Building the Next Generation
Earth System Grid Federation (ESGF2)
Forrest M. Hoffman (ORNL), Ian Foster (ANL), and Sasha Ames (LLNL)
DOE Data Days 2002 (All Virtual)
June 2, 2022
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.
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 researcher via ESGF.

Data are about the future of life on Earth!
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

As of August 22, 2021
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 based on the Globus platform.

- ESGF2 proposal was reviewed by panel of 8 scientists on August 30–31, 2021, and was selected for funding by the US Department of Energy in September.

- 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 DOE’s major computing facilities
- Replicating data from the global Federation
- Providing cloud indexing, automated migration, and tape archiving
Design and implementation principles

- **Open architecture and protocols**
  - Enable substitution of alternative implementations

- **Leverage highly available and scalable** central services from Globus
  - 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
Enabling a new level of research productivity

Logging in with her institutional credentials, Samantha is presented with new data, code, and papers relevant to her current research. Intrigued by a new report on extreme precipitation events, she examines a Jupyter notebook that implements the method used. Wondering how this method would work with higher-resolution E3SM data, she quickly locates required datasets and runs the notebook on a subset. Results are promising, so she shares them with collaborators via ESGF2 federated storage, and they agree that a larger ensemble analysis is called for. ESGF2 confirms that the full ensemble data are available at OLCF, so they submit a request to execute the analysis there. Within 24 hours, results have been published to ESGF2 for broader consumption, along with the notebook used to produce and validate the results.

Flood risk increases with water availability
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.

An ESnet representative is part of our Resource & Project Liaisons group.
Data Discovery Platform: Architecture
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
In the US, LLNL operates the primary ESGF node, which replicates much of the CMIP6 and related model output from around the globe.

Since the data at LLNL are contained only on spinning disk, we decided to replicate the **entire ~7.5 PB collection of data** to Argonne National Laboratory (ANL) and Oak Ridge National Laboratory (ORNL).

**Solution:** Use Globus to transfer all the data over ESnet.

We used custom Globus scripting (*thanks to Lukasz Lacinski*), ESnet network monitoring and diagnostics (*thanks to Eli Dart*), DTN and GPFS optimized configurations (*thanks to Cameron Harr and others*), and debugging and problem-solving (*thanks to Sasha Ames, Lee Liming, and others*).
Data transferred to ALCF

Replication to ALCF
ACTIVE, PAUSED and the latest SUCCEEDED transfers

<table>
<thead>
<tr>
<th>No</th>
<th>Datasets</th>
<th>From</th>
<th>Requested</th>
<th>Completed</th>
<th>Status</th>
<th>Directories</th>
<th>Files</th>
<th>Bytes Transferred</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>/cmip5_css02_data/cmip5/output1/NCAR/CESM4</td>
<td>LLNL</td>
<td>2022-05-02 01:53:03</td>
<td>2022-05-03 00:50:23</td>
<td>SUCCEEDED</td>
<td>52571</td>
<td>48925</td>
<td>3345543876668</td>
<td>11 405 MB/s</td>
</tr>
</tbody>
</table>

Data transferred to OLCF

Replication to OLCF
ACTIVE, PAUSED and the latest SUCCEEDED transfers

<table>
<thead>
<tr>
<th>No</th>
<th>Datasets</th>
<th>From</th>
<th>Requested</th>
<th>Completed</th>
<th>Status</th>
<th>Directories</th>
<th>Files</th>
<th>Bytes Transferred</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>/cmip5_css02_data/cmip5/output1/NASA-GISS/GISS-E2-R-CC</td>
<td>ALCF</td>
<td>2022-05-02 05:34:04</td>
<td>2022-05-02 05:44:32</td>
<td>SUCCEEDED</td>
<td>2098</td>
<td>9576</td>
<td>1087745690416</td>
<td>0 1.73 GB/s</td>
</tr>
</tbody>
</table>

As of May 4, 2022
7.5 PB transferred between mid-Feb and May 4
17,347,671 directories and 28,907,532 files
Cumulative Data Transferred Over Time

Progress of transfers
- Blue line: to ALCF
- Red line: to OLCF

- CMIP5
- CMIP6
- Big files
- Small files
- GPFS Errors & Unreadable Files
- GPFS Reconfigured
Transfer Rates Over Time

Transfer rate (on an hourly basis)

- LLNL→ALCF
- OLCF→ALCF
- LLNL→OLCF
- ALCF→OLCF

>7.5 GB/s OLCF ➔ ALCF
The next generation **Earth System Grid Federation (ESGF2)**
- Will be designed for an order of magnitude increase in data sizes
- Will be highly available, scalable, and fast
- Will automatically migrate data as needed
- Will have improved data discovery and sharing tools
- Will offer server-side computing for derived data
- Will offer user computing capabilities (e.g., JupyterHub/JupyterLab) near the data

The **Globus platform** is expected to provide many of the central services of the ESGF2 data backplane

We used Globus to make two redundant copies of the **7.5 PB of ESGF data** via ESnet in less than 3 months