



Integrating Globus into the NCAR Research Data Archive User Services

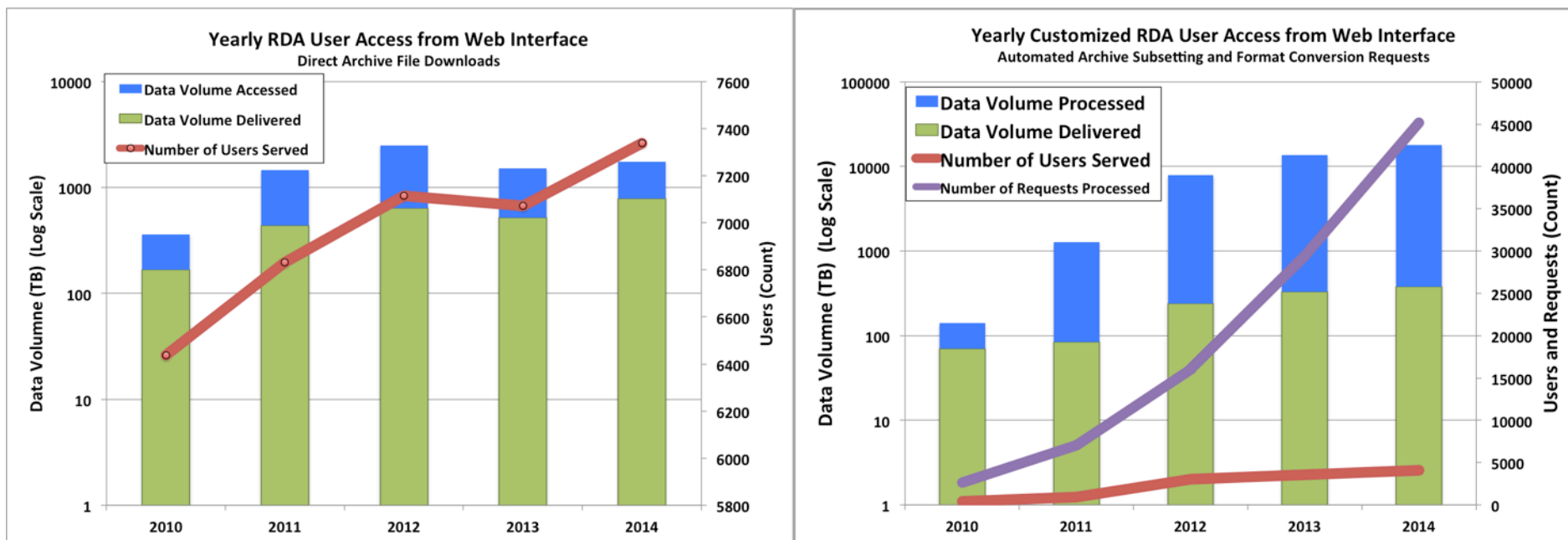
Thomas Cram
Doug Schuster
Steven Worley
NCAR, Boulder, CO



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)

RDA 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: 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

RDA example user interface

Subset Data Requested From DS083.2

Go to [DS083.2 Home Page](#).

The Subset Data requested by Thomas Cram (tcram@ucar.edu) from ds083.2 - 'NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999' are listed below. The data files will remain available online for 5 days (until 2014-12-13 11:59:07). If you need additional time to complete the download, please notify [Grace Peng](#). After you have completed the download, you may delete the request from our web server by clicking the "Purge request" button.

```
acl-add rda#data_request
--perm=r
--path=/download.auto/CRAM80279/
--email=tcram@ucar.edu
```

- Select multiple data files to download as a single tar file
- Currently 0 File selected [Clear Selection](#)

[Scroll to **END** of the Filelist]

<input type="checkbox"/>	INDEX	File Name	Size	Data Format	Date Online
<input type="checkbox"/>	1	fnl_20100901_12_00.grib2.spasub.cram99578	169.11K	GRIB2	12/08/2014
<input type="checkbox"/>	2	fnl_20100909_00_00.grib2.spasub.cram99578	154.99K	GRIB2	12/08/2014
<input type="checkbox"/>	3	fnl_20100916_12_00.grib2.spasub.cram99578	154.39K	GRIB2	12/08/2014
<input type="checkbox"/>	4	fnl_20100924_00_00.grib2.spasub.cram99578	158.90K	GRIB2	12/08/2014
<input type="checkbox"/>	5	fnl_20100901_18_00.grib2.spasub.cram99578	168.21K	GRIB2	12/08/2014
<input type="checkbox"/>	6	fnl_20100909_06_00.grib2.spasub.cram99578	155.75K	GRIB2	12/08/2014
<input type="checkbox"/>	7	fnl_20100916_18_00.grib2.spasub.cram99578	151.58K	GRIB2	12/08/2014

RDA Alternate Identity login

The screenshot shows the Globus Sign In interface. At the top left is the Globus logo. At the top right are 'Log In' and 'Sign Up' buttons. The main heading is 'Sign In', with a link for 'Sign Up with Globus'. Below the heading, there are two options: 'Using your Globus login.' and 'alternate login'. The 'alternate login' option is selected, and a modal window titled 'Select Identity Provider' is open. This modal lists various identity providers in two columns. The 'NCAR RDA' provider is highlighted with a red rectangular box. Other providers listed include Globus, Argonne LCF, Argonne MCS & LCRC, BIRN, CLI Transition, EGI, ESG ANL, Exeter, Google, InCommon / CILogon, IRZ, NCSA, NCSA Blue Waters, NERSC, Tuakiri, UChicago CI, UChicago iBi, UK NGS, WestGrid, and XSEDE. A link 'Forgot password?' is visible to the right of the modal.

globus

Log In Sign Up

Sign In Sign Up with Globus

Using your Globus login. alternate login

Select Identity Provider

Globus	IRZ
Argonne LCF	NCAR RDA
Argonne MCS & LCRC	NCSA
BIRN	NCSA Blue Waters
CLI Transition	NERSC
EGI	Tuakiri
ESG ANL	UChicago CI
Exeter	UChicago iBi
Google	UK NGS
InCommon / CILogon	WestGrid
	XSEDE

Forgot password?

RDA Alternate Identity login



NCAR Research Data Archive (RDA) MyProxy Client Authorization

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

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

RDA



Collecting usage metrics (*cont.*)

Metrics stored in RDA DB

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