Geophysical Research Abstracts
Vol. 13, EGU2011-8272, 2011
EGU General Assembly 2011

Access Control Architecture for the Earth System Grid Federation: Building an Infrastructure of Secured Data Access Services for the Climate Science Research Community

Philip Kershaw (1), Rachana Ananthakrishnan (2), Luca Cinquini (3), Estanislao Gonzalez (4), Dennis Heimbigner (5), and Bryan Lawrence (1)

(1) STFC Rutherford Appleton Laboratory, NCAS/British Atmospheric Data Centre, Didcot, United Kingdom (philip.kershaw@stfc.ac.uk), (2) Argonne National Laboratory, Argonne, IL, USA, (3) Jet Propulsion Laboratory, National Aeronautics and Space Administration, Pasadena, CA, USA, (4) Max-Planck-Institut für Meteorologie, Data Management, Hamburg, Germany, (5) University Corporation for Atmospheric Research, Boulder, CO, USA

The Earth System Grid Federation (ESGF) is deploying a software infrastructure developed as part of an international collaboration to facilitate access to data produced by the fifth Coupled Model Intercomparison Project (CMIP5). The analysis of this data will be a key component of the forthcoming fifth IPCC assessment on Climate Change. Many petabytes (PB) of climate data will be made available in a globally federated network, with approximately 1PB of key data replicated in a set of "long-term archives". Three sites will take long term responsibility for the "key data" archive (The Program for Climate Model Diagnosis and Intercomparison, PCMDI, the German Climate Computing Centre, DKRZ, and the British Atmospheric Data Centre, BADC), but other sites are expected to host replicants too.

The large scale and scope of ESGF provides strong driver for a federated approach to access control to services. The collective focus of a common goal, a willingness of partners to co-operate and the availability of resources to support this work were necessary conditions for the construction of the ESGF systems. Security is predicated on trust and inter-organisational trust is fundamental to a federation and its ability to function. Conversely, the collaboration effort required to create such a security infrastructure fosters that relationship of trust between the respective development communities involved. As a cross-cutting concern, security threads its way through all aspects of the system, its services and their interfaces and as such provides an important litmus test for the functioning health of the federation.

For the ESGF, the security system was developed from requirements including the ability to restrict access to registered users, keep these users up to date with changes to data and services, audit access, protect finite computing resources and where possible, maintain the ease of use for the user community with existing software tools and services with which they are already familiar. Significantly, PCMDI with its lead role for CMIP5, maintains authority for user registration authorizing users to access datasets that are part of the CMIP5 archive, while individual institutions retain control over other datasets. Whilst ease of use for end users was identified as a strong requirement, the development, deployment and integration processes have in turn highlighted the inherent complexities in rolling out such a system for the community of developers and deployers and the need to disseminate security knowledge and provide tools to support them in this task.

The ESGF security system makes use of OpenID, PKI (Public Key Infrastructure), SAML (Security Assertion Mark-up Language) and Grid based security solutions to make a complete solution for federated security. Here we provide an overview of the federated security system and how it is being deployed at the Centre for Environmental Data Archival (CEDA, the host organisation for the BADC). A key enabler in this context has been the extension to secure both OPeNDAP services and the NetCDF client API to enable authentication with Grid based authentication credentials over HTTP/HTTPS. Within CEDA, the ESGF security system is being deployed not only with vanilla ESG software, but also with python based middleware protecting both PyDAP and other webservices developed with previous projects.

The discussion highlights the challenges of bridging together independent security protocols and applying them within the climate science research application domain, as well as the heritage of the concepts and code (including the strong influence on both from the development of secure systems for the Natural Environment

Research Council DataGrid, NDG).