Leveraging Globus to support access and delivery of scientific data

Thomas Cram
National Center for Atmospheric Research, Boulder, CO

USGS Community for Data Integration Monthly Meeting
9 September 2015
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)
RDA Web Usage

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

• Globus Connect GridFTP server at NCAR
  – 4 IBM x3650 nodes
  – 16 cores each node
  – 5 Gb/s internal, 2 Gb/s external

• RDA shared endpoints
  – rda#datashare and rda#data_request
  – Data shared from subdirectories under endpoint source path
  – Access managed programatically (Globus CLI)
  – Data shares deleted after 5 days for custom orders
# RDA example user interface

## NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999
ds083.2

For assistance, contact Grace Peng (303-497-1218).

```
<table>
<thead>
<tr>
<th>Description</th>
<th>Data File Downloads</th>
<th>Customizable Data Requests</th>
<th>Other Access Methods</th>
<th>NCAR-Only Access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Web Server Holdings</td>
<td>Globus Transfer Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data File Downloads</td>
<td>Data Format Conversion</td>
<td>Subsetting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Customizable Data Requests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other Access Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NCAR-Only Access</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Union of Available Products

<table>
<thead>
<tr>
<th>PRODUCTS</th>
<th>Description</th>
<th>Data Description</th>
<th>Data File Downloads</th>
<th>Customizable Data Requests</th>
<th>Other Access Methods</th>
<th>NCAR-Only Access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grib1 6 Hourly Files 1999.07.30 to 2007.12.06</td>
<td>Web File Listing</td>
<td>Request Globus Invitation</td>
<td>Get Converted Files</td>
<td>Get a Subset</td>
<td>Central File System (GLADE) Holdings</td>
</tr>
<tr>
<td></td>
<td>Grib2 6 Hourly Files 2007.12.06 to current</td>
<td>Web File Listing</td>
<td></td>
<td>Get Converted Files</td>
<td>Get a Subset</td>
<td>TDS Aggregation</td>
</tr>
</tbody>
</table>
```
RDA example user interface

Globus CLI:

```
acl-add rda#datashare/ds083.2/
    --perm=r
    --email=tcrm@ucar.edu
```
Globus endpoint share invitation

Globus Online Notification <notify@globus.org> to me

You're invited to use "rda#data_request:/download.auto/CRAM99578/"

Globus user "rda" (rda@ucar.edu) has invited you to use the folder "/download.auto/CRAM99578/" on shared endpoint "rda#data_request"

Please click here to accept: https://www.globus.org/accept?type=share&shareId=2c11a38cdee1

Once you accept, you can access this and all following sub-folders from the endpoint, and if the Globus user has granted you appropriate permissions, you can also upload files to this endpoint.

The Globus Team support@globus.org
RDA user dashboard

View, re-send, and delete Globus shares

## Globus Endpoints Shared With You

8 RDA data paths have been shared with you.

<table>
<thead>
<tr>
<th>Endpoint:Path</th>
<th>Globus URL</th>
<th>Manage</th>
</tr>
</thead>
<tbody>
<tr>
<td>rda#datashare:/ds084.3/</td>
<td><a href="https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds084.3%2F">https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds084.3%2F</a></td>
<td></td>
</tr>
<tr>
<td>rda#datashare:/ds132.0/</td>
<td><a href="https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds132.0%2F">https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds132.0%2F</a></td>
<td></td>
</tr>
<tr>
<td>rda#datashare:/ds601.0/</td>
<td><a href="https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds601.0%2F">https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds601.0%2F</a></td>
<td></td>
</tr>
</tbody>
</table>
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: **********
Collecting usage metrics

Globus Management Console REST API

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

Usage (Nov 2014–current)

- 120 users
- 40+ Tb
- 3.8 M files
Upcoming

• Migration to “data cart” functionality
  • new shared endpoint rda#datacart

• Globus HTTP downloads

• Globus Auth service
  – single sign-on with RDA identity
  – no Globus identity requirement
Contact us

rda.ucar.edu
tcram@ucar.edu