



ESGF2 Project Overview

Forrest M. Hoffman, Ian Foster, and Sasha Ames



U.S. DEPARTMENT OF
ENERGY

Office of
Science





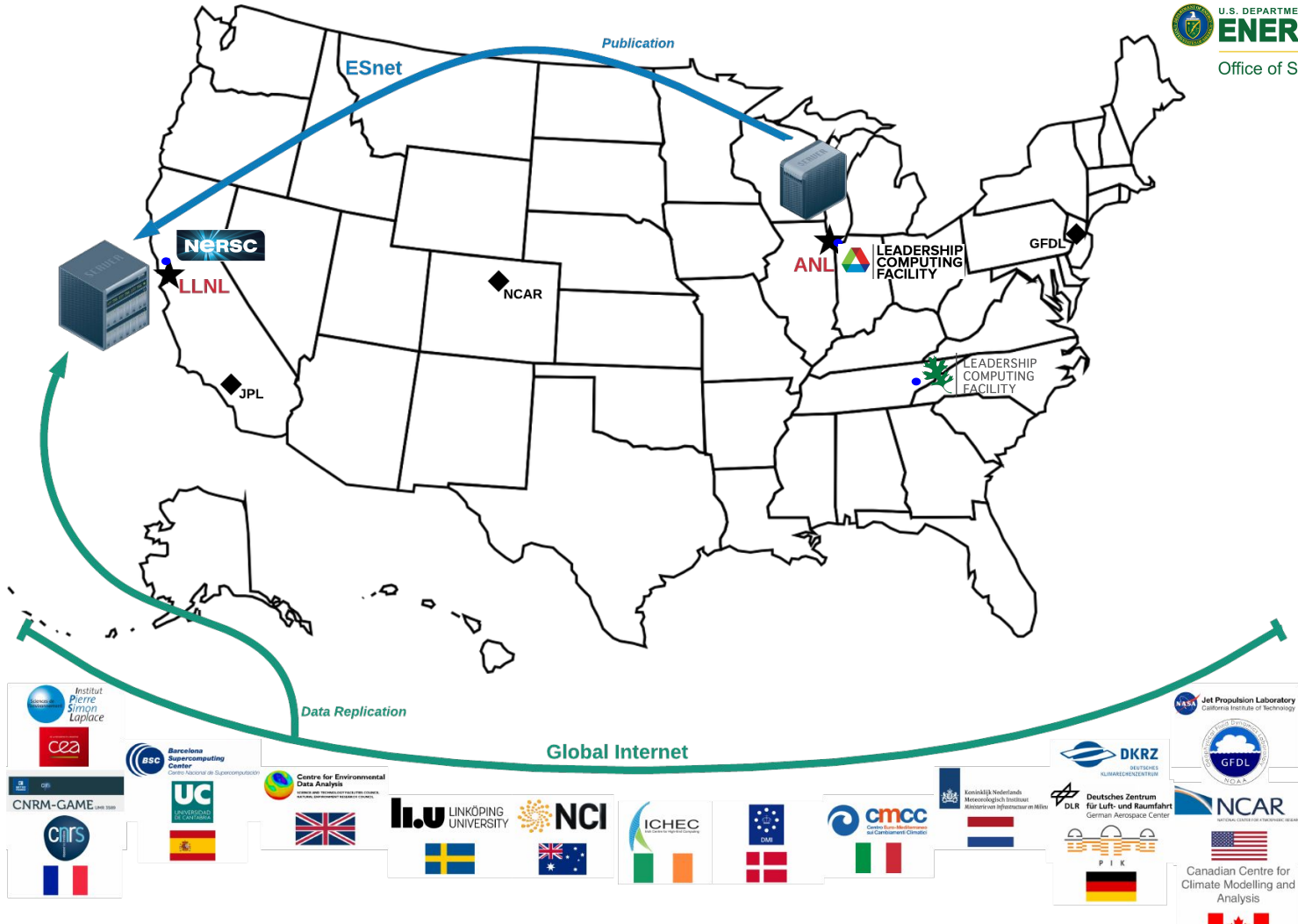
ESGF2 A New Consortium Project in the USA

- New team from **Oak Ridge National Laboratory**, **Argonne National Laboratory**, and **Lawrence Livermore National Laboratory** proposed ESGF2
- ESGF2 proposal reviewed by panel of 8 scientists on August 30–31, 2021
- Project was selected for funding by the US Department of Energy in September, and it will begin once the US Government has a budget for the fiscal year
- We plan to continue operating the existing platform at LLNL for CMIP6, etc.
- 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



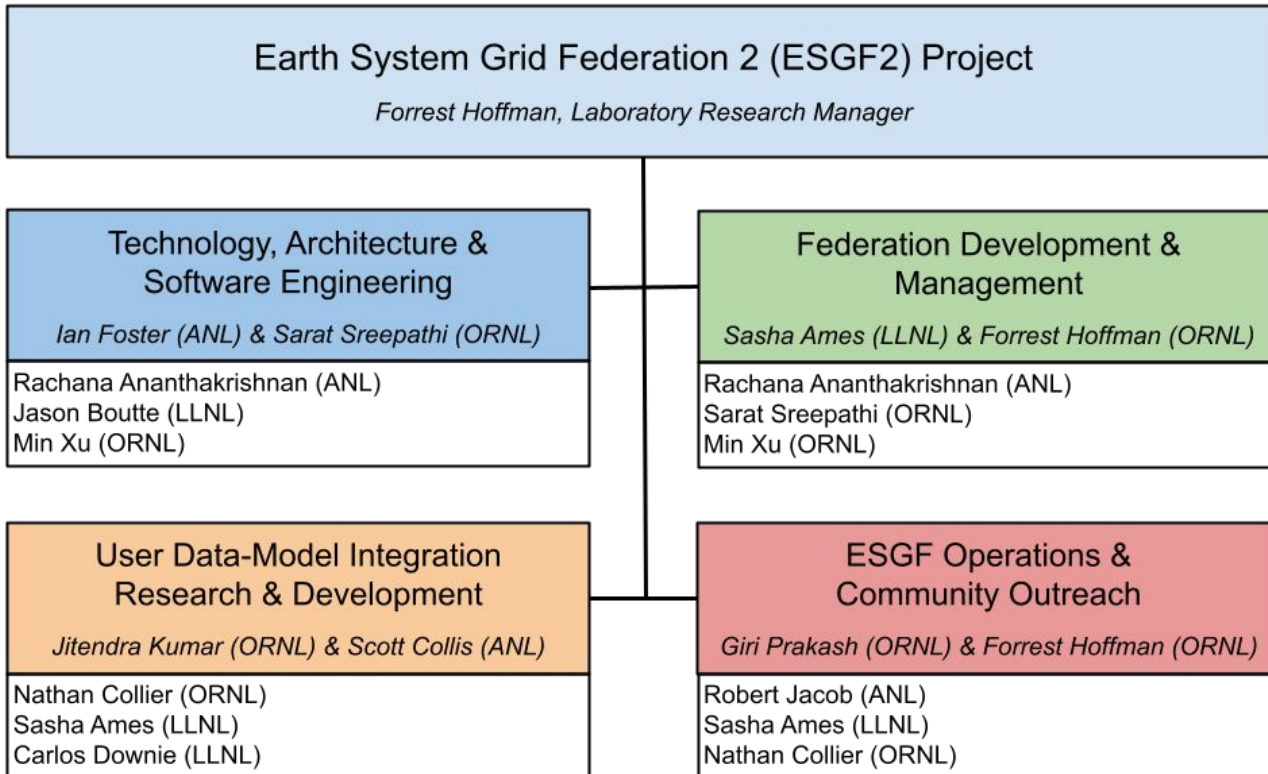
DOE's Current Earth System Grid Federation

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





ESGF2 Project Organization





ESGF2 Project Timeline

	Year 1				Year 2				Year 3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Technology evaluation and architecture planning	█	█	█	█								
Develop and deploy performant search & general-purpose publisher		█	█	█	█	█						
Develop and implement ESGF2 components					█	█	█	█	█	█		
Test integrated prototype of ESGF2 data system									█	█	█	█
Executive Committee Meetings	█	█	█	█	█	█	█	█	█	█	█	█
Steering Committee Meetings	█		█	█	█		█	█	█		█	█
Design specifications for third generation ESGF									█	█	█	█
Annual ESGF Meetings	█				█				█			
Develop and deploy new user interface for file retrieval			█	█	█	█						
Establish compute nodes for analysis	█	█	█	█								
Deploy integrated environment for diagnostics and benchmarking				█	█	█						
Develop new data access interfaces							█	█	█	█	█	█
Form stakeholder groups and setup help ticket system	█	█	█	█								
Establish self-replicating data nodes at ORNL and ANL	█	█	█	█								
Operate ESGF infrastructure at ORNL, ANL, and LLNL					█	█	█	█	█	█	█	█
ESGF tutorials and hackathons					█	█	█	█	█	█	█	█

- Each group has an aggressive set of tasks to carry out over the three years of the project
- Deliverables from these tasks include:
 - General-purpose publisher workflow
 - Fully integrated prototype of ESGF2 data backplane with metrics tracking
 - Performant search capability
 - New user interface that includes semantic search and recommendation capabilities
 - Server-side computing for spatio-temporal subsetting and statistical summaries
 - User computing capability in use by researchers
 - Help ticketing system deployed
 - Manage governance and outreach activities
 - Be on planned path to build out capacity for CMIP7
 - Draft specification for 3rd generation with partners

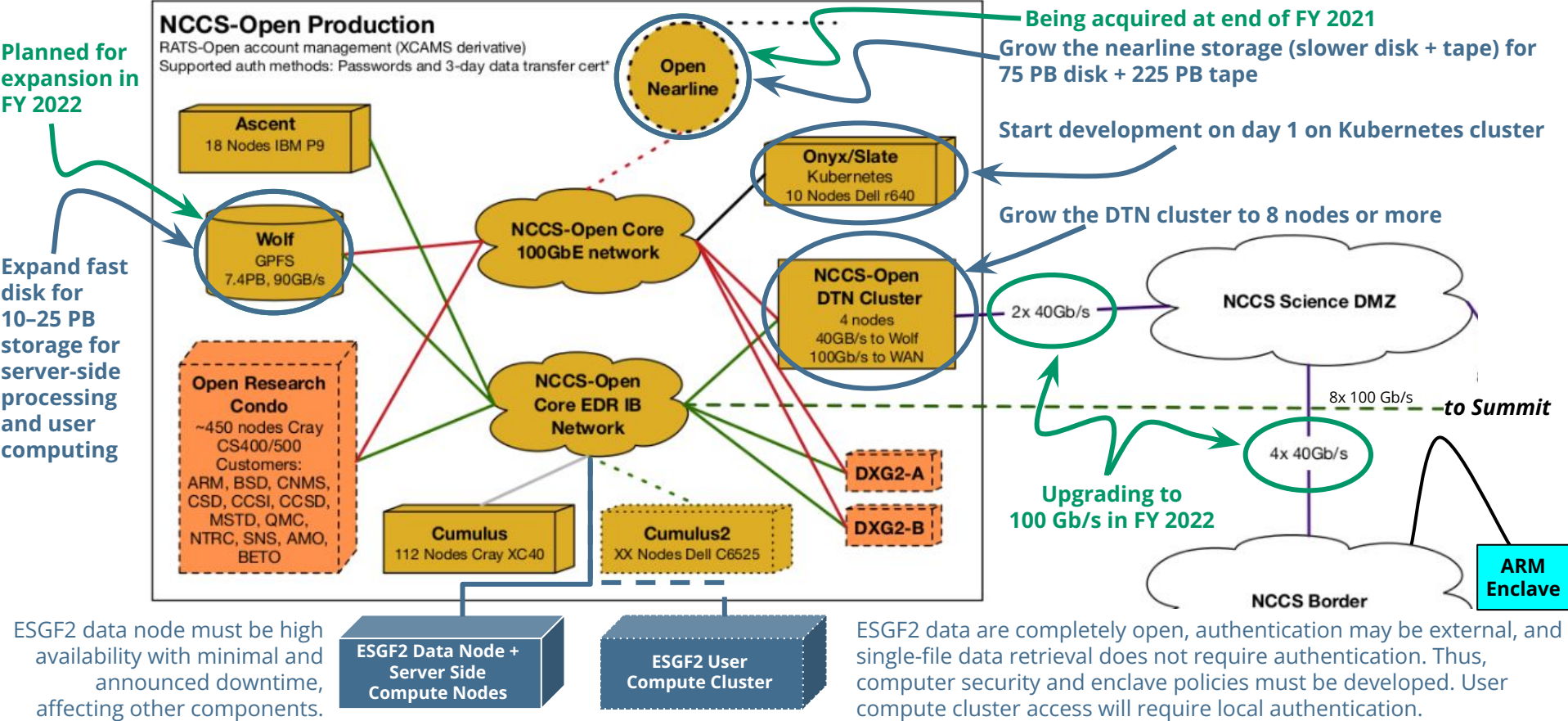


Design and implementation principles

- **Open architecture and protocols**
 - Enable substitution of alternative implementations
- Leverage **highly available and scalable** central services
 - Reduce complexity, increase reliability, provide economies of scale
- Use proven, modern **security technologies and practices**
 - Integrated access control; protect against attacks and intrusions
- **Use case approach** to design, implementation, and evaluation
 - Ensure that solutions meet real user needs
- Integrated **instrumentation**
 - Metrics drive data management, data access features, capability development
- Focus on **performance** to deal with big data
 - High-speed data transfer, search, server-side processing

ESGF2 Adding ESGF2 to ORNL's NCCS-Open Enclave

"Open" Enclave in ORNL's National Center for Computational Science (NCCS)



ESGF2 ESnet Global Connectivity

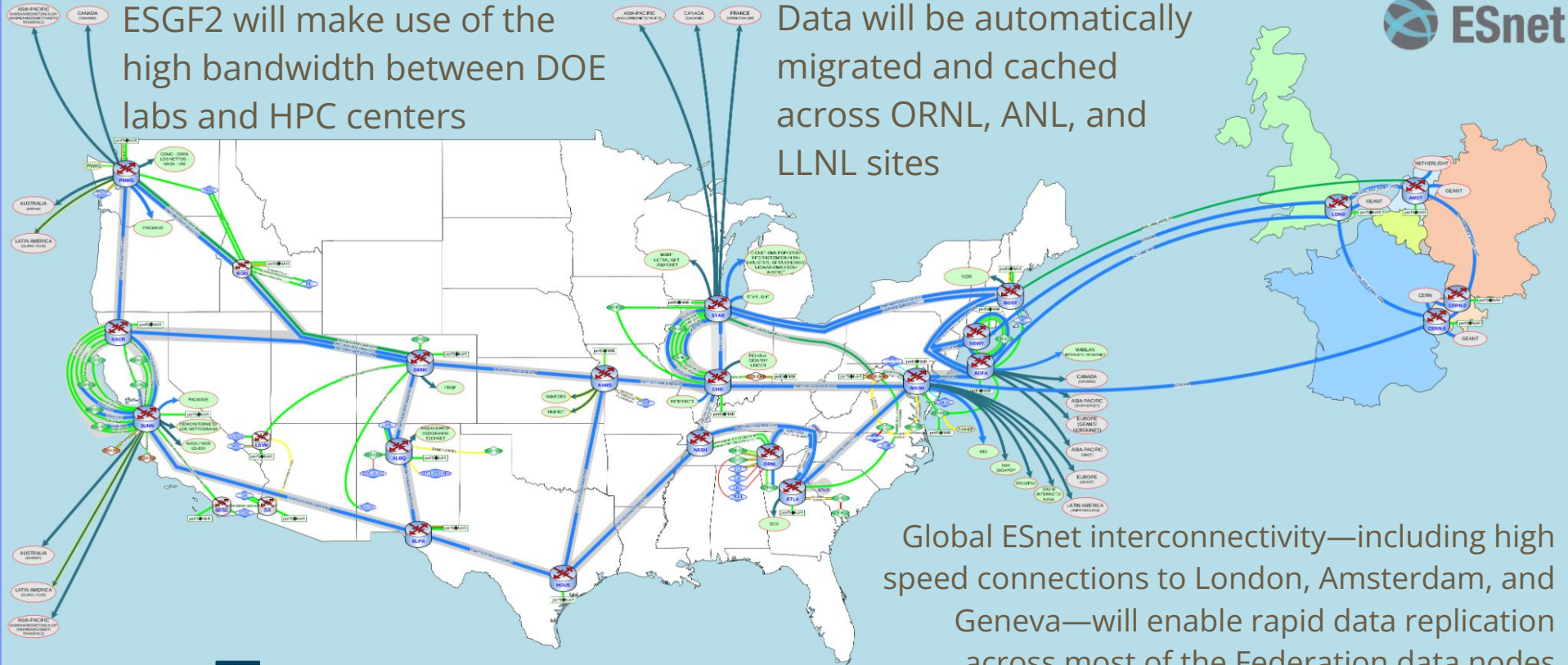


ESnet
ENERGY SCIENCES NETWORK

ESnet representative, Eli Dart, is part of our Resource & Project Liaisons group

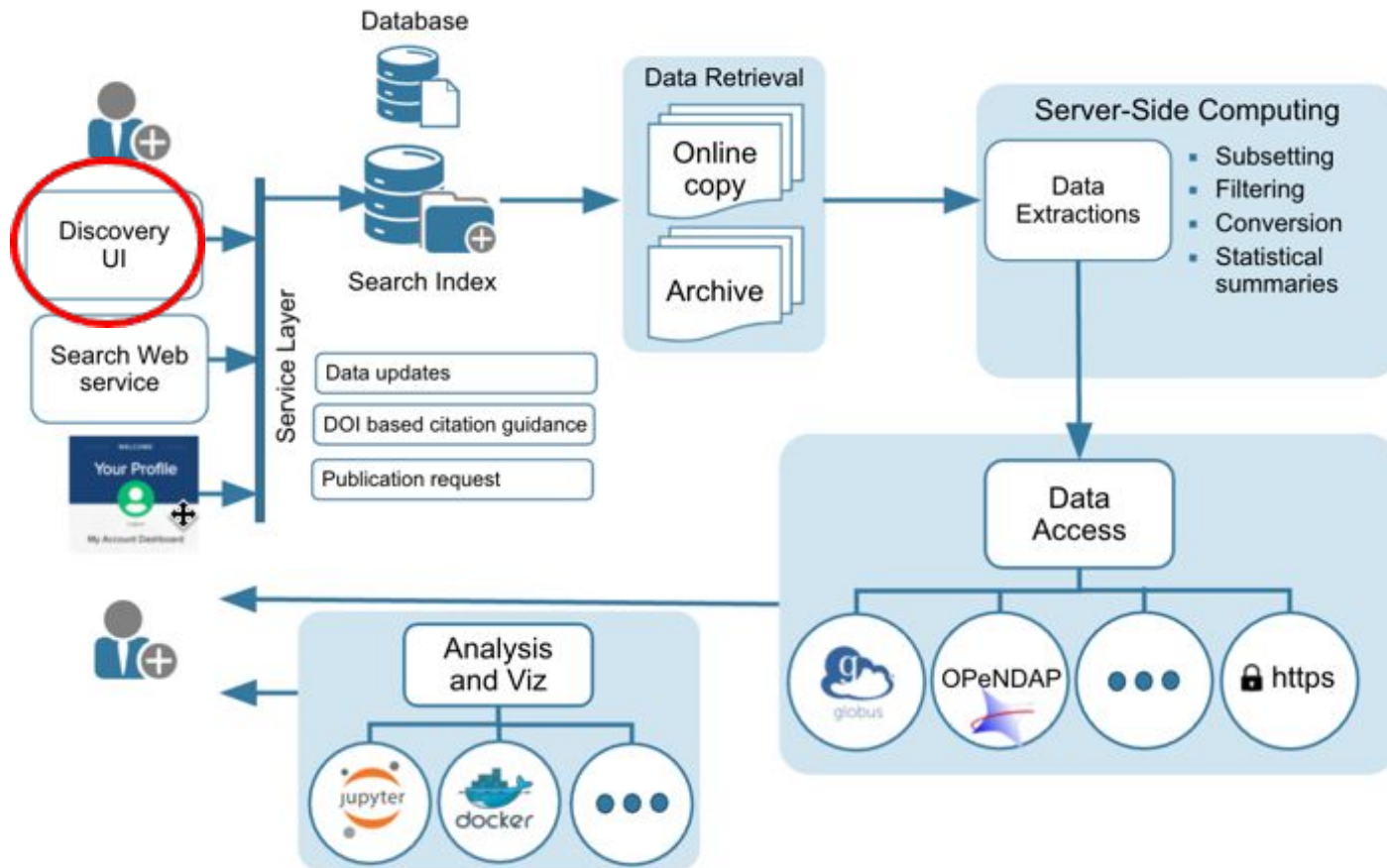
ESGF2 will make use of the high bandwidth between DOE labs and HPC centers

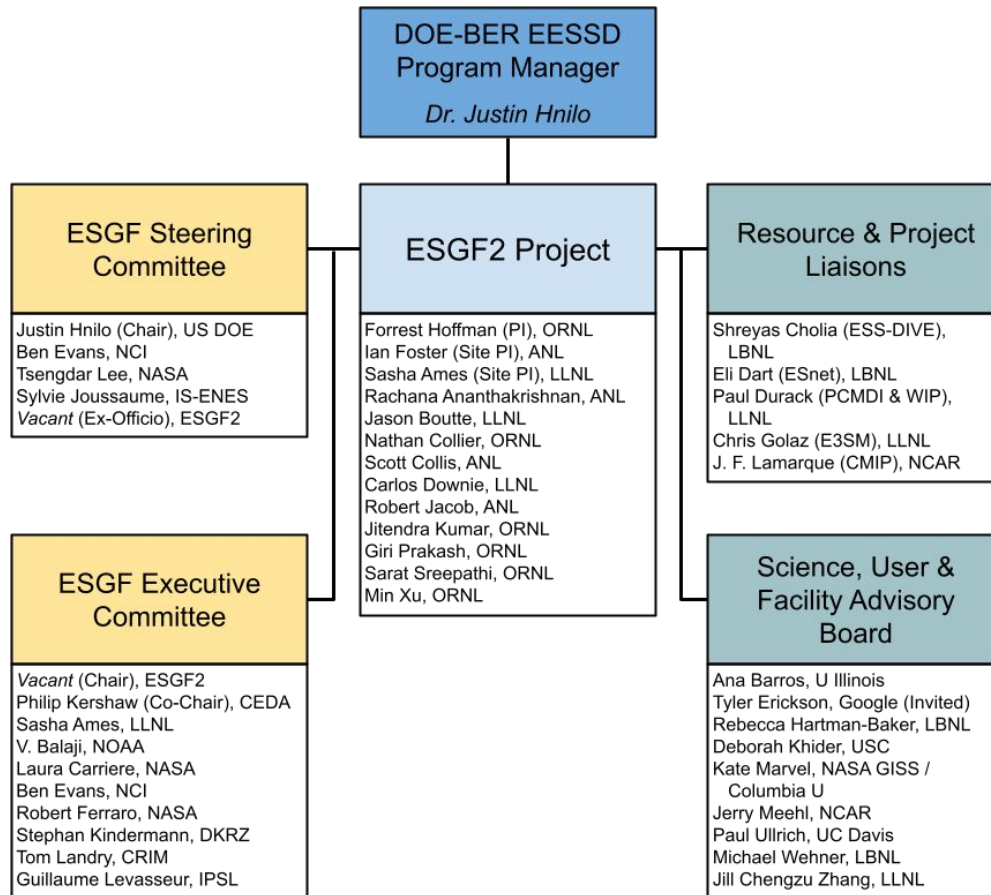
Data will be automatically migrated and cached across ORNL, ANL, and LLNL sites



Global ESnet interconnectivity—including high speed connections to London, Amsterdam, and Geneva—will enable rapid data replication across most of the Federation data nodes

ESGF2 Data Discovery Platform: Architecture





ESGF2 Outreach Activities

- Organize Webinars, Tutorials, and ESGF2 Bootcamps
 - Data management lessons learned
 - Ingest best practices
 - Data discovery and access
- Hackathons and Workshops
 - Data standards
 - Data node deployment
 - User compute resources
 - Hold at large relevant conferences, e.g., AGU Fall Meeting, EGU, and AMS Annual Meeting
- Organize and host an annual **ESGF Developer and User Conference**



ESGF2 User Support

- Support traditionally given via single email list
 - Volunteer basis from collaborating projects
- Need to formalize process for a “Helpdesk”
 - Introduce ticketing system for user inquires
 - e.g., ServiceNow or RT, leverage site resources
 - Assign dedicated support staff (need to triage)
 - **Get help from Federation partners for responses**
 - Maintain quality documentation for everything user-facing
 - Walkthroughs for GUI-based tools / webapps
 - Explore additional tools to provide support
 - Github Issues (when associated with software “bugs”)
 - Discourse - need subscription
 - Slack (invite single channel guests as needed)





ESGF2 How Can We Help Each Other?

- Engagement in development of CMIP7 protocol and plans
- Capacity planning for CMIP7: simulations, ensembles, users
- Help with outreach activities and user support
- ...