OAK RIDGE National Laboratory

OLCF's Constellation Project

DOI's at the OLCF (and ORNL in General)

ORNL is managed by UT-Battelle LLC for the US Department of Energy

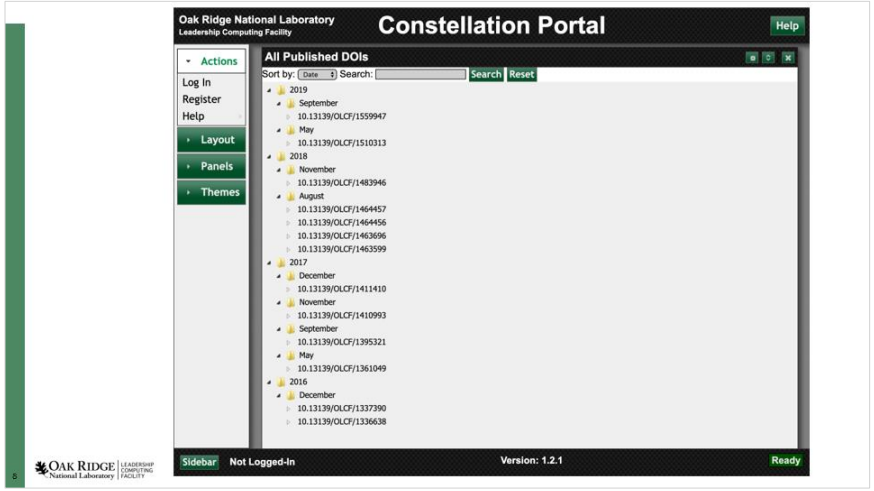
U.S. DEPARTMENT OF ENERGY

Overview

- Started as a means to provide OLCF users with a place to store and publish their simulation results
  - Plans to integrate tightly with OLCF resources (automatically scrape files for metadata, for example)
  - Not implemented yet…
  - Assumption was that we'd be handling large (multi-gigabyte) datasets
- Opened the service up to everyone at the lab Summer 2019

User's Perspective

- Custom web service
  - https://doi.ccs.ornl.gov
- File transfers via Globus Online
  - Globus is designed to handle large files: error checking, automatic restart after interruption, multiple streams, etc…
  - Most OLCF users are familiar with it. (Lab-wide users seem to be less familiar)

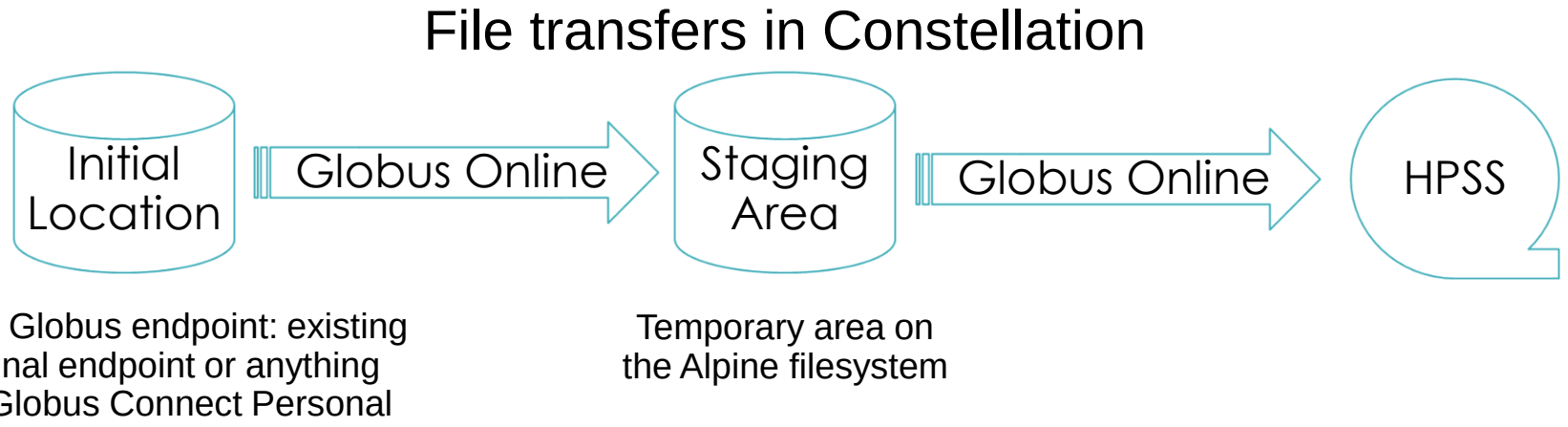


Behind The Scenes – Metadata Handling

- Standard MariaDB database
- Metadata manually entered (by users) via web interface
- Metadata (and data) is manually reviewed by admins
  - This isn't scalable, but we have plans in case the volume of requests increases significantly
- Metadata is sent to OSTI via the E-Link interface
  - OSTI issues the actual identifier string
- The "landing page" for the DOI is generated from the metadata

Behind The Scenes – Bulk Data

- Globus Online for all transfers
- Temporary staging area (OLCF parallel filesystem)
- HPSS for the final, long-term storage
- High speed network connections between staging area and HPSS



Thoughts And Musings On Operations

- The biggest operational issues haven't been technical; they've been people and policy issues
- Where are all the users? Where is all the data?
  - Uptake has been slow. "If you build it, they will come" apparently only works with baseball…
  - Less data than expected: one DOI was ~550GB, but most are < 10GB
- Upper management doesn't like the concept of forever
  - Can't get anyone to commit to resources for "forever". All I can officially get is a commitment for the next FY (sometimes with a verbal commitment to renew for another year).
  - Practically speaking, data in HPSS is forever, and we're such a small percentage that it's not worth trying to account for it

Questions?

Acknowledgement:

This work uses resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.