Integrating Globus into the NCAR Research Data Archive User Services

Thomas Cram
Doug Schuster
Steven Worley
NCAR, Boulder, CO
What is the RDA?

• 600+ datasets for climate and weather research
• Free and open access
• Worldwide usage
• Various data access pathways
  – HTTP (wget, cURL, etc.)
  – OPeNDAP, WCS, WMS
  – Web services (CLI, API)
  – Direct analysis on HPC systems (NCAR users)
2014:

- 17+ PB virtual processing
- Web downloads: 7300 users, 750 TB served
- Custom orders: 4000 users, 380 TB served, 45,000 orders
That’s a lot of data. How can we improve download efficiency and reliability?
Globus @ RDA: How we do it

• Single shared endpoint
  – Data copied to subdirectories under endpoint source path
  – ACLs managed programatically via Globus CLI
  – Allow read permission to subdirectories under the shared endpoint
acl-add rda#data_request
   --perm=r
   --path=/download.auto/CRAM80279/
   --email=tcram@ucar.edu
RDA Alternate Identity login
Welcome to the NCAR RDA OAuth for MyProxy Client Authorization Page. The Client below is requesting access to your account. If you approve, please sign in with your RDA email/username and RDA password.

Client Information
Name: Globus Online
URL: https://www.globusonline.org

NCAR RDA Email/Username: tcram@ucar.edu
NCAR RDA Password: ************

Sign In  Cancel
Collecting usage metrics

**Globus Endpoint Manager API**

- **Method** = GET
- **URL** = https://transfer.api.globusonline.org/v0.10/endpoint_manager/task_list
- **Headers** = 
  ```
  {'Authorization': 'Globus-Goauthtoken $GOTOKEN'}
  ```
- **Filters** = 
  ```
  {'filter_endpoint' : 'rda#data_request'}
  ```
What’s next / improvements

• Globus access for general file lists (not just custom orders)
  – Files are hard-linked to enable data shares
  – Globus groups to manage access to restricted datasets

• Migration to “data cart” functionality

• Map Globus user ID to RDA user in API metrics
  – Easier linkage to individual user activity
Upcoming

• GlobusWorld 2015 poster — this evening

• ESIP Summer Meeting
  • 14-17 July 2015, Pacific Grove, CA
  • Globus breakout session: Speakers wanted
  • http://esipfed.org/
Contact us

rda.ucar.edu
tcram@ucar.edu
Collecting usage metrics (cont.)

Metrics stored in RDA DB

- Globus user ID
- Transfer volume
- Timestamps
- Source & destination endpoints
- # files transferred/skipped