Using Globus to improve access and delivery of scientific data

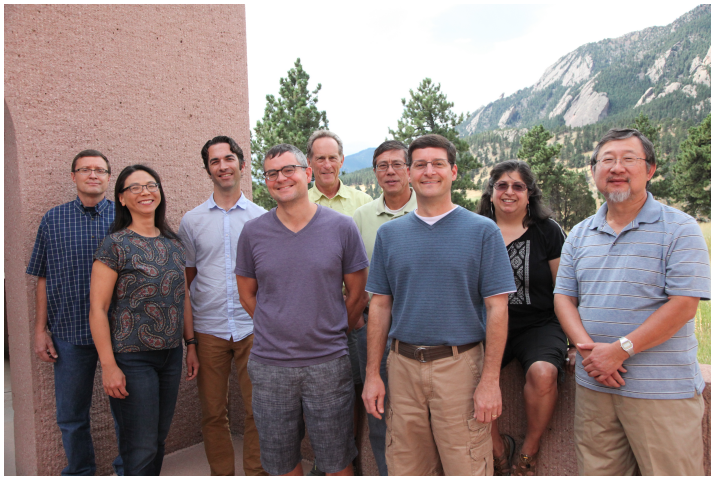
Thomas Cram

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

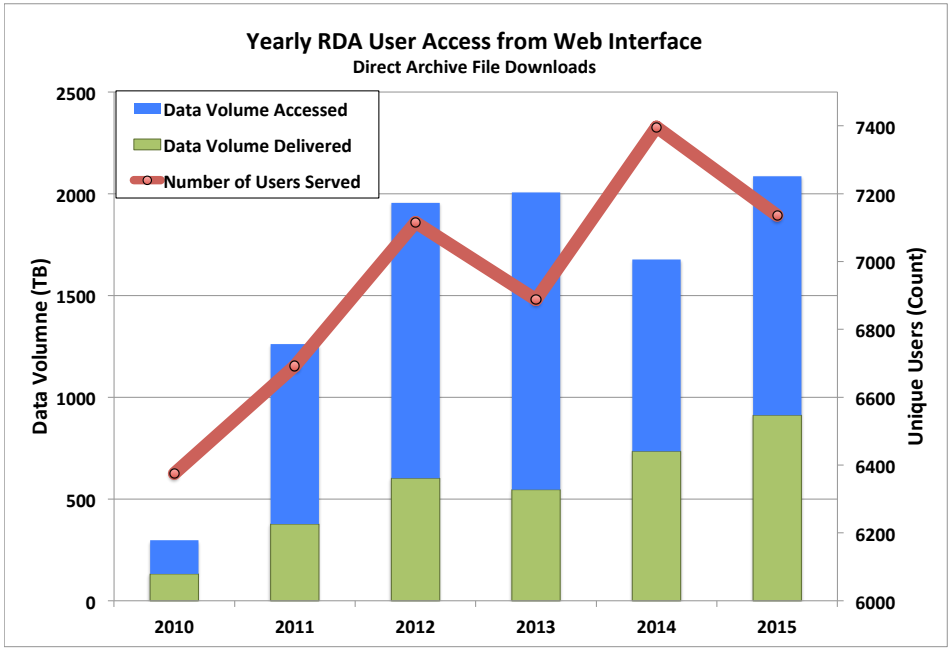
GlobusWorld 2016
April 22, 2016

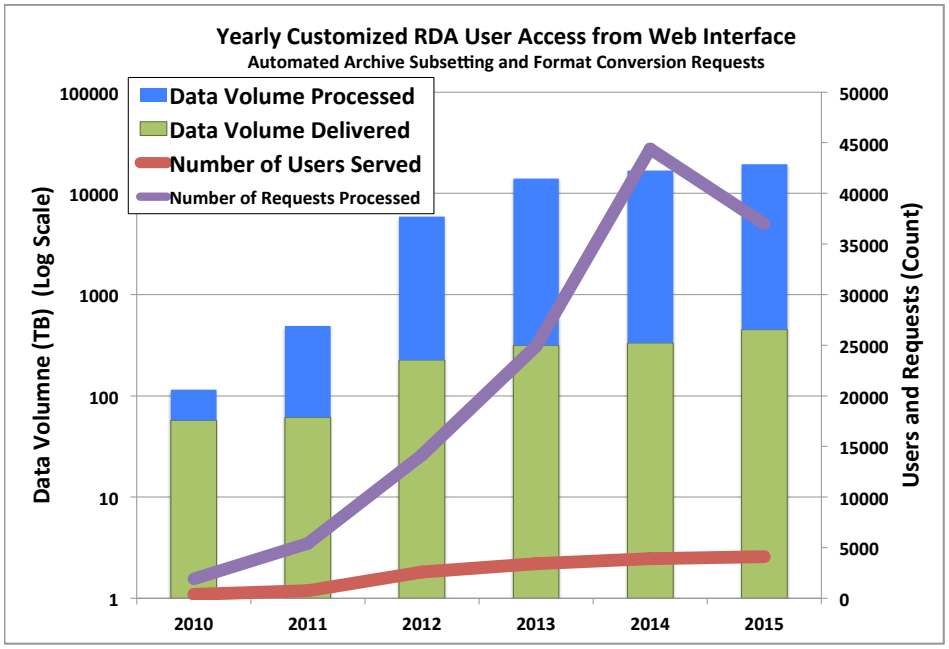
About the RDA

- ***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:

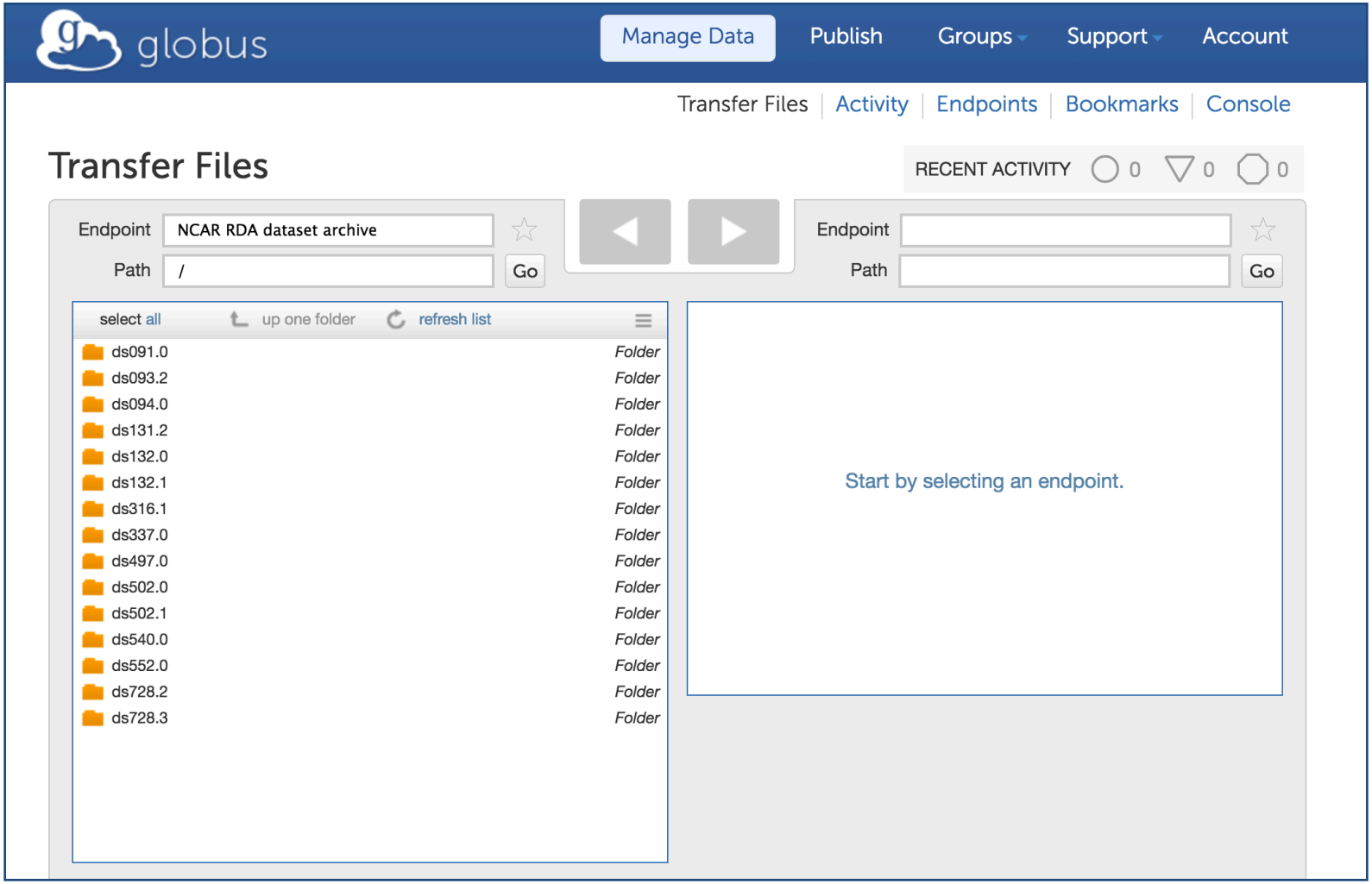
- 17+ PB virtual processing
- 11,000+ unique users
- 1.6 PB served

RDA Globus infrastructure

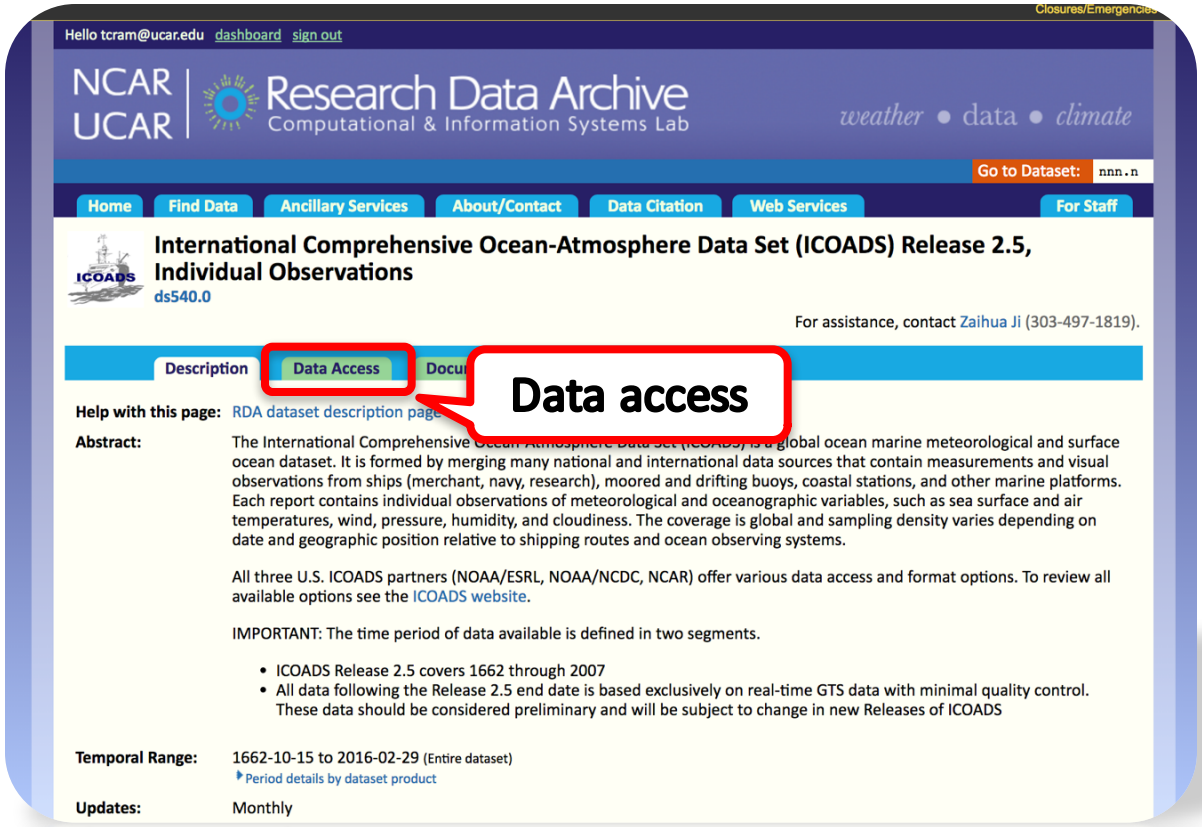
- Shared endpoints
  - **NCAR RDA dataset archive**
    (rda#datashare): permanent dataset access
  - **NCAR RDA data requests**
    (rda#data_request): custom data products

- Data access managed programatically
  - Globus REST API and CLI
  - acl-add & acl-remove
  - Initiated by user

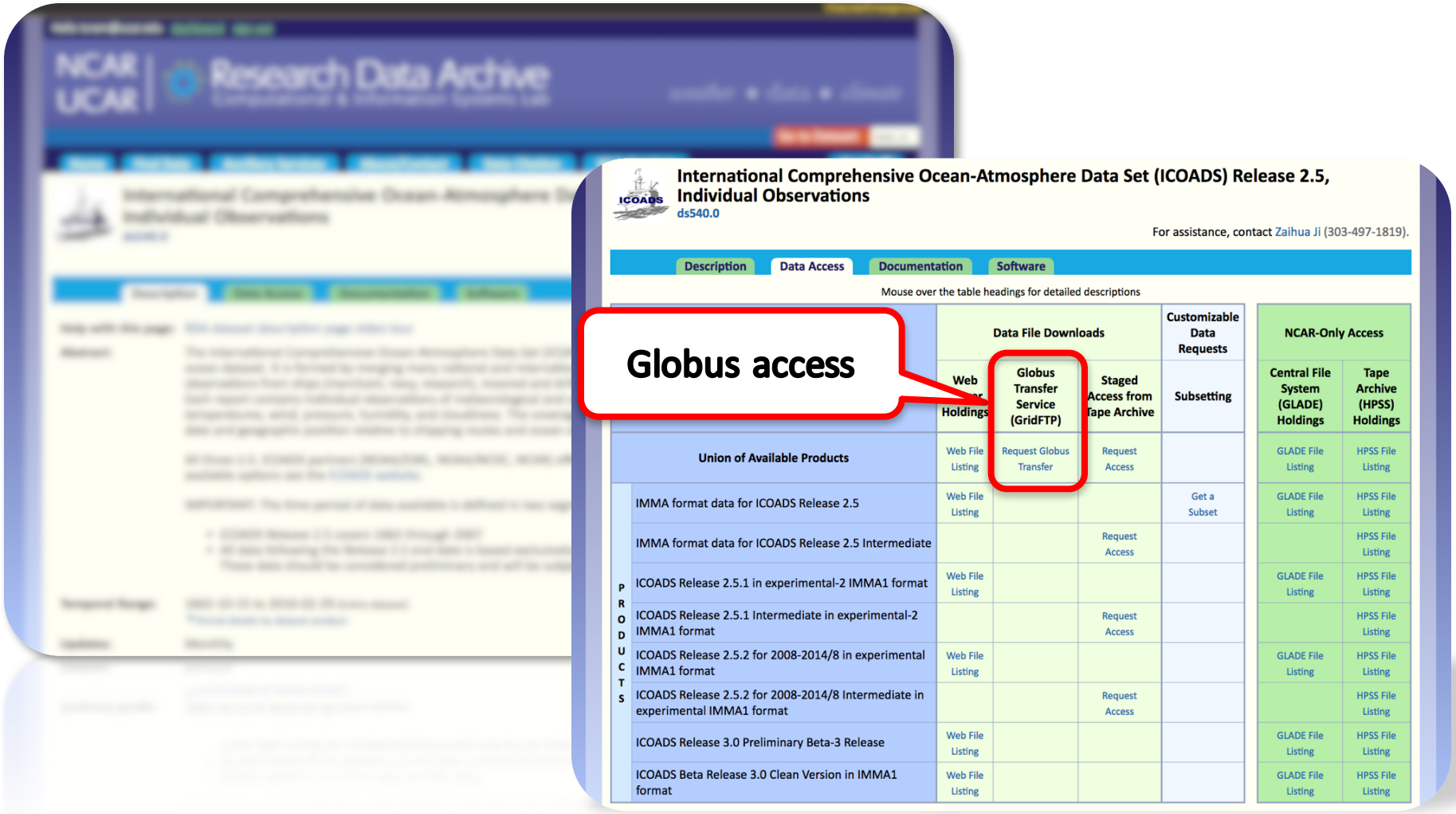
One ACL, One Dataset



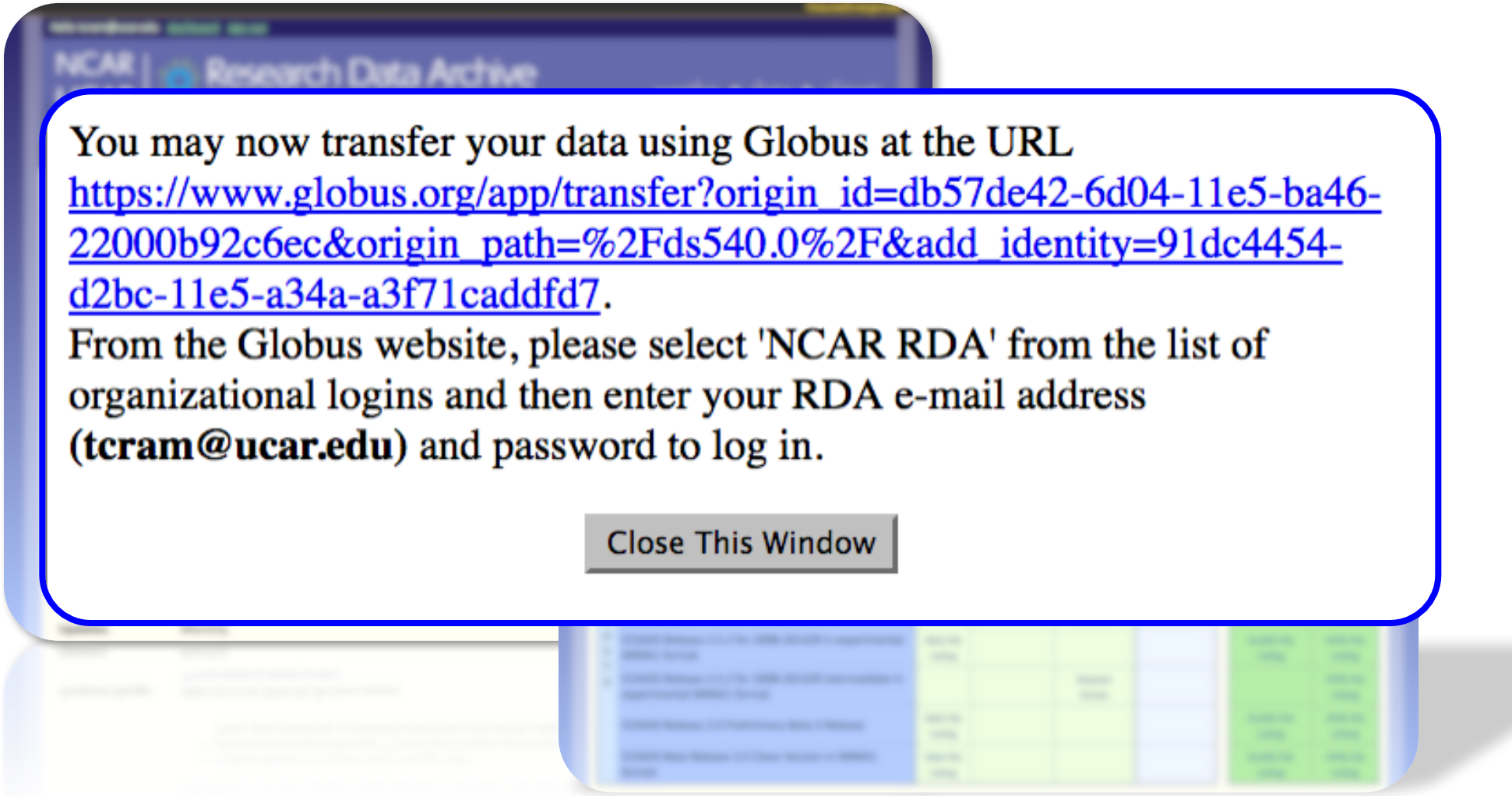
Example user interface



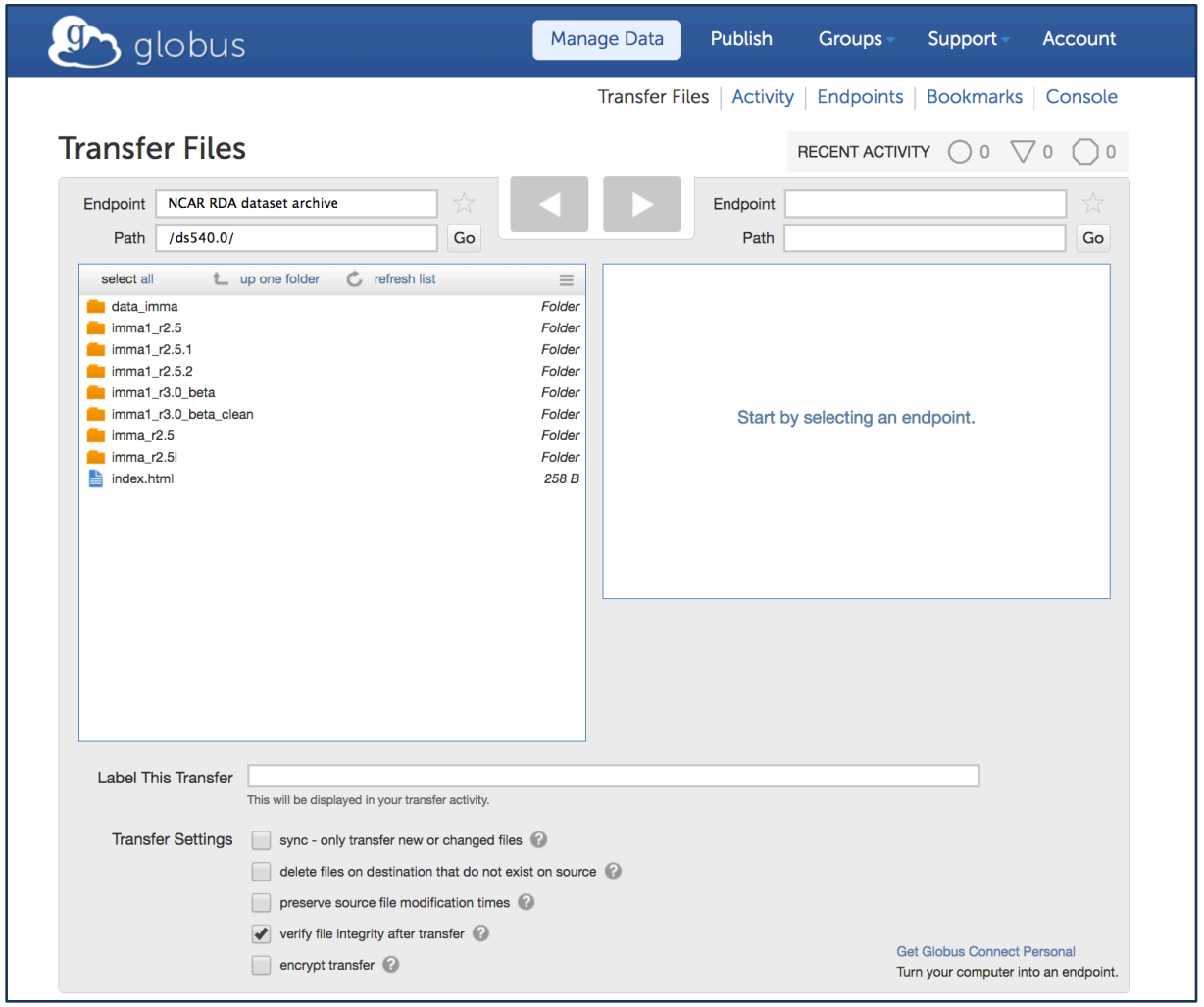
Example user interface



Example user interface



Example user interface



User dashboard

Dashboard

User Profile Edit/Change Profile

Thomas Cram
National Center for Atmospheric Research

Customized Data Requests

You don't have any customized data requests.

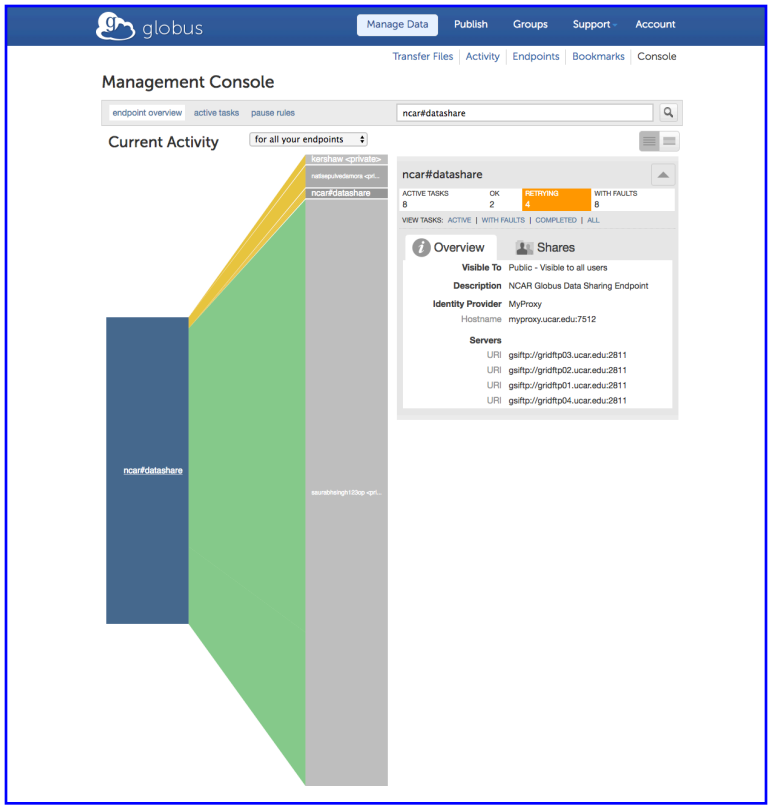
Globus Endpoints Shared With You Hide Shares

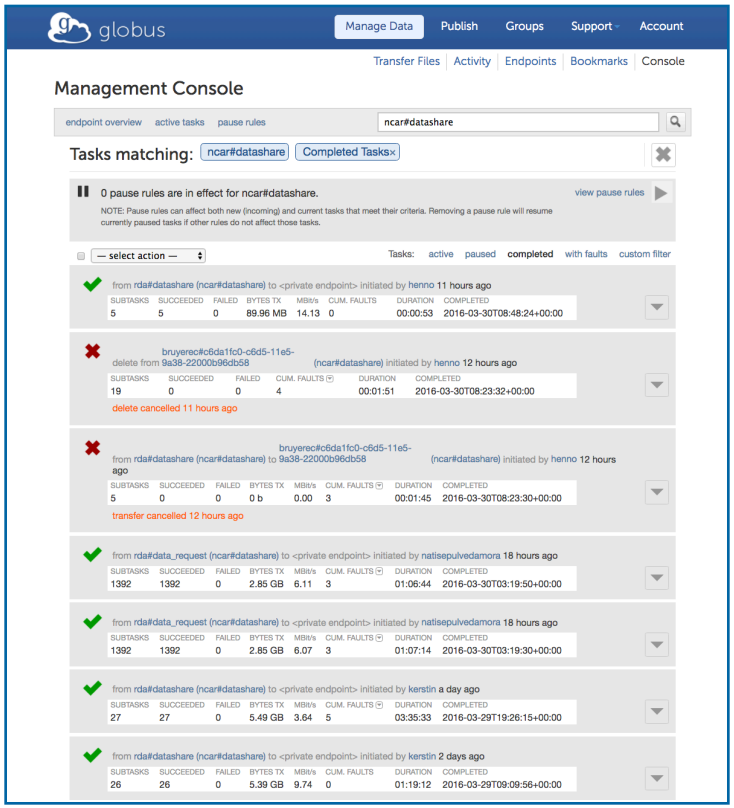
15 RDA data paths have been shared with you.

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





Advanced endpoint API
Useful resources

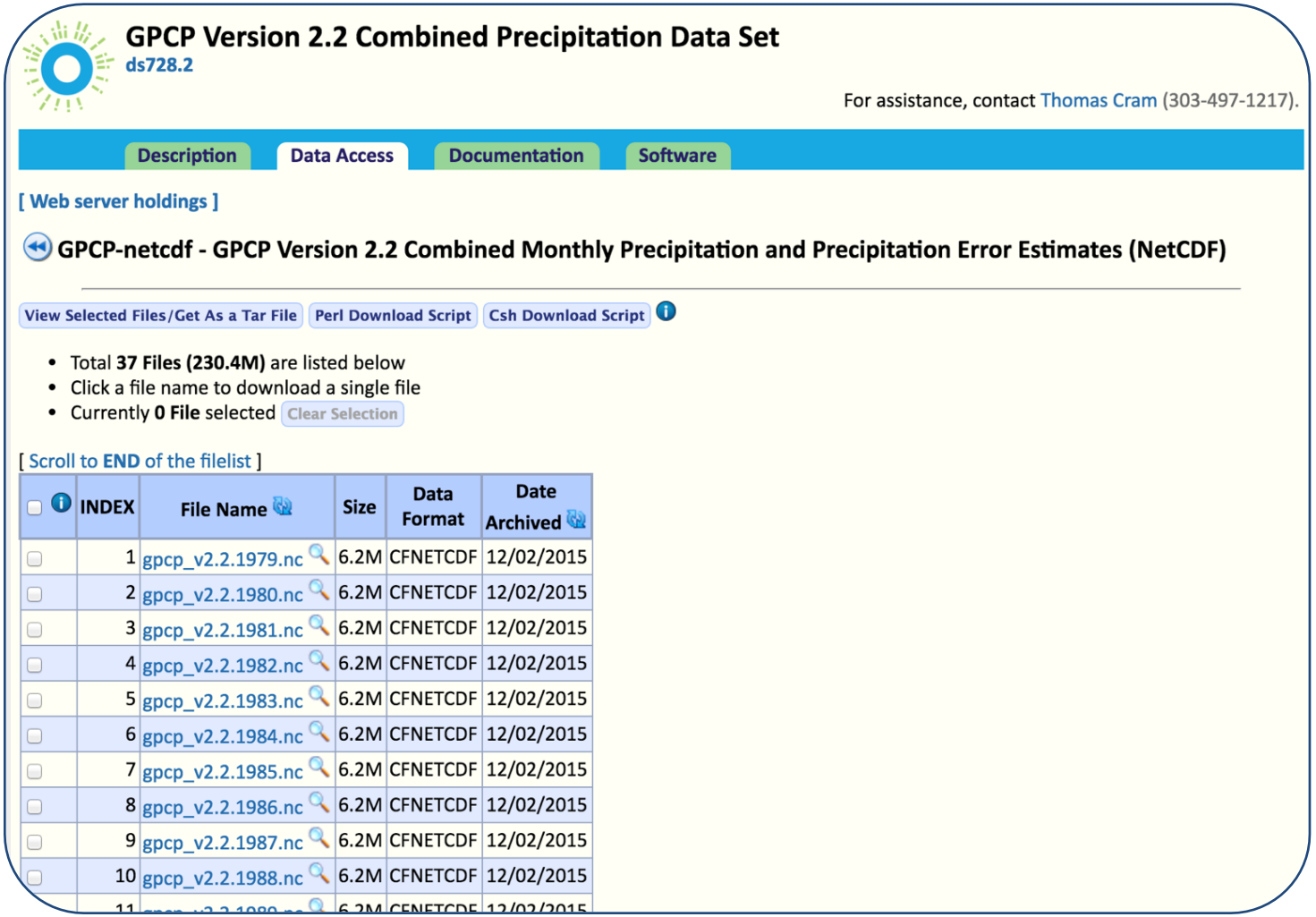
Base URL: ***https://transfer.api.globusonline.org/v0.10/***

- **/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]

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

Contact us

Research Data Archive

web: ***rda.ucar.edu***

@NCAR_RDA

NCAR CISL Research Data Archive

ncarrda.blogspot.com