Integration and validation of a data grid software

Nicolas Carenton-Madiec (1), Katharina Berger (2), and Antonio Cofino (3)

(1) CNRS/IPSL, Climate Data Management, Rouen, France (nicolas.carenton@ipsl.jussieu.fr), (2) DKRZ, Climate Data Management, Hamburg, Germany (berger@dkrz.de), (3) UNICAN, Climate Data Management, Santander, Spain (antonio.cofino@unican.es)

The Earth System Grid Federation (ESGF) Peer-to-Peer (P2P) is a software infrastructure for the management, dissemination, and analysis of model output and observational data. The ESGF grid is composed with several types of nodes which have different roles. About 40 data nodes host model outputs and datasets using thredds catalogs. About 25 compute nodes offer remote visualization and analysis tools. About 15 index nodes crawl data nodes catalogs and implement faceted and federated search in a web interface. About 15 Identity providers nodes manage accounts, authentication and authorization. Here we will present an actual size test federation spread across different institutes in different countries and a python test suite that were started in December 2013. The first objective of the test suite is to provide a simple tool that helps to test and validate a single data node and its closest index, compute and identity provider peer. The next objective will be to run this test suite on every data node of the federation and therefore test and validate every single node of the whole federation. The suite already implements nosetests, requests, myproxy-logon, subprocess, selenium and fabric python libraries in order to test both web front ends, back ends and security services. The goal of this project is to improve the quality of deliverable in a small developers team context. Developers are widely spread around the world working collaboratively and without hierarchy. This kind of working organization context en-lighted the need of a federated integration test and validation process.