



Ensuring a unified user experience in a federated portal architecture: experiences of the CMIP5 archive

Stephen Pascoe, Philip Kershaw, Martin Jukes, and Bryan Lawrence
(Stephen.Pascoe@stfc.ac.uk)

The 5th Coupled Model Intercomparison Project (CMIP5) is producing petabytes of data at multiple institutions around the world. These data are not only generated globally but download demand is also anticipated to be global given that CMIP5 will form the basis for many papers included in the IPCC's 5th Assessment Report.

The large data volume and global distribution of the CMIP5 archive naturally fits a federated architecture, however end-users expect a unified interface to the data which allows them to find and download what they want where ever it is located. This paper focuses on how the Earth System Grid Federation (ESGF) architecture, deployed for CMIP5, balances the restrictions imposed by distributed resources and the requirement for a unified user experience. We also aim to provide a practical introduction to users wishing to download CMIP5 data.

Users access the CMIP5 archive via one of several web Gateways which enable search and browse of datasets held on underlying Datanodes. Each Datanode publish the datasets they hold to a single Gateway, however inter-Gateway metadata exchange enables all datasets to be discoverable from all Gateways. Also some datasets are available as replicas on other Datanodes to mitigate bandwidth limitations of long-distance data transfer. Thus the user must be presented with multiple download locations once they have discovered the data of interest. The Gateway then generates download links directly to the relevant Datanodes and data download is carried out directly between client and Datanode, avoiding bottlenecks at the Gateway.

A federated security architecture is vital to providing a unified user experience. ESGF security supports single sign-on via both OpenID and X.509 PKI technologies with role-based authorisation facilitated by SAML and the OpenID attribute exchange protocol. This dual-technology approach allows ESGF security to be applied to access methods that are familiar to users of climate data but that have been traditionally been insecure. Thus users can access data through a browser, scripted HTTP client, OPeNDAP protocol or GridFTP using a single OpenID identity from any member of the federation. Where adaptation of the user's tools and workflow is required the Gateway helps by generating scripts containing security-enabled HTTP calls and by providing a web-executable MyProxy login application for retrieving PKI credentials.

Particular efforts have been made to enable OPeNDAP clients to access secured CMIP5 data through a collaboration with the NetCDF developers at Unidata. By linking to a snapshot of the NetCDF-4 API any NetCDF-enabled client should be able to access secured data without further client modifications.