

## **A data infrastructure for data-intensive climate research**

Dean Williams<sup>†</sup>;

<sup>†</sup> Lawrence Livermore National Laboratory, USA

Leading author: [williams13@llnl.gov](mailto:williams13@llnl.gov)

Through the ESG project, the ESG-CET team has developed and delivered a production environment for climate data from multiple climate model sources (e.g., CMIP (IPCC), CESM, ocean model data (e.g., Parallel Ocean Program), observation data (e.g., Atmospheric Infrared Sounder, Microwave Limb Sounder), and analysis and visualization tools) that serves a worldwide climate research community. Data holdings are distributed across multiple sites including LANL, LBNL, LLNL, NCAR, and ORNL as well as unfunded partners sites such as the Australian National University (ANU) National Computational Infrastructure (NCI), the British Atmospheric Data Center (BADC), the Geophysical Fluid Dynamics Laboratory/NOAA, the Max Planck Institute for Meteorology (MPI-M), the German Climate Computing Centre (DKRZ), and NASA/JPL. More recently, ESG has focused on broadening and generalizing the ESG system to support a more distributed, international, and diverse collection of archived sites and data types. In addition, ESG is extending services beyond access to raw data files by developing information products (scientific graphics, animations, etc.) and server-side analysis capabilities that will allow users to request output from commonly used analysis and intercomparison procedures. As we transition from development activities to production and operations, the ESG-CET team is tasked with making data available to all users who want to understand it, process it, extract value from it, visualize it, and/or communicate it to others. This ongoing effort is extremely large and complex, but it will be incredibly valuable for building science gateways to critical climate resources (such as CESM, CMIP5, ARM, NARCCAP, Atmospheric Infrared Sounder (AIRS), etc.) for processing the next IPCC assessment report. Continued ESG progress will result in a production-scale system that will empower scientists to attempt new and exciting data exchanges, which could ultimately lead to breakthrough climate science discoveries. The ESG-CET executive committee consists of Dean N. Williams, Lawrence Livermore National Laboratory (LLNL); Ian Foster, Argonne National Laboratory (ANL); and Don Middleton, National Center for Atmospheric Research (NCAR). The ESG-CET team is a group of researchers and scientists with diverse domain knowledge, whose home institutions include eight laboratories and two universities: ANL, Los Alamos National Laboratory (LANL), Lawrence Berkeley National Laboratory (LBNL), LLNL, NASA/Jet Propulsion Laboratory (JPL), NCAR, Oak Ridge National Laboratory (ORNL), Pacific Marine Environmental Laboratory (PMEL)/NOAA, Rensselaer Polytechnic Institute (RPI), and University of Southern California Information Sciences Institute (USC/ISI). All ESG-CET work is accomplished under DOE open-source guidelines and in close collaboration with the project's stakeholders, domain researchers, and scientists.