Enhancing Data Access of the Research Data Archive Through Full-Stack Development

Minh Le[1,2], Connie Chau[1,3], Riley Conroy[1], Harsha Hampapura[1], Nihanth Cherukuru[1]

[1]NSF National Center for Atmospheric Research, [2]Dickinson College, [3]Northwestern University

Background

- RDA serves global researchers with diverse technical skills
- Huge volume of data → Need for intuitive and efficient data access, both programmatic and web-based
- **Focus**: Discoverability, Usability, Transparency

Transparency: Cache Outage Map

- RDA data is now available via the Open Science Data Federation's (OSDF) network of caches
- Server status visualized with Leaflet.js (green = OK, red = down)
- Distance to nearest cache calculated using Haversine algorithm
- Django backend fetches and parses OSDF API data

Outage Map

This interactive map displays cache servers that are currently down or have unknown status. If you grant location access, the map will also show you the nearest working cache server.

- Down Cache
- Unknown Status
- Nearest Working Cache

Distance Information:

Nearest Down Cache:	Nearest Working Cache:
PSU-LIGO-CACHE - 1364.50 km away	NCAR_NRP_CACHE - 11.23 km away

Down Caches (1 total):	Unknown Status Caches (1 total):
• PSU-LIGO-CACHE	• CIT_LIGO_STASHCACHE

Discoverability: Documented APIs

- Developed RESTful API endpoints for dataset and data access metadata
- Auto-generated OpenAPI docs with drf-spectacular
- Interactive documentation via Swagger & ReDoc

```
path(r'datasets/<dsid>/abstract/', views.get_abstract),
path(r'datasets/<dsid>/acknowledgement/', views.get_acknowledgement ),
path(r'datasets/<dsid>/temporal/', views.get_temporal ),
path(r'datasets/<dsid>/variables/', views.get_variables ),
path(r'datasets/<dsid>/publications/', views.get_publications ),
path(r'datasets/<dsid>/data_license/', views.get_data_license ),
path(r'datasets/<dsid>/data_types/', views.get_data_types ),
path(r'datasets/<dsid>/data_formats/', views.get_data_formats ),
path(r'datasets/<dsid>/spatial_coverage/', views.get_spatial_coverage ),
path(r'datasets/<dsid>/contributors/', views.get_contributors ),
path(r'datasets/<dsid>/volume/', views.get_total_volume ),
path(r'datasets/<dsid>/resources/', views.get_related_resources ),
path(r'datasets/<dsid>/related_datasets/', views.get_related_datasets ),
path(r'datasets/', views.get_all_datasets )
```

```
HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

{
    "dataset_id": "d010000",
    "summary": {
        "has_groups": true,
        "total_groups": 3,
        "available_access_methods": [
            "web_files",
            "globus",
            "glade"
        ],
        "web_data_available": "0"
    },
    "access_methods": {
        "web_files": {
            "available": true,
            "url": "/datasets/d010000/filelist/",
            "description": "Direct web file access"
        },
        "globus": {
            "available": true,
            "url": "https://app.globus.org/file-manager?origin_id=be4aa6a8-9e35-11eb-8a8e-d7bd56a48c8a&origin_path=%2Fd010000%2F",
            "description": "Globus file transfer"
        },
        "glade": {
            "available": true,
            "url": "/datasets/d010000/filelist/",
            "description": "GLADE file system access"
        }
    },
    "groups": [
        {
            "index": 101,
            "title": "**** primaries on virtual volume CTDS0000 ****",
            "file_counts": {
                "web_accessible": 0,
                "total": 0
            },
            "access_methods": {}
        },
        {
            "index": 102,
            "title": "*****  monthly mean slp",
            "file_counts": {
                "web_accessible": 1,
                "total": 0
            },
            "access_methods": {}
        },
        {
            "index": 103,
            "title": "***** 500mb z, data ordered all years each month *****",
            "file_counts": {
                "web_accessible": 1,
                "total": 0
            },
            "access_methods": {}
        }
    ]
}
```

Usability: Dataset Page Redesign

DATA DESCRIPTION	DATA FILE DOWNLOADS		CUSTOMIZABLE DATA REQUESTS	OTHER ACCESS METHODS	NCAR-ONLY ACCESS
	Web Server Holdings — Powered by OSDF	Globus Transfer Service (GridFTP)	Subsetting	THREDDS Data Server	Central File System (GLADE) Holdings
UNION OF AVAILABLE PRODUCTS	Web File Listing	Globus Transfer		TDS Access	GLADE File Listing
Yearly Time Series 3-Hourly Analysis Fields (Gaussian 1-254)	Web File Listing		Get a Subset (login required)	TDS Access	GLADE File Listing
Yearly Time Series 3-Hourly First Guess Forecast Fields (Gaussian 1-254)	Web File Listing		Get a Subset (login required)	TDS Access	GLADE File Listing
Yearly Time Series 6-Hourly Analysis Fields (Gaussian 1-254)	Web File Listing		Get a Subset (login required)	TDS Access	GLADE File Listing
Observation Feedback	Web File Listing				GLADE File Listing
Invariant variables	Web File Listing				GLADE File Listing

Current Interface

- Table view with many access methods shown at once
- Redundancy across rows and columns
-> Need to eliminate redundancy and provide simplified starting points

Data Access

Choose your preferred method to access the dataset

- DOWNLOAD FULL FILES
- REQUEST A SUBSET

Additional Services

Alternative access methods and specialized tools

- THREDDS
- GLOBUS

Redesigned landing interface with clear download options

UI Prototypes

- Web File Listing | Subsetting | THREDDS Access | NCAR Access
- Web Server Holdings
- BROWSE ALL FILES

  - Yearly Time Series 3-Hourly Analysis Fields (Gaussian 1-254) — BROWSE FILES
  - Yearly Time Series 3-Hourly First Guess Forecast Fields (Gaussian 1-254) — BROWSE FILES
  - Yearly Time Series 6-Hourly Analysis Fields (Gaussian 1-254) — BROWSE FILES

Tab-based navigation

Temporal 1805-12-31 18:00 to 2016-01-01 00:00 (Initialization Time)

- TIME RANGE
- PARAMETER SELECTION
- CONTINUE
- RESET

Parameter Selection You must choose at least one parameter

Search parameters...

Loading parameters...

- 10 metre U wind component
- Plant canopy surface water
- 10 metre V wind component
- Potential evaporation rate
- 2 metre relative humidity
- Potential temperature

† This dataset has data files that can be subsetted using csls. ☐ If you make one or more parameter selections, you will get more information about this in your list of data files.

New faceted browse with:
- View-switching buttons
- Dynamic selection bar
- Integrated search within filters

Conclusion

This project laid the foundation for a more intuitive user experience on the RDA platform. However, improving usability is an iterative process that will require:
- Continuous user testing and feedback collection
- Further refinements to the UI/UX
- Expansion of API functionality to support broader use cases