U.S. DEPARTMENT OF ENERGY

Office of Science

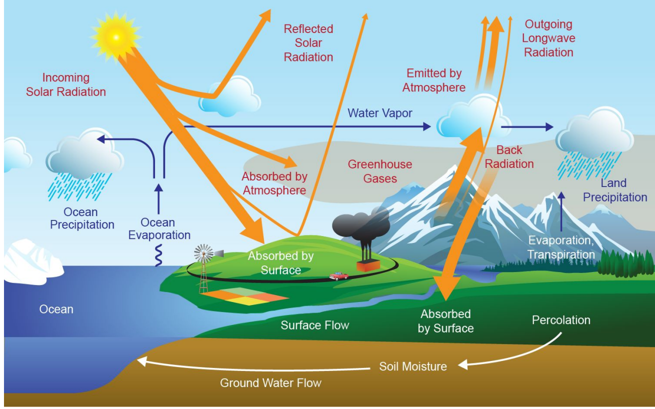
# The CMIP6 Data Lake at NERSC

*Forrest M. Hoffman (ORNL), Wilbert Weijer (LANL),
Paul A. Ullrich (UC Davis), and Michael Wehner (LBNL)*

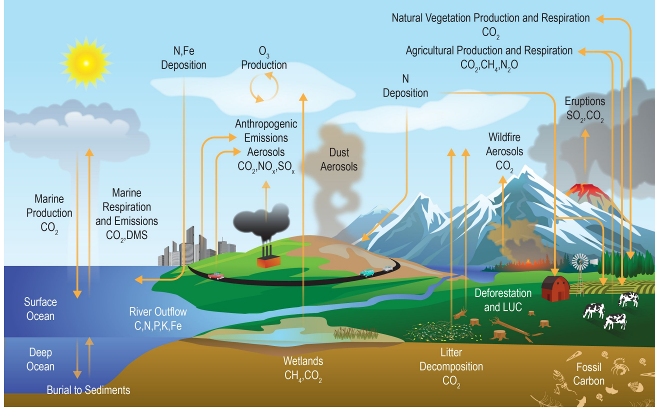
# What is an Earth System Model?

An **Earth System Model (ESM)** is a coupled model that

- Solves differential equations of fluid motion and thermodynamics to obtain time and space dependent values for temperature, winds and currents, moisture and/or salinity and pressure in the atmosphere and ocean
- Combines component models of the atmosphere, ocean, land surface, sea ice, and land ice
- Closes the global carbon cycle by simulating processes and feedbacks of vegetation and marine ecology and biogeochemistry, land use change, and (increasingly) human system processes

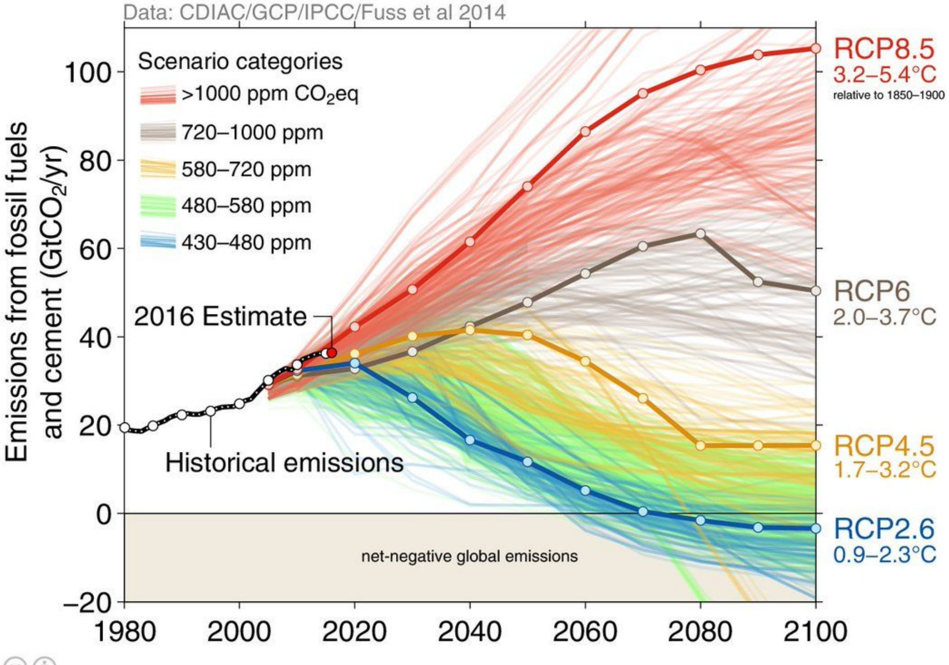


*Energy and Water Cycles*

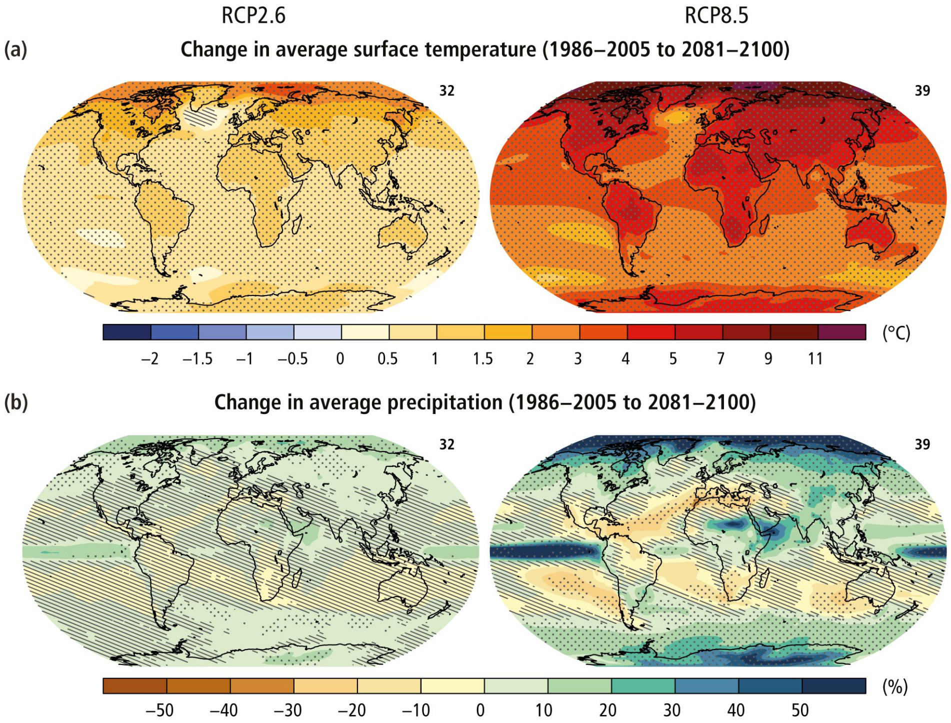


*Carbon and Biogeochemical Cycles*

# ESMs project future changes in the climate system



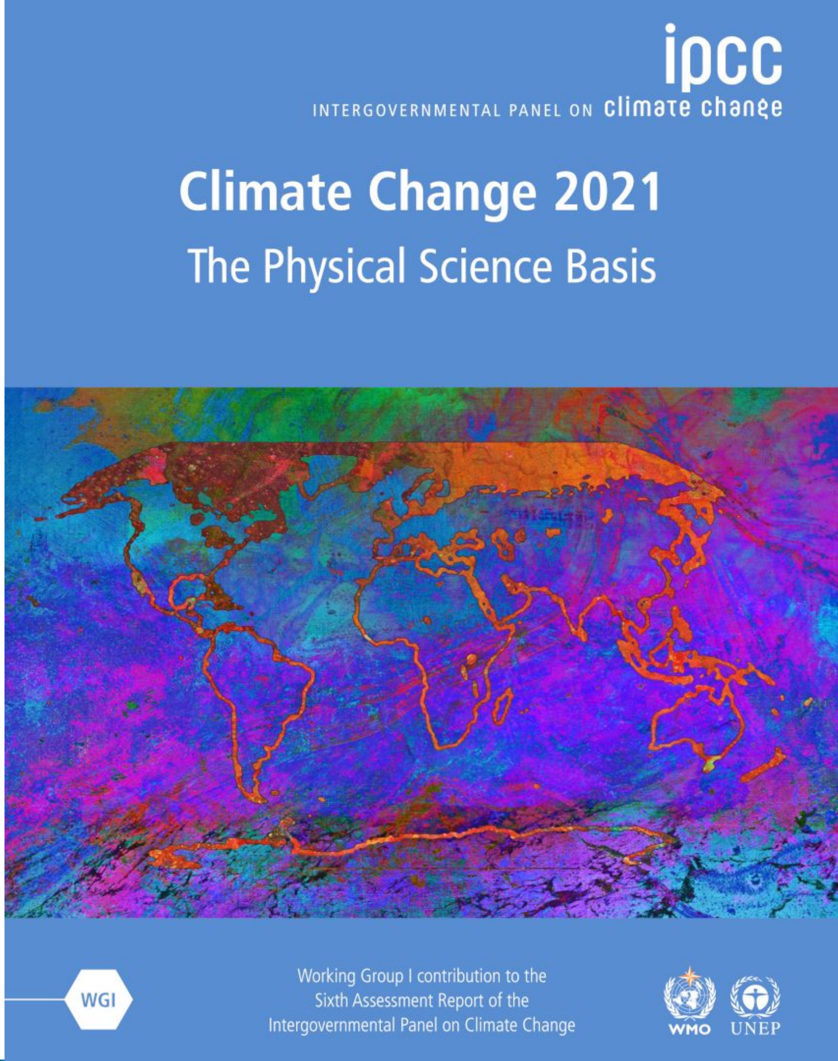
Global Carbon Project


RCP2.6
RCP8.5
(a)
Change in average surface temperature (1986–2005 to 2081–2100)
32
39
(°C)
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
3
4
5
7
9
11
(b)
Change in average precipitation (1986–2005 to 2081–2100)
32
39
(%)
−50
−40
−30
−20
−10
0
10
20
30
40
50

*From IPCC AR5 WG1 Summary for Policymakers (SPM)*

# IPCC Sixth Assessment Report

- The United Nations' Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) 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 the Earth System Grid Federation (ESGF)



ipcc

INTERGOVERNMENTAL PANEL ON climate change

Climate Change 2021

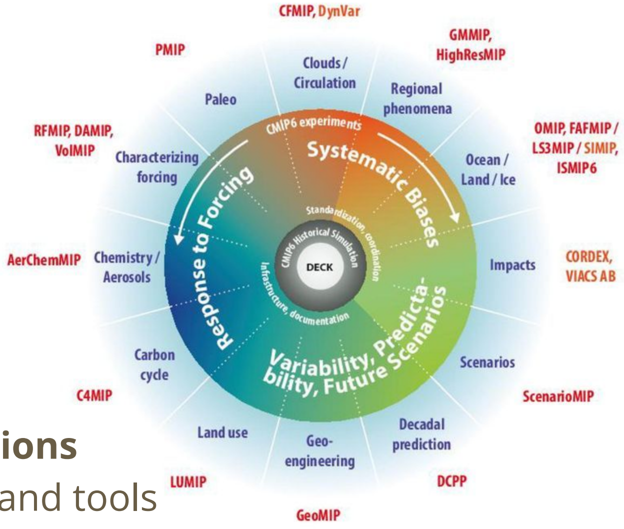
The Physical Science Basis

WGI

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

WMO UNEP

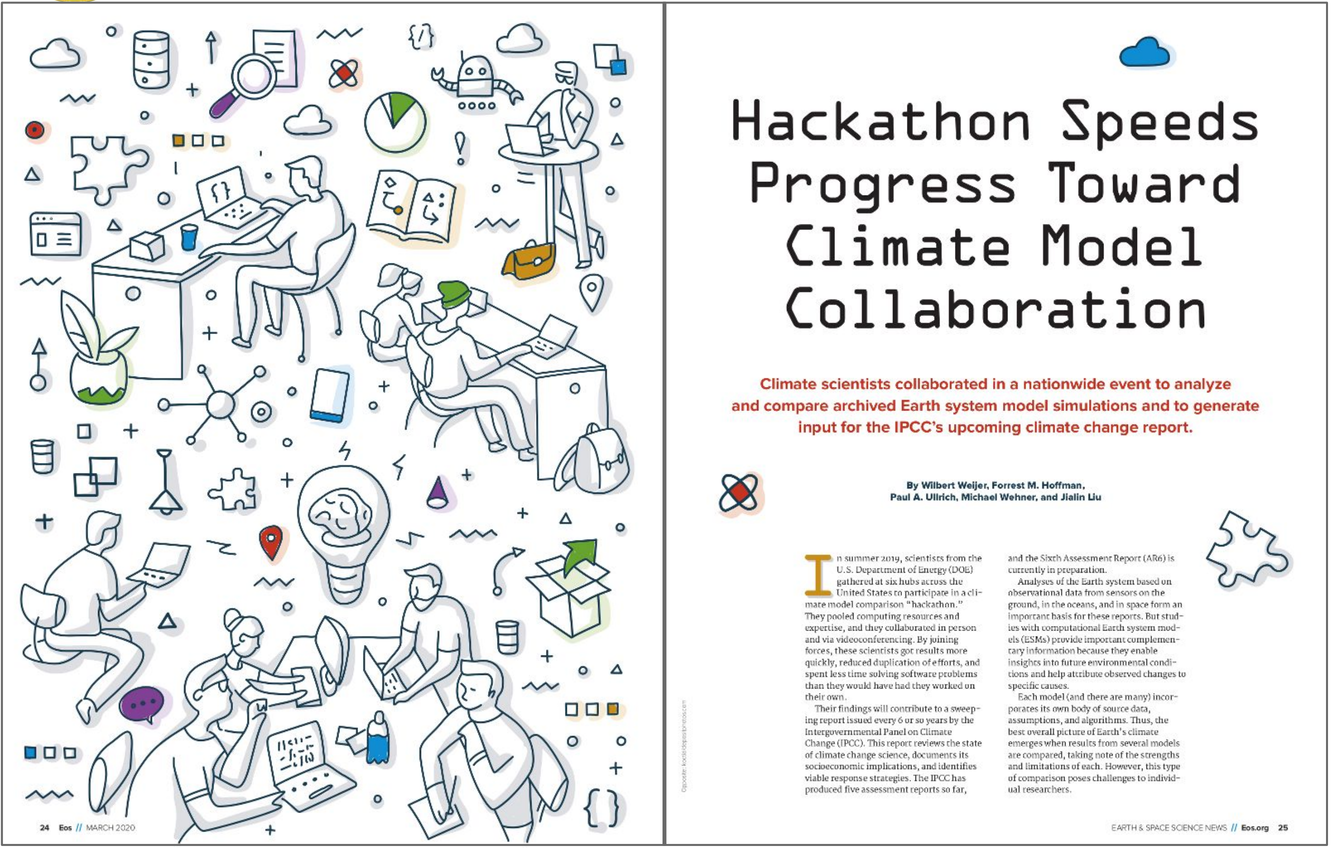
# Coupled Model Intercomparison Project



- **CMIP6** analysis papers that were accepted by September 30, 2020, could be cited in the **IPCC Sixth Assessment Report**
- In 2019, to support DOE-BER scientists doing multi-model research, **BER RGMA** & **Data Programs** coordinated and sponsored
  - Staging **CMIP6 output from ESGF** plus **reanalysis** & **observations**
  - Series of **tutorials** on CMIP6 organization, Jupyter notebooks, and tools
  - **RGMA CMIP6 Hackathon** via videoconferencing at multiple hubs
- Lab & university researchers co-organized activities
  - *Forrest Hoffman (ORNL, RUBISCO), Jialin Liu (NERSC), Paul Ullrich (UC Davis, HYPERFACETS), Michael Wehner (LBNL, CASCADE), Wilbert Weijer (LANL, HiLAT)*
- **NERSC provided multiple PB disk storage** and **interactive computing resources**

**CMIP6 Data Lake at NERSC**
/global/cfs/cdirs/m3522/cmip6

# RGMA CMIP6 Analysis and Hackathon



## Hackathon Speeds Progress Toward Climate Model Collaboration

Climate scientists collaborated in a nationwide event to analyze and compare archived Earth system model simulations and to generate input for the IPCC's upcoming climate change report.

By Wilbert Weijer, Forrest M. Hoffman, Paul A. Ullrich, Michael Wehner, and Jialin Liu

In summer 2019, scientists from the U.S. Department of Energy (DOE) gathered at six hubs across the United States to participate in a climate model comparison "hackathon." They pooled computing resources and expertise, and they collaborated in person and via videoconferencing. By joining forces, these scientists got results more quickly, reduced duplication of efforts, and spent less time solving software problems than they would have had they worked on their own.

Their findings will contribute to a sweeping report issued every 6 or so years by the Intergovernmental Panel on Climate Change (IPCC). This report reviews the state of climate change science, documents its socioeconomic implications, and identifies viable response strategies. The IPCC has produced five assessment reports so far, and the Sixth Assessment Report (AR6) is currently in preparation.

Analyses of the Earth system based on observational data from sensors on the ground, in the oceans, and in space form an important basis for these reports. But studies with computational Earth system models (ESMs) provide important complementary information because they enable insights into future environmental conditions and help attribute observed changes to specific causes.

Each model (and there are many) incorporates its own body of source data, assumptions, and algorithms. Thus, the best overall picture of Earth's climate emerges when results from several models are compared, taking note of the strengths and limitations of each. However, this type of comparison poses challenges to individual researchers.

24 Eos // MARCH 2020

EARTH & SPACE SCIENCE NEWS // Eos.org 25

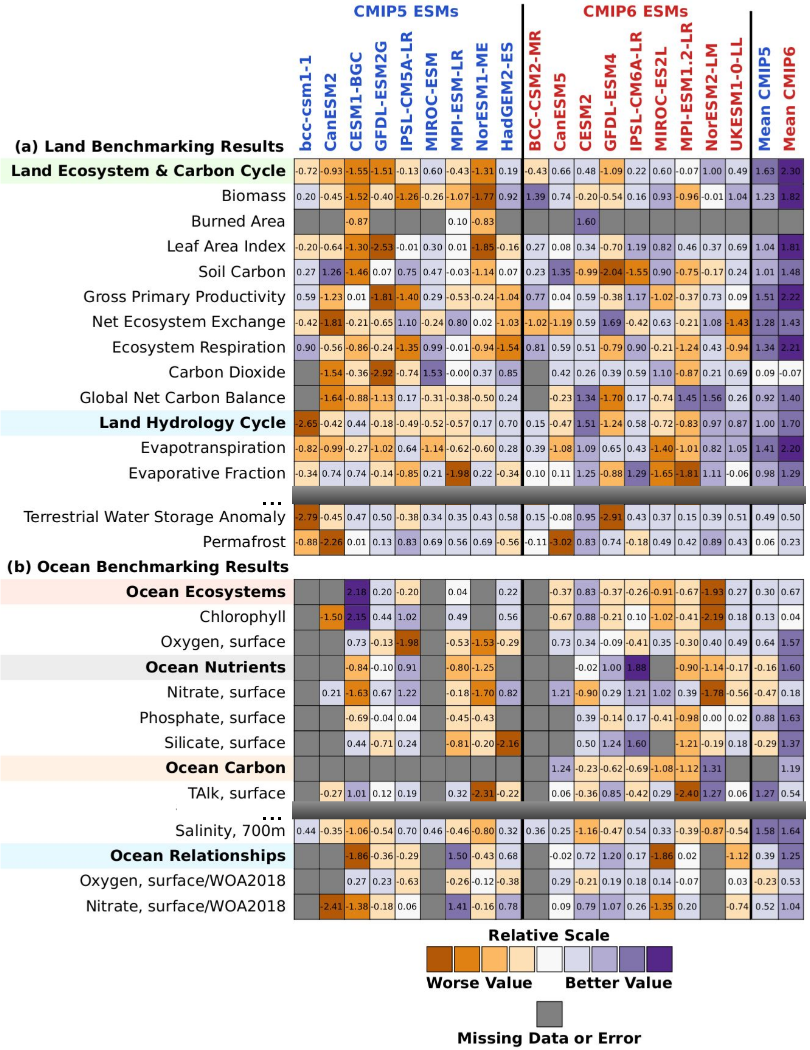
Weijer, Wilbert, Forrest M. Hoffman, Paul A. Ullrich, Michael Wehner, and Jialin Liu (2020), Hackathon Speeds Progress Toward Climate Model Collaboration, *Eos Trans. AGU*, 101(3):24–27, doi:10.1029/2019EO137735.

**RGMA CMIP6 Hackathon** held July 31–August 6, 2019

- **RGMA researchers** participated at LANL, LBNL, ORNL, PNNL, and U. Washington
- The data were loaded and analysis tools were installed at NERSC beforehand
- **Slack Workspace** and **GitHub** repository used for sharing tips, tricks, code, and Jupyter notebooks
- Hackathon fostered cross-institutional/project collaboration

# CMIP5 vs. CMIP6 Evaluation

- (a) **International Land Model Benchmarking (ILAMB)** and (b) **International Ocean Model Benchmarking (IOMB)** tools were used to evaluate how land and ocean model performance changed from CMIP5 to CMIP6
- Model fidelity is assessed through comparison of historical simulations with a wide variety of contemporary observational datasets
- The UN's **Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6)** from Working Group 1 (WG1) Chapter 5 contains the full ILAMB/IOMB evaluation as **Figure 5.22**



| | CMIP5 ESMs | | | | | | | | | CMIP6 ESMs | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | bcc-csm1-1 | CanESM2 | CESM1-BGC | GFDL-ESM2G | IPSL-CM5A-LR | MIROC-ESM | MPI-ESM-LR | NorESM1-ME | HadGEM2-ES | BCC-CSM2-MR | CanESM5 | CESM2 | GFDL-ESM4 | IPSL-CM6A-LR | MIROC-ES2L | MPI-ESM1.2-LR | NorESM2-LM | UKESM1-0-LL | Mean CMIP5 | Mean CMIP6 |
| **(a) Land Benchmarking Results** | | | | | | | | | | | | | | | | | | | | |
| **Land Ecosystem & Carbon Cycle** | -0.72 | -0.93 | -1.55 | -1.51 | -0.13 | 0.60 | -0.43 | -1.31 | 0.19 | -0.43 | 0.66 | 0.48 | -1.09 | 0.22 | 0.60 | -0.07 | 1.00 | 0.49 | 1.63 | 2.30 |
| Biomass | 0.20 | -0.45 | -1.52 | -0.40 | -1.26 | -0.26 | -1.07 | -1.77 | 0.92 | 1.39 | 0.74 | -0.20 | -0.54 | 0.16 | 0.93 | -0.96 | -0.01 | 1.04 | 1.23 | 1.82 |
| Burned Area | | | -0.87 | | | | 0.10 | -0.83 | | | | 1.60 | | | | | | | | |
| Leaf Area Index | -0.20 | -0.64 | -1.30 | -2.53 | -0.01 | 0.30 | 0.01 | -1.85 | -0.16 | 0.27 | 0.08 | 0.34 | -0.70 | 1.19 | 0.82 | 0.46 | 0.37 | 0.69 | 1.04 | 1.81 |
| Soil Carbon | 0.27 | 1.26 | -1.46 | 0.07 | 0.75 | 0.47 | -0.03 | -1.14 | 0.07 | 0.23 | 1.35 | -0.99 | -2.04 | -1.55 | 0.90 | -0.75 | -0.17 | 0.24 | 1.01 | 1.48 |
| Gross Primary Productivity | 0.59 | -1.23 | 0.01 | -1.81 | -1.40 | 0.29 | -0.53 | -0.24 | -1.04 | 0.77 | 0.04 | 0.59 | -0.38 | 1.17 | -1.02 | -0.37 | 0.73 | 0.09 | 1.51 | 2.22 |
| Net Ecosystem Exchange | -0.42 | -1.81 | -0.21 | -0.65 | 1.10 | -0.24 | 0.80 | 0.02 | -1.03 | -1.02 | -1.19 | 0.59 | 1.69 | -0.42 | 0.63 | -0.21 | 1.08 | -1.43 | 1.28 | 1.43 |
| Ecosystem Respiration | 0.90 | -0.56 | -0.86 | -0.24 | -1.35 | 0.99 | -0.01 | -0.94 | -1.54 | 0.81 | 0.59 | 0.51 | -0.79 | 0.90 | -0.21 | -1.24 | 0.43 | -0.94 | 1.34 | 2.21 |
| Carbon Dioxide | | -1.54 | -0.36 | -2.92 | -0.74 | 1.53 | -0.00 | 0.37 | 0.85 | | 0.42 | 0.26 | 0.39 | 0.59 | 1.10 | -0.87 | 0.21 | 0.69 | 0.09 | -0.07 |
| Global Net Carbon Balance | | -1.64 | -0.88 | -1.13 | 0.17 | -0.31 | -0.38 | -0.50 | 0.24 | | -0.23 | 1.34 | -1.70 | 0.17 | -0.74 | 1.45 | 1.56 | 0.26 | 0.92 | 1.40 |
| **Land Hydrology Cycle** | -2.65 | -0.42 | 0.44 | -0.18 | -0.49 | -0.52 | -0.57 | 0.17 | 0.70 | 0.15 | -0.47 | 1.51 | -1.24 | 0.58 | -0.72 | -0.83 | 0.97 | 0.87 | 1.00 | 1.70 |
| Evapotranspiration | -0.82 | -0.99 | -0.27 | -1.02 | 0.64 | -1.14 | -0.62 | -0.60 | 0.28 | 0.39 | -1.08 | 1.09 | 0.65 | 0.43 | -1.40 | -1.01 | 0.82 | 1.05 | 1.41 | 2.20 |
| Evaporative Fraction | -0.34 | 0.74 | 0.74 | -0.14 | -0.85 | 0.21 | -1.98 | 0.22 | -0.34 | 0.10 | 0.11 | 1.25 | -0.88 | 1.29 | -1.65 | -1.81 | 1.11 | -0.06 | 0.98 | 1.29 |
| ... | | | | | | | | | | | | | | | | | | | | |
| Terrestrial Water Storage Anomaly | -2.79 | -0.45 | 0.47 | 0.50 | -0.38 | 0.34 | 0.35 | 0.43 | 0.58 | 0.15 | -0.08 | 0.95 | -2.91 | 0.43 | 0.37 | 0.15 | 0.39 | 0.51 | 0.49 | 0.50 |
| Permafrost | -0.88 | -2.26 | 0.01 | 0.13 | 0.83 | 0.69 | 0.56 | 0.69 | -0.56 | -0.11 | -3.02 | 0.83 | 0.74 | -0.18 | 0.49 | 0.42 | 0.89 | 0.43 | 0.06 | 0.23 |
| **(b) Ocean Benchmarking Results** | | | | | | | | | | | | | | | | | | | | |
| **Ocean Ecosystems** | | | 2.18 | 0.20 | -0.20 | | 0.04 | | 0.22 | | -0.37 | 0.83 | -0.37 | -0.26 | -0.91 | -0.67 | -1.93 | 0.27 | 0.30 | 0.67 |
| Chlorophyll | | -1.50 | 2.15 | 0.44 | 1.02 | | 0.49 | | 0.56 | | -0.67 | 0.88 | -0.21 | 0.10 | -1.02 | -0.41 | -2.19 | 0.18 | 0.13 | 0.04 |
| Oxygen, surface | | | 0.73 | -0.13 | -1.98 | | -0.53 | -1.53 | -0.29 | | 0.73 | 0.34 | -0.09 | -0.41 | 0.35 | -0.30 | 0.40 | 0.49 | 0.64 | 1.57 |
| **Ocean Nutrients** | | | -0.84 | -0.10 | 0.91 | | -0.80 | -1.25 | | | | -0.02 | 1.00 | 1.88 | | -0.90 | -1.14 | -0.17 | -0.16 | 1.60 |
| Nitrate, surface | | 0.21 | -1.63 | 0.67 | 1.22 | | -0.18 | -1.70 | 0.82 | | 1.21 | -0.90 | 0.29 | 1.21 | 1.02 | 0.39 | -1.78 | -0.56 | -0.47 | 0.18 |
| Phosphate, surface | | | -0.69 | -0.04 | 0.04 | | -0.45 | -0.43 | | | | 0.39 | -0.14 | 0.17 | -0.41 | -0.98 | 0.00 | 0.02 | 0.88 | 1.63 |
| Silicate, surface | | | 0.44 | -0.71 | 0.24 | | -0.81 | -0.20 | -2.16 | | | 0.50 | 1.24 | 1.60 | | -1.21 | -0.19 | 0.18 | -0.29 | 1.37 |
| **Ocean Carbon** | | | | | | | | | | | 1.24 | -0.23 | -0.62 | -0.69 | -1.08 | -1.12 | 1.31 | | | 1.19 |
| TAlk, surface | | -0.27 | 1.01 | 0.12 | 0.19 | | 0.32 | -2.31 | -0.22 | | 0.06 | -0.36 | 0.85 | -0.42 | 0.29 | -2.40 | 1.27 | 0.06 | 1.27 | 0.54 |
| ... | | | | | | | | | | | | | | | | | | | | |
| Salinity, 700m | 0.44 | -0.35 | -1.06 | -0.54 | 0.70 | 0.46 | -0.46 | -0.80 | 0.32 | 0.36 | 0.25 | -1.16 | -0.47 | 0.54 | 0.33 | -0.39 | -0.87 | -0.54 | 1.58 | 1.64 |
| **Ocean Relationships** | | | -1.86 | -0.36 | -0.29 | | 1.50 | -0.43 | 0.68 | | -0.02 | 0.72 | 1.20 | 0.17 | -1.86 | 0.02 | | -1.12 | 0.39 | 1.25 |
| Oxygen, surface/WOA2018 | | | 0.27 | 0.23 | -0.63 | | -0.26 | -0.12 | -0.38 | | 0.29 | -0.21 | 0.19 | 0.18 | 0.14 | -0.07 | | 0.03 | -0.23 | 0.53 |
| Nitrate, surface/WOA2018 | | -2.41 | -1.38 | -0.18 | 0.06 | | 1.41 | -0.16 | 0.78 | | 0.09 | 0.79 | 1.07 | 0.26 | -1.35 | 0.20 | | -0.74 | 0.52 | 1.04 |

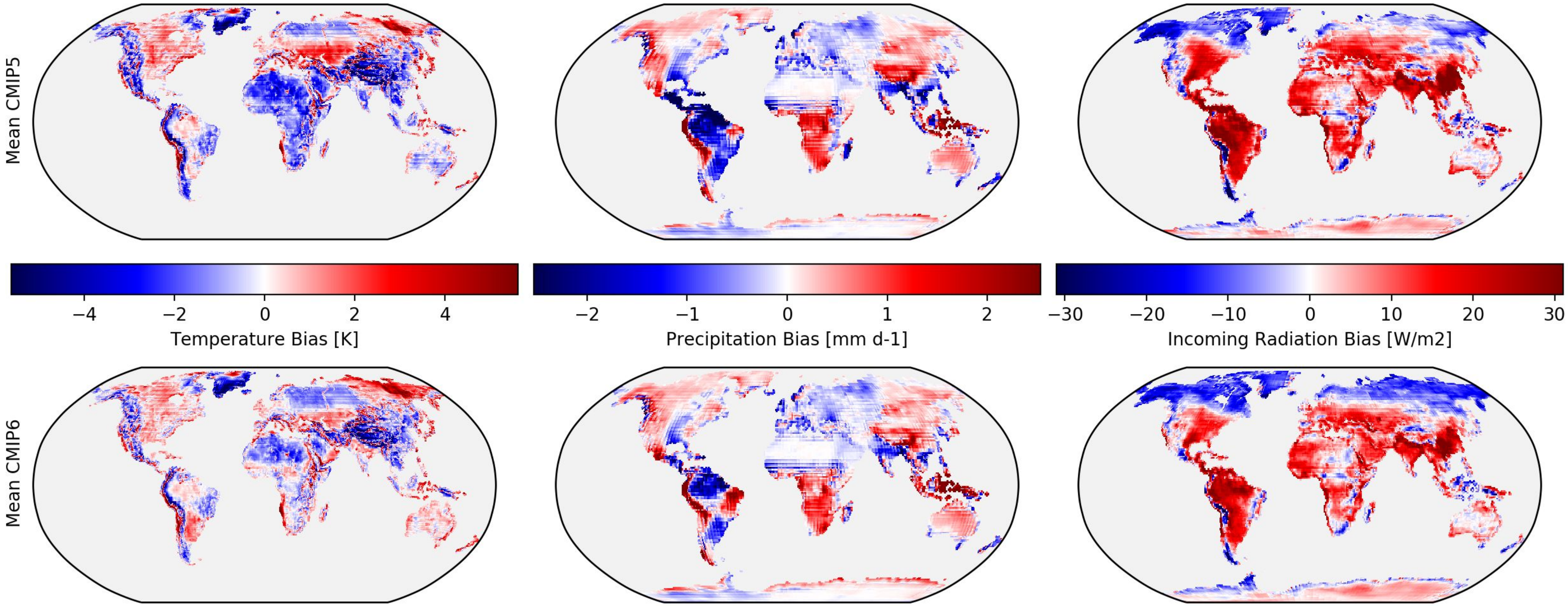
**Relative Scale**

**Worse Value** **Better Value**

**Missing Data or Error**

# Reasons for Land Model Improvements

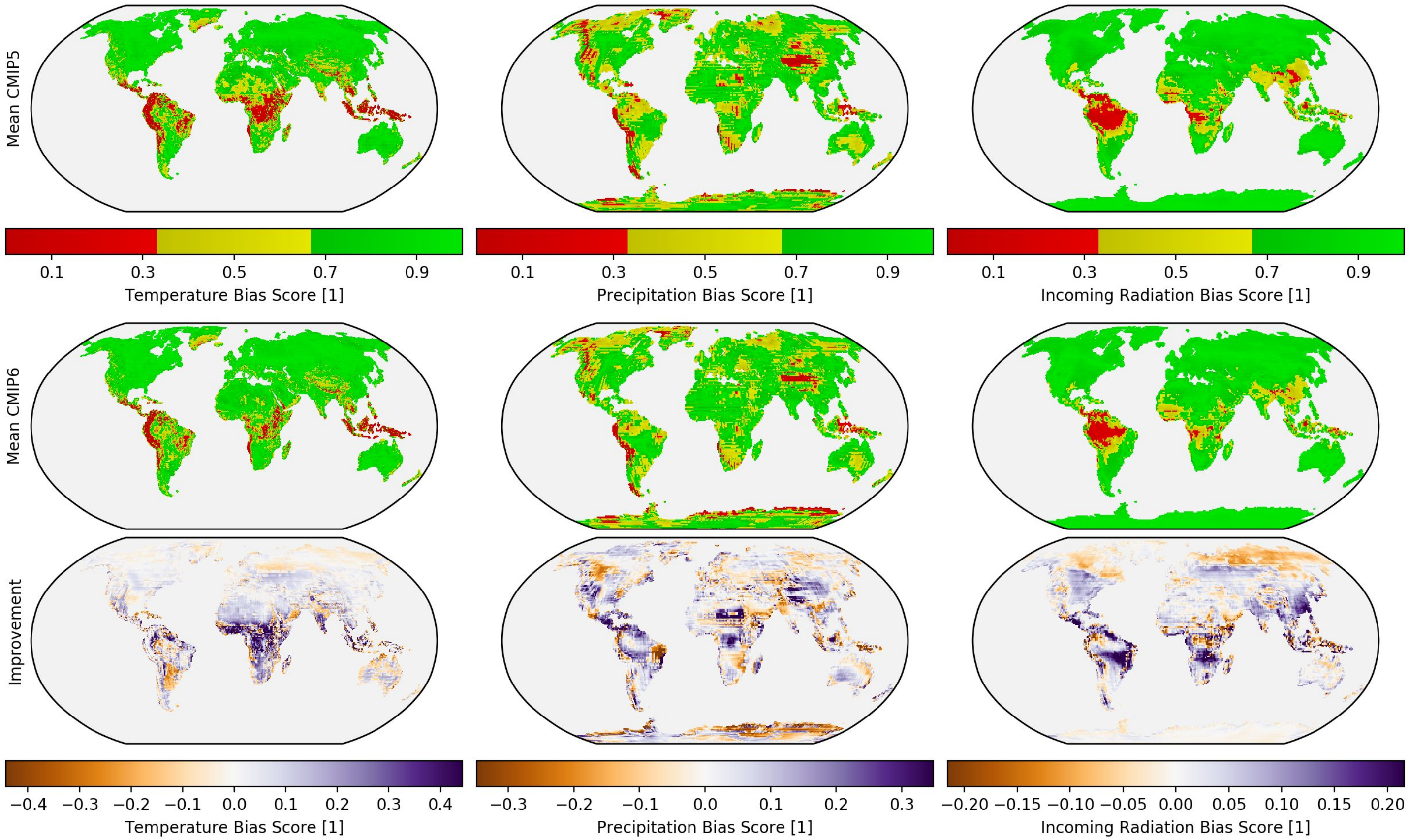
ESM improvements in **climate forcings** (temperature, precipitation, radiation) likely **partially drove improvements** exhibited by land carbon cycle models



(Hoffman et al., in prep)

# Reasons for Land Model Improvements

Differences in bias scores for temperature, precipitation, and incoming radiation were primarily positive, further indicating **more realistic climate representation**

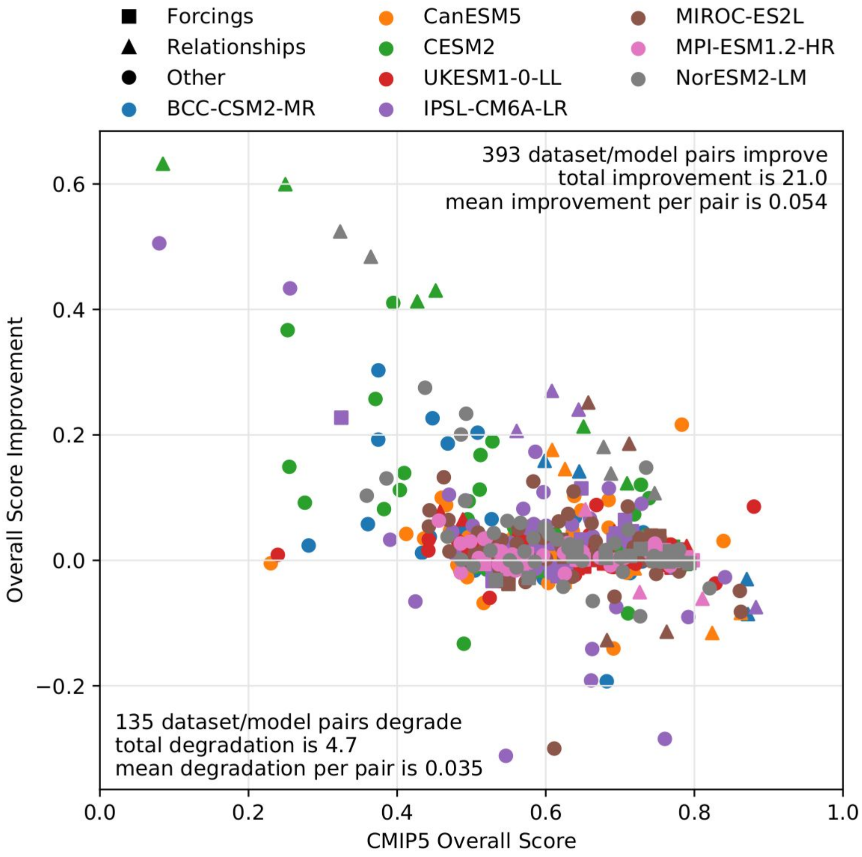


Mean CMIP5
0.1
0.3
0.5
0.7
0.9
Temperature Bias Score [1]
Precipitation Bias Score [1]
Incoming Radiation Bias Score [1]
Mean CMIP6
Improvement
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
Temperature Bias Score [1]
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
Precipitation Bias Score [1]
−0.20
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
Incoming Radiation Bias Score [1]

(Hoffman et al., in prep)

# Reasons for Land Model Improvements

- While forcings got better, the largest improvements were in **variable-to-variable relationships**, suggesting that increased land model complexity was also partially responsible for higher CMIP6 model scores
- These results suggest that **rigorous model evaluation & benchmarking** with tools like ILAMB and IOMB can lead to model improvements



(Hoffman et al., in prep)

# Coupled Earth System Model Analytics Consortium (CESMAC)

- We have maintained the **CMIP6 Data Lake at NERSC** for use by BER-funded researchers and collaborators
- Since 2019, we have added about **200 NERSC users** to the CMIP6 group for access to the data
- We are presently updating the contents of the data lake, prioritizing monthly to daily output from more simulation experiments and more models
- We are planning another **CMIP6 Hackathon** focused on the CMIP6 Data Lake and computing resources at NERSC for summer of 2023 to train early career researchers

**CMIP6 Data Lake at NERSC**
/global/cfs/cdirs/m3522/cmip6

ESGF2 EARTH SYSTEM GRID FEDERATION

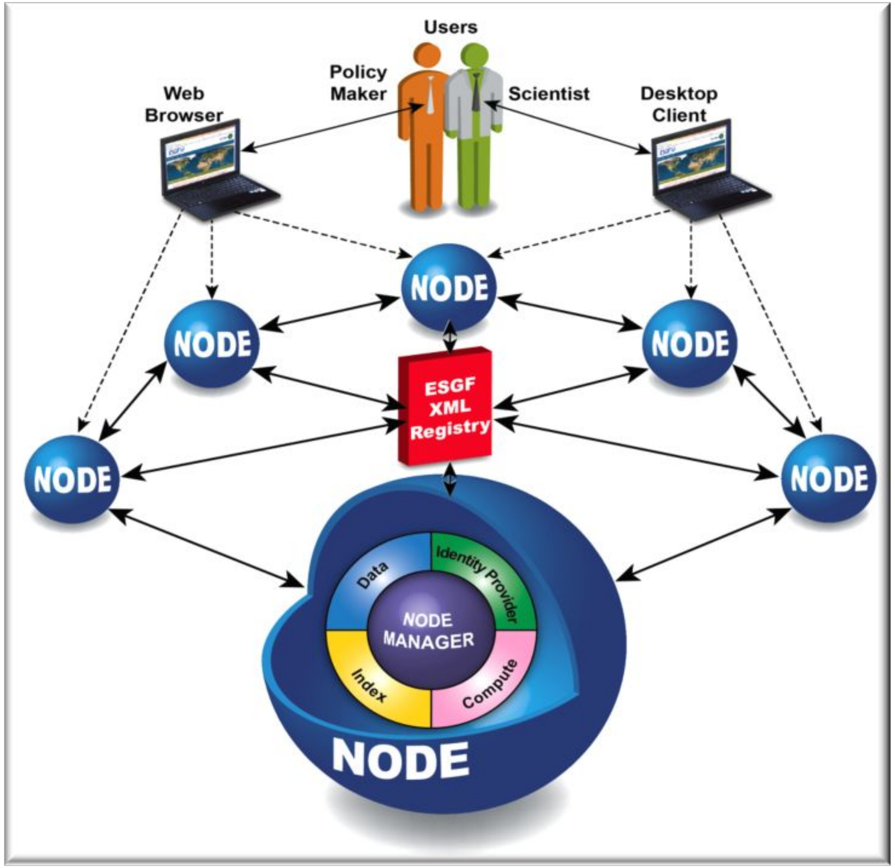
# Building the Next Generation Earth System Grid Federation (ESGF2)

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

U.S. DEPARTMENT OF ENERGY | Office of Science
OAK RIDGE National Laboratory
Argonne NATIONAL LABORATORY
Lawrence Livermore National Laboratory

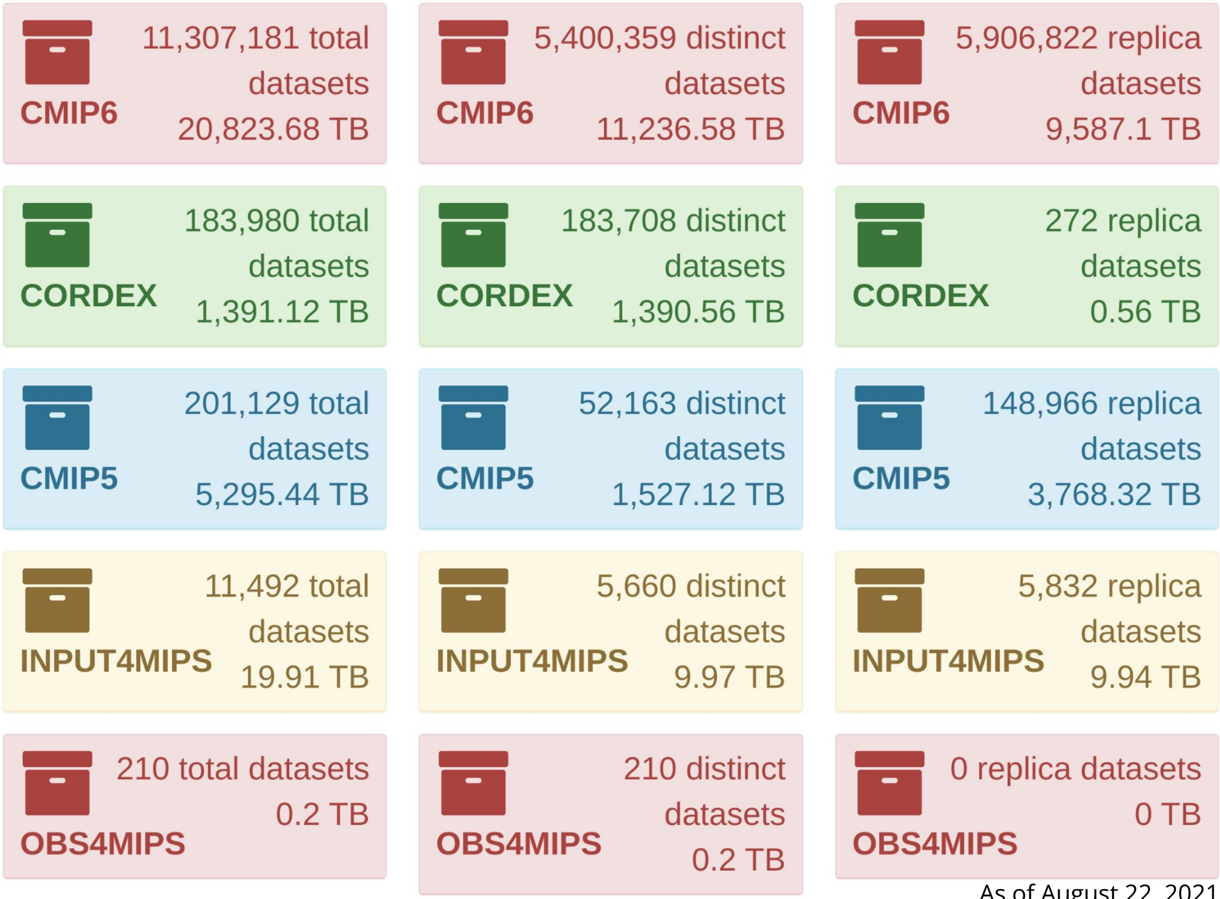
# What is the Earth System Grid Federation?

- 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



# 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



| | | |
|---|---|---|
| CMIP6<br>11,307,181 total datasets<br>20,823.68 TB | CMIP6<br>5,400,359 distinct datasets<br>11,236.58 TB | CMIP6<br>5,906,822 replica datasets<br>9,587.1 TB |
| CORDEX<br>183,980 total datasets<br>1,391.12 TB | CORDEX<br>183,708 distinct datasets<br>1,390.56 TB | CORDEX<br>272 replica datasets<br>0.56 TB |
| CMIP5<br>201,129 total datasets<br>5,295.44 TB | CMIP5<br>52,163 distinct datasets<br>1,527.12 TB | CMIP5<br>148,966 replica datasets<br>3,768.32 TB |
| INPUT4MIPS<br>11,492 total datasets<br>19.91 TB | INPUT4MIPS<br>5,660 distinct datasets<br>9.97 TB | INPUT4MIPS<br>5,832 replica datasets<br>9.94 TB |
| OBS4MIPS<br>210 total datasets<br>0.2 TB | OBS4MIPS<br>210 distinct datasets<br>0.2 TB | OBS4MIPS<br>0 replica datasets<br>0 TB |

As of August 22, 2021

# ESGF2 A New Consortium Project in the USA

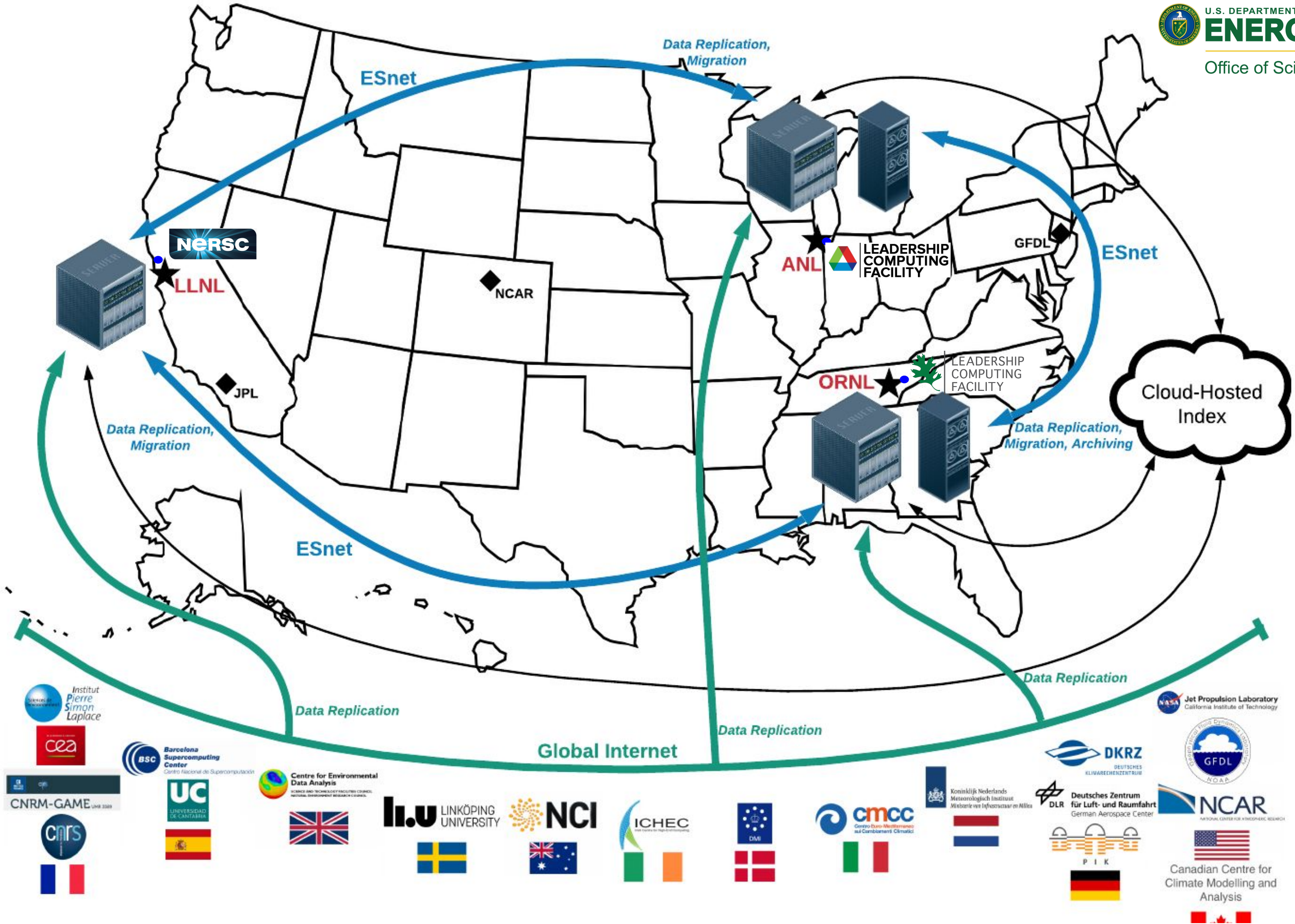
- New team from **Oak Ridge National Laboratory**, **Argonne National Laboratory**, and **Lawrence Livermore National Laboratory** proposed to modernize the data backplane based on the Globus platform
- ESGF2 proposal was **selected for funding** by DOE, starting in 2022
- In collaboration with the **ESGF 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

U.S. DEPARTMENT OF ENERGY
Office of Science

# 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


Data Replication, Migration
ESnet
NERSC
LLNL
NCAR
JPL
ANL
LEADERSHIP COMPUTING FACILITY
GFDL
ESnet
ORNL
LEADERSHIP COMPUTING FACILITY
Cloud-Hosted Index
Data Replication, Migration
Data Replication, Migration, Archiving
ESnet
Data Replication
Data Replication
Global Internet
Data Replication
Institut Pierre Simon Laplace
cea
CNRM-GAME
cnrs
Barcelona Supercomputing Center
UC Universidad de Cantabria
Centre for Environmental Data Analysis
LINKÖPING UNIVERSITY
NCI
ICHEC
DMI
cmcc
Koninklijk Nederlands Meteorologisch Instituut
DKRZ
Deutsches Zentrum für Luft- und Raumfahrt
DLR German Aerospace Center
P I K
Jet Propulsion Laboratory California Institute of Technology
GFDL
NCAR
Canadian Centre for Climate Modelling and Analysis

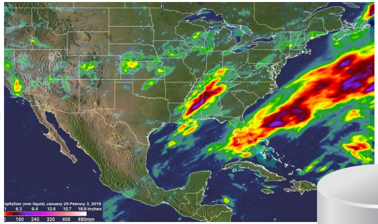
ESGF2

# 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

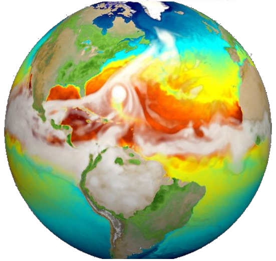
# ESGF2 Enabling a new level of research productivity

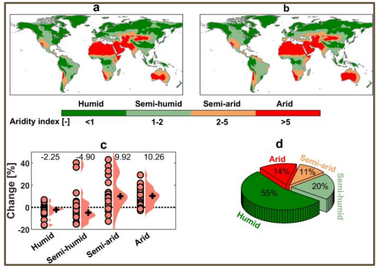
*Logging in with her* ***institutional credentials****, Samantha is presented with* ***new data, code, and papers*** *relevant to her current research. Intrigued by a new report on extreme precipitation events, she examines a* ***Jupyter notebook*** *that implements the method used. Wondering how this 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.*



E3SM
Energy Exascale Earth System Model

GitHub



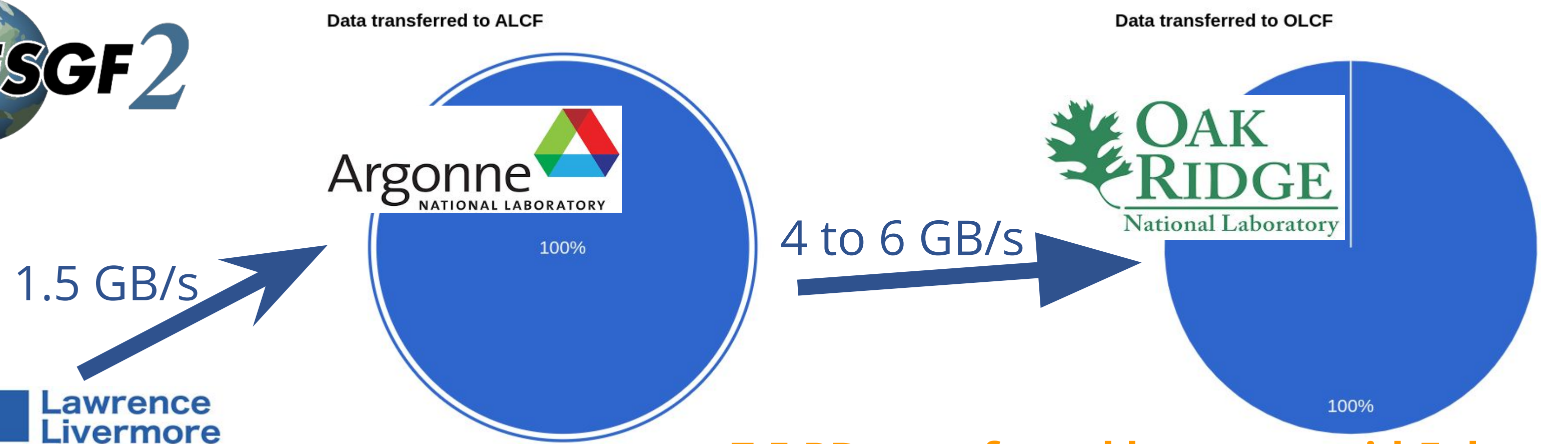


*Flood risk increases with water availability*

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

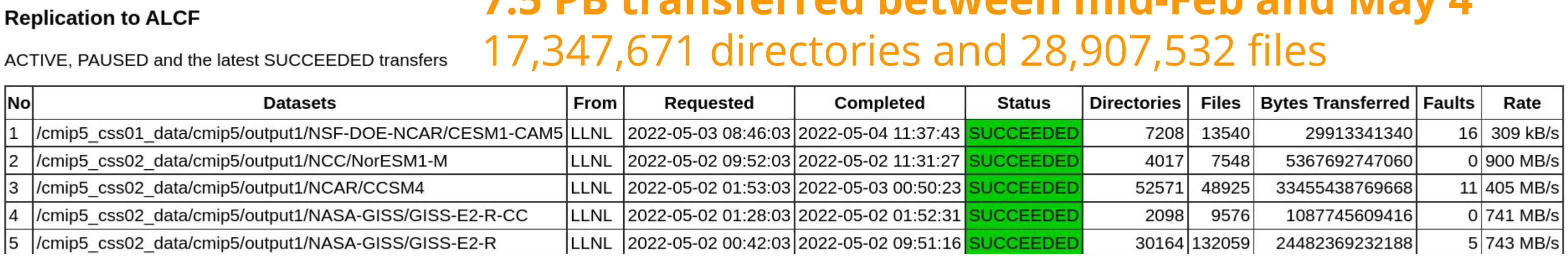
# ESGF2 ESGF Failsafe Data Replication

- **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*)



**7.5 PB transferred between mid-Feb and May 4**

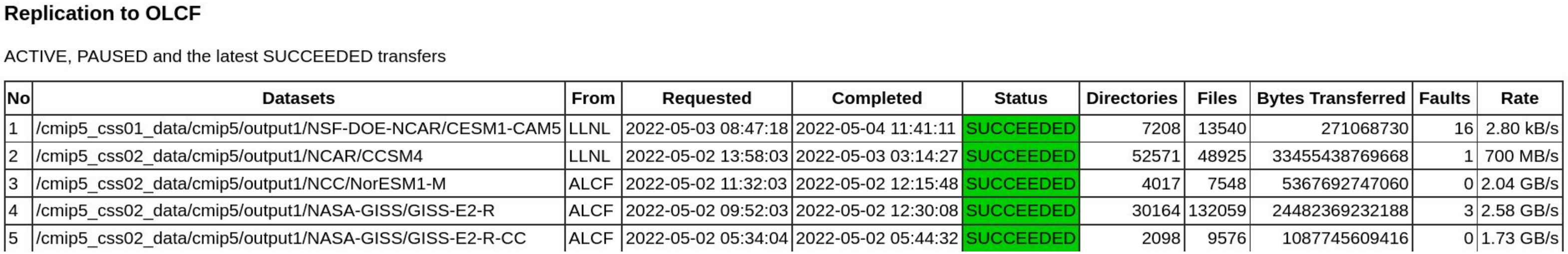
17,347,671 directories and 28,907,532 files



### Replication to ALCF

ACTIVE, PAUSED and the latest SUCCEEDED transfers

| No | Datasets | From | Requested | Completed | Status | Directories | Files | Bytes Transferred | Faults | Rate |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | /cmip5_css01_data/cmip5/output1/NSF-DOE-NCAR/CESM1-CAM5 | LLNL | 2022-05-03 08:46:03 | 2022-05-04 11:37:43 | SUCCEEDED | 7208 | 13540 | 29913341340 | 16 | 309 kB/s |
| 2 | /cmip5_css02_data/cmip5/output1/NCC/NorESM1-M | LLNL | 2022-05-02 09:52:03 | 2022-05-02 11:31:27 | SUCCEEDED | 4017 | 7548 | 5367692747060 | 0 | 900 MB/s |
| 3 | /cmip5_css02_data/cmip5/output1/NCAR/CCSM4 | LLNL | 2022-05-02 01:53:03 | 2022-05-03 00:50:23 | SUCCEEDED | 52571 | 48925 | 33455438769668 | 11 | 405 MB/s |
| 4 | /cmip5_css02_data/cmip5/output1/NASA-GISS/GISS-E2-R-CC | LLNL | 2022-05-02 01:28:03 | 2022-05-02 01:52:31 | SUCCEEDED | 2098 | 9576 | 1087745609416 | 0 | 741 MB/s |
| 5 | /cmip5_css02_data/cmip5/output1/NASA-GISS/GISS-E2-R | LLNL | 2022-05-02 00:42:03 | 2022-05-02 09:51:16 | SUCCEEDED | 30164 | 132059 | 24482369232188 | 5 | 743 MB/s |



### Replication to OLCF

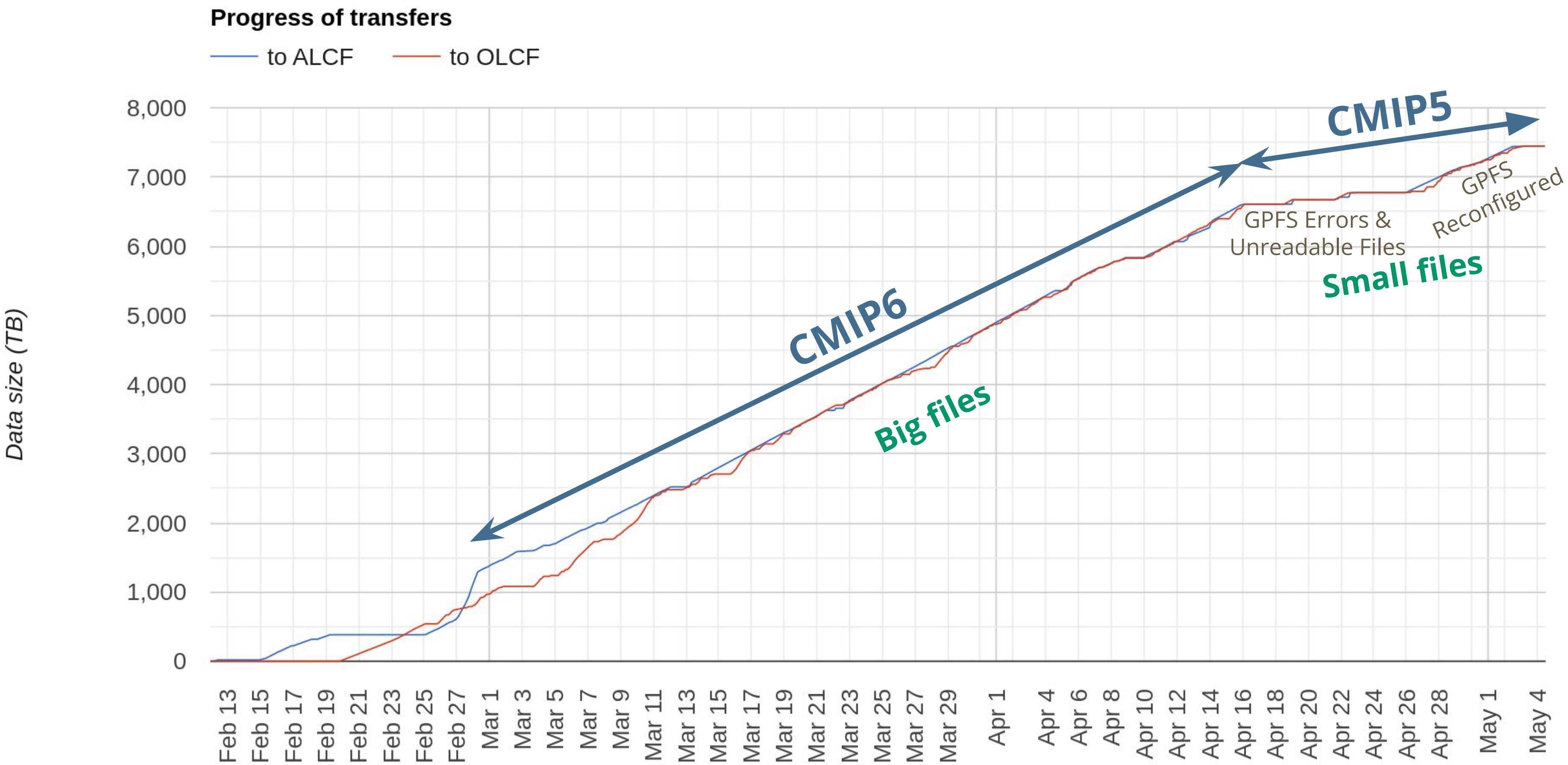
ACTIVE, PAUSED and the latest SUCCEEDED transfers

| No | Datasets | From | Requested | Completed | Status | Directories | Files | Bytes Transferred | Faults | Rate |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | /cmip5_css01_data/cmip5/output1/NSF-DOE-NCAR/CESM1-CAM5 | LLNL | 2022-05-03 08:47:18 | 2022-05-04 11:41:11 | SUCCEEDED | 7208 | 13540 | 271068730 | 16 | 2.80 kB/s |
| 2 | /cmip5_css02_data/cmip5/output1/NCAR/CCSM4 | LLNL | 2022-05-02 13:58:03 | 2022-05-03 03:14:27 | SUCCEEDED | 52571 | 48925 | 33455438769668 | 1 | 700 MB/s |
| 3 | /cmip5_css02_data/cmip5/output1/NCC/NorESM1-M | ALCF | 2022-05-02 11:32:03 | 2022-05-02 12:15:48 | SUCCEEDED | 4017 | 7548 | 5367692747060 | 0 | 2.04 GB/s |
| 4 | /cmip5_css02_data/cmip5/output1/NASA-GISS/GISS-E2-R | ALCF | 2022-05-02 09:52:03 | 2022-05-02 12:30:08 | SUCCEEDED | 30164 | 132059 | 24482369232188 | 3 | 2.58 GB/s |
| 5 | /cmip5_css02_data/cmip5/output1/NASA-GISS/GISS-E2-R-CC | ALCF | 2022-05-02 05:34:04 | 2022-05-02 05:44:32 | SUCCEEDED | 2098 | 9576 | 1087745609416 | 0 | 1.73 GB/s |

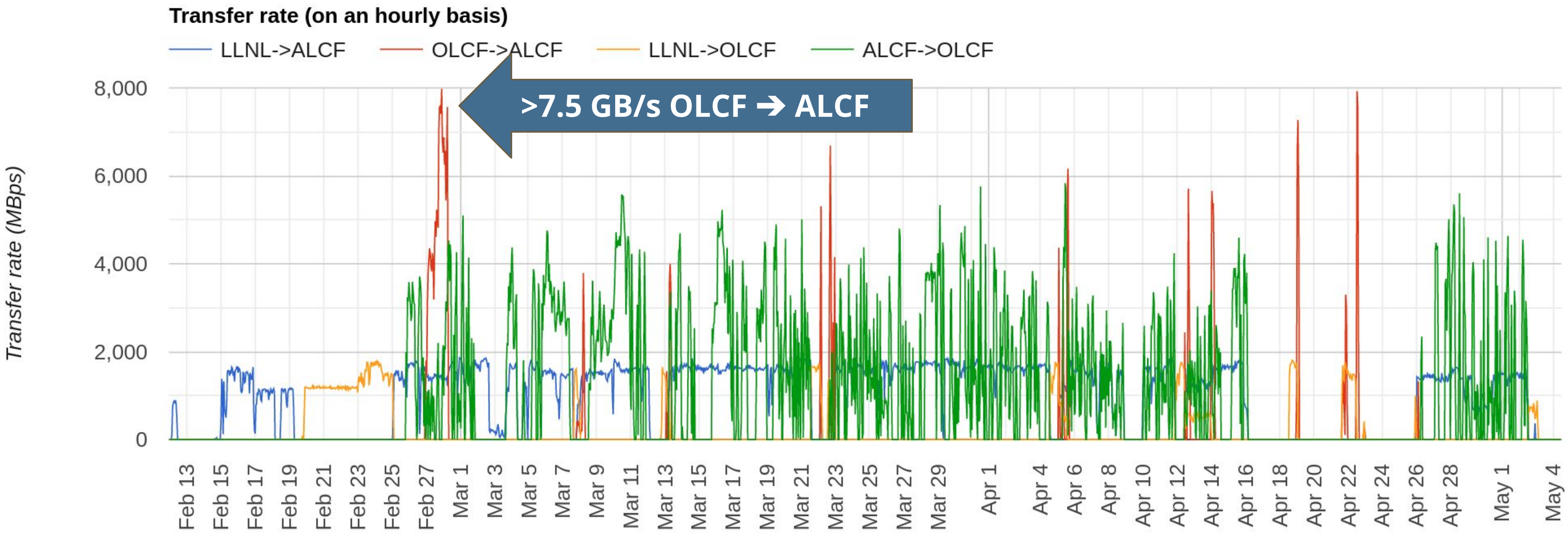
https://dashboard.globus.org/esgf

As of May 4, 2022

# ESGF2 Cumulative Data Transferred Over Time



# ESGF2 Transfer Rates Over Time



>7.5 GB/s OLCF ➔ ALCF

# ESGF2 Summary

- 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