Integrating Globus into the NCAR Research Data Archive User Services

Thomas Cram
Doug Schuster
Steven Worley

Zaihua Ji
Bob Dattore

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)
**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 programmatically via Globus CLI
  - Data shares deleted after 5 days for custom orders
RDA example user interface

<p>| NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999 ds083.2 |</p>
<table>
<thead>
<tr>
<th>Description</th>
<th>Data Access</th>
<th>Documentation</th>
<th>Software</th>
</tr>
</thead>
</table>

Mouse over the table headings for detailed descriptions

<table>
<thead>
<tr>
<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>Web Server Holdings</td>
<td>Globus Transfer Service</td>
<td>Subsetting</td>
<td>Central File System (GLADE) Holdings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data Format Conversion</td>
<td></td>
<td>GLADE File Listing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HPSS File Listing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Products</th>
<th>Web File Listing</th>
<th>Get Converted Files</th>
<th>Get a Subset</th>
<th>TDS Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRIB1 6 HOURLY FILES 1999.07.30 to 2007.12.06</td>
<td>Request Globus Invitation</td>
<td></td>
<td></td>
<td>GLADE File Listing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HPSS File Listing</td>
</tr>
<tr>
<td>GRIB2 6 HOURLY FILES 2007.12.06 to current</td>
<td>Web File Listing</td>
<td>Get Converted Files</td>
<td>Get a Subset</td>
<td>TDS Aggregation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GLADE File Listing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HPSS File Listing</td>
</tr>
</tbody>
</table>
RDA example user interface

```
ac1-add rda#datashare/ds083.2/
    --perm=r
    --email=tcram@ucar.edu
```
# 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:/ds083.2/</td>
<td><a href="https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds083.2%2F">https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds083.2%2F</a></td>
<td></td>
</tr>
<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:/ds131.2/</td>
<td><a href="https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds131.2%2F">https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds131.2%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:/ds316.1/</td>
<td><a href="https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds316.1%2F">https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds316.1%2F</a></td>
<td></td>
</tr>
<tr>
<td>rda#datashare:/ds502.0/</td>
<td><a href="https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds502.0%2F">https://www.globus.org/xfer/StartTransfer?origin=rda%23datashare%2Fds502.0%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
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 Management Console REST API

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

Usage (Nov 2014–current)

– 84 users
– 20+ Tb
Challenges

• Globus identity required to link to RDA account
Challenges (cont.)

• Server response timeout for large dir listings – current solution: use CLI or API instead
Upcoming

• Migration to “data cart” functionality
• Globus HTTP downloads
• Globus Auth service
  – single sign-on with RDA identity
  – no Globus identity requirement
Contact us

rda.ucar.edu
tcram@ucar.edu