



## **ESGF Node - A data infrastructure for data-intensive science**

Estanislao Gonzalez (1), Gavin M. Bell (2), Luca Cinquini (3), Philip Kershaw (4), Stephan Kindermann (5), Michael Lautenschlager (5), Stephen Pascoe (4), and Dean Williams (2)

(1) Max-Planck-Institut für Meteorologie, Data Management, Hamburg, Germany (estanislao.gonzalez@zmaw.de), (2) Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Labs, (3) Jet Propulsion Laboratory, National Aeronautics and Space Administration, (4) NCAS/British Atmospheric Data Centre, Rutherford Appleton Laboratory, STFC, (5) Deutsches Klimarechenzentrum (DKRZ)

The ESGF is a community driven effort to produce open-source data infrastructures to support data-intensive science by guaranteeing performant and transparent access to large amounts of data in a geographically distributed federation. The primary application is the ESGF Node. The ESGF Node will be used to support the CMIP5/IPCC-AR5 data infrastructure having 1 PB of replicated data and 10 PB in total across the federation. The ESGF Node allows an institution to check, publish, replicate and administer data as well as to secure access to its resources. The modularity of the ESGF Node is central to its conception and it already provides means for file serving (HTTP, GridFTP, OpenDAP), browsing (Thredds), visualization, access metrics, as well as many others.