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

Minh Le
SIParCS Intern
Mentors: Riley Conroy, Harsha Hampapura, Nihanth Cherukuru

July 29, 2025

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Background

- RDA serves a diverse user base with different levels of technical background
- Tens of thousands of users in 206 countries
- 7.5 PB of data holdings
- Serve ~4PB/year
- Tech stack: Django, HTML, Bootstrap, JavaScript, PostgreSQL

Engineering Contributions

1. **Discoverability – API Endpoints**
   Exposed metadata via documented APIs

2. **Usability – Dataset Page Redesign**
   Simplified UI

3. **Transparency – Cache Outage Map**
   Helps users understand data access issues

Engineering Contributions

1

Discoverability – API Endpoints

Exposed metadata via documented APIs

ABSTRACT:
To better understand the rainfall climatology and its impacts on hydrological cycle over the Hawaiian Islands under historical and future climates, regional climate simulations over the main Hawaiian islands have been conducted for two 10-year periods using the Weather Research and Forecasting (WRF) model in a configuration of two nested domains. The historical 10-year simulation was driven by the ERA-Interim global reanalysis data and observed sea surface temperature from October. 2002 to September 2012 (historical simulation). A high-resolution vertical coordinate was employed to better resolve the trade wind inversion (TWI). Results show that the historical simulation reproduces the mean surface temperature, relative humidity and winds with low biases (+/- 1 degree C, +/- 4 % and +/- 1 m/s, respectively) and high spatial correlations (correlation coefficient over 0.80). Additionally, for the historical simulation WRF accurately reproduced aggregated daily and hourly rainfall probability density functions (PDFs) and rainfall spatial-temporal distributions, likely because WRF captured the TWI properties well. The historical simulation outputs are available at hourly resolution for near surface (2-dimensional) fields and for the 3-dimensional atmosphere.

TEMPORAL RANGE:
2002-10-01 00:00 +0000 to 2012-10-01 00:00 +0000

VARIABLES:

Air Temperature	Downwelling Longwave Radiation	Downwelling Shortwave Radiation	Precipitation Amount
Upwelling Longwave Radiation	U/V Wind Components	Water Vapor Mixing Ratio Profiles	

DATA TYPES:
Grid

DATA CONTRIBUTORS:
UCAR/NCAR/RAL

TOTAL VOLUME:
62.5 TB (Entire dataset)

VOLUME DETAILS BY DATASET PRODUCT

DATA FORMATS:
HDF5/NetCDF4

METADATA RECORD:
Display in choose from the list format

DATA LICENSE:
This work is licensed under a Creative Commons Attribution 4.0 International License.

Dataset description page

- Limited machine-readable format available
- Requires scraping

Metadata API

Designed a set of RESTful endpoints using Django REST Framework

```
path(r'datasets/<dsid>/abstract/', views.get_abstract),
path(r'datasets/<dsid>/acknowledgment/', 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 )
```

For datasets


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

{
    "dataset_id": "d018000",
    "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/d018000/filelist/",
            "description": "Direct web file access"
        },
        "globus": {
            "available": true,
            "url": "https://app.globus.org/file-manager?origin_id=be4aa6a8-9e35-11eb-8a8e-d70d98a40c8d&origin_path=%2Fd018000%2F",
            "description": "Globus file transfer"
        },
        "glade": {
            "available": true,
            "url": "/datasets/d018000/filelist/",
            "description": "GLADE file system access"
        }
    },
    "groups": [
        {
            "index": 101,
            "title": "**** primaries are on virtual volume CTDS0006 ****",
            "file_counts": {
                "web_accessible": 6,
                "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": {}
        }
    ]
}
```


For data access

API Documentation

drf-spectacular

```
path('api/schema/', SpectacularAPIView.as_view(patterns=[path('api/', include('api.urls'))]), name='schema'),
path('api/schema/documentation/', SpectacularSwaggerView.as_view(url_name='schema'), name='swagger-ui'),
path('api/schema/redoc/', SpectacularRedocView.as_view(url_name='schema'), name='redoc'),
```

OpenAPI Schema Construction

- Generates a full OpenAPI 3.0 JSON/YAML schema

Swagger / ReDoc Integration

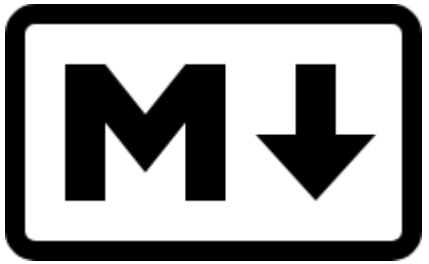
Comes with built-in views to serve:

- **Swagger UI**: interactive explorer and tester
- **ReDoc**: clean, scrollable documentation interface

Why not other tools?

Postman

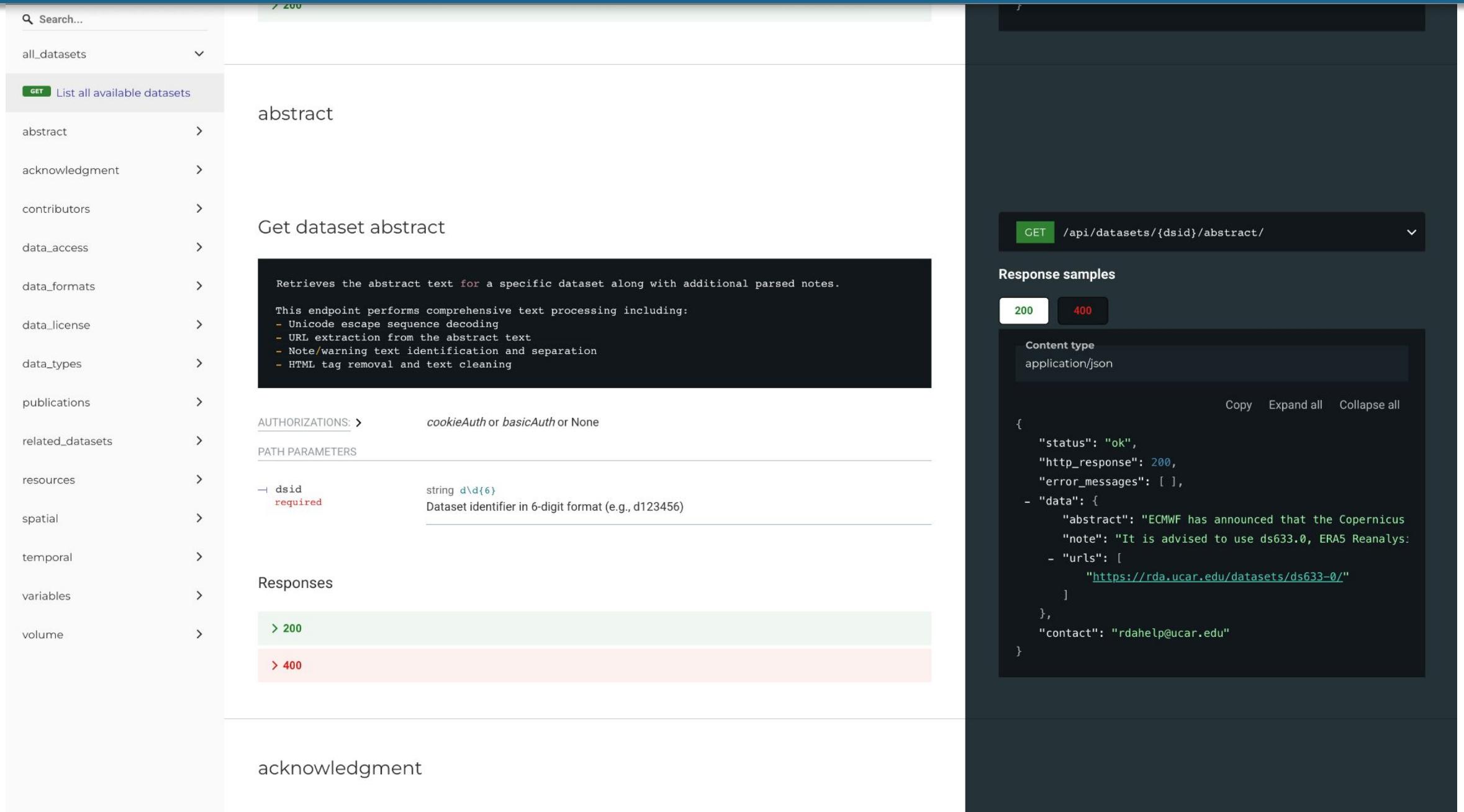
- Docs are not directly synced with code
- Better suited for testing than for public-facing docs



Manual docs (e.g, Markdown)

- Easy to go out of sync with implementation
- High maintenance, not interactive

ReDoc UI


Search...
all_datasets
GET List all available datasets
abstract
acknowledgment
contributors
data_access
data_formats
data_license
data_types
publications
related_datasets
resources
spatial
temporal
variables
volume
abstract
Get dataset abstract
Retrieves the abstract text for a specific dataset along with additional parsed notes.
This endpoint performs comprehensive text processing including:
- Unicode escape sequence decoding
- URL extraction from the abstract text
- Note/warning text identification and separation
- HTML tag removal and text cleaning
AUTHORIZATIONS:
cookieAuth or basicAuth or None
PATH PARAMETERS
dsid
required
string d\d{6}
Dataset identifier in 6-digit format (e.g., d123456)
Responses
200
400
GET /api/datasets/{dsid}/abstract/
Response samples
200
400
Content type
application/json
Copy
Expand all
Collapse all
"status": "ok",
"http_response": 200,
"error_messages": [ ],
"data": {
"abstract": "ECMWF has announced that the Copernicus
"note": "It is advised to use ds633.0, ERA5 Reanalys
"urls": [
"https://rda.ucar.edu/datasets/ds633-0/"
],
"contact": "rdahelp@ucar.edu"
acknowledgment

Swagger UI

- Swagger is widely adopted
- Allows live interaction:
  - Make API requests directly in the browser
  - View responses, test parameters, and see errors in real-time

RDA Datasets API 1.0.0 OAS 3.0

/api/schema/

Research Data Archive API for accessing and discovering scientific datasets.

Authorize

all_datasets

GET /api/datasets/ List all available datasets

abstract

GET /api/datasets/{dsid}/abstract/ Get dataset abstract

acknowledgment

GET /api/datasets/{dsid}/acknowledgment/ Get dataset acknowledgement text

contributors

GET /api/datasets/{dsid}/contributors/ Get dataset contributors

data_access

GET /api/datasets/{dsid}/data_access/root

data_formats

GET /api/datasets/{dsid}/data_formats/ Get dataset data formats

Engineering Contributions

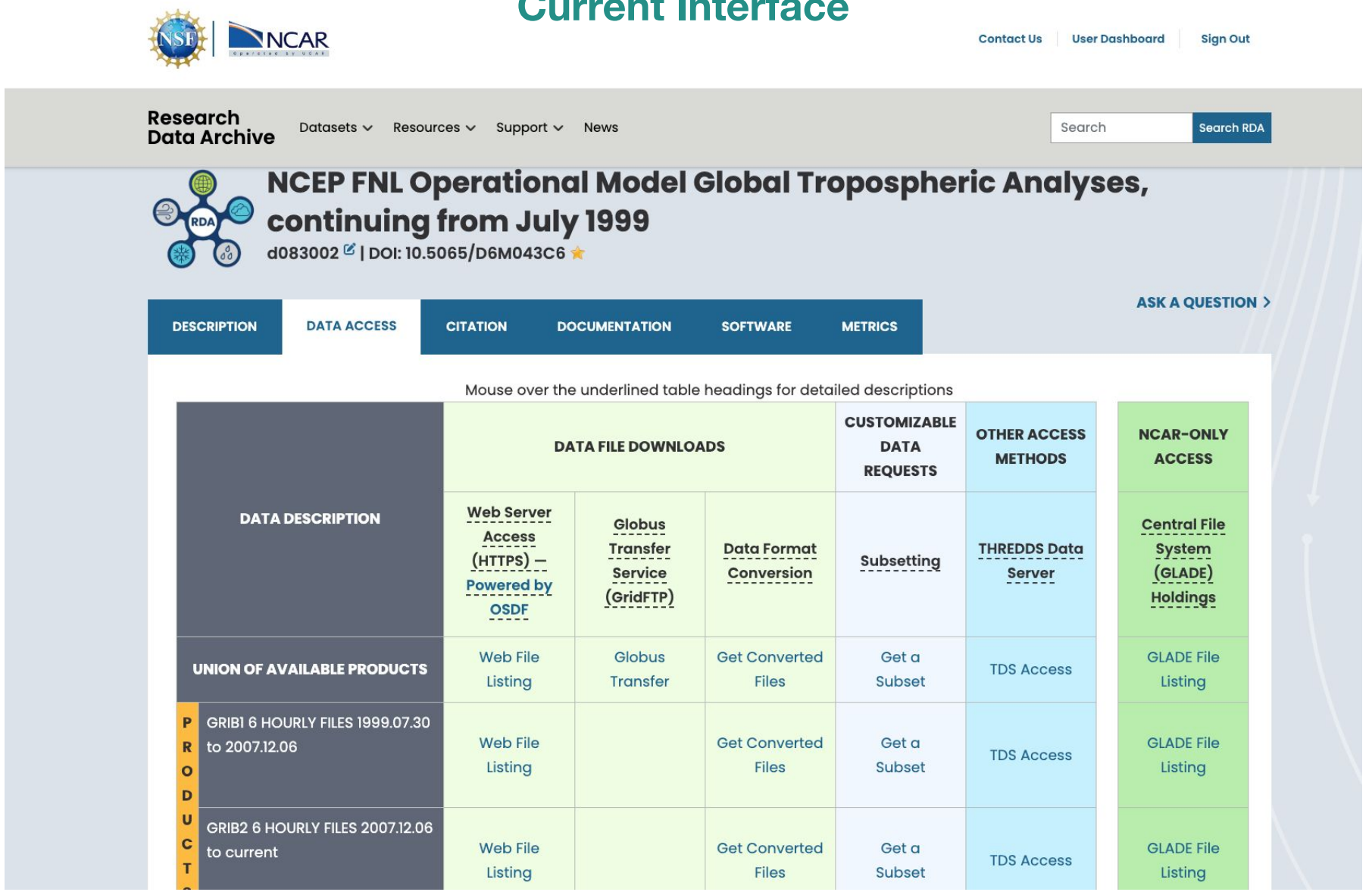
2

Usability – Dataset Page Redesign

Simplified UI

UI Prototypes

Current Interface



Contact Us | User Dashboard | Sign Out

Research Data Archive — Datasets — Resources — Support — News — Search — Search RDA

NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999

d083002 | DOI: 10.5065/D6M043C6

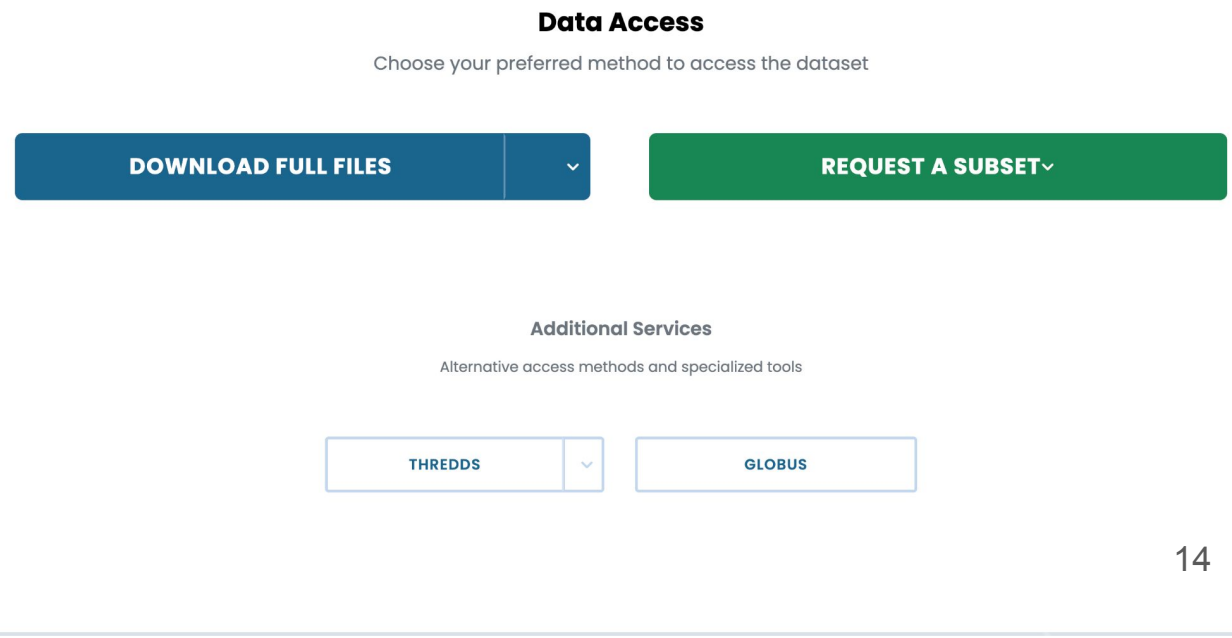
ASK A QUESTION >

DESCRIPTION | DATA ACCESS | CITATION | DOCUMENTATION | SOFTWARE | METRICS

Mouse over the underlined table headings for detailed descriptions

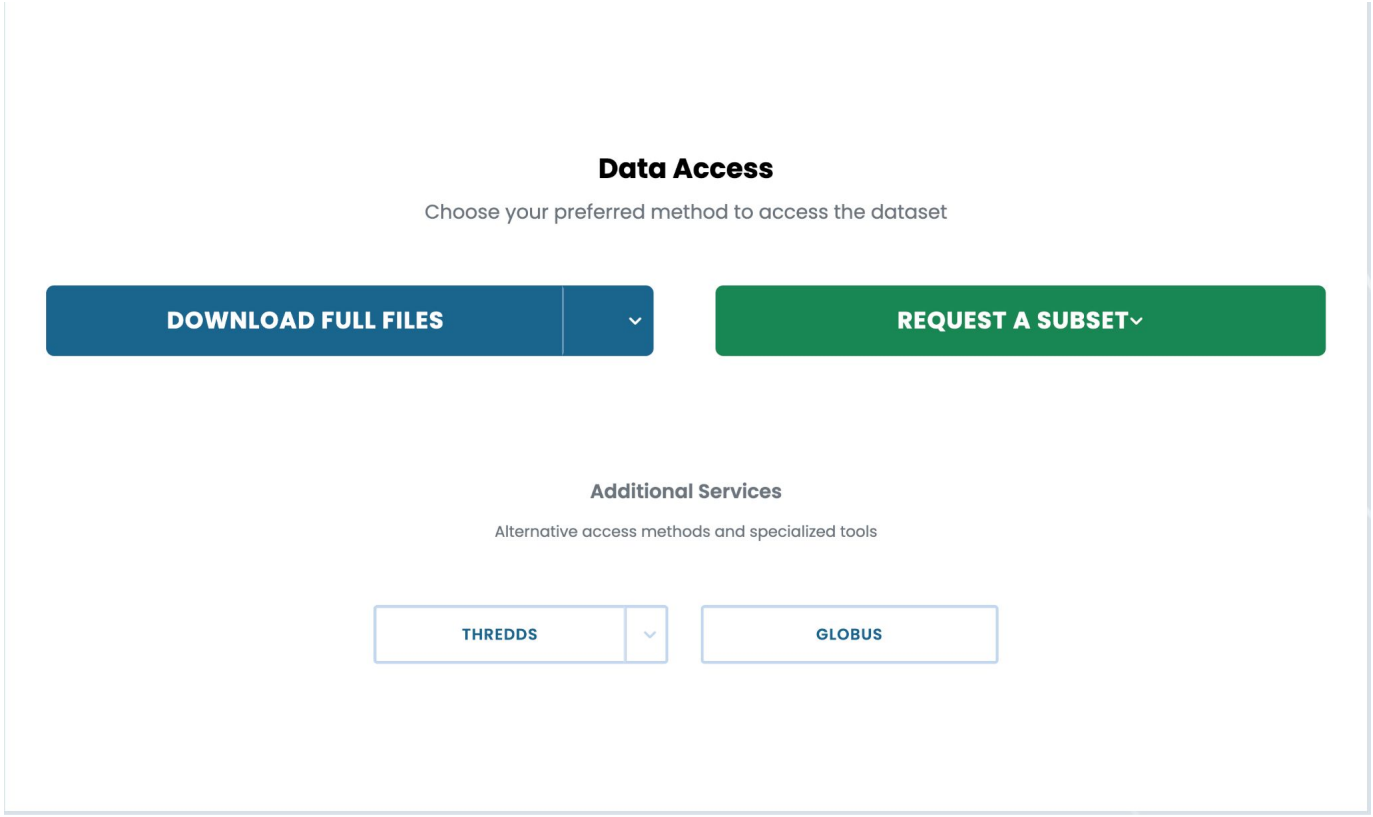
DATA DESCRIPTION	DATA FILE DOWNLOADS			CUSTOMIZABLE DATA REQUESTS	OTHER ACCESS METHODS	NCAR-ONLY ACCESS
	Web Server Access (HTTPS) — Powered by OSDF	Globus Transfer Service (GridFTP)	Data Format Conversion	Subsetting	THREDDS Data Server	Central File System (GLADE) Holdings
UNION OF AVAILABLE PRODUCTS	Web File Listing	Globus Transfer	Get Converted Files	Get a Subset	TDS Access	GLADE File Listing
PRODUCT: GRIB1 6 HOURLY FILES 1999.07.30 to 2007.12.06	Web File Listing		Get Converted Files	Get a Subset	TDS Access	GLADE File Listing
PRODUCT: GRIB2 6 HOURLY FILES 2007.12.06 to current	Web File Listing		Get Converted Files	Get a Subset	TDS Access	GLADE File Listing

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
PRODUCTS	Yearly Time Series 3-Hourly Analysis Fields (Gaussian T-254)	Web File Listing		Get a Subset *(login required)*	TDS Access	GLADE File Listing
	Yearly Time Series 3-Hourly First Guess Forecast Fields (Gaussian T-254)	Web File Listing		Get a Subset *(login required)*	TDS Access	GLADE File Listing
	Yearly Time Series 6-Hourly Analysis Fields (Gaussian T-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



UI Prototypes

Simplified entry points



Guide users to choose between:

- Full file download
- Custom subset request
- Other services

Data Access

Choose your preferred method to access the dataset

DOWNLOAD FULL FILES

REQUEST A SUBSET

Additional Services

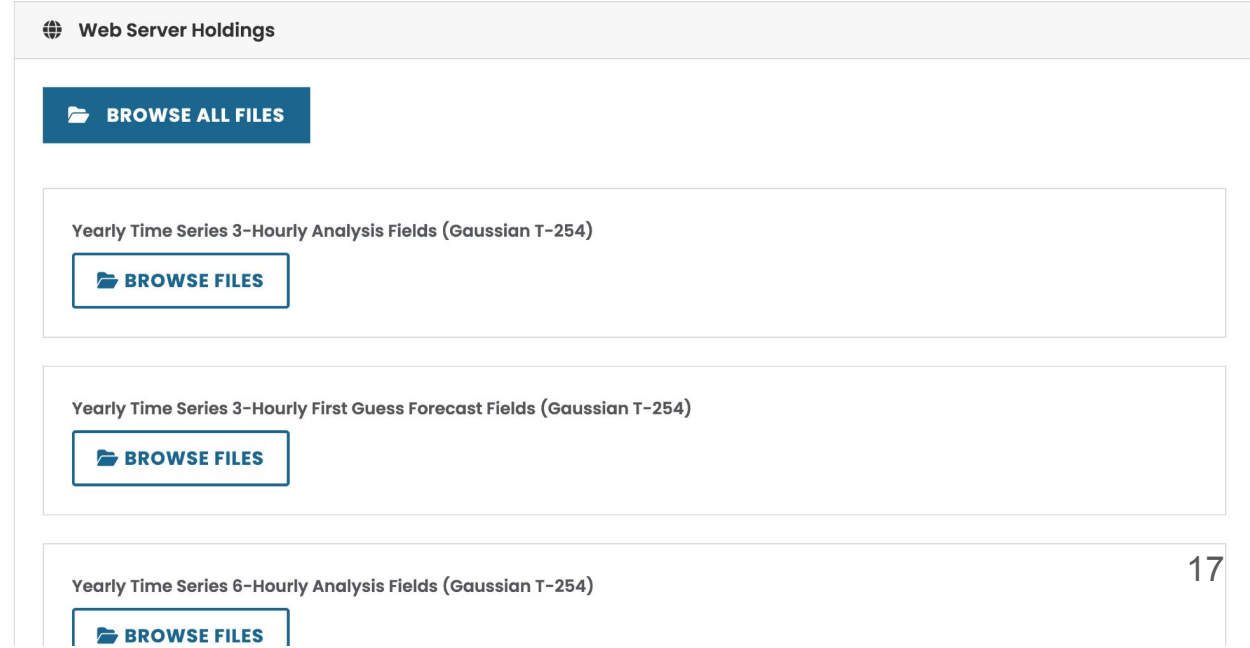
Alternative access methods and

High-speed, reliable data transfer using Globus for large datasets

THREDDS

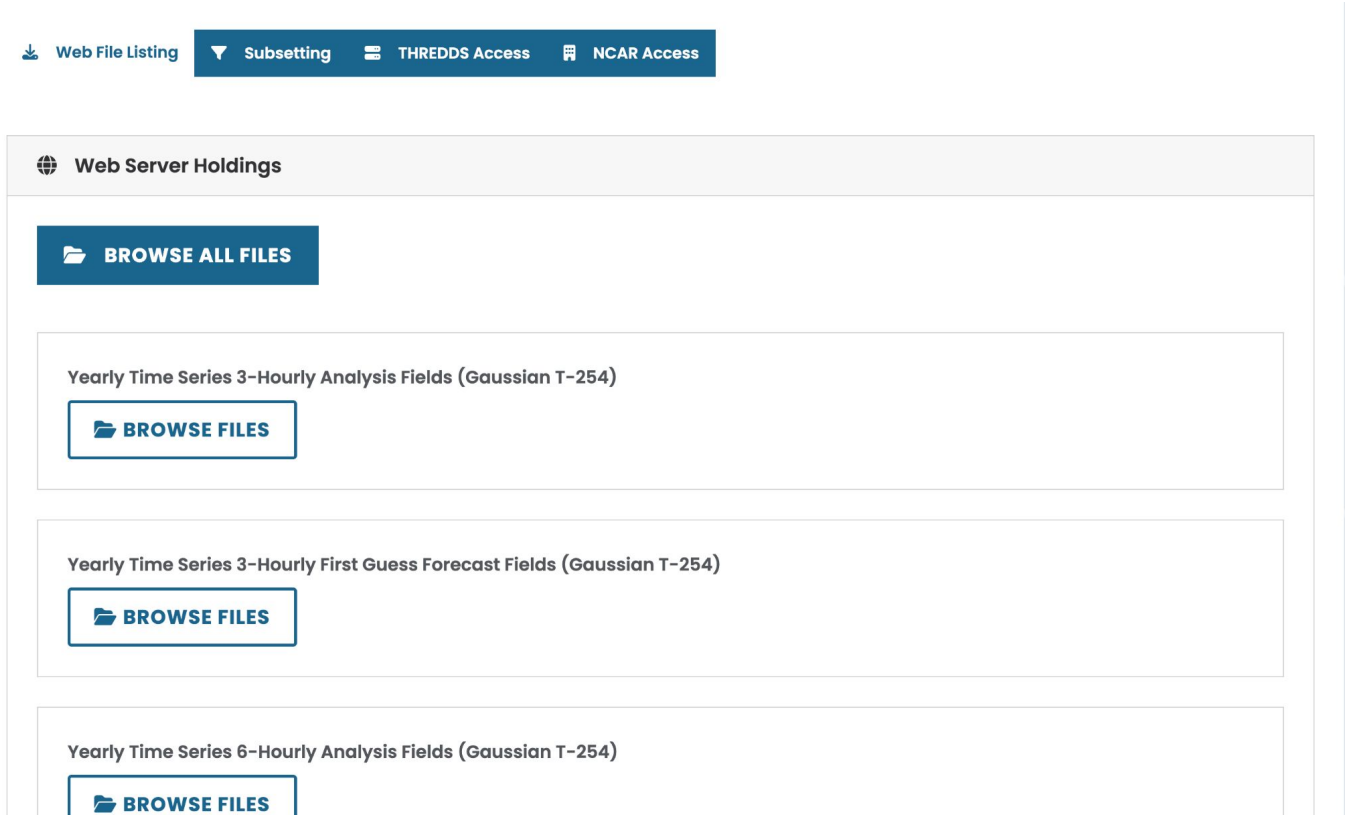
GLOBUS

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
PRODUCTS: Yearly Time Series 3-Hourly Analysis Fields (Gaussian T-254)	Web File Listing		Get a Subset *(login required)*	TDS Access	GLADE File Listing
PRODUCTS: Yearly Time Series 3-Hourly First Guess Forecast Fields (Gaussian T-254)	Web File Listing		Get a Subset *(login required)*	TDS Access	GLADE File Listing
PRODUCTS: Yearly Time Series 6-Hourly Analysis Fields (Gaussian T-254)	Web File Listing		Get a Subset *(login required)*	TDS Access	GLADE File Listing
PRODUCTS: Observation Feedback	Web File Listing				GLADE File Listing
PRODUCTS: Invariant variables	Web File Listing				GLADE File Listing



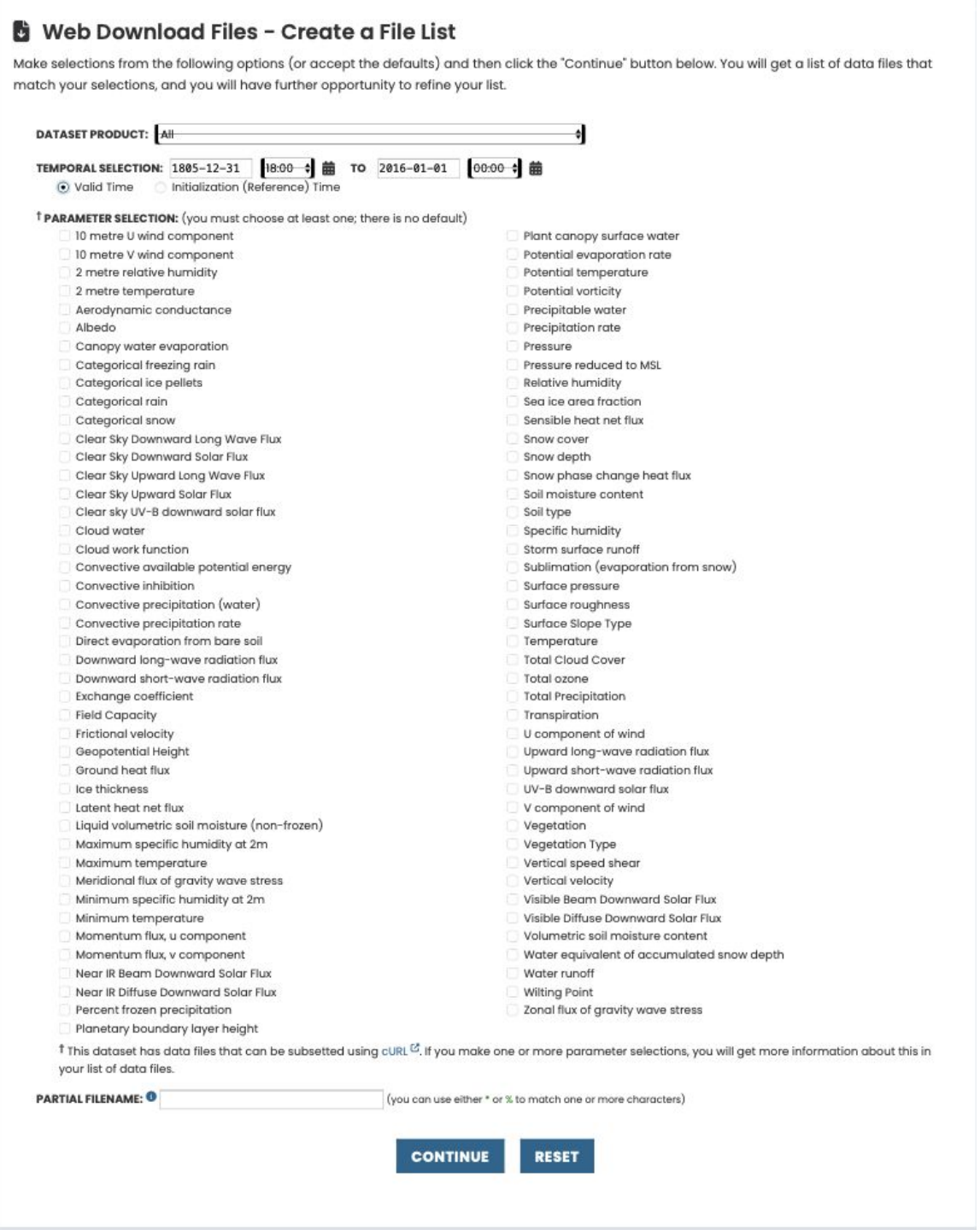
UI Prototypes

Tab-based access



Introduced tab-based navigation for organizing access options:

- Separation of access methods
- Users land directly in a browsable file structure tailored to their selected method



Web Download Files - Create a File List

Make selections from the following options (or accept the defaults) and then click the "Continue" button below. You will get a list of data files that match your selections, and you will have further opportunity to refine your list.

DATASET PRODUCT: All

TEMPORAL SELECTION: 1805-12-31 18:00 TO 2016-01-01 00:00

Valid Time Initialization (Reference) Time

† **PARAMETER SELECTION:** (you must choose at least one; there is no default)

- 10 metre U wind component
- 10 metre V wind component
- 2 metre relative humidity
- 2 metre temperature
- Aerodynamic conductance
- Albedo
- Canopy water evaporation
- Categorical freezing rain
- Categorical ice pellets
- Categorical rain
- Categorical snow
- Clear Sky Downward Long Wave Flux
- Clear Sky Downward Solar Flux
- Clear Sky Upward Long Wave Flux
- Clear Sky Upward Solar Flux
- Clear sky UV-B downward solar flux
- Cloud water
- Cloud work function
- Convective available potential energy
- Convective inhibition
- Convective precipitation (water)
- Convective precipitation rate
- Direct evaporation from bare soil
- Downward long-wave radiation flux
- Downward short-wave radiation flux
- Exchange coefficient
- Field Capacity
- Frictional velocity
- Geopotential Height
- Ground heat flux
- Ice thickness
- Latent heat net flux
- Liquid volumetric soil moisture (non-frozen)
- Maximum specific humidity at 2m
- Maximum temperature
- Meridional flux of gravity wave stress
- Minimum specific humidity at 2m
- Minimum temperature
- Momentum flux, u component
- Momentum flux, v component
- Near IR Beam Downward Solar Flux
- Near IR Diffuse Downward Solar Flux
- Percent frozen precipitation
- Planetary boundary layer height
- Plant canopy surface water
- Potential evaporation rate
- Potential temperature
- Potential vorticity
- Precipitable water
- Precipitation rate
- Pressure
- Pressure reduced to MSL
- Relative humidity
- Sea ice area fraction
- Sensible heat net flux
- Snow cover
- Snow depth
- Snow phase change heat flux
- Soil moisture content
- Soil type
- Specific humidity
- Storm surface runoff
- Sublimation (evaporation from snow)
- Surface pressure
- Surface roughness
- Surface Slope Type
- Temperature
- Total Cloud Cover
- Total ozone
- Total Precipitation
- Transpiration
- U component of wind
- Upward long-wave radiation flux
- Upward short-wave radiation flux
- UV-B downward solar flux
- V component of wind
- Vegetation
- Vegetation Type
- Vertical speed shear
- Vertical velocity
- Visible Beam Downward Solar Flux
- Visible Diffuse Downward Solar Flux
- Volumetric soil moisture content
- Water equivalent of accumulated snow depth
- Water runoff
- Wilting Point
- Zonal flux of gravity wave stress

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

PARTIAL FILENAME: (you can use either * or % to match one or more characters)

CONTINUE RESET

Current faceted browse view

- Allows interactive filtering of data files by time range and parameters
- Returns a list of matching whole files

Web Download Files - Create a File List

Make selections from the following options (or accept the defaults) and then click the "Continue" button below. You will get a list of data files that match your selections, and you will have further opportunity to refine your list.

DATASET PRODUCT: All

TEMPORAL SELECTION: 1805-12-31 18:00 TO 2016-01-01 00:00
Valid Time / Initialization (Reference) Time

† **PARAMETER SELECTION:** (you must choose at least one; there is no default)

- 10 metre U wind component
- 10 metre V wind component
- 2 metre relative humidity
- 2 metre temperature
- Aerodynamic conductance
- Albedo
- Canopy water evaporation
- Categorical freezing rain
- Categorical ice pellets
- Categorical rain
- Categorical snow
- Clear Sky Downward Long Wave Flux
- Clear Sky Downward Solar Flux
- Clear Sky Upward Long Wave Flux
- Clear Sky Upward Solar Flux
- Clear sky UV-B downward solar flux
- Cloud water
- Cloud work function
- Convective available potential energy
- Convective inhibition
- Convective precipitation (water)
- Convective precipitation rate
- Direct evaporation from bare soil
- Downward long-wave radiation flux
- Downward short-wave radiation flux
- Exchange coefficient
- Field Capacity
- Frictional velocity
- Geopotential Height
- Ground heat flux
- Ice thickness
- Latent heat net flux
- Liquid volumetric soil moisture (non-frozen)
- Maximum specific humidity at 2m
- Maximum temperature
- Meridional flux of gravity wave stress
- Minimum specific humidity at 2m
- Minimum temperature
- Momentum flux, u component
- Momentum flux, v component
- Near IR Beam Downward Solar Flux
- Near IR Diffuse Downward Solar Flux
- Percent frozen precipitation
- Planetary boundary layer height
- Plant canopy surface water
- Potential evaporation rate
- Potential temperature
- Potential vorticity
- Precipitable water
- Precipitation rate
- Pressure
- Pressure reduced to MSL
- Relative humidity
- Sea ice area fraction
- Sensible heat net flux
- Snow cover
- Snow depth
- Snow phase change heat flux
- Soil moisture content
- Soil type
- Specific humidity
- Storm surface runoff
- Sublimation (evaporation from snow)
- Surface pressure
- Surface roughness
- Surface Slope Type
- Temperature
- Total Cloud Cover
- Total ozone
- Total Precipitation
- Transpiration
- U component of wind
- Upward long-wave radiation flux
- Upward short-wave radiation flux
- UV-B downward solar flux
- V component of wind
- Vegetation
- Vegetation Type
- Vertical speed shear
- Vertical velocity
- Visible Beam Downward Solar Flux
- Visible Diffuse Downward Solar Flux
- Volumetric soil moisture content
- Water equivalent of accumulated snow depth
- Water runoff
- Wilting Point
- Zonal flux of gravity wave stress

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

PARTIAL FILENAME: (you can use either * or % to match one or more characters)

CONTINUE RESET

Download Full or Partial Files

Make your selections below

TIME RANGE
PARAMETER SELECTION
CONTINUE
RESET

Time Range

Start Date & Time: 1805-12-31 18:00
End Date & Time: 2016-01-01 00:00

Date Type
Initialization (Reference) Time
Valid Time

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

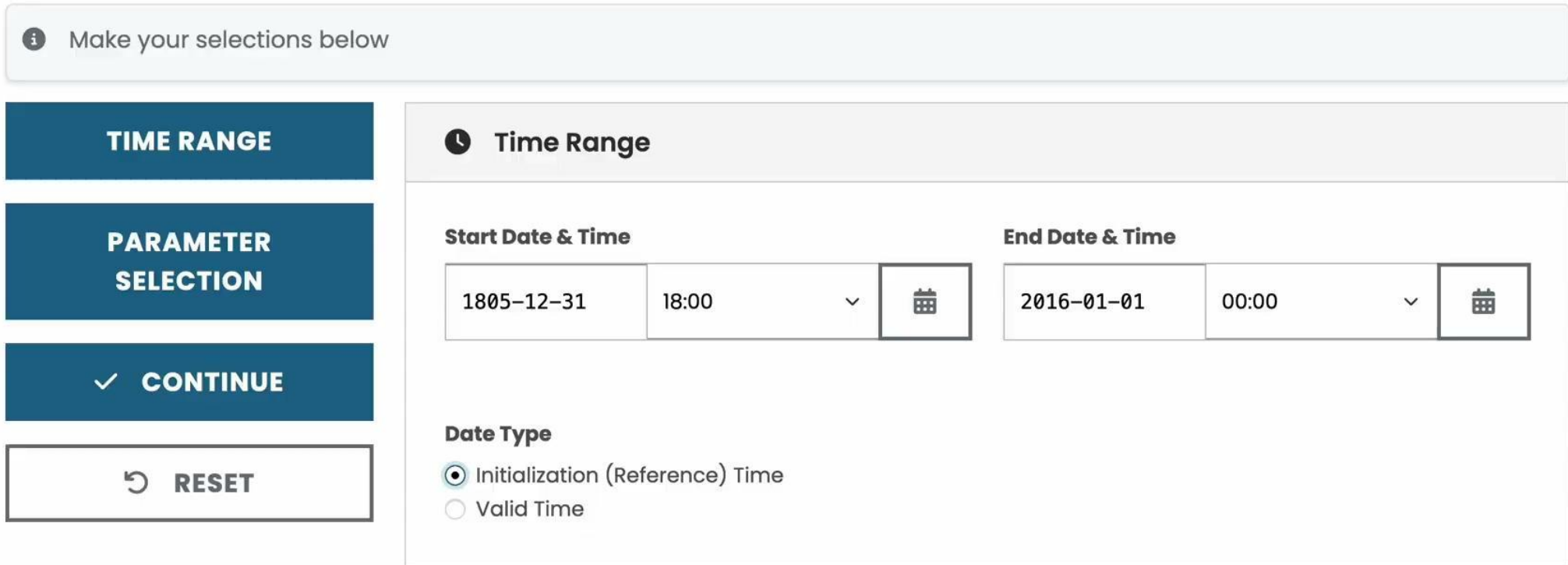
Search parameters...

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 cURL. If you make one or more parameter selections, you will get more information about this in your list of data files.

Download Full or Partial Files



Engineering Contributions

3

Transparency – Cache Outage Map

Helps users understand data access issues

Open Science Data Federation

Outage Map

Front end

Key Features:

- Visual indicators – red markers for down caches, green for healthy
- Distance calculation shows nearest available/unavailable servers
- Leaflet.js for interactive mapping
- Haversine algorithm calculates distances

Back end

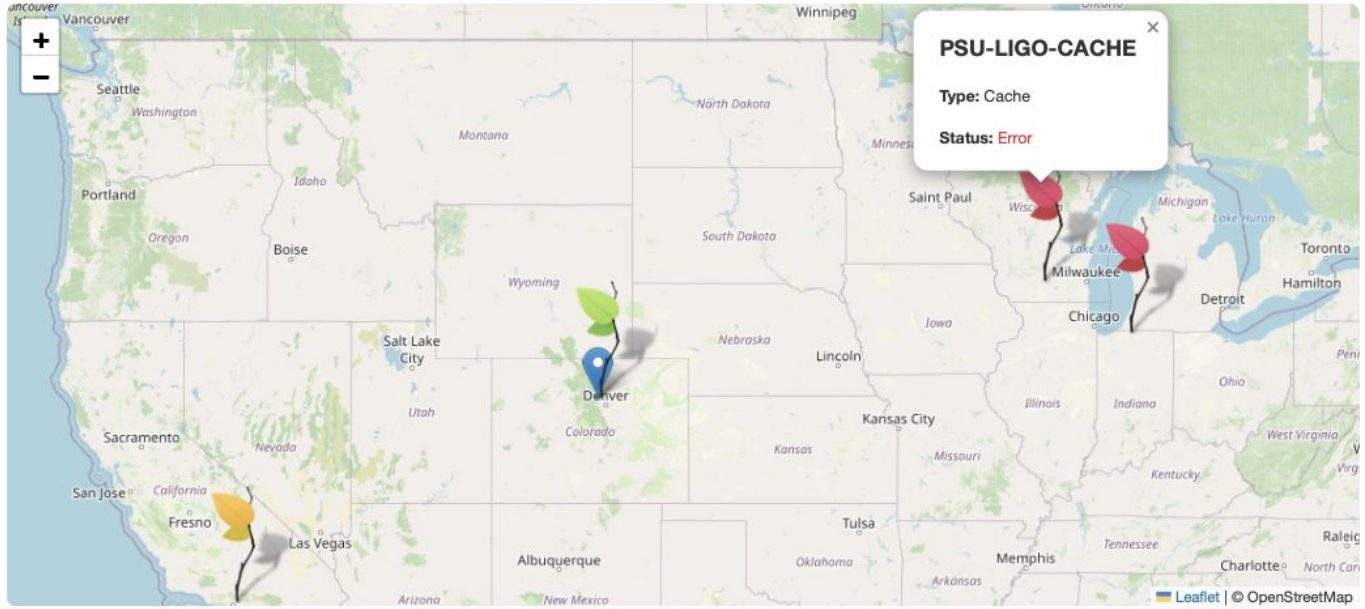
Django views fetch live server data from OSDF API

Parsing:

- extracts Cache servers and NCAR nodes only
- separates healthy (OK) vs down/error caches
- filters out invalid (0,0) locations
- JSON embedding passes server data to frontend

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:
PSU-LIGO-CACHE - 1364.50 km away

Nearest Working Cache:
NCAR_NRP_CACHE_OSDF - 11.24 km away

Down Caches (2 total):
- NOTRE_DAME_OSDF_CACHE
- PSU-LIGO-CACHE

Unknown Status Caches (1 total):
- CIT_LIGO_STASHCACHE

Future Work

Improving usability and access 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

Acknowledgements

I’m especially grateful to my mentors Riley Conroy, Harsha Hampapura, Nihanth Cherukuru, and to my collaborator Connie Chau for their guidance and support throughout this internship. Thank you to the incredible SIParCS team and administrative staff, particularly Virginia Do and Jessica Wang for making this program possible. Special thanks to the NSF, NCAR, CISL, the DECS team, and my fellow interns for a memorable summer

This material is based upon work supported by the U.S. National Science Foundation National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1755088. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation.