ESnet

ENERGY SCIENCES NETWORK

Petascale Data Architectures for Portals and Computing Centers

Eli Dart, Science Engagement

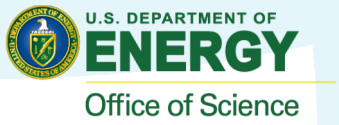
Energy Sciences Network (ESnet)

Lawrence Berkeley National Laboratory

TNC18

Trondheim, NO

12 June, 2018

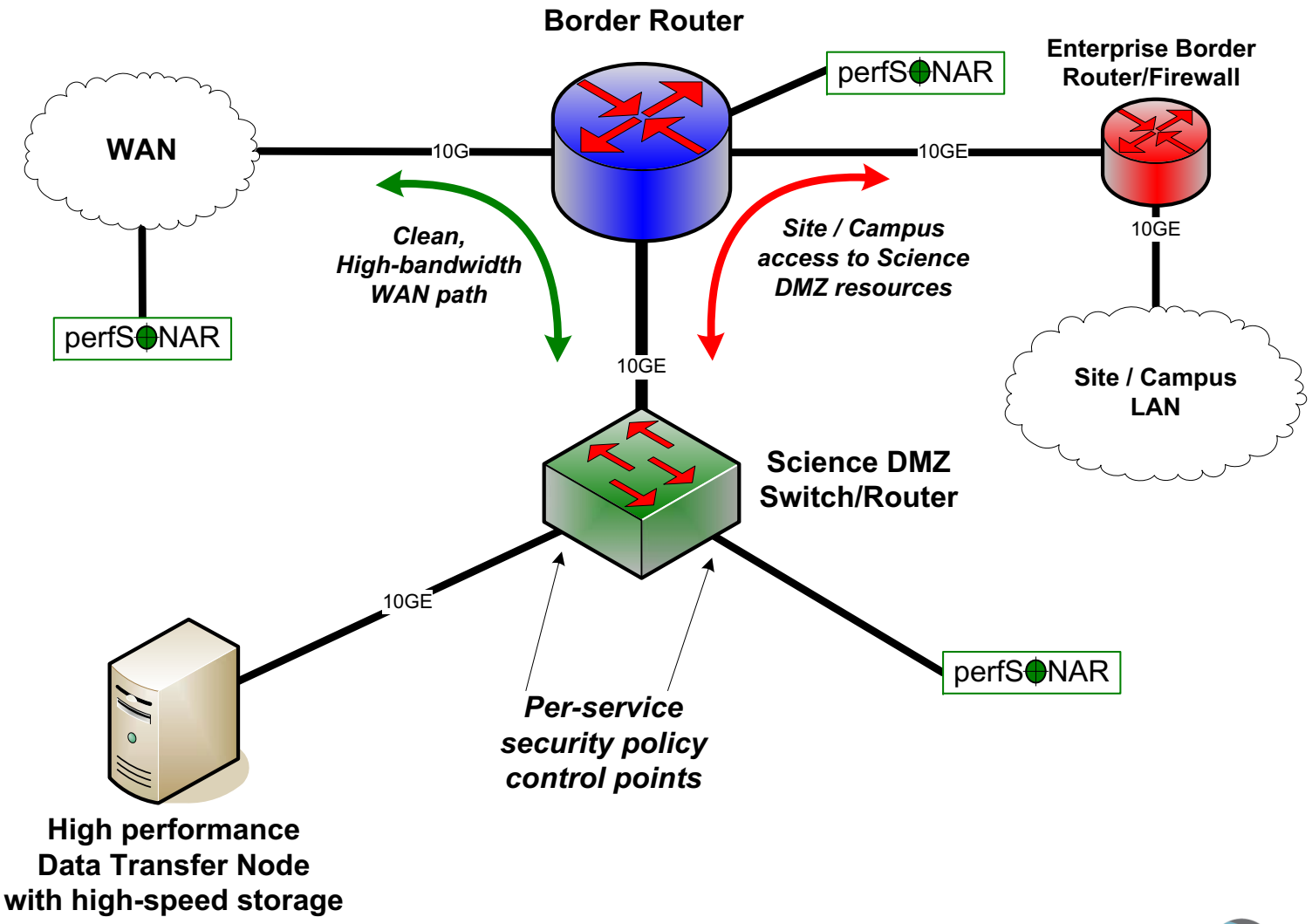


Outline

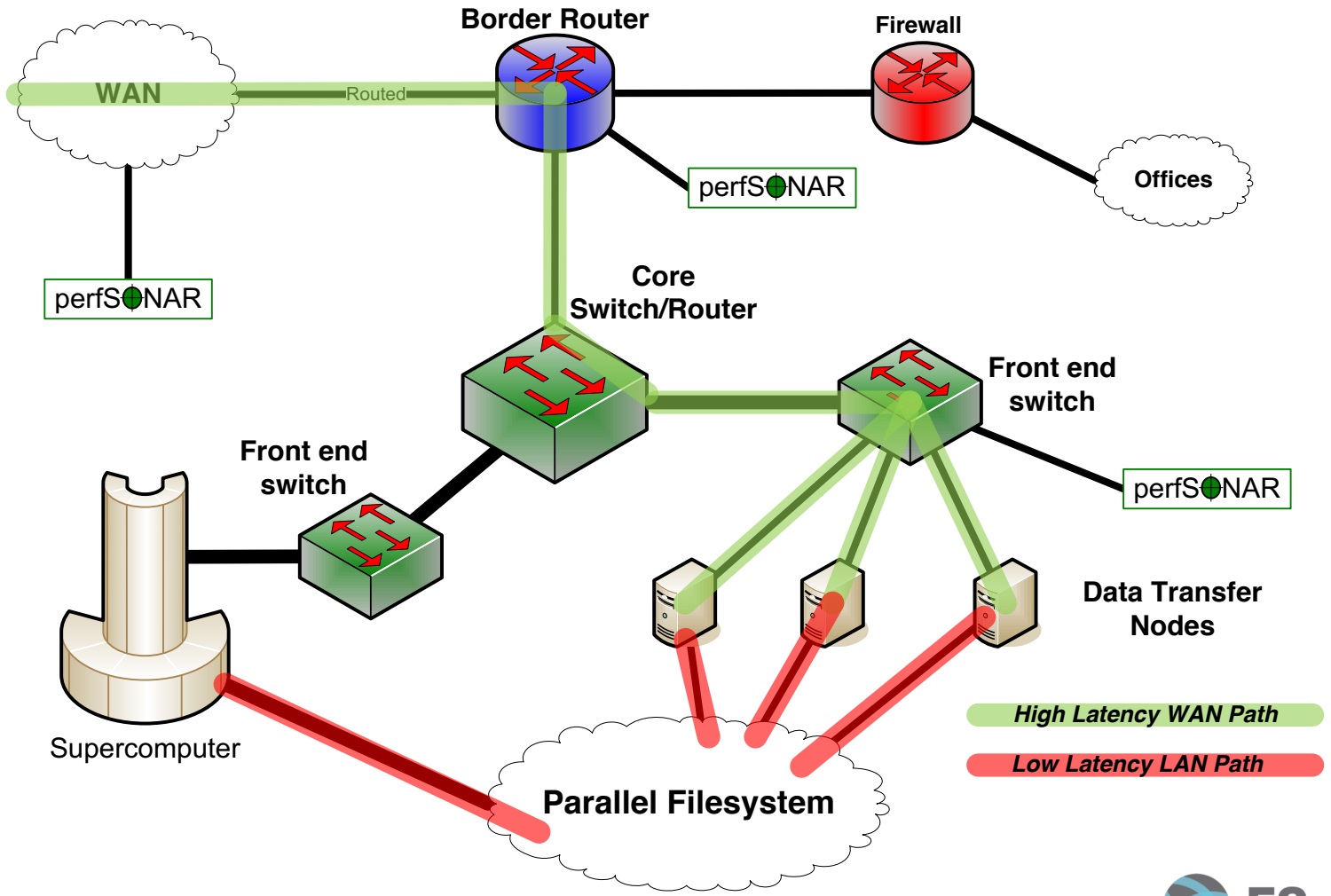
- Context
- Petascale DTN Project
- Modern Research Data Portal
- Long Term Vision

ESnet

Science DMZ Design Pattern (Abstract)



DMZ – Supercomputer Center DTN Cluster



Context: Science DMZ Adoption

- DOE National Laboratories
  - Supercomputer centers, LHC sites, experimental facilities
  - Both large and small sites
- NSF CC* programs have funded many Science DMZs
  - Large investments across the US university complex: over $100M
  - Significant strategic importance
- Outside the USA
  - Australia
  - Brazil
  - UK
  - More…

ESnet

Strategic Impacts

- What does this mean?
  - We are in the midst of a significant cyberinfrastructure upgrade
  - Enterprise networks need not be unduly perturbed ☺
- Significantly enhanced capabilities compared to 5 years ago
  - Terabyte-scale data movement is much easier
  - Petabyte-scale data movement possible outside the LHC experiments
    - ~3.1Gbps = 1PB/month
    - ~14Gbps = 1PB/week
  - Widely-deployed tools are much better (e.g. Globus)
- Metcalfe's Law of Network Utility
  - Value of Science DMZ proportional to the number of DMZs
    - n^2 or $n(\log_n)$ doesn't matter – the effect is real
  - Cyberinfrastructure value increases as we all upgrade

ESnet

Next Steps – Building On The Science DMZ

- Enhanced cyberinfrastructure substrate now exists
  - Wide area networks (ESnet, GEANT, NRENs, Internet2, Regionals)
  - Science DMZs connected to those networks
  - DTNs in the Science DMZs
- What does the scientist see?
  - Scientist sees a science application
    - Data transfer
    - Data portal
    - Data analysis
  - Science applications are the user interface to networks and DMZs
- Large-scale data-intensive science requires that we build science applications on top of the substrate components

ESnet

NCAR RDA Data Portal

- Let's say I have a nice compute allocation at NERSC – climate science
- Let's say I need some data from NCAR for my project

- https://rda.ucar.edu/

- Data sets (there are many more, but these are two):
- https://rda.ucar.edu/datasets/ds199.1/
- https://rda.ucar.edu/datasets/ds313.0/

- Download to NERSC (could also do ALCF or NCSA or OLCF)

ESnet

HPC Centers Matter

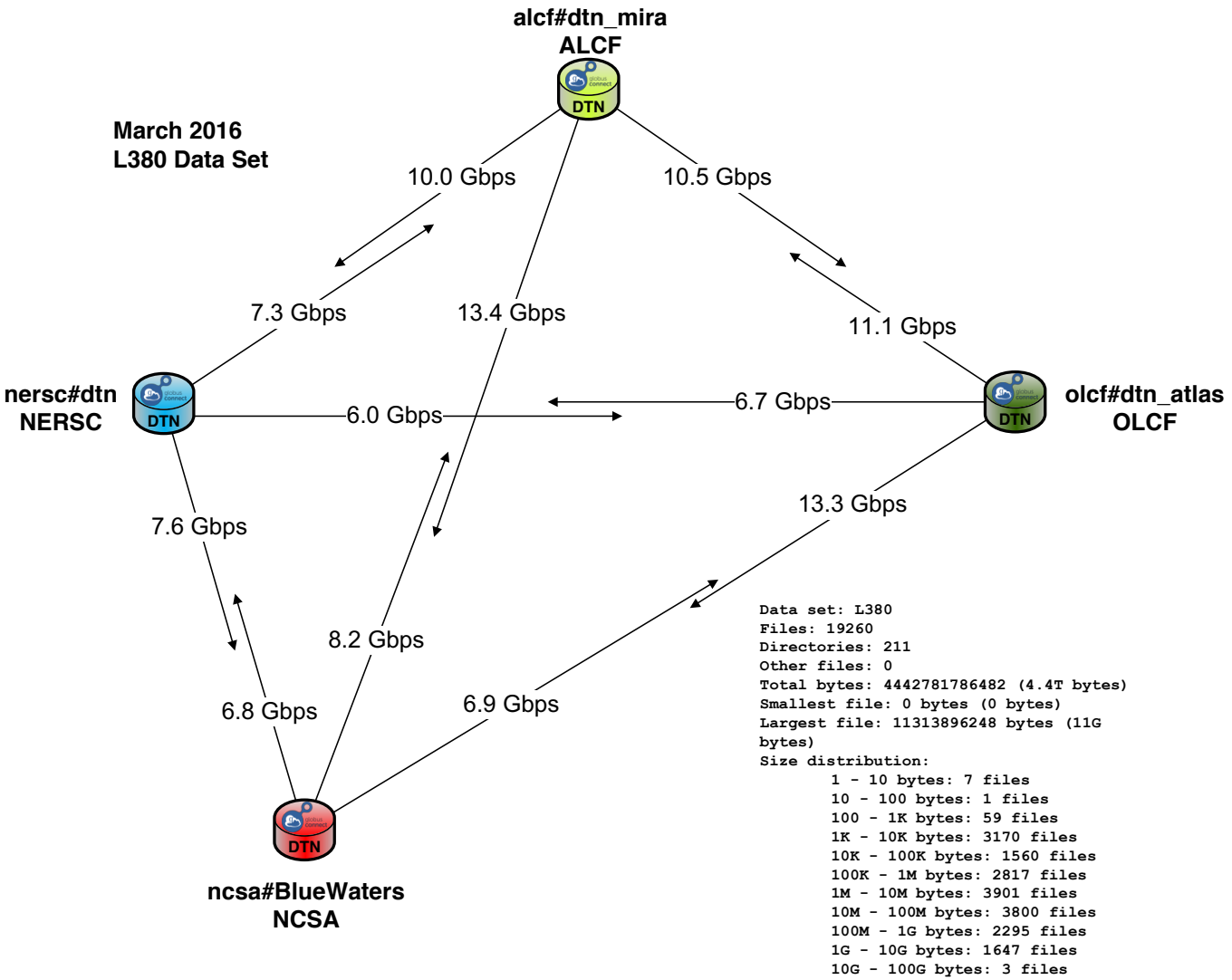
- Computing centers are special
  - Centers of excellence / expertise
  - Data repositories
  - Computing for simulation, data analysis
- Really though – the people + cyberinfrastructure combo is special
  - People who know how computers, networking, and storage work
  - Enough resources to make things happen
- Computing facilities are anchors for many collaborations
  - Common pattern: multi-institution team with access to one HPC center
  - Shared data, analysis, simulation platform

ESnet

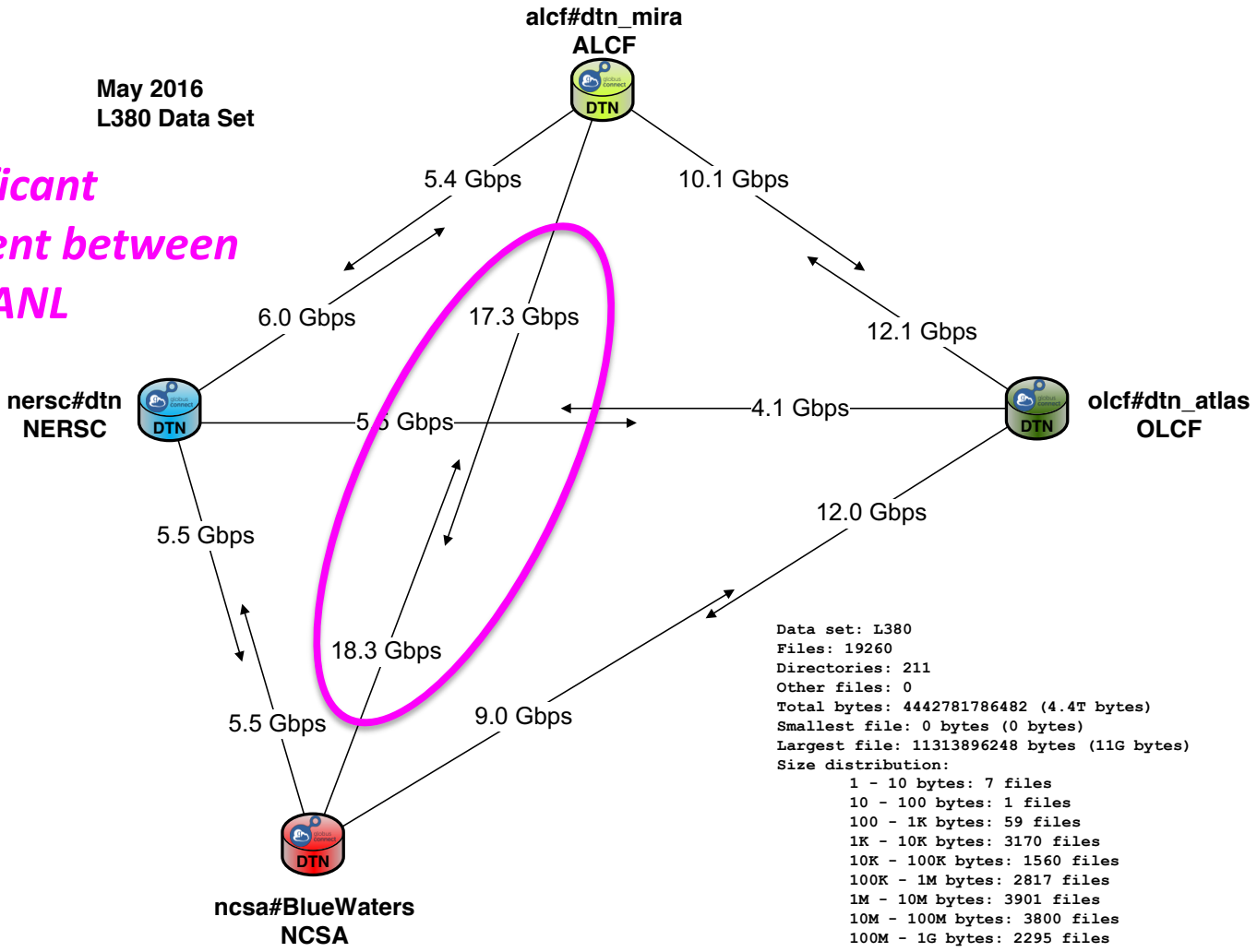
Data And HPC: The Petascale DTN Project

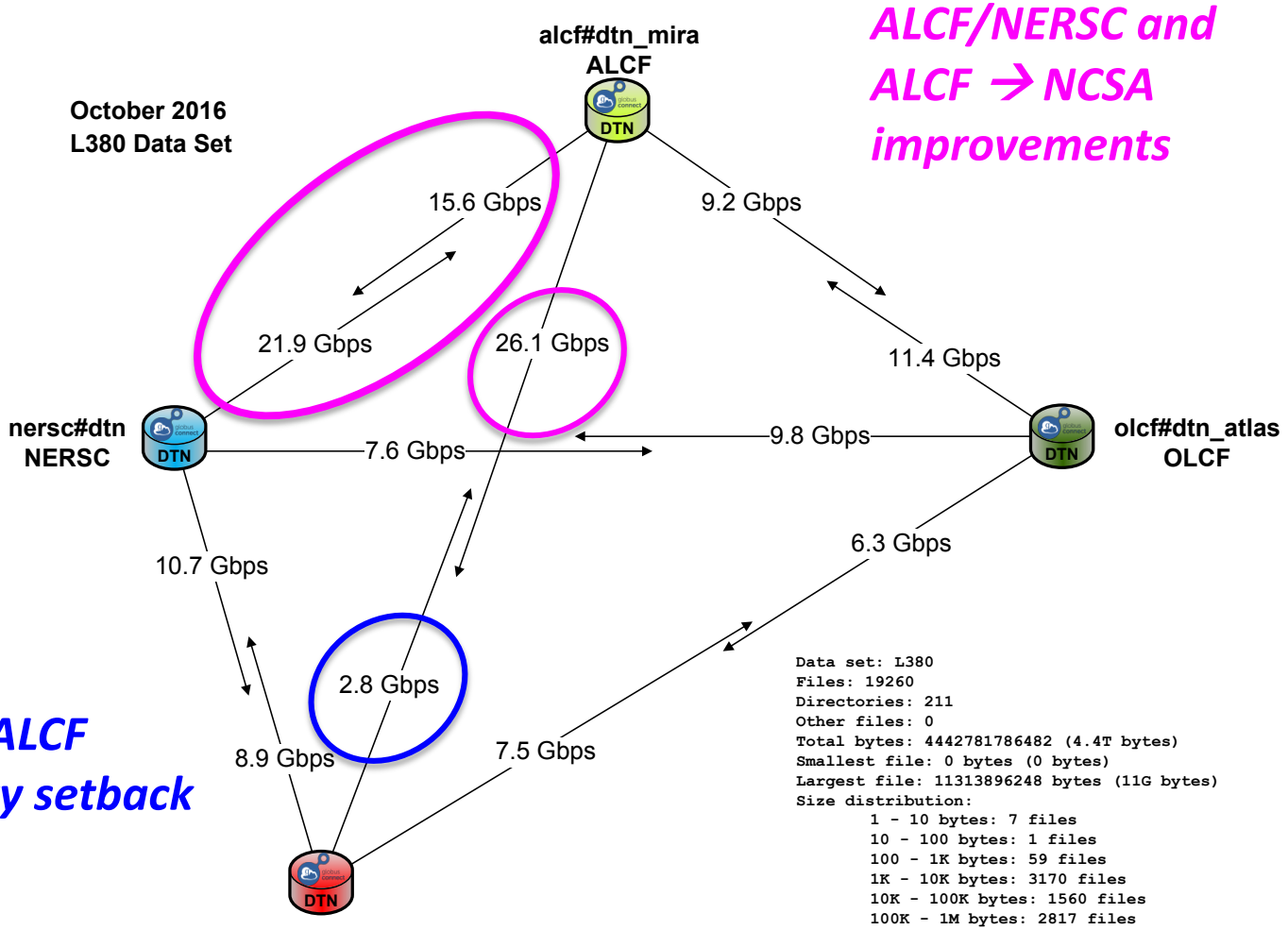
- Built on top of the Science DMZ
- Effort to improve data transfer performance between the DOE ASCR HPC facilities at ANL, LBNL, and ORNL, and also NCSA.
  - Multiple current and future science projects need to transfer data between HPC facilities
  - Performance was slow, configurations inconsistent
  - Performance goal of 15 gigabits per second (equivalent to 1PB/week)
  - Realize performance goal for routine Globus transfers without special tuning
- Reference data set is 4.4TB of cosmology simulation data
- Use performant, easy-to-use tools with production options on
  - Globus Transfer service (previously Globus Online)
  - Use GUI just like a user would, with default options
    - E.g. integrity checksums enabled, as they should be
    - No arcane magic!

ESnet


alcf#dtn_mira
ALCF
DTN
March 2016
L380 Data Set
10.0 Gbps
10.5 Gbps
7.3 Gbps
13.4 Gbps
11.1 Gbps
nersc#dtn
NERSC
DTN
6.0 Gbps
6.7 Gbps
DTN
olcf#dtn_atlas
OLCF
13.3 Gbps
7.6 Gbps
8.2 Gbps
6.8 Gbps
6.9 Gbps
DTN
ncsa#BlueWaters
NCSA
Data set: L380
Files: 19260
Directories: 211
Other files: 0
Total bytes: 4442781786482 (4.4T bytes)
Smallest file: 0 bytes (0 bytes)
Largest file: 11313896248 bytes (11G bytes)
Size distribution:
1 - 10 bytes: 7 files
10 - 100 bytes: 1 files
100 - 1K bytes: 59 files
1K - 10K bytes: 3170 files
10K - 100K bytes: 1560 files
100K - 1M bytes: 2817 files
1M - 10M bytes: 3901 files
10M - 100M bytes: 3800 files
100M - 1G bytes: 2295 files
1G - 10G bytes: 1647 files
10G - 100G bytes: 3 files

Note significant improvement between NCSA and ANL





ALCF/NERSC and ALCF → NCSA improvements

NCSA → ALCF temporary setback

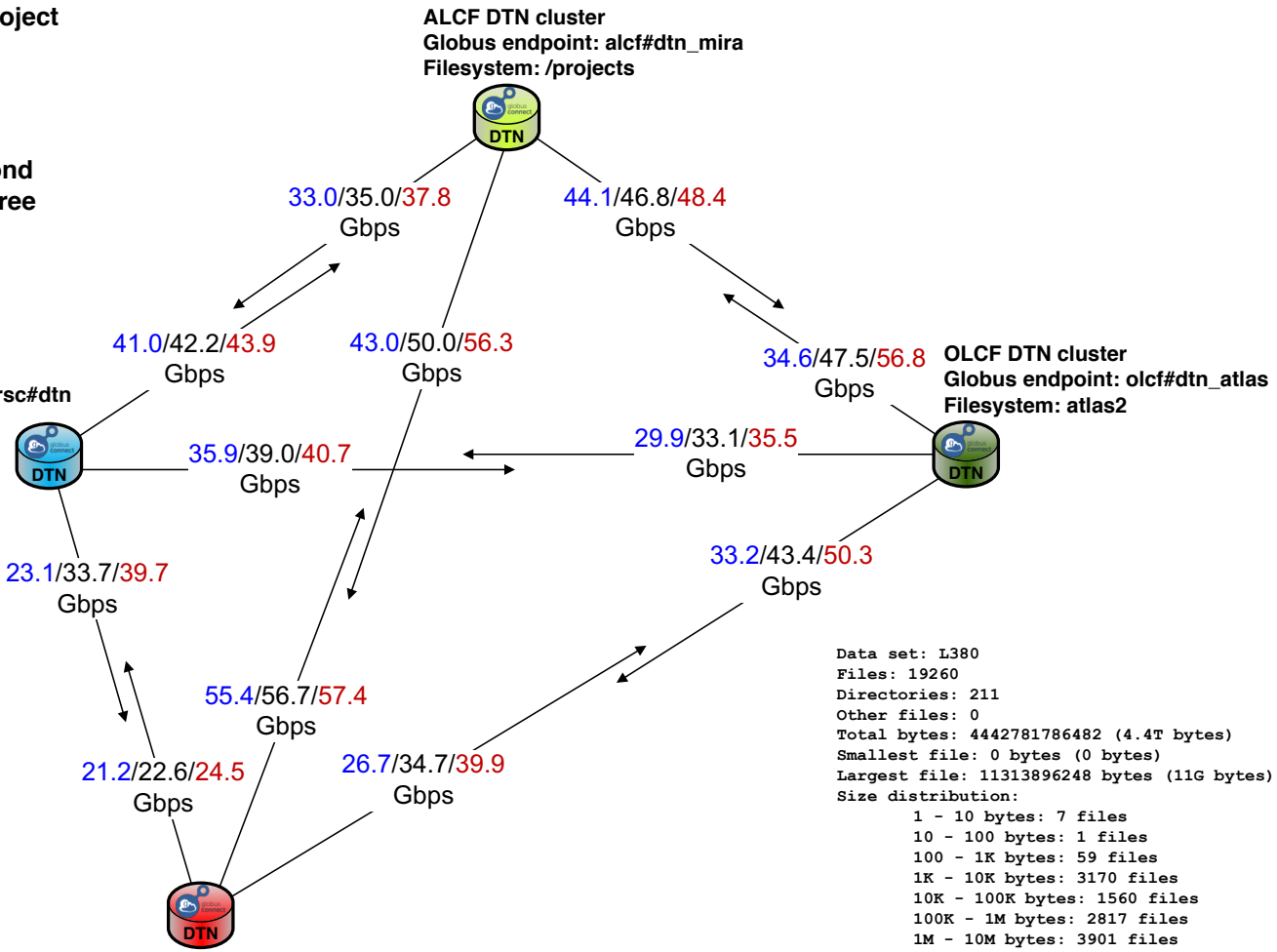
```
Data set: L380
Files: 19260
Directories: 211
Other files: 0
Total bytes: 4442781786482 (4.4T bytes)
Smallest file: 0 bytes (0 bytes)
Largest file: 11313896248 bytes (11G bytes)
Size distribution:
        1 - 10 bytes: 7 files
        10 - 100 bytes: 1 files
        100 - 1K bytes: 59 files
        1K - 10K bytes: 3170 files
        10K - 100K bytes: 1560 files
        100K - 1M bytes: 2817 files
        1M - 10M bytes: 3901 files
        10M - 100M bytes: 3800 files
        100M - 1G bytes: 2295 files
        1G - 10G bytes: 1647 files
        10G - 100G bytes: 3 files
```

DTN Cluster Performance – HPC Facilities (2017)

Petascale DTN Project

November 2017
L380 Data Set

Gigabits per second (min/avg/max), three transfers




```
Data set: L380
Files: 19260
Directories: 211
Other files: 0
Total bytes: 4442781786482 (4.4T bytes)
Smallest file: 0 bytes (0 bytes)
Largest file: 11313896248 bytes (11G bytes)
Size distribution:
        1 - 10 bytes: 7 files
        10 - 100 bytes: 1 files
        100 - 1K bytes: 59 files
        1K - 10K bytes: 3170 files
        10K - 100K bytes: 1560 files
        100K - 1M bytes: 2817 files
        1M - 10M bytes: 3901 files
        10M - 100M bytes: 3800 files
        100M - 1G bytes: 2295 files
        1G - 10G bytes: 1647 files
        10G - 100G bytes: 3 files
```

ESnet

Improvements In Multiple Places

- DTN clusters expanded
  - ALCF: 8 DTNs to 12
  - NERSC: 4 DTNs to 8
  - ORNL: 8 DTNs to 16 (for Globus pool)
- Network Upgrades
  - ORNL: Nx10G to Nx40G
  - NCSA: Nx10G to Nx100G
  - NERSC: Nx10G to Nx40G
- Unrelated Globus change gave a big performance boost
  - Change in behavior of transfers with large file counts (batch size increased)

ESnet

Petascale DTN Lifts All Boats

- Petascale DTN project benefits all projects which use the HPC facility DTNs
- Modern science data portal architecture
  - Data portals which use modern architecture benefit from DTN improvements
  - DTN scaling/improvements benefit all data portals which use the same pool
- Globus API supports this – see Globus World Tour
  - https://www.globusworld.org/tour/

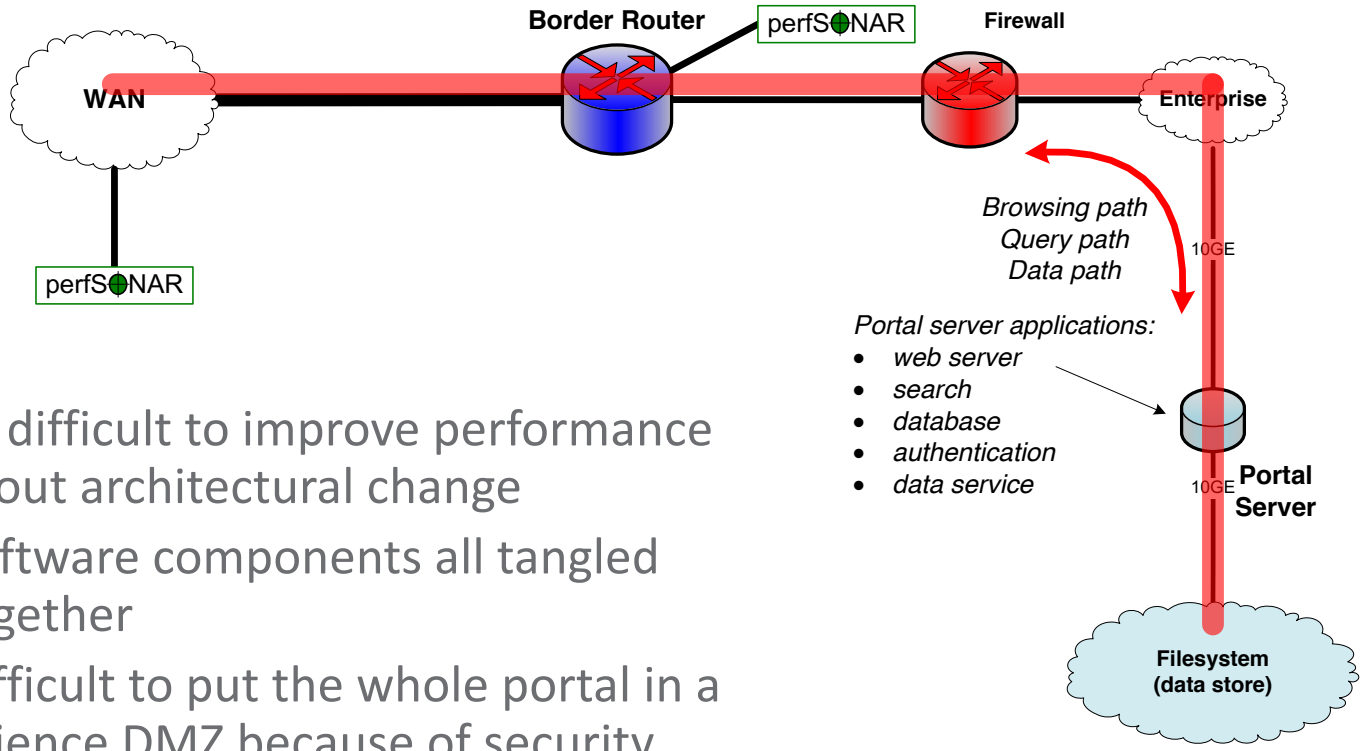
ESnet

Science Data Portals

- Large repositories of scientific data
  - Climate data
  - Sky surveys (astronomy, cosmology)
  - Many others
  - Data search, browsing, access
- Many scientific data portals were designed 15+ years ago
  - Single-web-server design
  - Data browse/search, data access, user awareness all in a single system
  - All the data goes through the portal server
    - In many cases by design
    - E.g. embargo before publication (enforce access control)
  - Better than old command-line FTP, but outdated by today's standards

ESnet

Legacy Portal Design



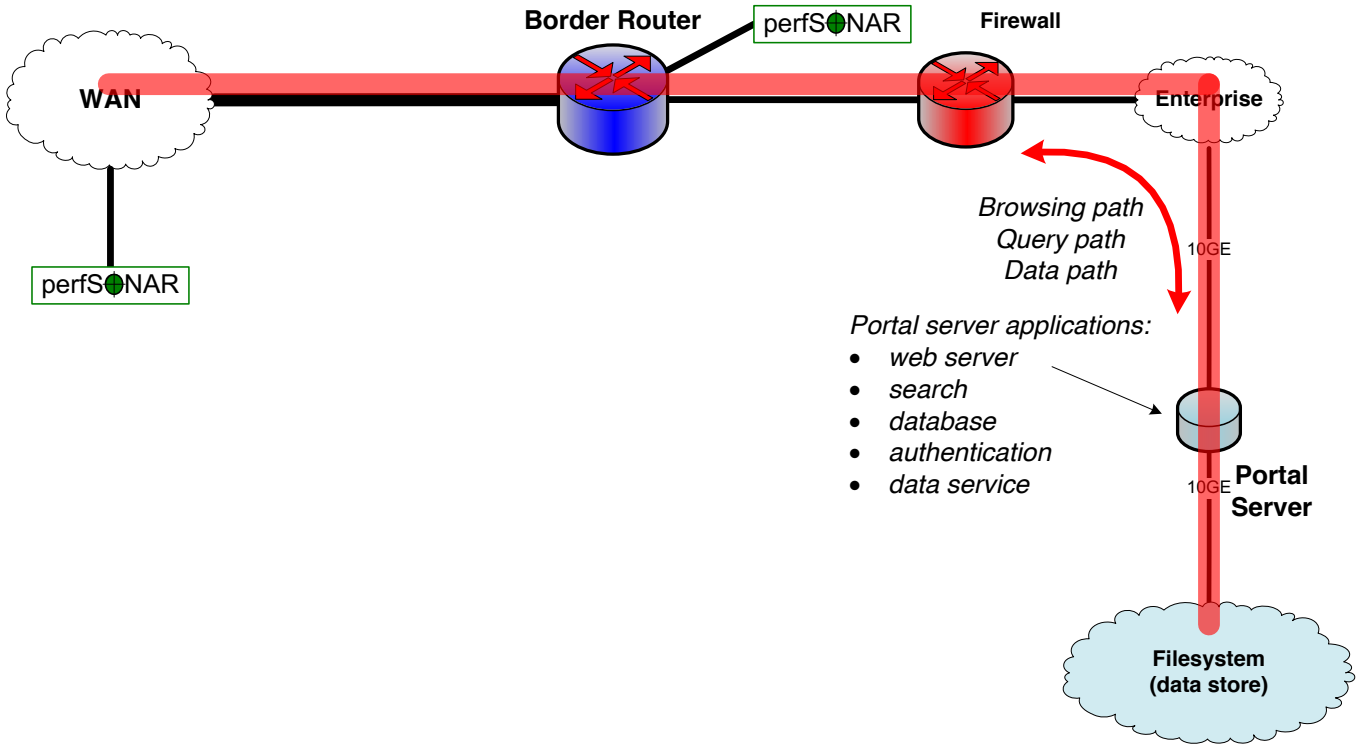
- Very difficult to improve performance without architectural change
  - Software components all tangled together
  - Difficult to put the whole portal in a Science DMZ because of security
  - Even if you could put it in a DMZ, many components aren't scalable
- What does architectural change mean?

Architectural Examination of Data Portals

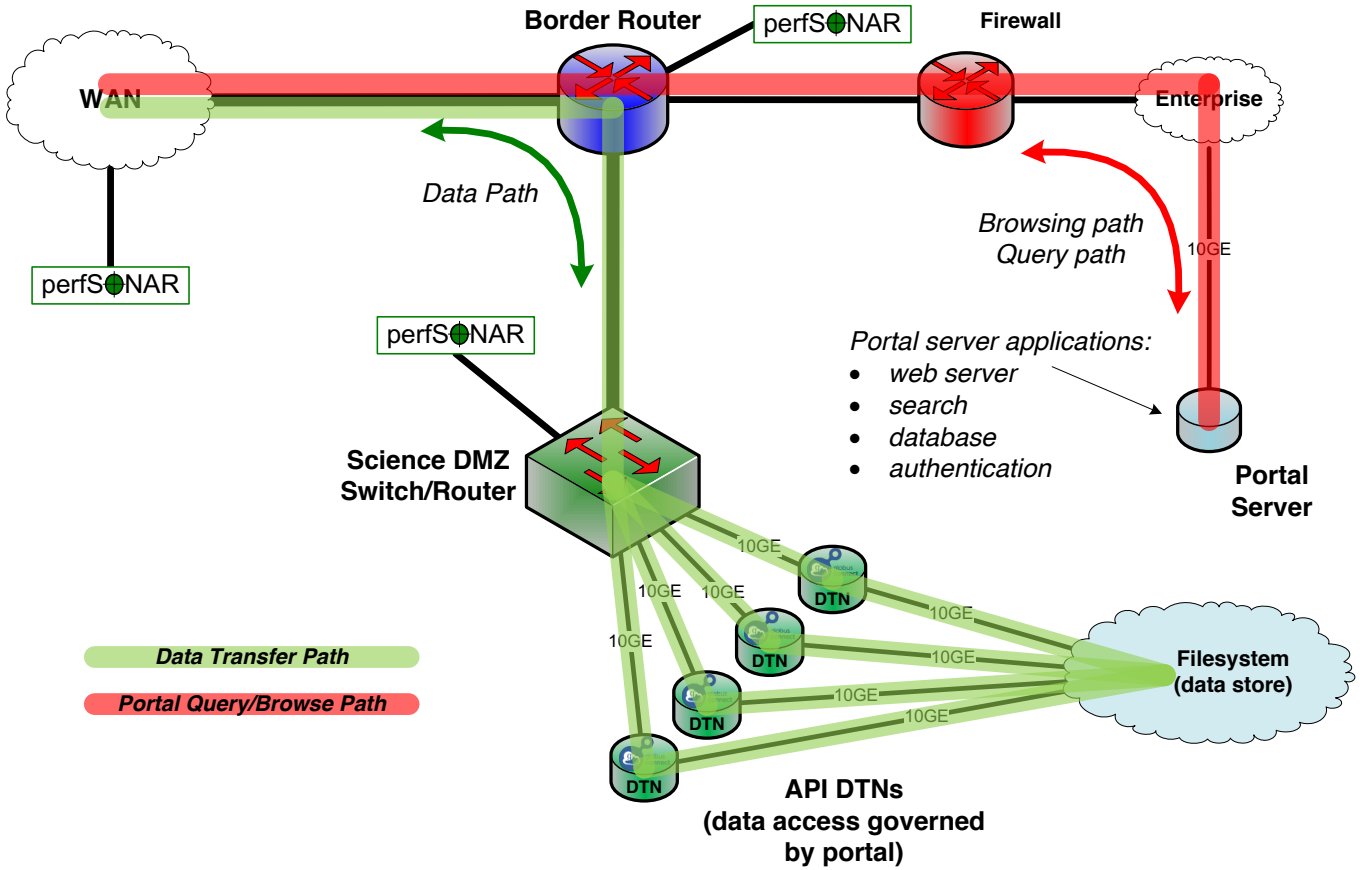
- Not necessarily advocating CDNs for science data (not really a good fit)
- Common data portal functions (most portals have these)
  - Search/query/discovery
  - Data download method for data access
  - GUI for browsing by humans
  - API for machine access – ideally incorporates search/query + download
- Performance pain is primarily in the data handling piece
  - Rapid increase in data scale eclipsed legacy software stack capabilities
  - Portal servers often stuck in enterprise network
- Can we “disassemble” the portal and put the pieces back together better?
  - Use Science DMZ as a platform for the data piece
  - Avoid placing complex software in the Science DMZ

ESnet

Legacy Portal Design



Next-Generation Portal Leverages Science DMZ



https://peerj.com/articles/cs-144/

ESnet

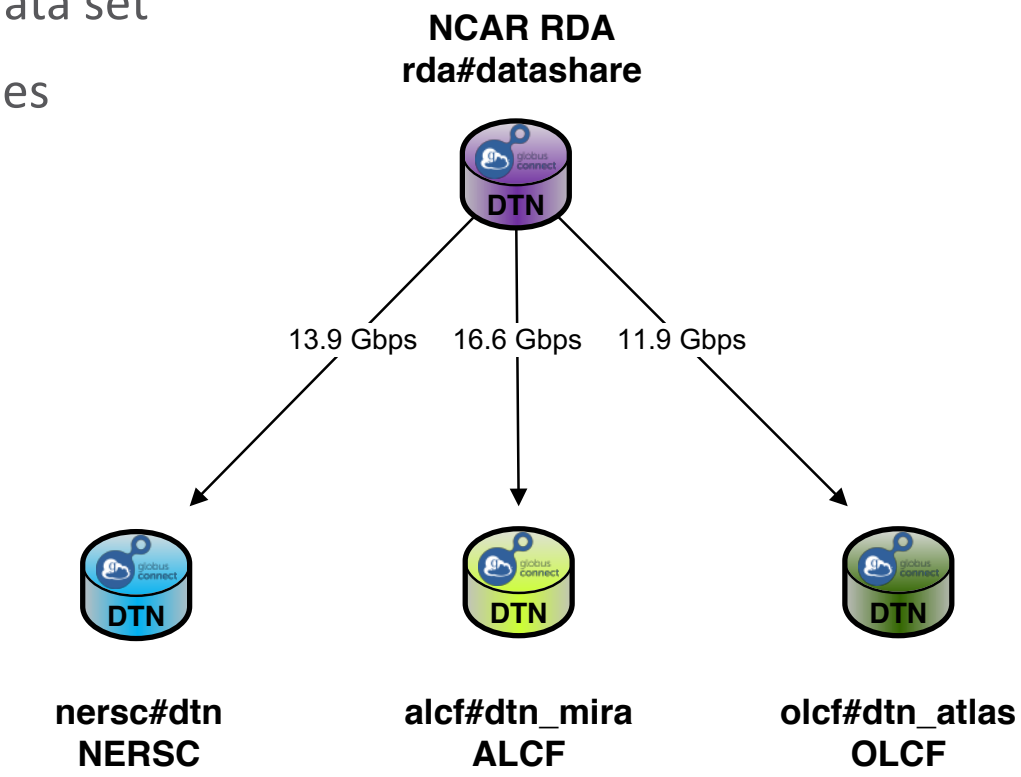
Put The Data On Dedicated Infrastructure

- We have separated the data handling from the portal logic
- Portal is still its normal self, but enhanced
  - Portal GUI, database, search, etc. all function as they did before
  - Query returns pointers to data objects in the Science DMZ
  - Portal is now freed from ties to the data servers (run it on Amazon if you want!)
- Data handling is separate, and scalable
  - High-performance DTNs in the Science DMZ
  - Scale as much as you need to without modifying the portal software
- Outsource data handling to computing centers
  - Computing centers are set up for large-scale data
  - Let them handle the large-scale data, and let the portal do the orchestration of data placement
- https://peerj.com/articles/cs-144/ - Modern Research Data Portal paper

ESnet

NCAR RDA Performance to DOE HPC Facilities

- 1.5TB data set
- 1121 files

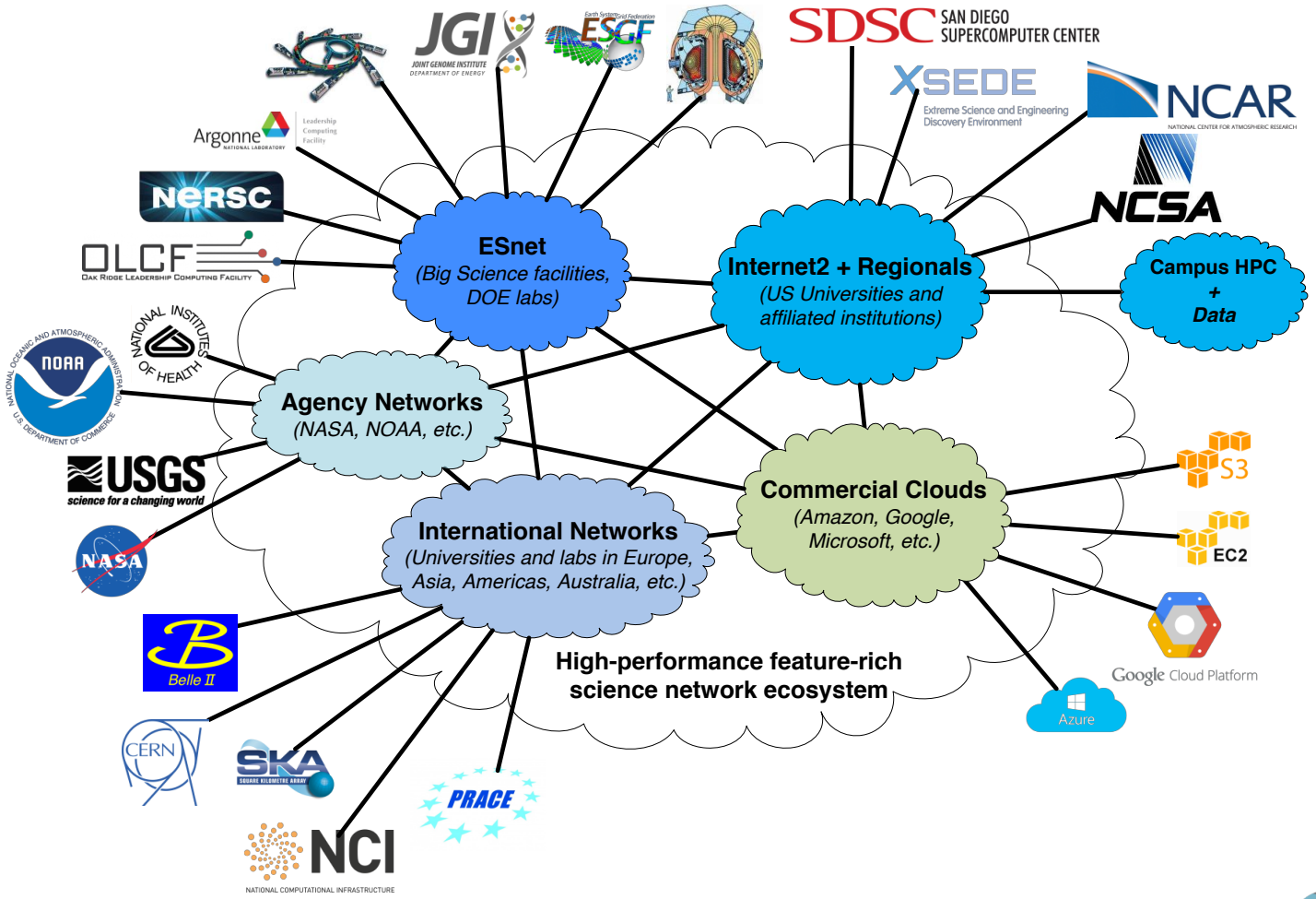


Larger Strategic Picture

- Across the scientific community, larger structures are being built
  - HPC facilities combined with experiments
  - DTNs between campuses
  - These create the platform for future scientific discoveries.
- Building DMZs, DTNs, and similar things for scientists puts the power of modern cyberinfrastructure in the hands of the people who will make the discoveries that change our world for the better.
- By doing this work, we help bring about the future that we all want - better medicine, better technology, more energy, a cleaner environment, etc.

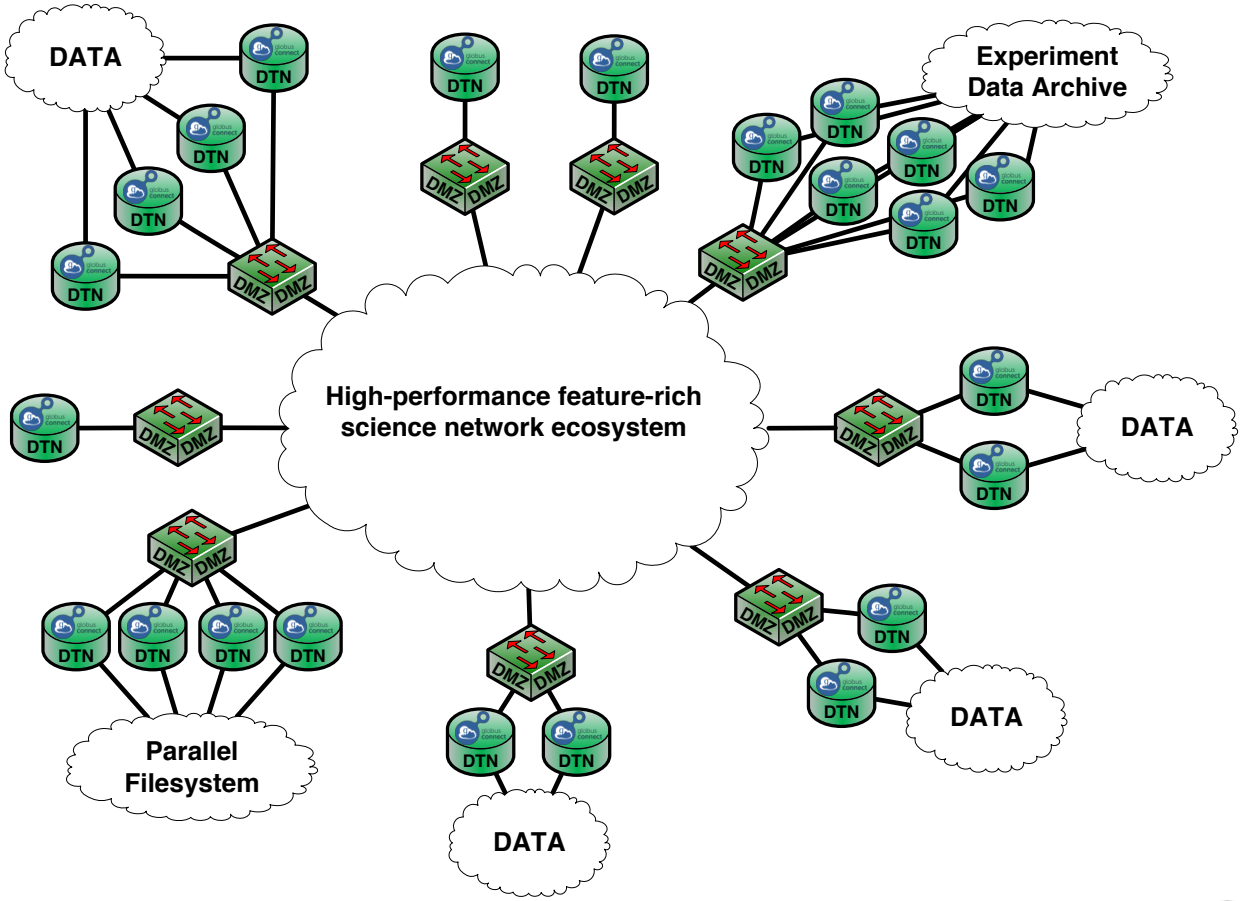
ESnet

Long-Term Vision



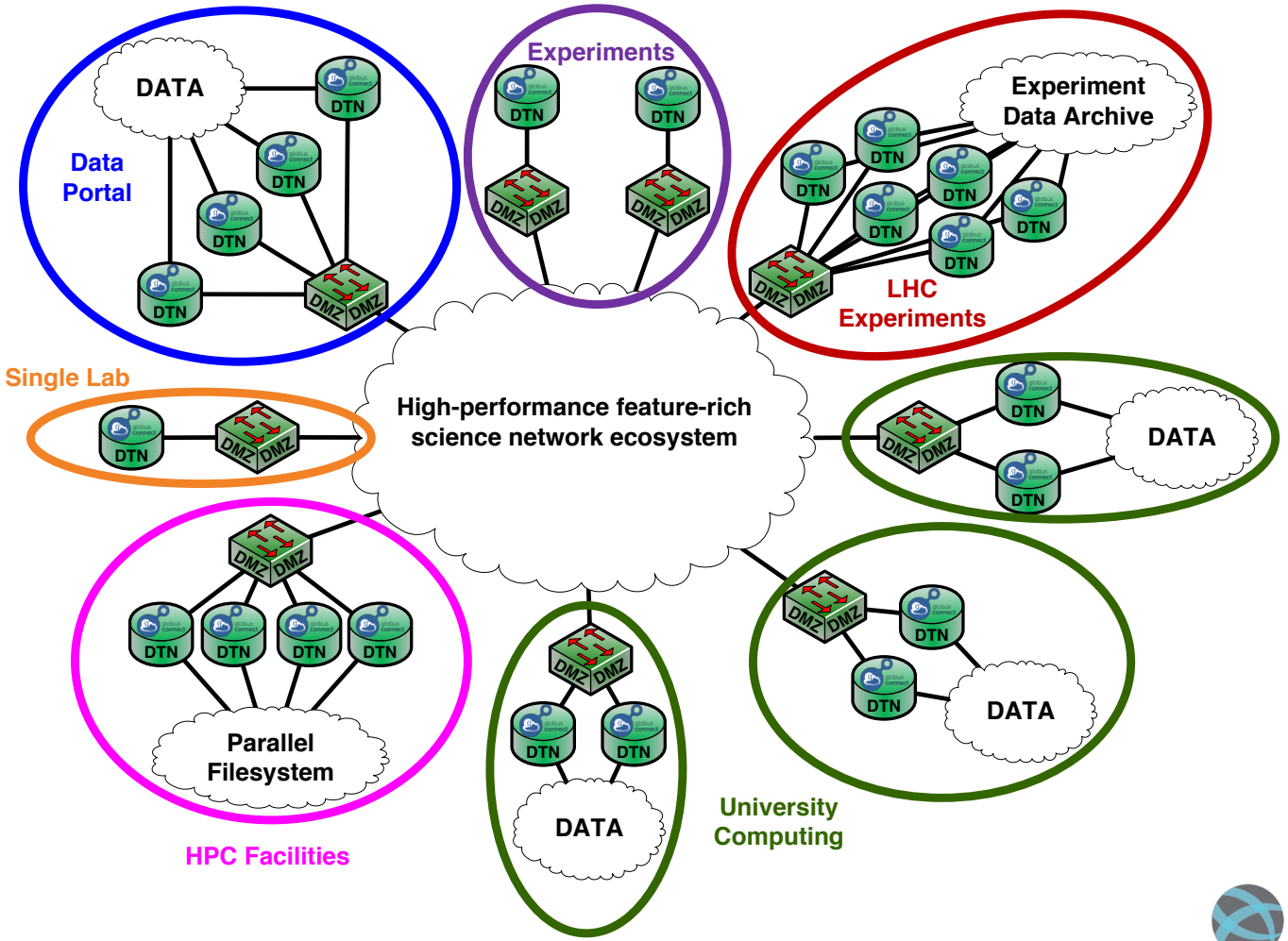
ESnet

It's All A Bunch Of Science DMZs



ESnet

It's All A Bunch Of Science DMZs



ESnet

ESnet

ENERGY SCIENCES NETWORK

Thanks!

Eli Dart

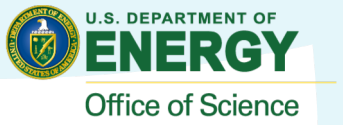
Energy Sciences Network (ESnet)

Lawrence Berkeley National Laboratory

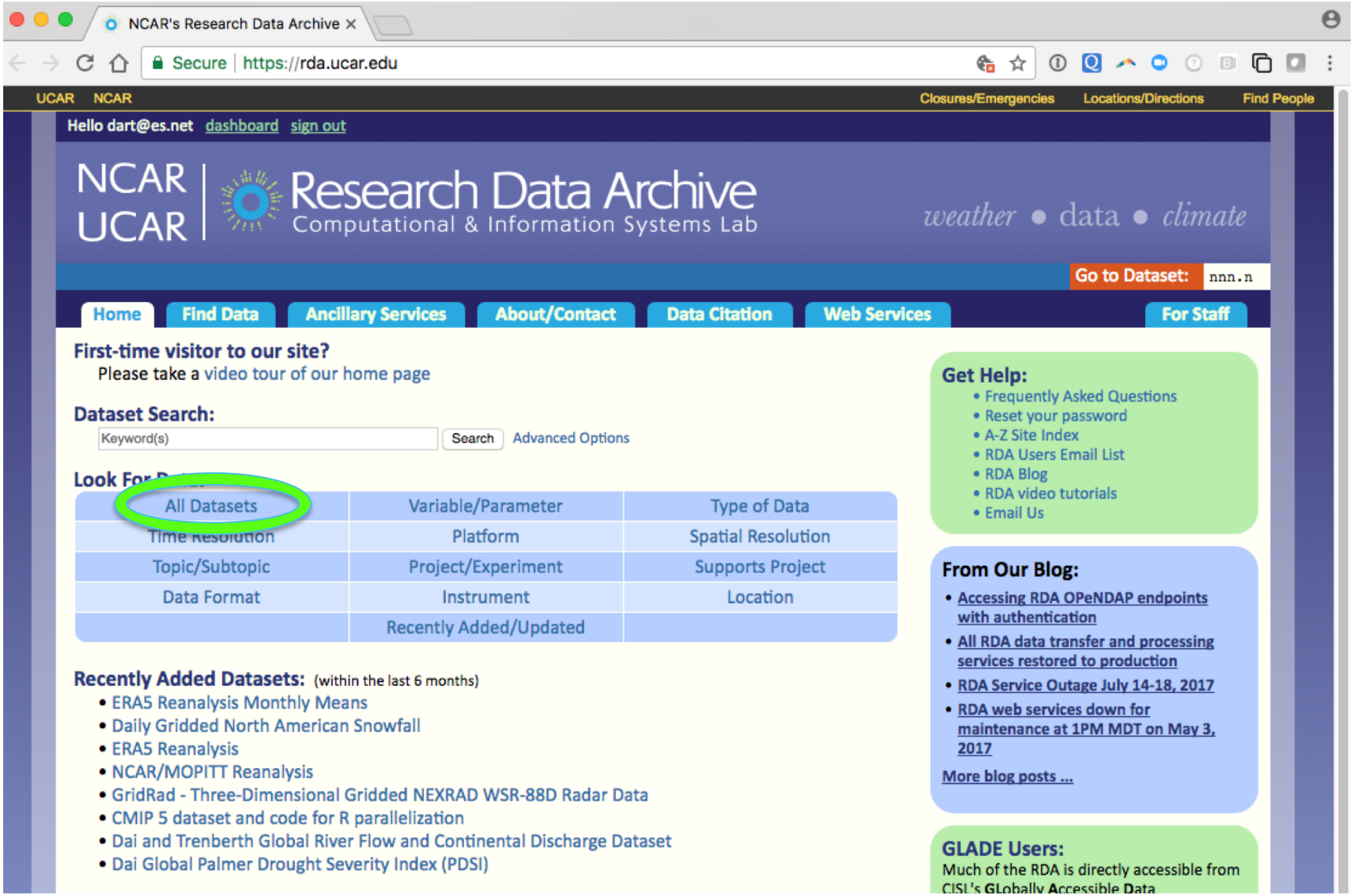
http://my.es.net/

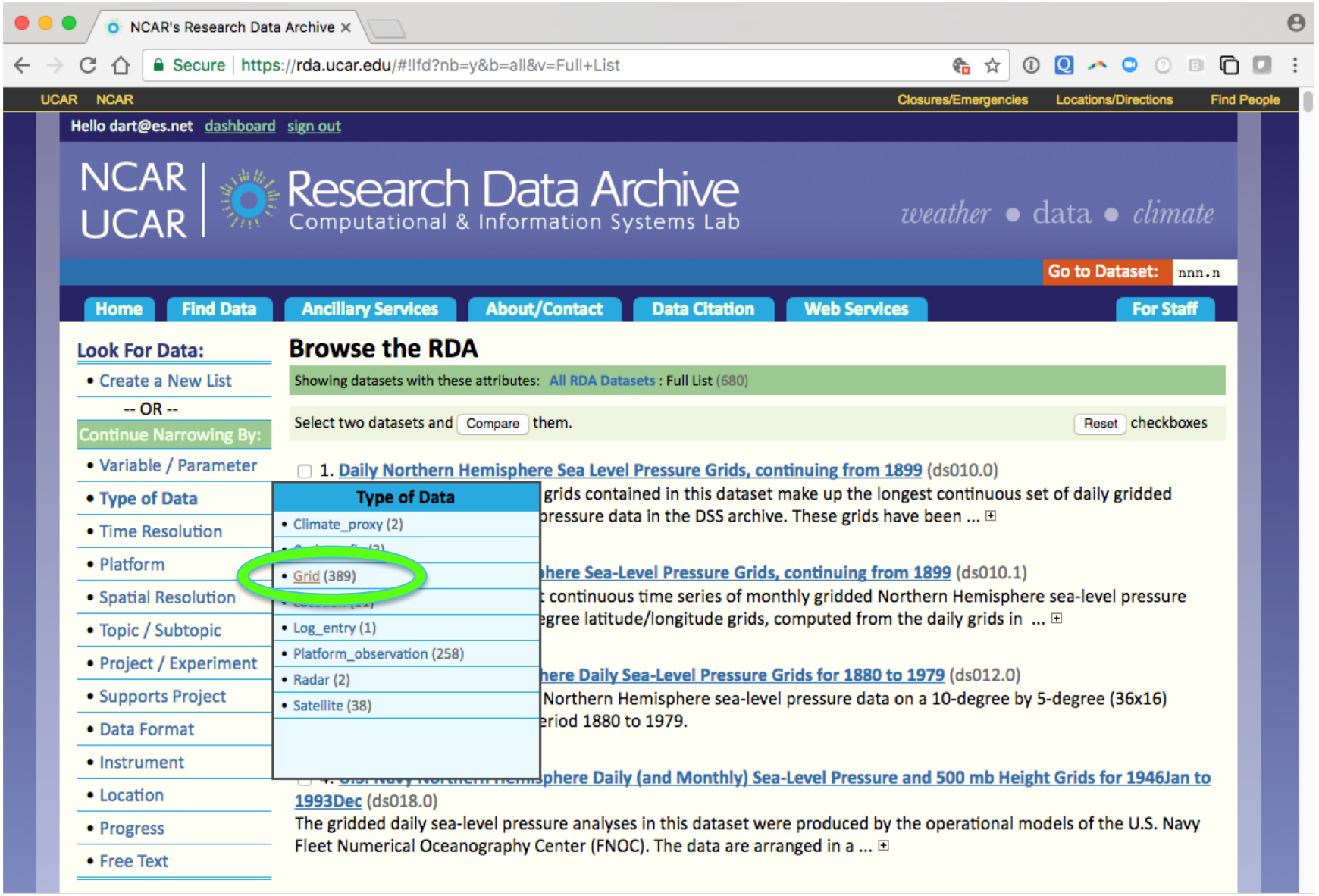
http://www.es.net/

http://fasterdata.es.net/



Extra slides – data download from portal


NCAR's Research Data Archive
Secure | https://rda.ucar.edu
UCAR NCAR
Closures/Emergencies
Locations/Directions
Find People
Hello dart@es.net dashboard sign out
NCAR UCAR
Research Data Archive
Computational & Information Systems Lab
weather • data • climate
Go to Dataset: nnn.n
Home
Find Data
Ancillary Services
About/Contact
Data Citation
Web Services
For Staff
First-time visitor to our site?
Please take a video tour of our home page
Dataset Search:
Keyword(s)
Search
Advanced Options
Look For
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
Recently Added Datasets: (within the last 6 months)
• ERA5 Reanalysis Monthly Means
• Daily Gridded North American Snowfall
• ERA5 Reanalysis
• NCAR/MOPITT Reanalysis
• GridRad - Three-Dimensional Gridded NEXRAD WSR-88D Radar Data
• CMIP 5 dataset and code for R parallelization
• Dai and Trenberth Global River Flow and Continental Discharge Dataset
• Dai Global Palmer Drought Severity Index (PDSI)
Get Help:
• Frequently Asked Questions
• Reset your password
• A-Z Site Index
• RDA Users Email List
• RDA Blog
• RDA video tutorials
• Email Us
From Our Blog:
• Accessing RDA OPeNDAP endpoints with authentication
• All RDA data transfer and processing services restored to production
• RDA Service Outage July 14-18, 2017
• RDA web services down for maintenance at 1PM MDT on May 3, 2017
More blog posts ...
GLADE Users:
Much of the RDA is directly accessible from


NCAR's Research Data Archive
Secure | https://rda.ucar.edu/#!lfd?nb=y&b=all&v=Full+List
UCAR NCAR
Closures/Emergencies
Locations/Directions
Find People
Hello dart@es.net dashboard sign out
NCAR UCAR
Research Data Archive
Computational & Information Systems Lab
weather • data • climate
Go to Dataset: nnn.n
Home
Find Data
Ancillary Services
About/Contact
Data Citation
Web Services
For Staff
Look For Data:
• Create a New List
-- OR --
Continue Narrowing By:
• Variable / Parameter
• Type of Data
• Time Resolution
• Platform
• Spatial Resolution
• Topic / Subtopic
• Project / Experiment
• Supports Project
• Data Format
• Instrument
• Location
• Progress
• Free Text
Browse the RDA
Showing datasets with these attributes: All RDA Datasets : Full List (680)
Select two datasets and Compare them.
Reset checkboxes
1. Daily Northern Hemisphere Sea Level Pressure Grids, continuing from 1899 (ds010.0)
grids contained in this dataset make up the longest continuous set of daily gridded
pressure data in the DSS archive. These grids have been ...
Type of Data
• Climate_proxy (2)
• Grid (389)
• Log_entry (1)
• Platform_observation (258)
• Radar (2)
• Satellite (38)
here Sea-Level Pressure Grids, continuing from 1899 (ds010.1)
continuous time series of monthly gridded Northern Hemisphere sea-level pressure
egree latitude/longitude grids, computed from the daily grids in ...
here Daily Sea-Level Pressure Grids for 1880 to 1979 (ds012.0)
Northern Hemisphere sea-level pressure data on a 10-degree by 5-degree (36x16)
eriod 1880 to 1979.
sphere Daily (and Monthly) Sea-Level Pressure and 500 mb Height Grids for 1946Jan to 1993Dec (ds018.0)
The gridded daily sea-level pressure analyses in this dataset were produced by the operational models of the U.S. Navy Fleet Numerical Oceanography Center (FNOC). The data are arranged in a ...

Go to Dataset: nnn.n

Home | Find Data | Ancillary Services | About/Contact | Data Citation | Web Services | For Staff

GEOS5 Global Atmosphere Forcing Data

ds313.0

For assistance, contact Chi-Fan Shih (303-497-1833).

Description | Data Access

Help with this page: RDA dataset description page video tour

Abstract: GEOS5 Atmospheric Forcing data, regridded and prepared as meteorological variables to run CESM and WRF simulations.

Temporal Range: 2004-01-02 00:00 +0000 to 2017-10-19 21:00 +0000 (Entire dataset)
Period details by dataset product

Updates: Irregularly

Variables: Surface Pressure | Upper Level Winds
Variables by dataset product

Vertical Levels: See the detailed metadata for level information

Data Types: Grid

Spatial Coverage: Longitude Range: Westernmost=180W Easternmost=180E
Latitude Range: Southernmost=90S Northernmost=90N
Detailed coverage information

Data Contributors: UCAR/NCAR/ACD | UCAR/NCAR/CGD

How to Cite This Dataset:
RIS
BibTeX

Tilmes, S.. 2016. *GEOS5 Global Atmosphere Forcing Data*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://rda.ucar.edu/datasets/ds313.0/. Accessed† dd mmm yyyy.
†Please fill in the "Accessed" date with the day, month, and year (e.g. - 5 Aug 2011) you last accessed the data from the RDA.
Bibliographic citation shown in Federation of Earth Science Information Partners (ESIP) style
Get a customized data citation

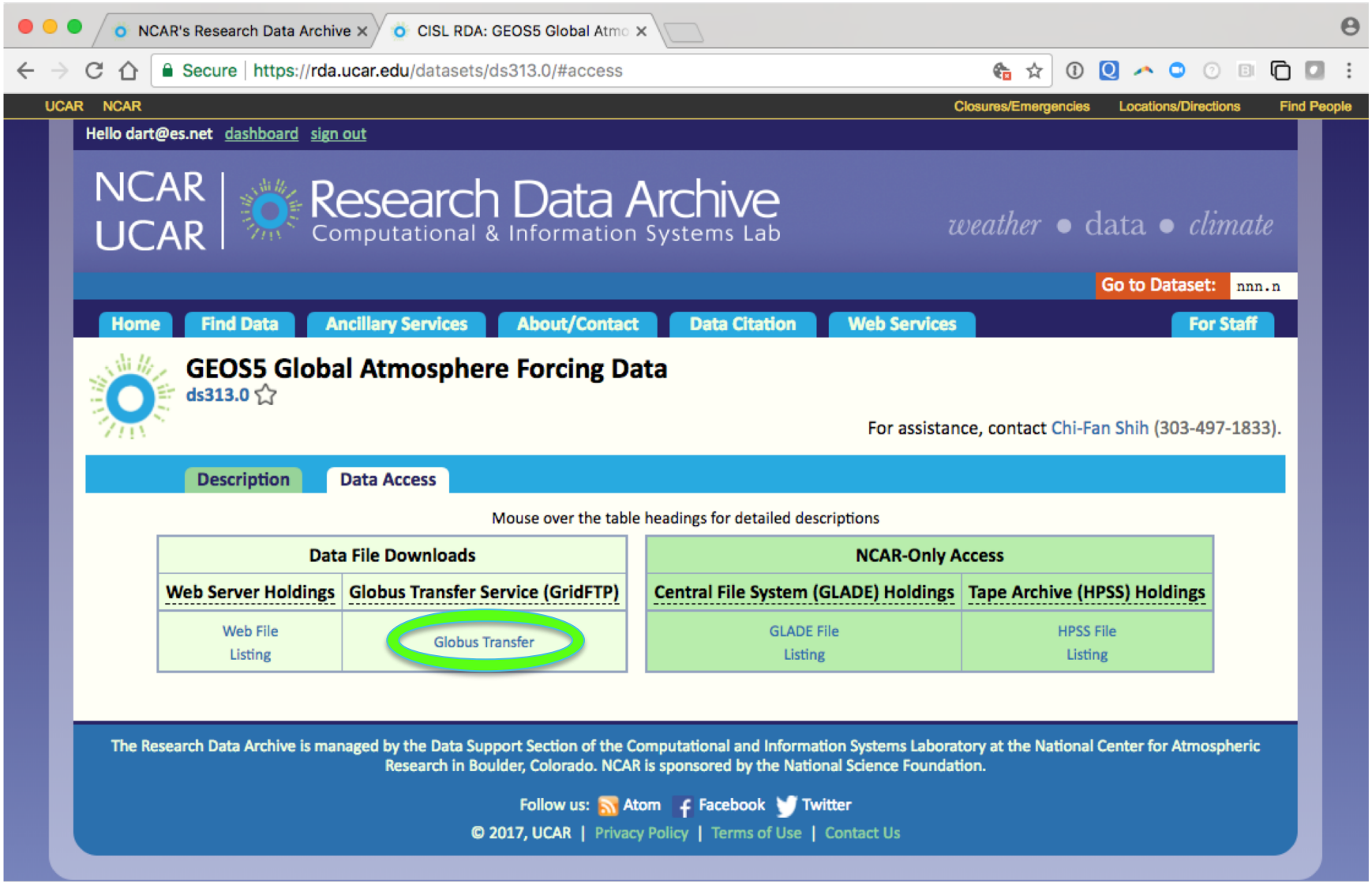
Total Volume: 449.28 GB

Data Formats: *netCDF*

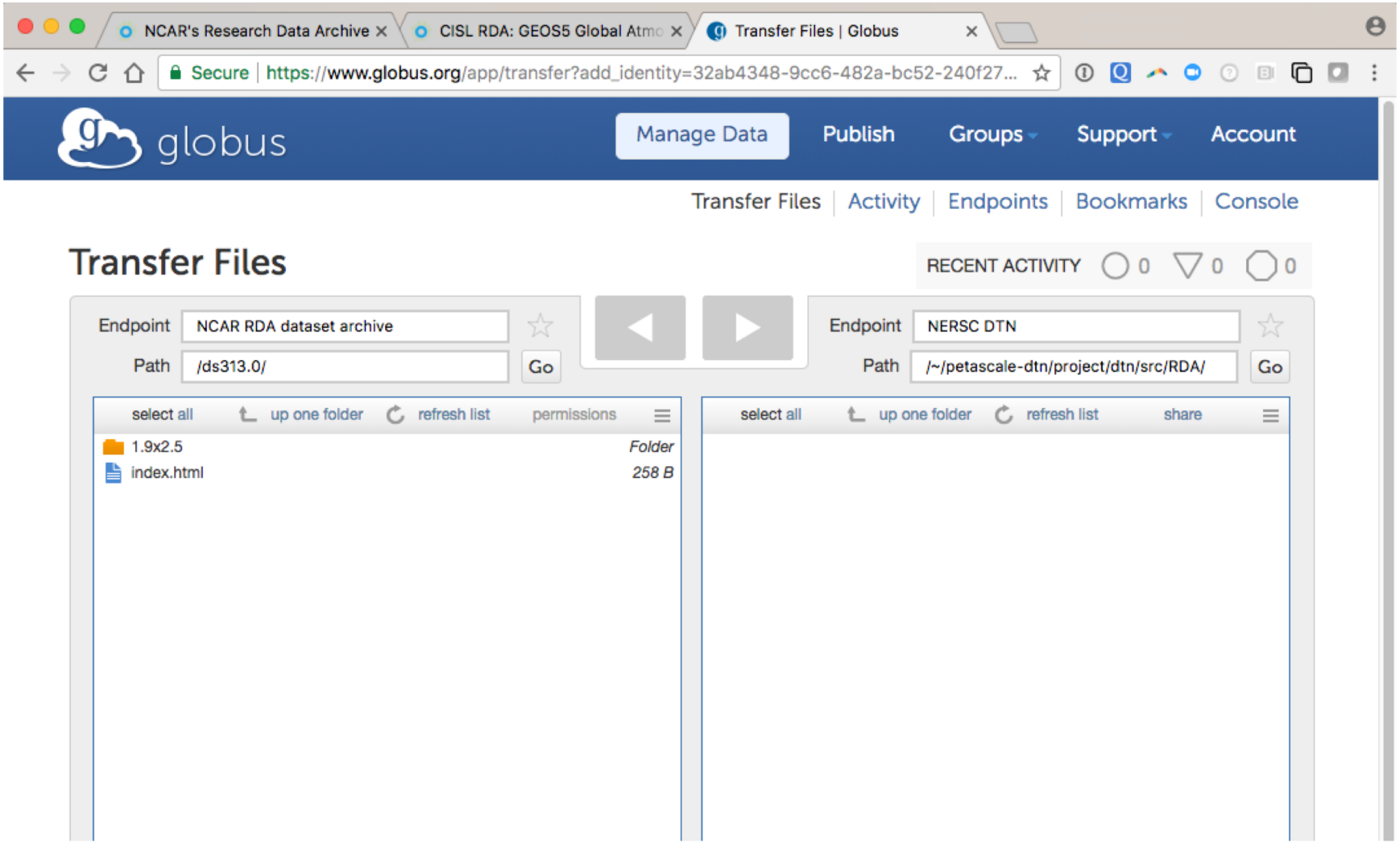
More Details: View more details for this dataset, including dataset citation, data contributors, and other detailed metadata

Data Access: Click the Data Access tab here or in the navigation bar near the top of the page

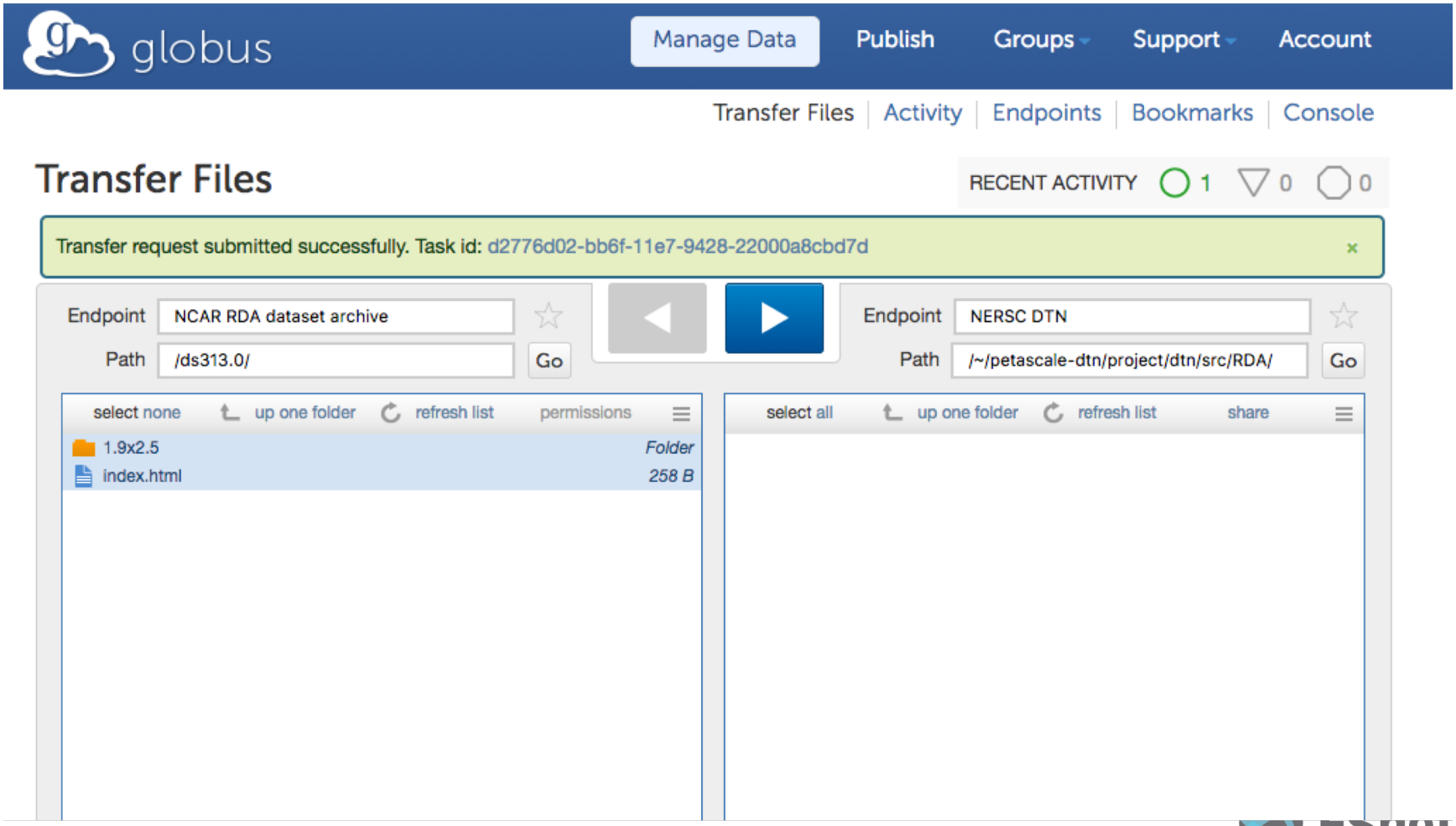
Metadata Record: Display in choose from the list format


NCAR's Research Data Archive
CISL RDA: GEOS5 Global Atmo
Secure | https://rda.ucar.edu/datasets/ds313.0/#access
UCAR NCAR
Closures/Emergencies
Locations/Directions
Find People
Hello dart@es.net dashboard sign out
NCAR UCAR
Research Data Archive
Computational & Information Systems Lab
weather • data • climate
Go to Dataset: nnn.n
Home
Find Data
Ancillary Services
About/Contact
Data Citation
Web Services
For Staff
GEOS5 Global Atmosphere Forcing Data
ds313.0
For assistance, contact Chi-Fan Shih (303-497-1833).
Description
Data Access
Mouse over the table headings for detailed descriptions
Data File Downloads
Web Server Holdings
Globus Transfer Service (GridFTP)
Web File Listing
Globus Transfer
NCAR-Only Access
Central File System (GLADE) Holdings
Tape Archive (HPSS) Holdings
GLADE File Listing
HPSS File Listing
The Research Data Archive is managed by the Data Support Section of the Computational and Information Systems Laboratory at the National Center for Atmospheric Research in Boulder, Colorado. NCAR is sponsored by the National Science Foundation.
Follow us: Atom Facebook Twitter
© 2017, UCAR | Privacy Policy | Terms of Use | Contact Us

Portal creates a Globus transfer job for us



Submit the transfer job, go about our business



Data Transfer from RDA Portal – Results