

# ESGF Publication, Security, and Data Management for CMIP6

Reporting period: July 1 2016 to September 30 2016

This is a report of work completed for third quarter of Year 2 for the subaward to University of Chicago: ESGF Publication, Security, and Data Management for CMIP6.

## Installation and admin guide

Updated ESGF installer to make Globus registration optional. This meets requirements from the European sites that do not want to register the server as a Globus endpoint by default.

Updated the ESGF administrator guide published at earthsystemcog.org to include instructions on setting up CoG enabled for Globus download, both for public and restricted datasets. This has been successfully tested at JPL instance, and these instructions will allow other sites to enable it.

There have been several DTNs dedicated for high performance data transfer that have been set up at various sites. ESGF data nodes leveraging these well tuned and dedicated DTNs will allow for best performance with file transfers. But this requires enforcement of ESGF security for such data access. To that end, we documented and tested processes for configuring Globus Connect Server with SAML authorization callout, such that it can be setup on an independent DTN.

## Ongoing maintenance

Continue to upgrade and maintain the ESGF services at Argonne. Two installations, one for development and another in production, were upgraded to pull in new features in CoG and esgf-search. Further maintenance activities to keep the servers operational, including fixes to operate the services post a storage issue and power outage at Argonne. Specific upgrades include, ESGF 2.3.8, ESGF publisher to 4.0.1 and ESGF search to 4.10.0.

## ESGF security

Publication in ESGF requires a specific role to be assigned to the user: that grants them the right to write data to the node and invoke publication scripts. In addition to support wildcards in authorization, ESGF used custom extensions to the server code. Extended the ESGF authorization callout to subsume the wildcard capability, allow write by users with the publisher role and also enabled use of sharing with this callout. This removes the need for multiple GridFTP servers on the node to support data download (with or without use of Globus sharing), and publication. With this change, an administrator can setup a single server and based on the authorization publication or download options are allowed.

Created the Argonne ESGF Certificate Authority to start issuing certificates for ESGF sites. Certificate Security Policy document is currently in progress, and discussions are ongoing to

determine level of assurance requirements for the CA. Certificate requests have been handled and a few certificates have been issued. Ongoing effort to plan maintenance and user support for this CA. As part of this effort, updated the ESGF trust roots to include this new CA.

## Prototype Globus as Identity Provider in ESGF

The initial prototype work for feasibility for using Globus as alternate identity provider in ESGF was expanded to support end to end data access, including using Globus Transfer for downloads, and HTTP/S downloads. This includes use of Globus Sharing and that required integration with Globus groups for permission. The prototype used the ESGF group management to manage groups, but update changes to Globus groups to support sharing. This allows for users with campus logins to use ESGF infrastructure, without removing the existing OpenID support. Further design work is ongoing to explore u.

OAuth 2.0 has been identified as a protocol of choice to improve integration across services and better delegation. Completed a prototype of integration of OAuth 2.0 protocol with the ESGF identity provider, and that meets the requirements for this effort also.

## Publication

Using the script that was developed for adding Globus URLs to published datasets, worked with various sites to include Globus URLs so users can use Globus Transfer to move the data of choice.

An Ingestion REST API was complete to provide programmatic interface to publication. This API has been tested and used any one project. A client script that uses the API to publish data was also completed. To make this generally useful to the ESGF users, and applicable to broader use than the single project, completed detailed documentation on installation and use of the service.

## Data transfer and replication

Added support for Globus URLs in the Synda tool so in addition to HTTP and GridFTP transfers, Globus Transfer service can be used. This has been tested and integrated with Synda.

In partnership with ICNWG, we continue to support installation, tuning and use of GridFTP servers and required security configuration, for basic testing of transfer speeds between Tier 1 sites. Documentation improvements and debugging of specific site configuration to support the ICNWG mission is ongoing.