

Globus integration in the NCAR RDA data portal: Recent enhancements

Thomas Cram

NCAR, Computational Information Systems
Laboratory (CISL)

GlobusWorld 2018

April 25, 2018



About the RDA

- Purpose
 - Support climate & weather research at NCAR and UCAR universities with reference datasets
- Data collections
 - Ocean & atmospheric observations, analyses, reanalyses, operational NWP products
 - 700+ datasets, 8M files, 1.8 PB
- Free and open access
- Worldwide usage
- Science educated staff

A screenshot of the Research Data Archive website homepage. The page features a navigation bar with links for Home, Find Data, Ancillary Services, About/Contact, Data Citation, Web Services, and For Staff. The main content area includes a search bar, a "Look For Data" table, and several informational sections like "Recently Added Datasets" and "Other Ways to Explore".

All Datasets	Variable/Parameter	Type of Data
Time Resolution	Platform	Spatial Resolution
Topic/Subtopic	Project/Experiment	Supports Project
Data Format	Instrument	Location
Recently Added/Updated		

Supported data access pathways

- HTTPS (wget, cURL, REST API, etc.)
- OPeNDAP
- **GridFTP (Globus)**
- HPSS staged-to-disk
- Direct access on GLADE filesystem (NCAR users)
- Subsetting (spatial, parameter, temporal, etc.)
- Data format conversion

NOAA/CIRES Twentieth Century Global Reanalysis Version 2c
ds131.2 ☆

For assistance, contact Doug Schuster (303-497-1216).

Description Data Access Documentation Software

Mouse over the table headings for detailed descriptions

Data Description	Data File Downloads				Customizable Data Requests	Other Access Methods	NCAR-Only Access	
	Web Server Holdings	Globus Transfer Service (GridFTP)	Data Format Conversion	Staged Access from Tape Archive	Subsetting	THREDDS Data Server	Central File System (GLADE) Holdings	Tape Archive (HPSS) Holdings
Union of Available Products	Web File Listing	Request Globus Transfer	Get Converted Files	Request Access		TDS Access	GLADE File Listing	HPSS File Listing
P R O D U C T S	Yearly Time Series Analysis Fields (2x2 degree Lat/Lon)	Web File Listing	Get Converted Files		Get a Subset	TDS Access	GLADE File Listing	HPSS File Listing
	Yearly Time Series First Guess Forecast Fields (Gaussian T-62)	Web File Listing			Get a Subset	TDS Access	GLADE File Listing	HPSS File Listing
	Hxens files			Request Access				HPSS File Listing
	Prepbufrbs_assim files			Request Access				HPSS File Listing
	High priority data tarfiles			Request Access				HPSS File Listing

How does Globus simplify data management and improve our user services?



Use cases:

1. Curated file lists
2. Automated transfers of delayed mode products

* Facilitated by:

- ***Globus Auth***
- ***Browse Endpoint helper***
- ***Python SDK***

Use case #1: Curated file lists

* BEFORE GLOBUS Integration *

- Data discovery on RDA web app
- **Redirect to Globus UI**
- **Blind search for files**

The screenshot displays the Globus Transfer Files interface. At the top, there are navigation links for 'Manage Data', 'Publish', 'Groups', 'Support', and 'Account'. Below this, the 'Transfer Files' section is active, showing two endpoints: 'NCAR RDA dataset archive' and 'cisi-goliad'. The 'NCAR RDA dataset archive' endpoint shows a list of files with columns for file names and sizes. The 'cisi-goliad' endpoint shows a folder named '2017'. Below the file lists, there are options to 'Label This Transfer' and 'Transfer Settings'.

File Name	Size
gfs.Dp25.2017041000.f000.grib2	206.74 MB
gfs.Dp25.2017041000.f003.grib2	228.52 MB
gfs.Dp25.2017041000.f006.grib2	234.43 MB
gfs.Dp25.2017041000.f009.grib2	235.36 MB
gfs.Dp25.2017041000.f012.grib2	239.10 MB
gfs.Dp25.2017041000.f015.grib2	238.23 MB
gfs.Dp25.2017041000.f018.grib2	240.47 MB
gfs.Dp25.2017041000.f021.grib2	239.83 MB
gfs.Dp25.2017041000.f024.grib2	241.53 MB
gfs.Dp25.2017041000.f027.grib2	239.95 MB
gfs.Dp25.2017041000.f030.grib2	241.81 MB
gfs.Dp25.2017041000.f033.grib2	240.20 MB
gfs.Dp25.2017041000.f036.grib2	243.09 MB
gfs.Dp25.2017041000.f039.grib2	241.67 MB
gfs.Dp25.2017041000.f042.grib2	243.96 MB
gfs.Dp25.2017041000.f045.grib2	241.01 MB
gfs.Dp25.2017041000.f048.grib2	241.19 MB
gfs.Dp25.2017041000.f051.grib2	241.19 MB
gfs.Dp25.2017041000.f054.grib2	243.43 MB
gfs.Dp25.2017041000.f057.grib2	241.63 MB
gfs.Dp25.2017041000.f060.grib2	241.97 MB

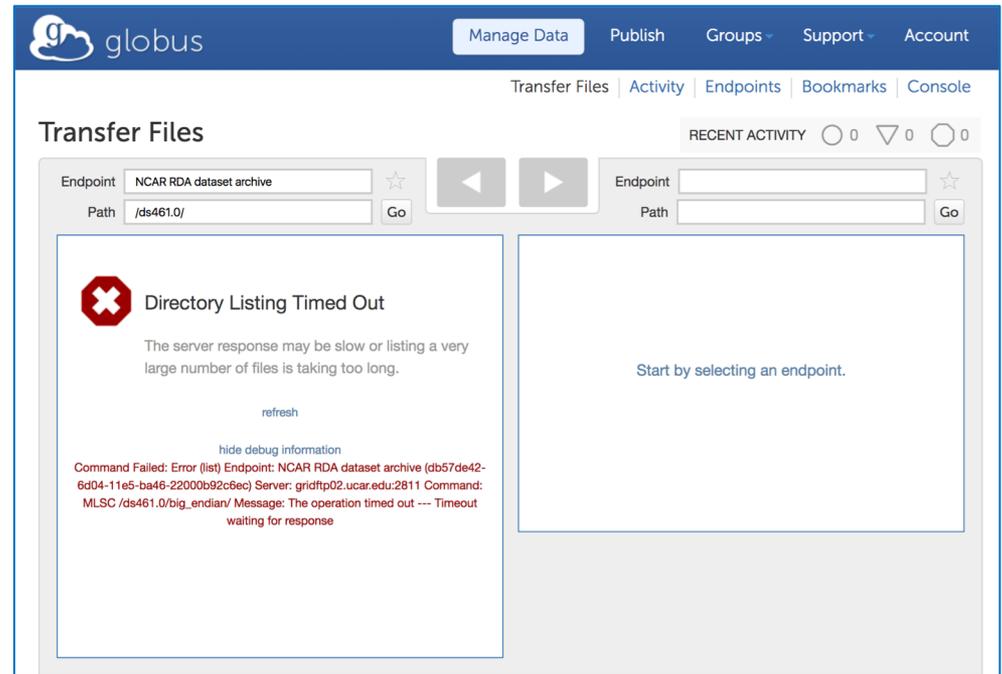
Transfer Settings

- 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
- encrypt transfer

Use case #1: Curated file lists

* BEFORE GLOBUS Integration *

- Data discovery on RDA web app
- Redirect to Globus UI
- Blind search for files
- **Request timeouts for long dir listings**



The screenshot shows the Globus web interface for 'Transfer Files'. The top navigation bar includes 'Manage Data', 'Publish', 'Groups', 'Support', and 'Account'. Below the navigation, 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 empty circular icons. Two endpoint selection boxes are visible, each with a 'Go' button. The left endpoint is set to 'NCAR RDA dataset archive' and the path is '/ds461.0/'. A red error message is displayed in the center: 'Directory Listing Timed Out'. The message states: 'The server response may be slow or listing a very large number of files is taking too long.' Below the message are buttons for 'refresh' and 'hide debug information'. The debug information shows a command failure: 'Command Failed: Error (list) Endpoint: NCAR RDA dataset archive (db57de42-6d04-11e5-ba46-22000b92c6ec) Server: gridftp02.ucar.edu:2811 Command: MLSC /ds461.0/big_endian/ Message: The operation timed out --- Timeout waiting for response'. The right endpoint selection box is empty, with the instruction 'Start by selecting an endpoint.' below it.

Use case #1: Curated file lists

* AFTER GLOBUS Integration *

- Discovery, file list curation (RDA portal)

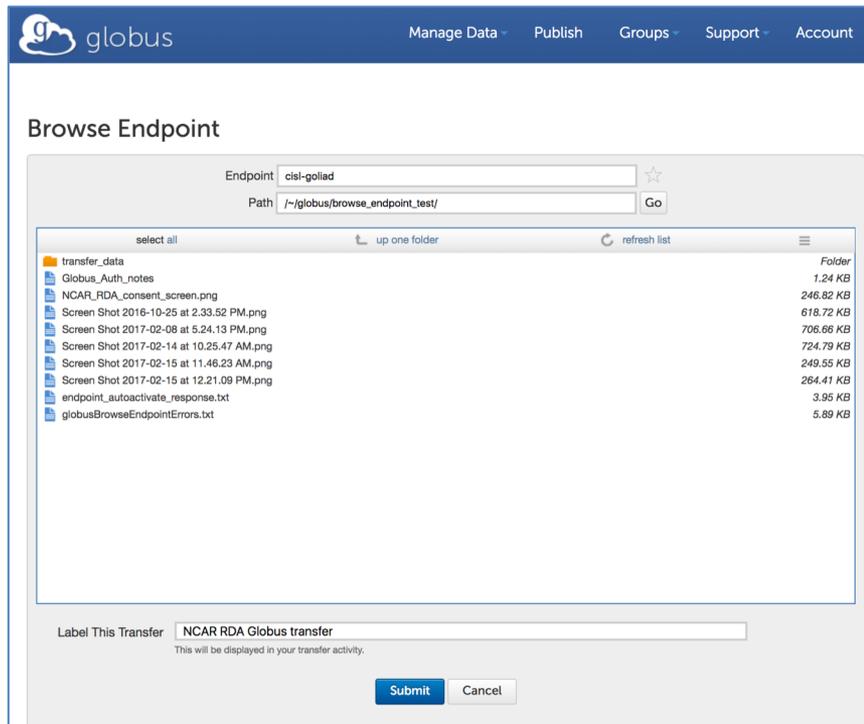
The screenshot displays the RDA portal interface for the dataset 'TS_AN-TMP_pres — Temperature at Pressure Levels'. At the top, there are buttons for 'View Selected Files/Get As a Tar File', 'Perl Download Script', and 'Csh Download Script'. Below these is a 'Globus download' button and a link 'What is Globus?'. The main content area shows 'Files selected for RDA dataset ds131.2' and states 'You have selected 8 files (5.46G)'. It provides instructions on how to use the Globus data transfer service, including a note that a Globus user account is not required and that users can sign in with their RDA credentials. A 'Globus transfer' button is highlighted with a red box. Below the instructions, a table lists the selected files with columns for Index, File name, and Size.

Index	File name	Size
1	pgrbanl/pgrbanl_mean_1851_TMP_pres.grib	680.44M
2	pgrbanl/pgrbanl_mean_1853_TMP_pres.grib	681.53M
3	pgrbanl/pgrbanl_mean_1854_TMP_pres.grib	683.09M
4	pgrbanl/pgrbanl_mean_1855_TMP_pres.grib	681.91M
5	pgrbanl/pgrbanl_mean_1857_TMP_pres.grib	681.6M
6	pgrbanl/pgrbanl_mean_1860_TMP_pres.grib	684.2M
7	pgrbanl/pgrbanl_mean_1862_TMP_pres.grib	681.6M
8	pgrbanl/pgrbanl_mean_1864_TMP_pres.grib	682.86M

Use case #1: Curated file lists

* AFTER GLOBUS Integration *

- Discovery, file list curation (RDA portal)
- **Select destination endpoint (Browse Endpoint helper API)**



The screenshot displays the Globus web interface for browsing an endpoint. The top navigation bar includes the Globus logo and links for 'Manage Data', 'Publish', 'Groups', 'Support', and 'Account'. The main heading is 'Browse Endpoint'. Below this, there are input fields for 'Endpoint' (set to 'cisl-gollad') and 'Path' (set to '/~/globus/browse_endpoint_test/'). A 'Go' button is next to the path field. The central area shows a file browser view with a list of files and folders. At the bottom, there is a 'Label This Transfer' field containing 'NCAR RDA Globus transfer' and 'Submit' and 'Cancel' buttons.

File/Folder	Size
transfer_data	Folder
Globus_Auth_notes	1.24 KB
NCAR_RDA_consent_screen.png	246.82 KB
Screen Shot 2016-10-25 at 2:33:52 PM.png	618.72 KB
Screen Shot 2017-02-08 at 5:24:13 PM.png	706.66 KB
Screen Shot 2017-02-14 at 10:25:47 AM.png	724.79 KB
Screen Shot 2017-02-15 at 11:46:23 AM.png	249.55 KB
Screen Shot 2017-02-15 at 12:21:09 PM.png	264.41 KB
endpoint_autoactivate_response.txt	3.95 KB
globusBrowseEndpointErrors.txt	5.89 KB

Use case #1: Curated file lists

* AFTER GLOBUS Integration *

- Discovery, file list curation (RDA portal)
- Select destination endpoint (Browse Endpoint helper API)
- **Receive Auth code, exchange for token**
- **Submit transfer on user's behalf**

```
code = form.getvalue('code')
tokens = client.oauth2_exchange_code_for_tokens(code)
```

```
...
...
tas
```

Transfer request submitted successfully. Task ID: c3a6334e-260a-11e7-bc68-22000b9a448b

Globus transfer details

Task ID: c3a6334e-260a-11e7-bc68-22000b9a448b

Source endpoint: NCAR RDA dataset archive

Destination Endpoint: cisl-goliad

Request Time: 2017-04-20 20:48:54+00:00

Status: ACTIVE

Files transferred: 0

Faults: 0

Overview and event log: <https://www.globus.org/app/activity/c3a6334e-260a-11e7-bc68-22000b9a448b>

Refresh

Example 2: Delayed mode request transfers

BEFORE GLOBUS Integration:

- User submits request
- Server side processing (delayed mode)
- E-mail download link to user
- Request Globus share on download page

Completed: RDA Subset Data request 292418 from ds502.1 Inbox x

Subset Data Requested From DS502.1

Go to [DS502.1 Home Page](#).

The Subset Data requested by Thomas Cram (tcram@ucar.edu) from RDA dataset ds502.1 - 'NOAA CPC Morphing Technique (CMORPH) Global Precipitation Analyses Version 0.x' are listed below. The data files will remain available online for 5 days (until 2018-04-13 16:35:16). If you need additional time to complete the download, please notify [Thomas Cram](#). After you have completed the download, you may delete the request from our web server by clicking the "Purge request" button.

[Purge request](#)

- Start date: 2017-02-01 00:00
- End date: 2017-02-14 21:00
- Parameter(s):
 CMORPH precipitation estimate
- Spatial subsetting (bounding box):
 Latitudes (top/bottom): 45 / 25
 Longitudes (left/right): -118 / -73

Download options

[Show selected files \(Download tar file\)](#) [Perf download script](#) [Csh download script](#)

[Globus download](#) [What is Globus?](#)

- A total of **14 Files (6.50M)** are listed below
- Click a file name to download a single file
- Select one or multiple files to get a download script
- Select multiple data files to download as a single tar file
- Currently **0 File** selected [Clear Selection](#)

<input type="checkbox"/>	INDEX	File Name	Size	Data Format	Date Online
<input type="checkbox"/>	1	289687.cmorph_precip.CMORPH_V0.x_0.25deg-3HLY_20170201.nc	464.36K	NetCDF	04/08/2018
<input type="checkbox"/>	2	289687.cmorph_precip.CMORPH_V0.x_0.25deg-3HLY_20170202.nc	464.36K	NetCDF	04/08/2018
<input type="checkbox"/>	3	289687.cmorph_precip.CMORPH_V0.x_0.25deg-3HLY_20170203.nc	464.36K	NetCDF	04/08/2018
<input type="checkbox"/>	4	289687.cmorph_precip.CMORPH_V0.x_0.25deg-3HLY_20170204.nc	464.36K	NetCDF	04/08/2018
<input type="checkbox"/>	5	289687.cmorph_precip.CMORPH_V0.x_0.25deg-3HLY_20170205.nc	464.36K	NetCDF	04/08/2018

Example 2: Delayed mode request transfers

With GLOBUS integration:

- User requests Globus transfer during form submission
- User selects destination endpoint
- Receive Auth code, exchange for token, store in session
- Finish processing, e-mail notification sent to user
 - Data already on its way

Completed: RDA Subset Data request 289687 from ds502.1

Inbox x

 **tcram@ucar.edu** 4:36 PM (7 minutes ago) ☆

to me ▾

Processing of the Subset Data from RDA dataset ds502.1 - 'NOAA CPC Morphing Technique (CMORPH) Global Precipitation Analyses Version 0.x' that you requested has completed successfully. As requested, we have submitted a Globus transfer of your data to the destination endpoint specified in your data request. You may monitor the status of your Globus transfer at the following URL:

<https://www.globus.org/app/activity/1e035e4c-3b7d-11e8-b9a8-0ac6873fc732>

You will receive an e-mail notification from Globus when your transfer has completed. If any problem occurs with the transfer, you may request a new Globus transfer or download the data directly at the following link:

<https://rda.ucar.edu/#dsrqst/CRAM289687/>

Request Detail:

- Start date: 2017-02-01 00:00
- End date: 2017-02-14 21:00
- Parameter(s):
 - CMORPH precipitation estimate
- Spatial subsetting (bounding box):
 - Latitudes (top/bottom): 45 / 25
 - Longitudes (left/right): -118 / -73

Looking ahead

- HTTPS access
- Big data challenges
 - Campaign storage
 - Cloud-hosted environments
 - Support required for a large community of NCAR users



Contact us

 rda.ucar.edu
 rdahelp@ucar.edu



 [@NCAR_RDA](https://twitter.com/NCAR_RDA)

 [@NCAR.RDA](https://www.facebook.com/NCAR.RDA)

 ncarrda.blogspot.com