Data Transfer Working Team

ESGF F2F Workshop,

Washington DC, December 2016

Working team leads: Rachana Ananthakrishnan, Lukasz Lacinski
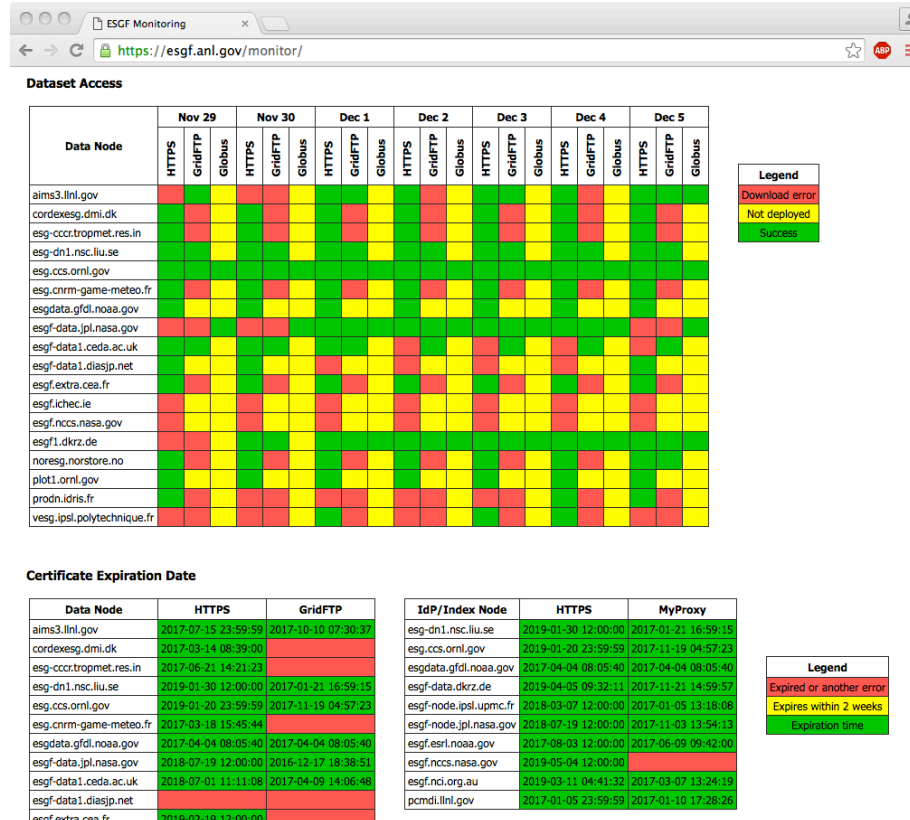
Members: Sasha Ames, Luca Cinquini, Katharina Berger, Georgi Kostov, Hans Ramthun, Ben Evans, Joseph Anthony, Dean N. Williams

- Clean up all Globus Toolkit libraries from custom code and move it to the esgsaml authorization callout library
  - Support for regex in a grid map file was in the standard Globus library globus-gss-assist and was moved to globus-authz-esgsaml-callout to avoid conflicts with packages provided by the Globus YUM repository
- Add support for shared Globus endpoints to the esgsaml authorization callout
  - JPL/NASA migrated to the esgsaml authorization callout
- Implement 'Globus download' in CoG for restricted datasets
- Implement Globus access to datasets
  - Required changes to the publisher (esgcet module, esg.ini) and CoG
  - DKRZ, JPL/NASA, ORNL, PCMDI/LLNL republished and NCI is publishing all their datasets with Globus URLs
- Provide documentation how to setup Globus Connect Server with the esgsaml authorization callout on DTN nodes
  - Deployed on PCMDI DTNs
- Modify 'Globus download' implementation to leverage OAuth 2.0
- Implement write access in the esgsaml authorization callout for publishers

2016 Progress & Accomplishments

- Develop and deploy the monitoring system, http://esgf.anl.gov/monitor.
  - The system checks data access through HTTPS with SSL authentication, GridFTP server, Globus service. It also checks certificate expiration (HTTPS, GridFTP, MyProxy servers)

- Updates needed to support data format transformation (e.g. data in netCDF4 but downloaded as netCDF3)

1. Roll out 'Globus download' with OAuth 2.0 support

2. Modify the installer to support write access in the esgsaml authorization callout

3. Improve the monitoring system
   1. Add new tests (DTN nodes, WGET scripts, etc.)
   2. Email error notifications to the admins

4. Support data format transformation (e.g. netCDF4 to netCDF3)

Community Feedback from Survey

- Problems with moving data from compute resources to the ESGF data nodes

- Problems with downloading datasets

Collaborations Needed

- **All ESGF sites**
  - Fix issues with HTTPS/GridFTP servers
  - Update ESGF software/services
  - Add Globus URLs to already published datasets

- **IdEA, UI Working Teams**
  - Roll out 'Globus download' with OAuth 2.0

- **Publication Working Team**
  - Integrate write access in the esgsaml authz callout with the installer, decide about directories with write access, etc.

- **User interface improvements**
  - Need feedback from users