

ESGF2 Project Overview

Forrest M. Hoffman, Ian Foster, and Sasha Ames



Office of Science







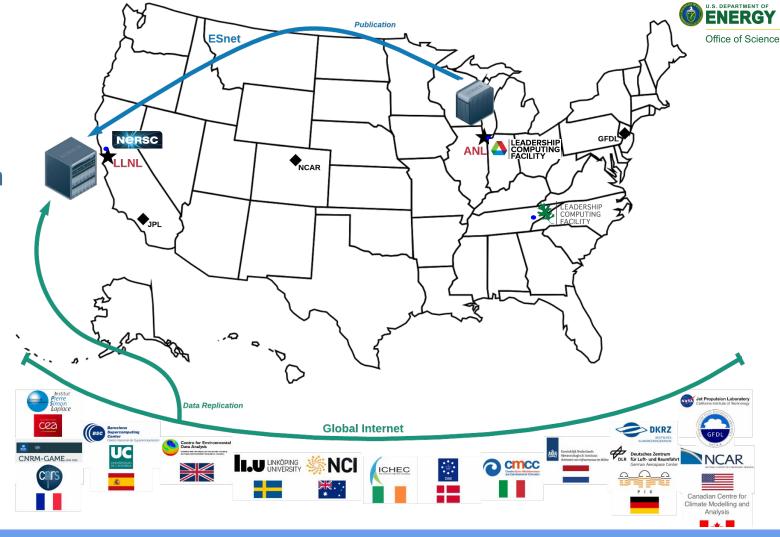
ESGF2 A New Consortium Project in the USA

- 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

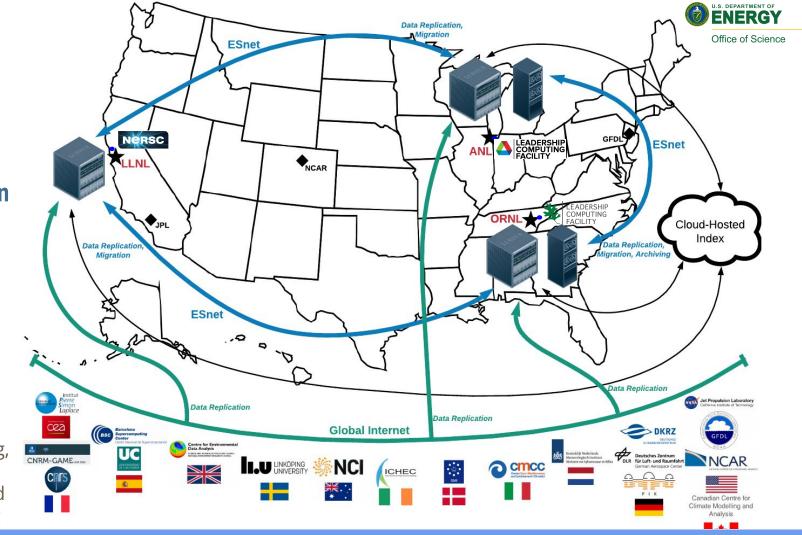
- Primary server at LLNL
- Replicating data from the global Federation
- Independent data node at ANL





DOE's Next Generation Earth System Grid Federation

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





Earth System Grid Federation 2 (ESGF2) Project

Forrest Hoffman, Laboratory Research Manager

Technology, Architecture & Software Engineering

Ian Foster (ANL) & Sarat Sreepathi (ORNL)

Rachana Ananthakrishnan (ANL) Jason Boutte (LLNL) Min Xu (ORNL)

User Data-Model Integration Research & Development

Jitendra Kumar (ORNL) & Scott Collis (ANL)

Nathan Collier (ORNL) Sasha Ames (LLNL) Carlos Downie (LLNL)

Federation Development & Management

Sasha Ames (LLNL) & Forrest Hoffman (ORNL)

Rachana Ananthakrishnan (ANL) Sarat Sreepathi (ORNL) Min Xu (ORNL)

ESGF Operations & Community Outreach

Giri Prakash (ORNL) & Forrest Hoffman (ORNL)

Robert Jacob (ANL) Sasha Ames (LLNL) Nathan Collier (ORNL)

ESGF2 Project Timeline

	Year 1				Year 2				Year 3					_
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4		E
Technology evaluation and architecture planning														C
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

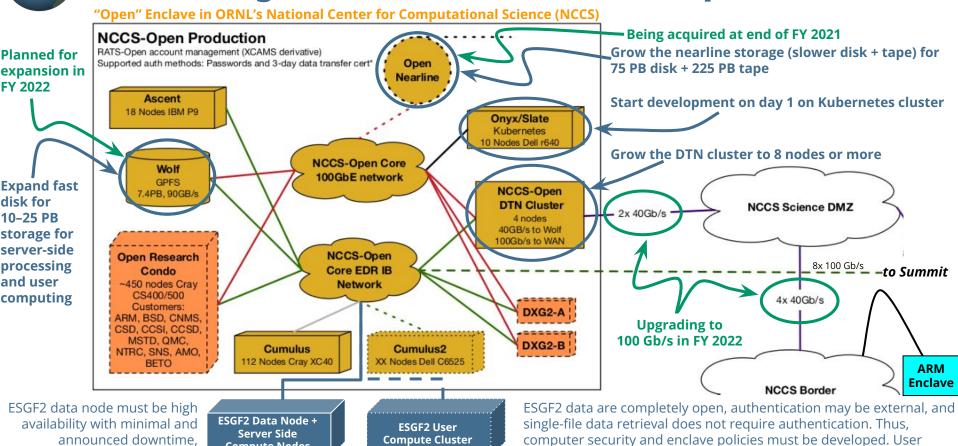
- Each group has an aggressive set of tasks to carry out over the three years of the project
- Deliverables from these tasks include:
 - General-purpose publisher workflow
 - Fully integrated prototype of ESGF2 data backplane with metrics tracking
 - Performant search capability
 - New user interface 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
 - o Drait specification for Srd generation with partners

ESGF2

Design and implementation principles

- Open architecture and protocols
 - Enable substitution of alternative implementations
- Leverage highly available and scalable central services
 - o Reduce complexity, increase reliability, provide economies of scale
- Use proven, modern security technologies and practices
 - o 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



compute cluster access will require local authentication.

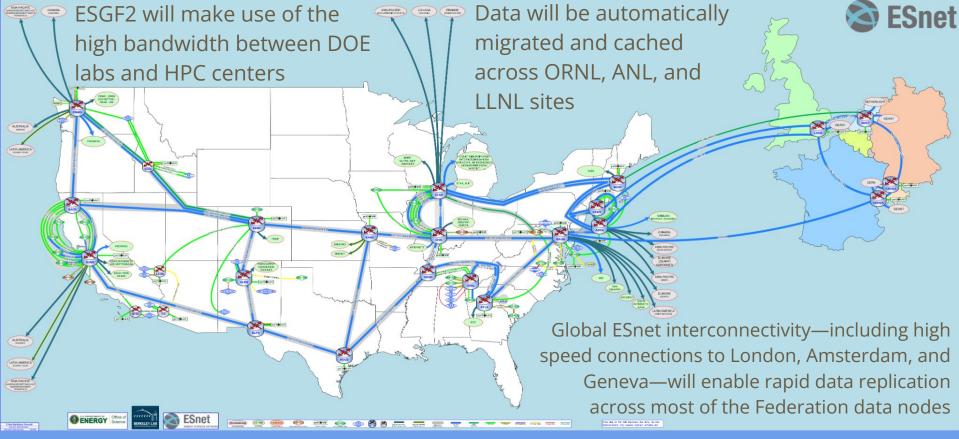
Compute Nodes

affecting other components.

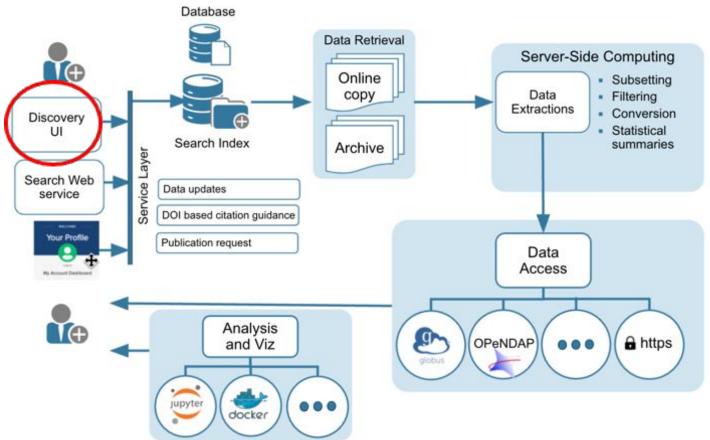
ESGF2 ESnet Global Connectivity

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





ESGF2 Data Discovery Platform: Architecture





DOE-BER EESSD Program Manager

Dr. Justin Hnilo

ESGF Steering Committee

Justin Hnilo (Chair), US DOE Ben Evans, NCI Tsengdar Lee, NASA Sylvie Joussaume, IS-ENES Vacant (Ex-Officio), ESGF2

ESGF Executive Committee

Vacant (Chair), ESGF2 Philip Kershaw (Co-Chair), CEDA Sasha Ames, LLNL V. Balaji, NOAA Laura Carriere, NASA Ben Evans, NCI Robert Ferraro, NASA Stephan Kindermann, DKRZ Tom Landry, CRIM Guillaume Levasseur, IPSL

ESGF2 Project

Ian Foster (Site PI), ANL Sasha Ames (Site PI), LLNL Rachana Ananthakrishnan, ANL Jason Boutte, LLNL Nathan Collier, ORNL Scott Collis, ANL Carlos Downie, LLNL Robert Jacob, ANL Jitendra Kumar, ORNL Giri Prakash, ORNL Sarat Sreepathi, ORNL Min Xu, ORNL

Forrest Hoffman (PI), ORNL

Resource & Project Liaisons

Shreyas Cholia (ESS-DIVE), LBNL Eli Dart (ESnet), LBNL Paul Durack (PCMDI & WIP), LLNL Chris Golaz (E3SM), LLNL J. F. Lamarque (CMIP), NCAR

Science, User & **Facility Advisory** Board

Ana Barros, U Illinois Tyler Erickson, Google (Invited) Rebecca Hartman-Baker, LBNL Deborah Khider, USC Kate Marvel, NASA GISS / Columbia U Jerry Meehl, NCAR Paul Ullrich, UC Davis Michael Wehner, LBNL Jill Chengzu Zhang, LLNL

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

• ...