



U.S. DEPARTMENT OF  
**ENERGY**

Office of Science



# Enabling GeoAI for Advancing Earth System Science

*Forrest M. Hoffman*

Oak Ridge National Laboratory

**THE**  
**TRILLION-PIXEL**  
**CHALLENGE**

Session 5:  
HPC Hardware and  
Software Architectures

*June 22, 2023*

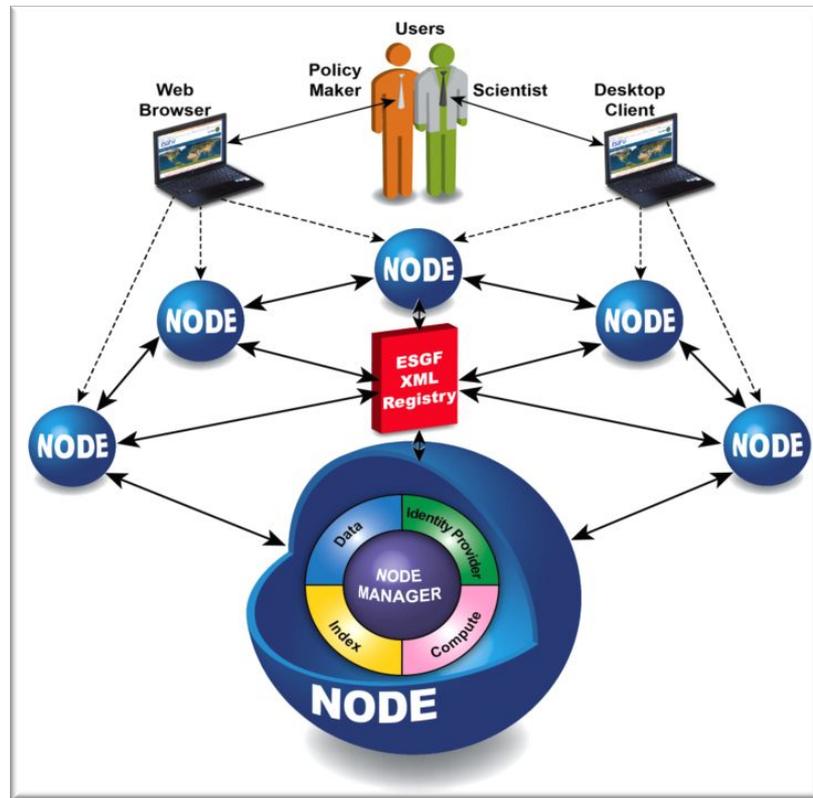
# Forrest M. Hoffman, Computational Earth System Scientist

- Group Leader for the ORNL Computational Earth Sciences Group
- 34 years at ORNL in Environmental Sciences Division, then Computer Science and Mathematics Division, and now Computational Sciences and Engineering Division
- Develop and apply Earth system models to study global biogeochemical cycles, including terrestrial & marine carbon cycle
- Investigate methods for reconciling uncertainties in carbon-climate feedbacks through comparison with observations
- Apply artificial intelligence methods (machine learning and data mining) to environmental characterization, simulation, & analysis
- Joint Faculty, University of Tennessee, Knoxville, Department of Civil & Environmental Engineering

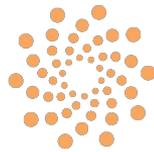


# ESGF US 2 What is the Earth System Grid Federation?

- The **Earth System Grid Federation (ESGF)** is an *international consortium* and a *globally distributed peer-to-peer network of data servers* using a common set of protocols and interfaces to archive and distribute climate and Earth system model output and related input, observational, and reanalysis data
- These **Open Science data** are used by scientists all over the world to investigate consequences of possible climate change scenarios and the resulting Earth system feedbacks

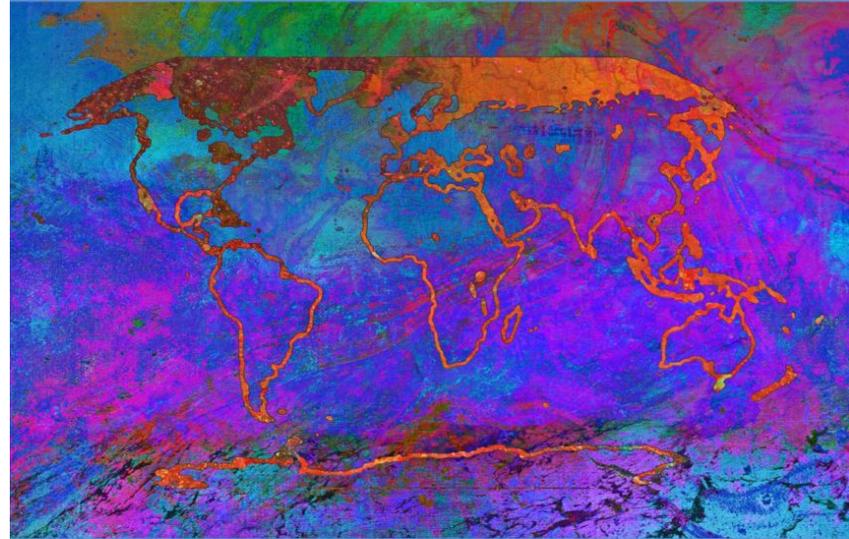


Logos represent primary international contributors: US Department of Energy, NASA, NOAA, NSF, European IS-ENES Project, and Australian NCI



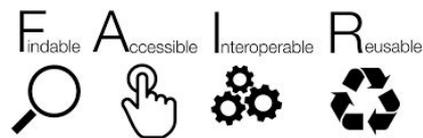
- The United Nations' Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report from Working Group I was released on Monday, August 9, 2021
- All of the climate and Earth system model simulation output underpinning this report was produced by modeling centers participating in the World 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!**

## Climate Change 2021 The Physical Science Basis

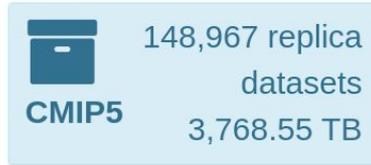
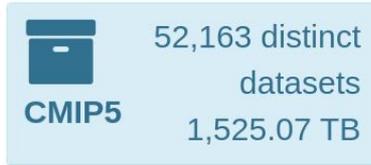
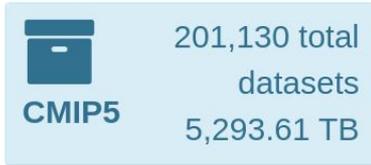
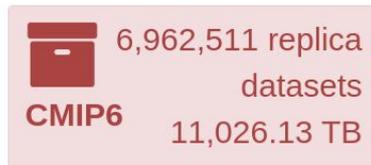
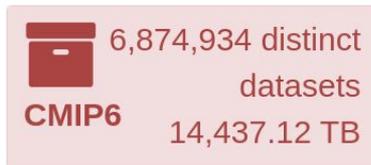
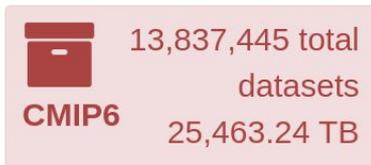




# ESGF Holdings are Large



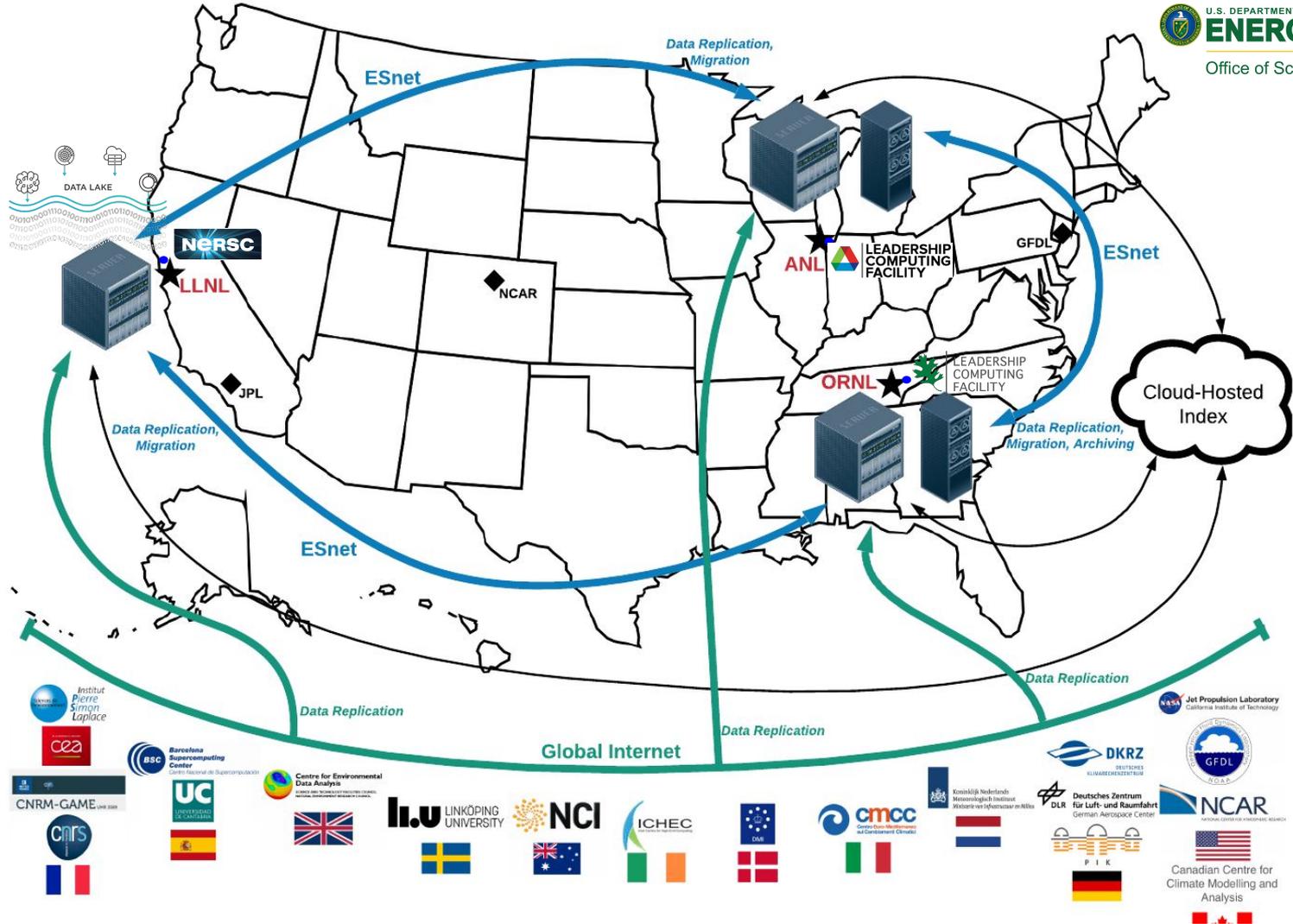
- **CMIP5** totals >5 PB  
**CMIP6** totals >25 PB (including replicas)  
**CMIP7** will total ~100 PB?
- ESGF is concerned with the full stack security and the integrity of the data, but we are **not** concerned about controlling access to the data
- Analysis is enabled by global data distribution and an enforced data format, metadata convention & controlled vocabulary





# DOE's Next Generation Earth System Grid Federation

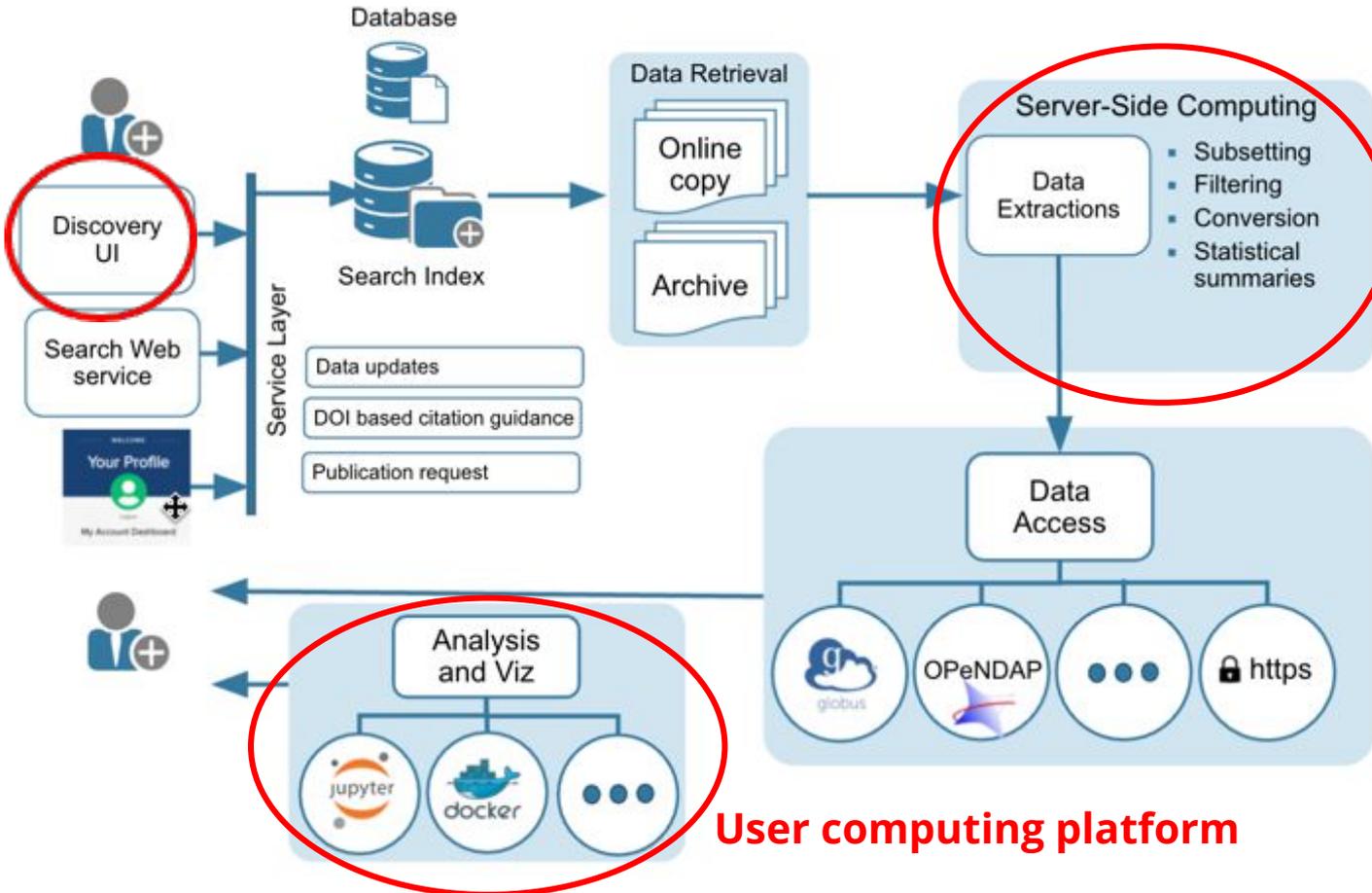
- As many as three nodes co-located at DOE's major computing facilities
- Replicating data from the global Federation
- Providing cloud indexing and tape archiving





# Data Discovery Platform: Architecture

Friendly user interface



Server-side computing platform

User computing platform