

A data statistics service for data publication and usage metrics in the climate domain

Tuesday, 1 October 2024 16:10 (20 minutes)

In the climate domain, the Coupled Model Intercomparison Project (CMIP) represents a collaborative framework designed to improve knowledge of climate change with the important goal of collecting output from global coupled models and making them publically available in a standardized format. CMIP has led to the development of the Earth System Grid Federation (ESGF), one of the largest-ever collaborative data efforts in earth system science involving a large set of data providers and modelling centres around the globe.

ESGF manages a huge distributed and decentralized database for accessing multiple petabytes of science data at dozens of federated sites. In this context, providing an in-depth understanding about the data published and exploited across the federation is of paramount importance in order to get useful insights on the long tail of research.

To this end, the ESGF infrastructure includes a specific software component, named ESGF Data Statistics, deployed at the CMCC SuperComputing Center. More specifically, the service takes care of collecting, storing, and analyzing data usage logs (prior filtering out sensitive information) sent by the ESGF data nodes on a daily basis. A set of relevant usage metrics and data archive information are then visualized on an analytics user interface including a rich set of charts, maps and reports, allowing users and system managers to visualize the status of the infrastructure through smart and attractive web gadgets.

Further insights relevant to the research infrastructure managers could come through the application of a data-driven approach applied to download information in order to identify changes in the download patterns and predict possible issues at the infrastructural level.

Topic

Needs and solutions in scientific computing: Platforms and gateway

Primary author: NUZZO, Alessandra (CMCC Foundation)

Co-authors: ANTONIO, Fabrizio (CMCC); NASSISI, Paola (CMCC); MIRTO, Maria (CMCC); FIORE, Sandro (University of Trento); ELIA, Donatello (CMCC); ALOISIO, Giovanni (CMCC)

Presenter: NUZZO, Alessandra (CMCC Foundation)

Session Classification: Unlocking the Potential of Environmental Data