

Using Globus to improve access and delivery of scientific data

Thomas Cram

National Center for Atmospheric Research
Computational & Information Systems Laboratory
Boulder, CO

GlobusWorld 2016

April 22, 2016

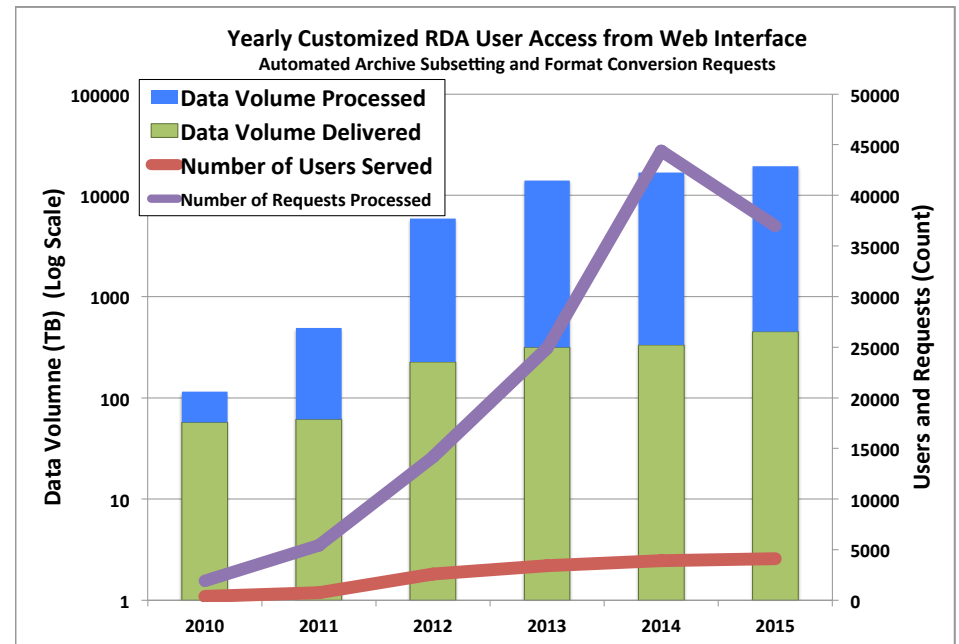
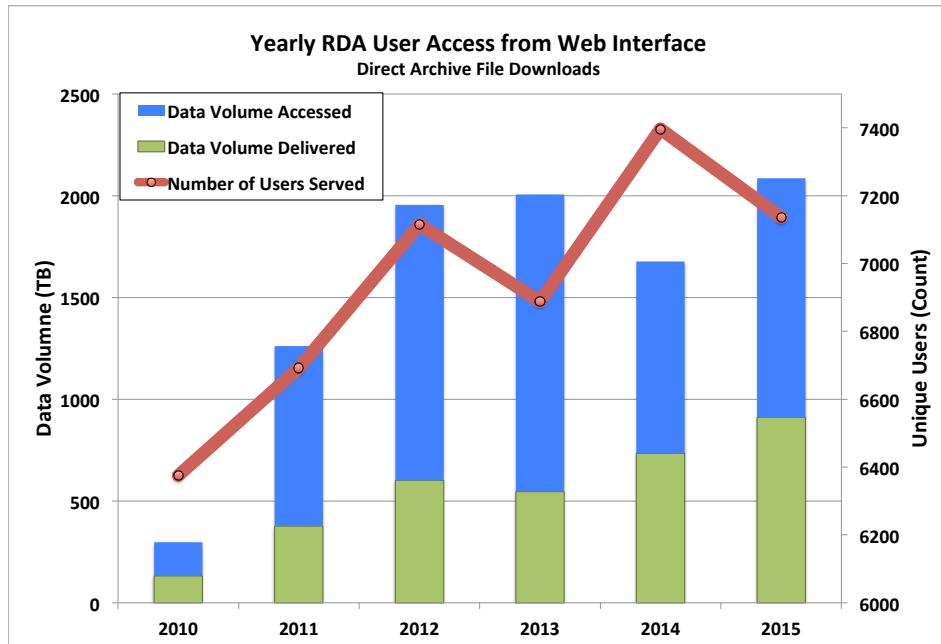
About the RDA



- <http://rda.ucar.edu>
- **Purposes**
 - Support climate & weather research at NCAR and UCAR universities with reference datasets
- **Collections**
 - Ocean & atmospheric observations, analyses, reanalyses, operational NWP products
 - 700+ datasets, 8M files, 1.8 PB
- **Free and open access**
- **Worldwide usage**
- **Various data access pathways**
 - HTTPS (wget, cURL, etc.)
 - OPeNDAP, WCS, WMS
 - Web services (API, CLI)
 - Direct analysis on HPC systems (NCAR users)



RDA Web Usage



FY 2015:

- 17+ PB virtual processing
- 11,000+ unique users
- 1.6 PB served

RDA Globus infrastructure

- Shared endpoints
 - **NCAR RDA dataset archive**
(rda#datashare): permanent dataset access
 - **NCAR RDA data requests**
(rda#data_request): custom data products
- Data access managed programmatically
 - Globus REST API and CLI
 - **acl-add** & **acl-remove**
 - Initiated by user

One ACL, One Dataset

The screenshot displays the Globus web interface. At the top, the Globus logo is on the left, and navigation links for 'Manage Data', 'Publish', 'Groups', 'Support', and 'Account' are on the right. Below this, a secondary navigation bar includes 'Transfer Files', 'Activity', 'Endpoints', 'Bookmarks', and 'Console'. The main content area is titled 'Transfer Files' and features a 'RECENT ACTIVITY' section with three circular indicators. Two endpoint selection panels are visible. The left panel is active, showing the endpoint 'NCAR RDA dataset archive' and a path of '/'. Below this, a list of folders is displayed, each with a folder icon and the label 'Folder'. The right panel is currently empty, displaying the instruction 'Start by selecting an endpoint.'

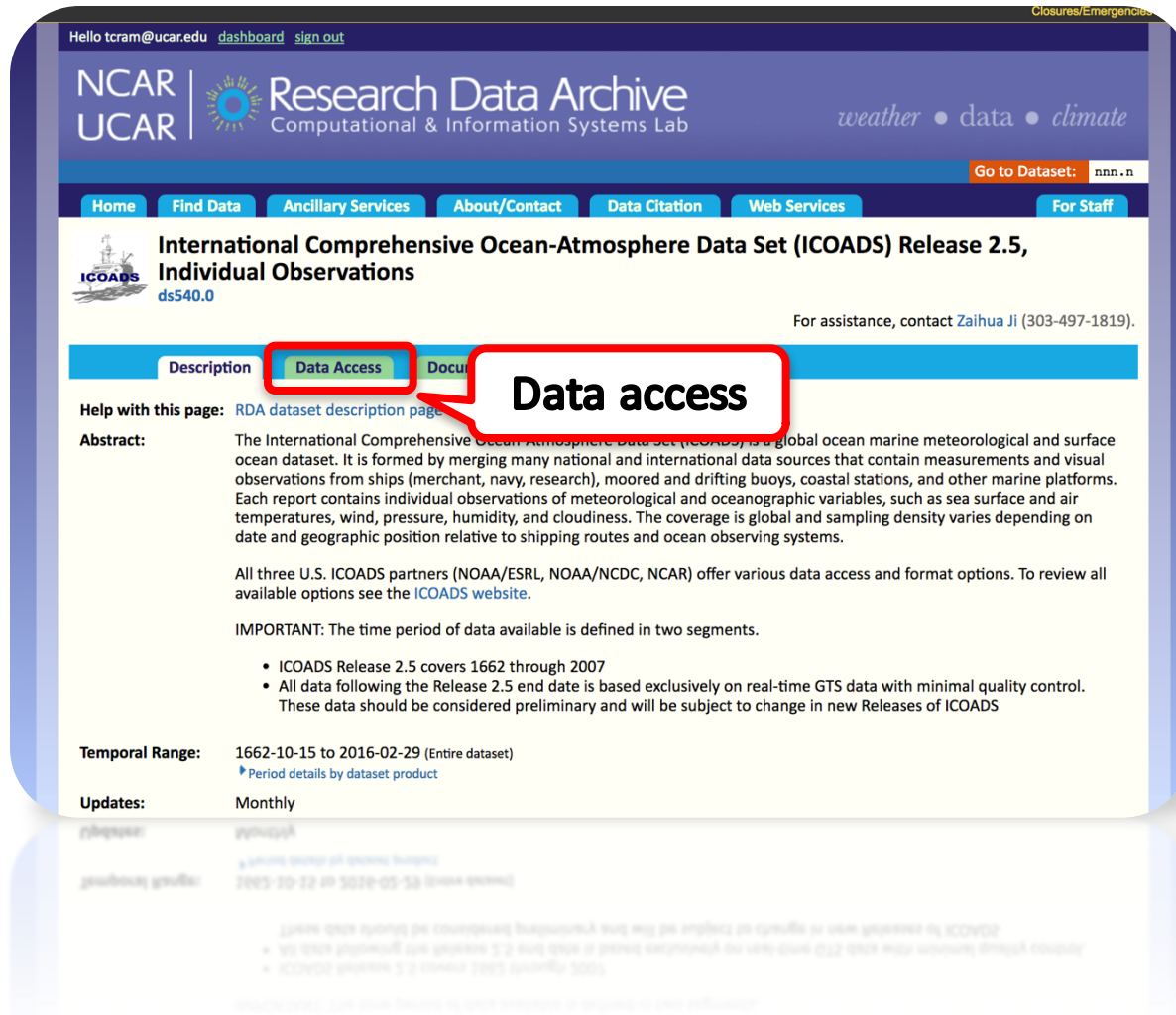
Endpoint	Path	Go
NCAR RDA dataset archive	/	Go

Endpoint	Path	Go
		Go

Folder	Type
ds091.0	Folder
ds093.2	Folder
ds094.0	Folder
ds131.2	Folder
ds132.0	Folder
ds132.1	Folder
ds316.1	Folder
ds337.0	Folder
ds497.0	Folder
ds502.0	Folder
ds502.1	Folder
ds540.0	Folder
ds552.0	Folder
ds728.2	Folder
ds728.3	Folder

Start by selecting an endpoint.

Example user interface



Example user interface

International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Release 2.5, Individual Observations
ds540.0

For assistance, contact Zaihua Ji (303-497-1819).

Description | **Data Access** | Documentation | Software

Mouse over the table headings for detailed descriptions

	Data File Downloads			Customizable Data Requests	NCAR-Only Access	
	Web File Holdings	Globus Transfer Service (GridFTP)	Staged Access from Tape Archive	Subsetting	Central File System (GLADE) Holdings	Tape Archive (HPSS) Holdings
Union of Available Products	Web File Listing	Request Globus Transfer	Request Access		GLADE File Listing	HPSS File Listing
IMMA format data for ICOADS Release 2.5	Web File Listing			Get a Subset	GLADE File Listing	HPSS File Listing
IMMA format data for ICOADS Release 2.5 Intermediate			Request Access			HPSS File Listing
ICOADS Release 2.5.1 in experimental-2 IMMA1 format	Web File Listing				GLADE File Listing	HPSS File Listing
ICOADS Release 2.5.1 Intermediate in experimental-2 IMMA1 format			Request Access			HPSS File Listing
ICOADS Release 2.5.2 for 2008-2014/8 in experimental IMMA1 format	Web File Listing				GLADE File Listing	HPSS File Listing
ICOADS Release 2.5.2 for 2008-2014/8 Intermediate in experimental IMMA1 format			Request Access			HPSS File Listing
ICOADS Release 3.0 Preliminary Beta-3 Release	Web File Listing				GLADE File Listing	HPSS File Listing
ICOADS Beta Release 3.0 Clean Version in IMMA1 format	Web File Listing				GLADE File Listing	HPSS File Listing

Example user interface

You may now transfer your data using Globus at the URL

https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds540.0%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7.

From the Globus website, please select 'NCAR RDA' from the list of organizational logins and then enter your RDA e-mail address (**tcram@ucar.edu**) and password to log in.

Close This Window

Example user interface

The screenshot shows the Globus user interface for file transfers. At the top, there is a navigation bar with the Globus logo and links for Manage Data, Publish, Groups, Support, and Account. Below this, there are links for Transfer Files, Activity, Endpoints, Bookmarks, and Console. The main content area is titled "Transfer Files" and features a "RECENT ACTIVITY" section with three circular indicators. The interface is split into two main panels. The left panel shows a file browser for the endpoint "NCAR RDA dataset archive" at the path "/ds540.0/". It lists several folders (data_imma, imma1_r2.5, imma1_r2.5.1, imma1_r2.5.2, imma1_r3.0_beta, imma1_r3.0_beta_clean, imma_r2.5, imma_r2.5i) and a file named "index.html". The right panel is currently empty, displaying the message "Start by selecting an endpoint." Below the file browser, there is a "Label This Transfer" field and a "Transfer Settings" section with several checkboxes: "sync - only transfer new or changed files", "delete files on destination that do not exist on source", "preserve source file modification times", "verify file integrity after transfer" (checked), and "encrypt transfer". At the bottom right, there is a link to "Get Globus Connect Personal" with the text "Turn your computer into an endpoint."

User dashboard

Dashboard

User Profile

Thomas Cram
National Center for Atmospheric Research

[Edit/Change Profile](#)

Customized Data Requests

You don't have any customized data requests.

Globus Endpoints Shared With You

15 RDA data paths have been shared with you.

[Hide Shares](#)

Endpoint: Path	Globus URL	Manage
rda#datashare: /ds091.0/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds091.0%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	
rda#datashare: /ds093.2/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds093.2%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	
rda#datashare: /ds094.0/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds094.0%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	
rda#datashare: /ds131.2/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds131.2%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	
rda#datashare: /ds132.0/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds132.0%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	
rda#datashare: /ds132.1/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds132.1%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	
rda#datashare: /ds316.1/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds316.1%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	
rda#datashare: /ds337.0/	https://www.globus.org/app/transfer?origin_id=db57de42-6d04-11e5-ba46-22000b92c6ec&origin_path=%2Fds337.0%2F&add_identity=91dc4454-d2bc-11e5-a34a-a3f71caddfd7	

Globus Management Console

- Monitor activity
- Retrieve transfer metrics via REST API

Management Console

endpoint overview active tasks pause rules ncar#datashare

Current Activity for all your endpoints

ncar#datashare

ACTIVE TASKS	OK	RETURNED	WITH FAULTS
8	2	4	8

VIEW TASKS: ACTIVE | WITH FAULTS | COMPLETED | ALL

Overview

Visible To Public - Visible to all users

Description NCAR Globus Data Sharing Endpoint

Identity Provider MyProxy

Hostname myproxy.ucar.edu:7512

Servers

- URI gftp://griftp0.ucar.edu:2811
- URI gftp://griftp02.ucar.edu:2811
- URI gftp://griftp01.ucar.edu:2811
- URI gftp://griftp04.ucar.edu:2811

Management Console

Tasks matching: ncar#datashare Completed Tasks

0 pause rules are in effect for ncar#datashare.

NOTE: Pause rules can affect both new (incoming) and current tasks that meet their criteria. Removing a pause rule will resume currently paused tasks if other rules do not affect those tasks.

select action Tasks: active paused completed with faults custom filter

STATUS	DESCRIPTION	INITIATED BY	AGE
✓	from rda#datashare (ncar#datashare) to <private endpoint>	henno	11 hours ago
✗	delete from 9a38-22000b96db58 (ncar#datashare)	henno	11 hours ago
✗	from rda#datashare (ncar#datashare) to 9a38-22000b96db58 (ncar#datashare)	henno	12 hours ago
✓	from rda#data_request (ncar#datashare) to <private endpoint>	natieepulvedamora	18 hours ago
✓	from rda#data_request (ncar#datashare) to <private endpoint>	natieepulvedamora	18 hours ago
✓	from rda#datashare (ncar#datashare) to <private endpoint>	kenstin	a day ago
✓	from rda#datashare (ncar#datashare) to <private endpoint>	kenstin	2 days ago

Advanced endpoint API

Useful resources

Base URL: <https://transfer.api.globusonline.org/v0.10/>

- **/endpoint_manager/task_list**
 - Listing of transfer tasks from a single endpoint
- **/endpoint_manager/task/<task_id>**
 - Details of a specific transfer task
- **/endpoint_manager/task/<task_id>/successful_transfers**
 - Complete listing of files successfully transferred
 - 30 day history
- **/endpoint/<endpoint_xid>/access_list**
 - Listing of ACL rules

Wishlist/Upcoming/Known issues

- HTTPS access
- HPSS gateway
- Truncated directory file lists on web interface
 - Globbing
- Sharing sublist of files within a directory
 - Browse Endpoint helper page

Example dataset file list



GPCP Version 2.2 Combined Precipitation Data Set ds728.2

For assistance, contact [Thomas Cram](#) (303-497-1217).

Description

Data Access

Documentation

Software

[Web server holdings]

GPCP-netcdf - GPCP Version 2.2 Combined Monthly Precipitation and Precipitation Error Estimates (NetCDF)

[View Selected Files/Get As a Tar File](#)

[Perl Download Script](#)

[Csh Download Script](#)



- Total **37 Files (230.4M)** are listed below
- Click a file name to download a single file
- Currently **0 File** selected [Clear Selection](#)

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

<input type="checkbox"/>		INDEX	File Name	Size	Data Format	Date Archived
<input type="checkbox"/>		1	gpcp_v2.2.1979.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		2	gpcp_v2.2.1980.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		3	gpcp_v2.2.1981.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		4	gpcp_v2.2.1982.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		5	gpcp_v2.2.1983.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		6	gpcp_v2.2.1984.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		7	gpcp_v2.2.1985.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		8	gpcp_v2.2.1986.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		9	gpcp_v2.2.1987.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		10	gpcp_v2.2.1988.nc	6.2M	CFNETCDF	12/02/2015
<input type="checkbox"/>		11	gpcp_v2.2.1989.nc	6.2M	CFNETCDF	12/02/2015

Contact us



web: rda.ucar.edu



@NCAR_RDA



NCAR CISL Research Data Archive



ncarrda.blogspot.com