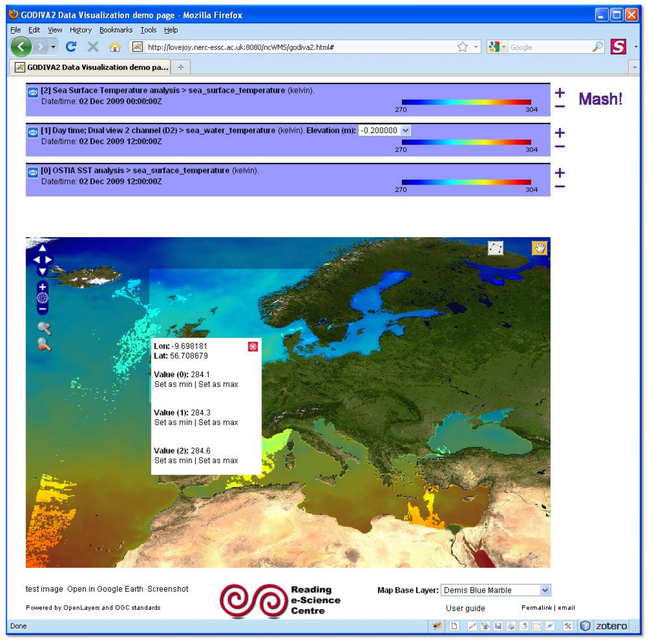
Science & Technology Facilities Council
Rutherford Appleton Laboratory

Building a Web Based Environment for the Intercomparison of Distributed Environmental Science Data: Challenges in Access Control and Security

Philip Kershaw [philip.kershaw@stfc.ac.uk] (1), Jon Blower (2), Alastair Gemmell (2), Stephen Pascoe (1), and Ag Stephens (1)
(1) STFC Rutherford Appleton Laboratory, NCAS/British Atmospheric Data Centre, Didcot, United Kingdom, (2) Reading e-Science Centre, Environmental Systems Science Centre, University of Reading, United Kingdom

The MashMyData Project

MashMyData[1] is a **NERC** (Natural Environment Research Council) funded *Technology Proof of Concept* project whose aim is to enable environmental scientists to combine and overlay geospatial data quickly and easily in the context of a web style mash-up. **Reading e-Science Centre** are developing a Web Portal to meet this task which enables users to upload their own data and compare it with data pulled from other sources. This is made possible via services hosted over web based data access protocols such as **OPeNDAP** and the **OGC (Open Geospatial Consortium) W*S** standards. MashMyData explores a use case of multiple services working together to perform retrievals and intercomparison of datasets. The **BADC**'s OGC Web Processing Service (WPS) implementation[2] is employed as a means to avoid large data transfers and perform calculations on data in situ. The BADC is also providing OPeNDAP services and expertise in access control and security.



MashMyData user interface with three layers presented on the map: OSTIA sea surface temperature layer (bottom), AATSR sea temperature (middle), and a model prediction of sea surface temperature from the Danish Meterological Institute (top). These are all for 2nd December 2009.

Federated Access to Restricted Datasets

Included in the remit for MashMyData is the ability to access *restricted* datasets from *independent administrative domains*. In this, it leverages **data services hosted at the BADC** which are secured using the access control system developed for the **Earth System Grid Federation**[3][4]. ESGF is a federated infrastructure of sites from around the world providing access to earth science data. Users can gain access to such data across the federation through *OpenID* and PKI (Public Key Infrastructure) based single sign-on technologies, and VO-wide (Virtual Organisation) management of access rights. Implementations have been made in both Java and Python (BADC's NDG Security system).

ESGF Security Extensions for NetCDF

Working in collaboration with Unidata, the makers of NetCDF, the NetCDF C libraries have been modified to support calls to secured ESGF OPeNDAP services (as of NetCDF 4.1.2 release). To make a secure client call, the library must supports the ability to follow HTTP redirects and the setting of SSL connection parameters including a client X.509 certificate. The client certificate can be an *End Entity Certificate* or a *Proxy Certificate*. **This opens up OPeNDAP services to access via the Grid based authentication paradigm.**

These additional security settings do not impact on the existing API, so client software and tools that build on NetCDF can incorporate the enhancements simply by relinking with the latest version of the libraries. No change to source code is necessary. The MashMyData development team are working with Unidata to incorporate similar enhancements to the Java NetCDF libraries.

The User Delegation Problem

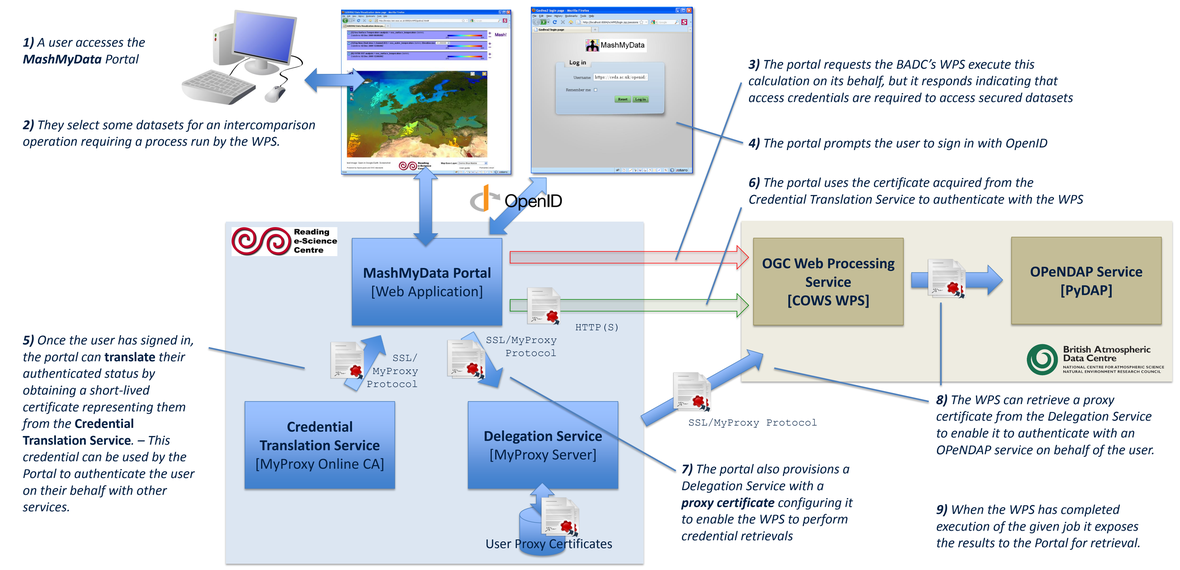
The MashMyData use case taken in the context of secured data services inevitably invokes the classical problem of user delegation. The Portal cannot access restricted data as an entity in itself since restricted data are secured on the basis of user identities not Portals or other abstract entities. Instead, the access control system requires a means to delegate the user's access rights to the Portal and the WPS so that these services can act on the user's behalf.

Two Alternative Solutions

The Grid provides the tried and tested solution of RFC3820 **Proxy Certificates**[5] but **OAuth**[6] also provides an attractive alternative. OAuth is an open Web based protocol which enables trusted parties authorisation to access secured resources on behalf of a user. Used together with OpenID, it has been demonstrated as part of the **OGC OWS-6 Testbed**[7]. Significantly though, the ESGF Security model already supports PKI based authentication to services with short-lived user X.509 Certificates. By altering data services' SSL middleware, they can support the consumption of Proxy certificates.

In the first instance, a Proxy certificate based approach has been adopted for MashMyData. This makes use of the MyProxy Credential Management Service available from the Globus Toolkit as a workhorse to perform two distinct functions: as a means to translate from the users' authenticated status at the Portal to an X.509 certificate which the Portal can use to authenticate with other services and as a repository for trusted services such as the WPS to retrieve delegated credentials to act on behalf of the user. This is illustrated in the diagram opposite. This approach has initially proved the most straightforward but OAuth based solutions deserve further attention as they would eliminate the need for a Credential Translation Service and specialist Grid middleware.

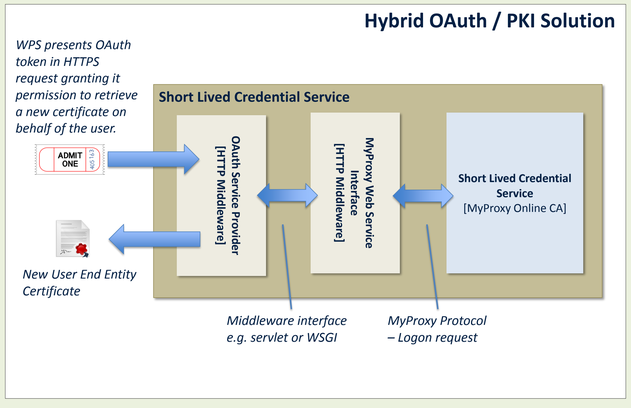
OGC and OPeNDAP Services in a Secured Workflow



The *OAuth* Option

A conventional OAuth based solution would see secure services hosting protected data – in OAuth terms *Service Providers* – fronted with OAuth filters. This is a viable option for MashMyData: WPS and OPeNDAP services would be secured with OAuth based middleware.

However, for the **CILogon**[8] project, a novel approach has been applied whereby the Service Provider is not a data server. It is actually a Credential Delegation Service. This replaces the more conventional approach to delegation adopted in the solution described above: delegation by means of proxy certificates issued from a MyProxy Server. Instead, the client presents an OAuth token over a HTTP interface. This delegates them rights to a new user certficate to use on their behalf. MyProxy is still used in this scenario, but rather than issuing proxy certificates, it issues so called *End Entity Certificates* (conventional X.509 Certificates). This is important because EECs can be consumed by other services using standard SSL middleware. No specialist GSI (Grid Security Infrastructure) based middleware is required.



References

[1] MashMyData: a Gateway for Scientific Visualization and Intercomparison of Secure, Distributed Data, Alastair Gemmell et al., EGU2011-9280, Session ESSI12, Poster XL236, Thurs 7 Apr 2011 [author in attendance: 17:30-19:00]
[2] Useful extensions to the OGC Web Processing Service based on a Python client/server implementation, Ag Stephens et al., EGU2011-7777, Session ESSI11, Poster XL220, Tue 5 Apr 2011 [17:30–19:00]
[3] ESGF Node - A data infrastructure for data-intensive science, Estanislao Gonzalez et al., EGU2011-4797, Session ESSI16, Poster XL173, Tue 5 Apr 10:30–11:15
[4] 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 et al., EGU2011-8272, Session NH1.6/HS12.8, Room 10 Wed 06 Apr 2011 14:15
[5] Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile, http://www.ietf.org/rfc/rfc3820.txt
[6] OAuth, http://oauth.net/
[7] OGC® OWS-6 Security Engineering Report, OGC 09-035, Rüdiger Gartmann, Lewis Leinenweber, 0.3.0, 9 Sept 2009
[8] CILogon, NCSA, http://www.cilogon.org/

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Reading e-Science Centre

NATURAL ENVIRONMENT RESEARCH COUNCIL

RAL Space