Using Globus' Shared Endpoints for Data Publication

Mitch Griffith
Constellation

• Constellation is a digital object identifier (DOI) based science network for supercomputing data

• Open Access to Metadata and Data

• Requirement to interface with OSTI
  – OSTI issues DOI and registers with DataCite

• DOIs tie data with the associated resources and processes that went into the production of the data (e.g., jobs, collaborators, projects).
  – Annotate scientific output with rich metadata

• Data is stored in our archival storage system, HPSS
  – HPSS-DSI allows direct HPSS access from Globus
DOI Review Process

• All DOIs must be reviewed
  – Already have a Resource Utilization Council (RUC) that peer reviews projects
  – DOI data review is handled by the RUC

• Required to store data for at least 5 years.
  – Archival storage already has 20+ year old data
  – Protection with RAIT in HPSS

• Data cannot be updated
Development Effort

• Originally not tied with Globus
  – Adding Globus support was quick and easy

• Separate tool for data transfer
  – Maintenance issues
  – Users always find unique ways to break tools
  – Infrastructure would ‘expire’ ports

• Limited number of developers
  – Never enough resources
Globus Shared Endpoints

• Current Implementation
  – Copy data from HPSS to an NFS area.
  – NFS area is the Globus Shared Endpoint

• Desired Implementation (coming soon)
  – Just use an HPSS Shared Endpoint
  – Eliminate a lot of code to avoid data duplication
  – Eliminate cache manager code
    • Let HPSS manage its own cache.
Next Steps

• Testing HPSS-DSI on production machines
  – HPSS-DSI does a great job of throttling requests
  – Currently in the process of updating from HPSS 7.4.3 to 7.5.1
  – HPSS-DSI will be production for all users with HPSS 7.5.1 (about 1 -2 months out)
    – Create a shared endpoint on HPSS

• Additional Metadata

• Data Discovery
This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.