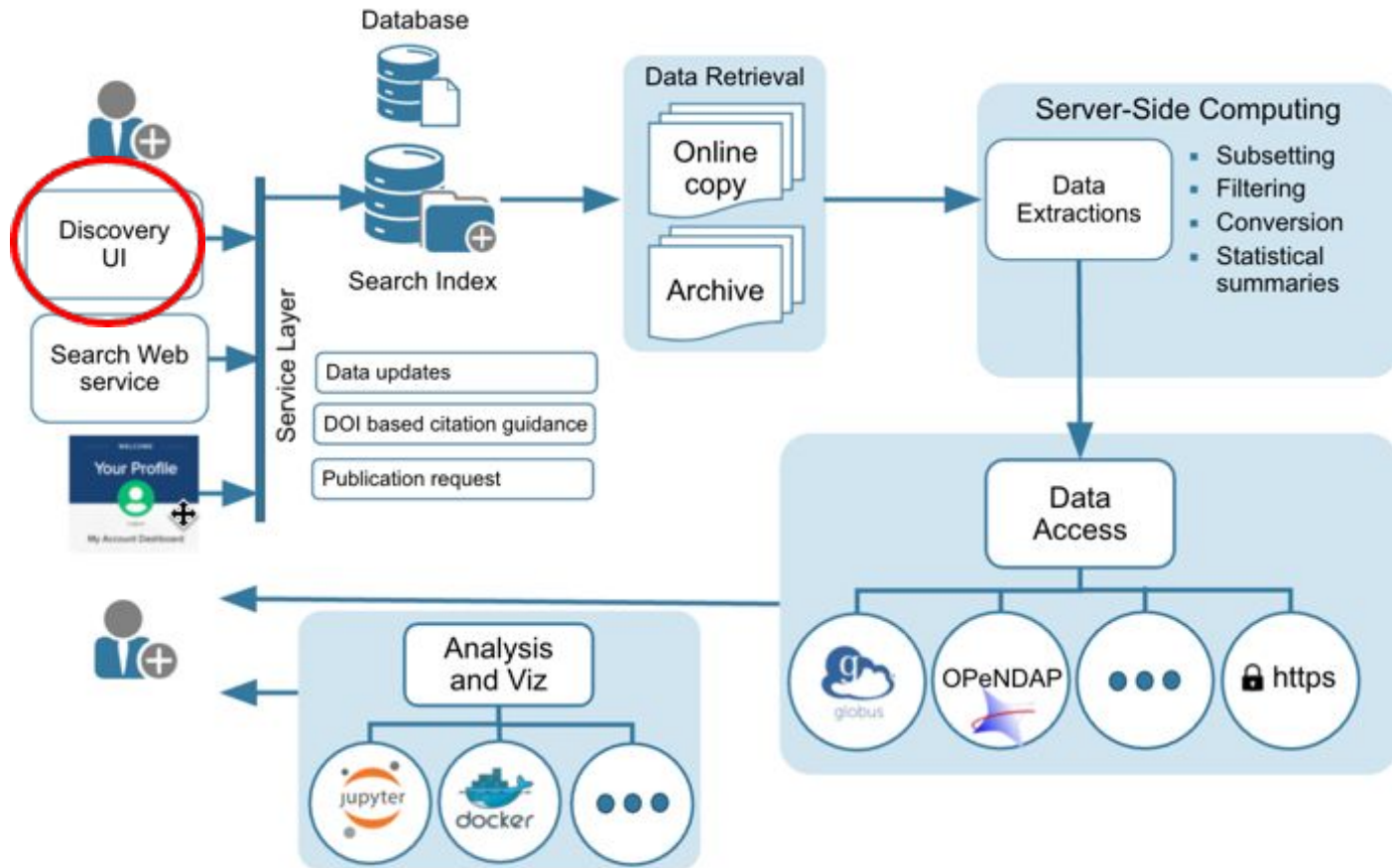


DOE's Next Generation Earth System Grid Federation

- Co-located at DOE's major computing facilities
- Replicating data from the global Federation
- Providing cloud indexing, automated migration, and tape archiving



ESGF2 Data Discovery Platform: Architecture



ESGF2 Summary

- **Designed to store and distribute large model (and related observational) data**, ESGF should be considered as a platform for distributing such data from projects to the science community and the global public
- We will soon be providing **Digital Object Identifiers (DOIs)** for data collections, suitable for citing in papers
- Currently, ESGF is designed to support **netCDF format data**, typical of model output, and we are considering methods for supporting other, project-specific data formats
- The next generation ESGF2-US technology will be **well integrated with Globus**, so large data movement and access will be faster and more reliable
- For more information, see the **ESGF website** at <https://esgf.llnl.gov/> and contact **Forrest Hoffman** <hoffmanfm@ornl.gov>

