Using ESS-DIVE Data

data.ess-dive.lbl.gov

Presenters

Emily Nagamoto

Research Associate

ESS-DIVE Data Management Support and Services

Welcome!
Audience Introductions

Goals

Learn how to use ESS-DIVE's tools to search and reuse public data on ESS-DIVE

1. **Search for Data -** Identify datasets using the dataset metadata
2. **Exploring inside Datasets -** Investigate specific files using the API tools

BONUS: Start to visualize your data and make a file download log

Key Takeaway: *Become aware of how you can investigate your science questions on ESS-DIVE*

Who is this for?

Anyone who wants to know what features ESS-DIVE has to offer.

- PI/Data Managers – bring this information back to your team
- Projects putting together synthesis product – this can help make your search faster
- Data publishers – make your data discoverable

There will be live demonstrations (programmatic experience helpful, but not required), and we will also talk about Reporting Formats.

Data Discovery Workflow

1. **Search for Data**
    a. Data Search Webpage
    b. Dataset API Service (Jupyter Notebook)

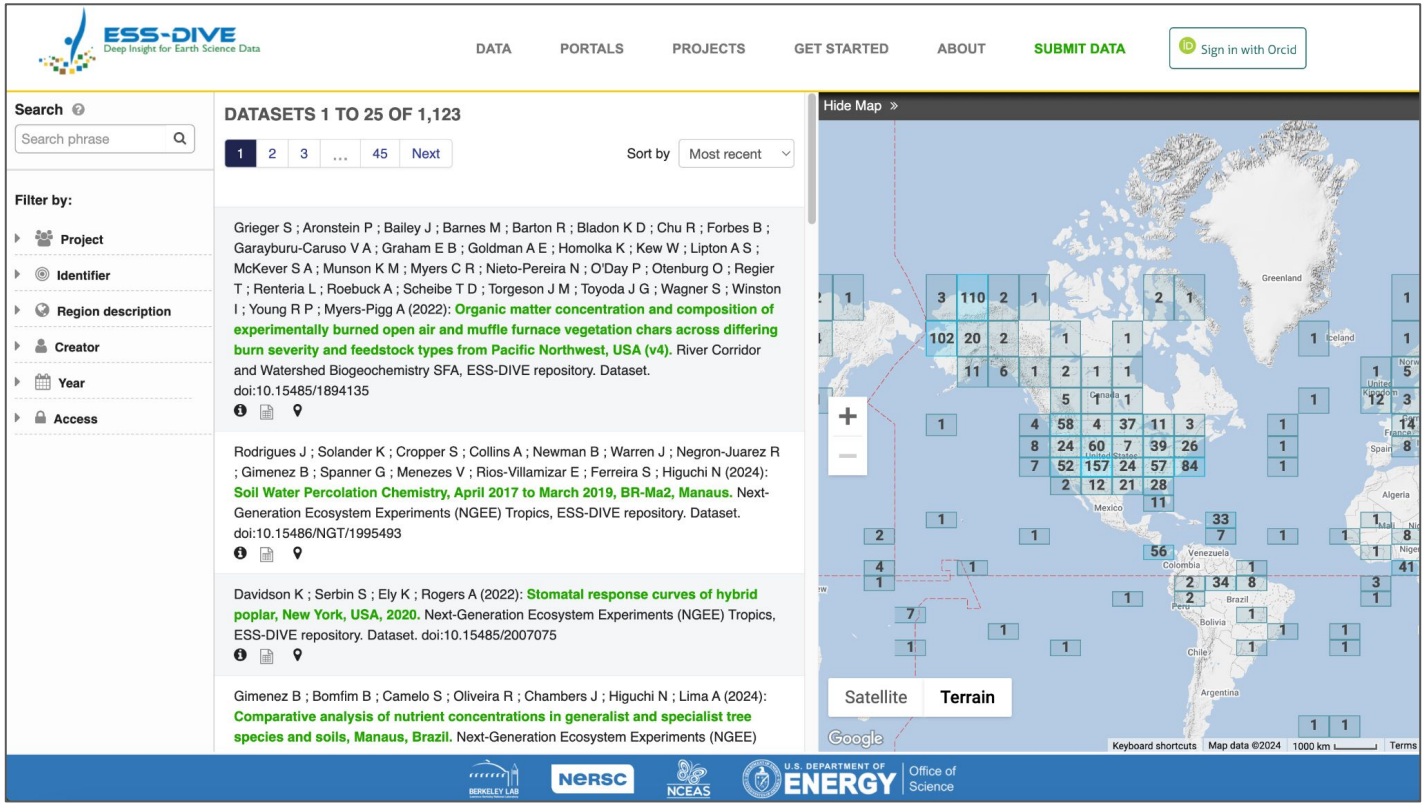
Let's discover datasets related to our example research interest by walking through this workflow

Example Research Interest for Data Discovery

Topic:	Water quality
Find a type of measured data:	DO, temperature, geochemistry data
Observed in a particular place:	Yakima River Basin, WA
During a particular time period:	2020 - now

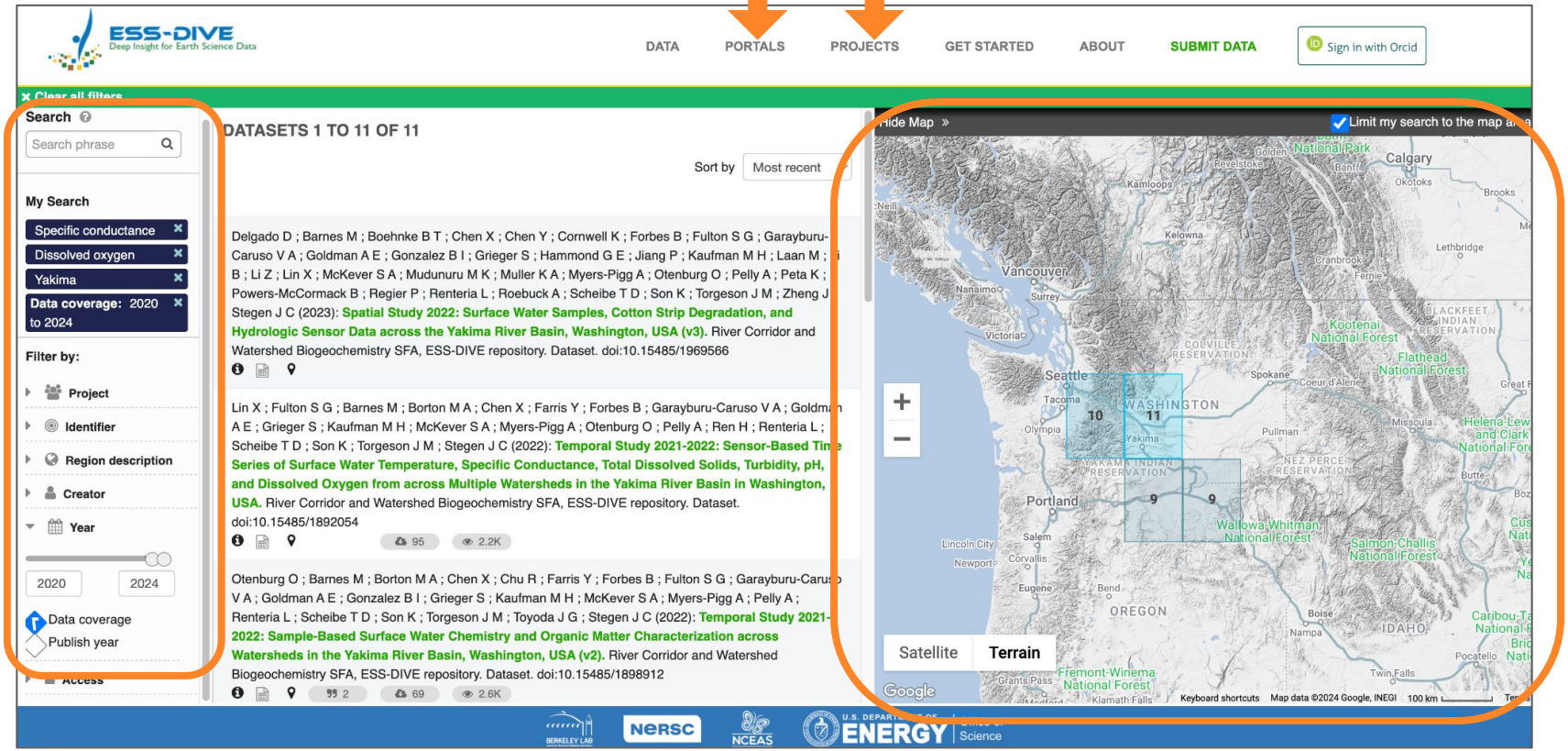
How to use ESS-DIVE's Web Portal to Search for Data

Data Search: Main Search Webpage



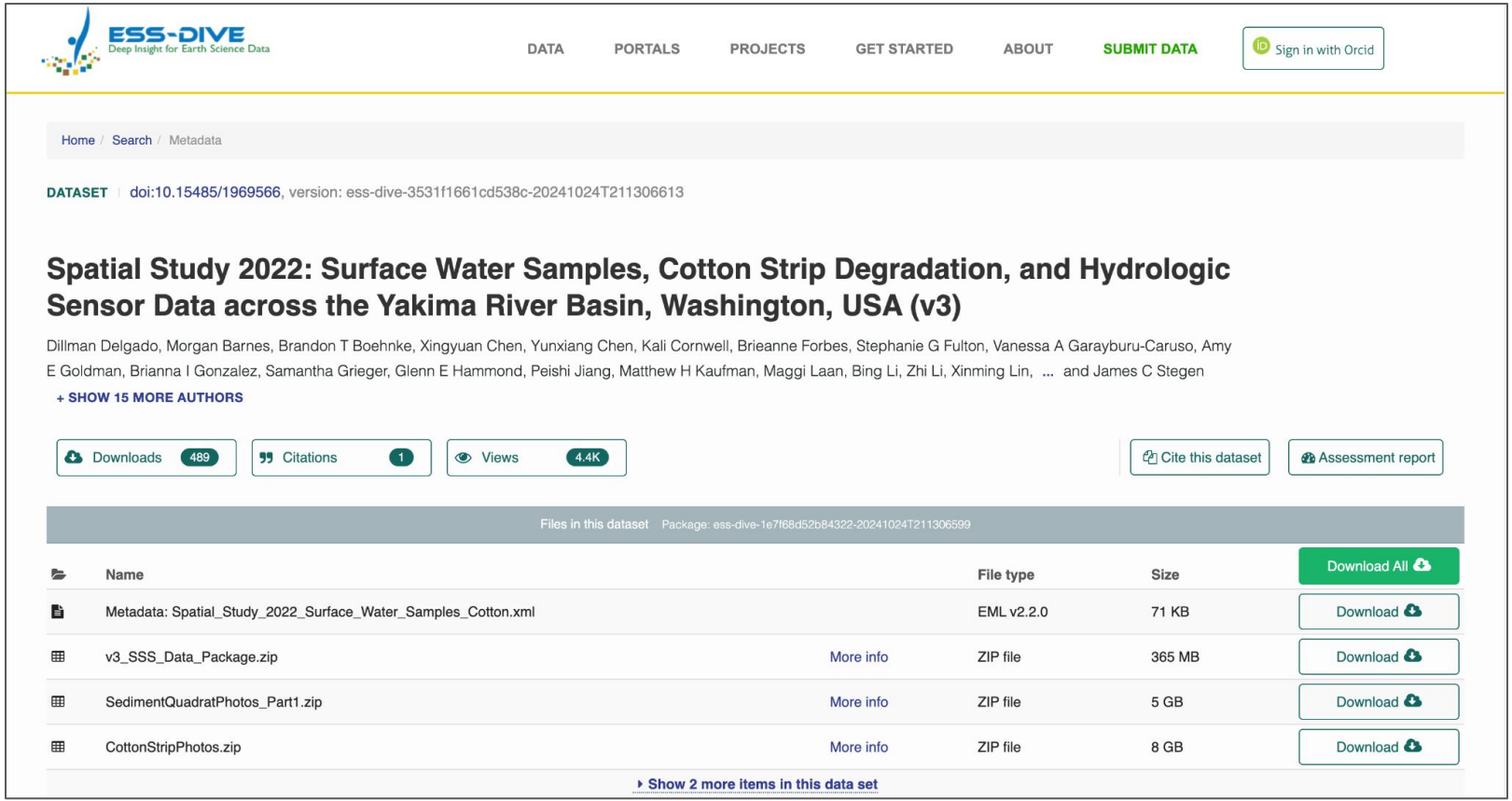
https://data.ess-dive.lbl.gov/data

Data Search: Main Search Webpage



https://data.ess-dive.lbl.gov/data

Dataset Landing Page



TOOL: Data Search Webpage

DETAILS:

- Search dataset metadata
- Fullest set of metadata fields
- Visual map-based search
- Portals (collections)
- Manual inspection

What did we find?: 11 datasets that I need to read through that may or may not have the variables I'm interested in

Questions?

What is the Dataset API?

What is an API?

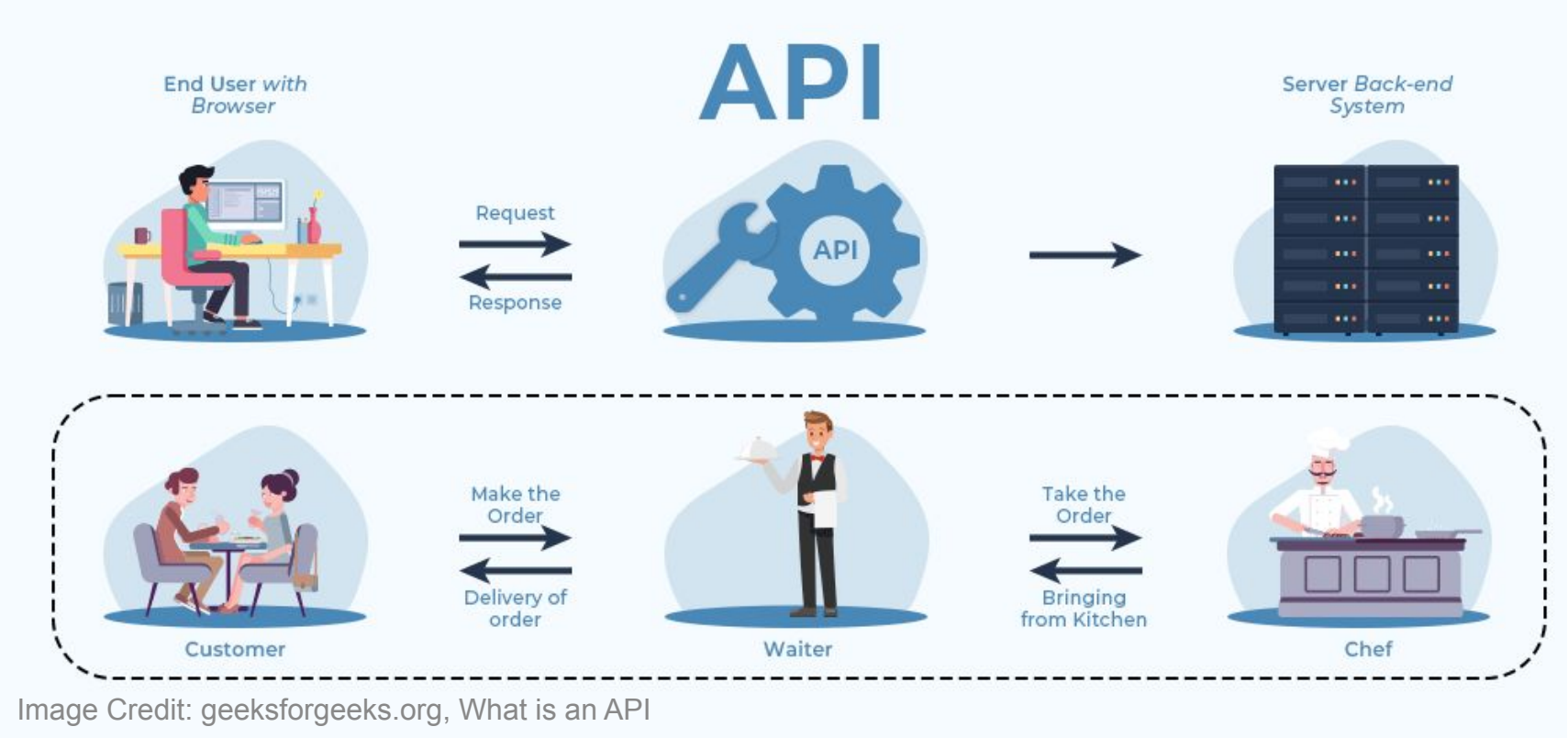
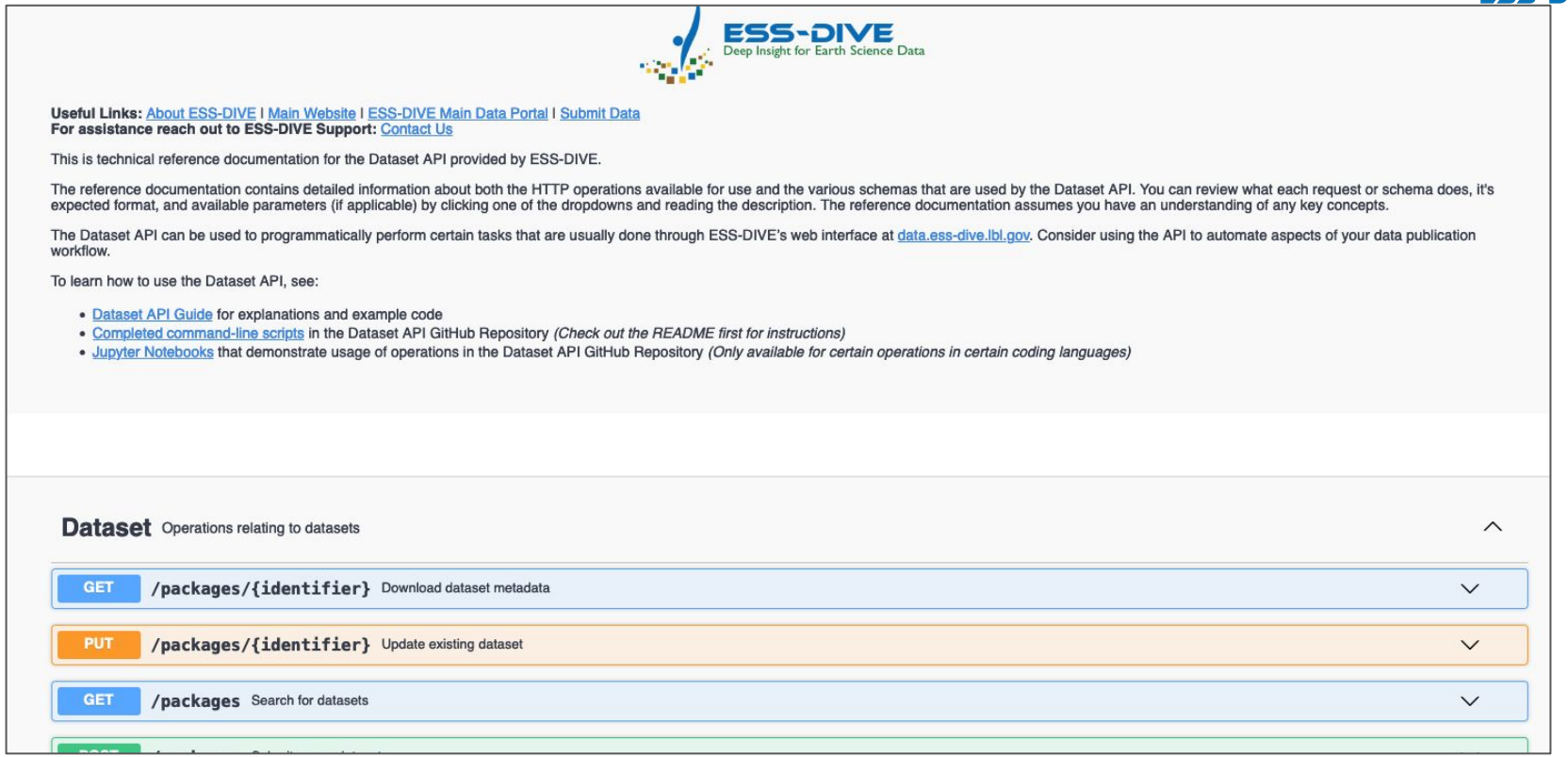


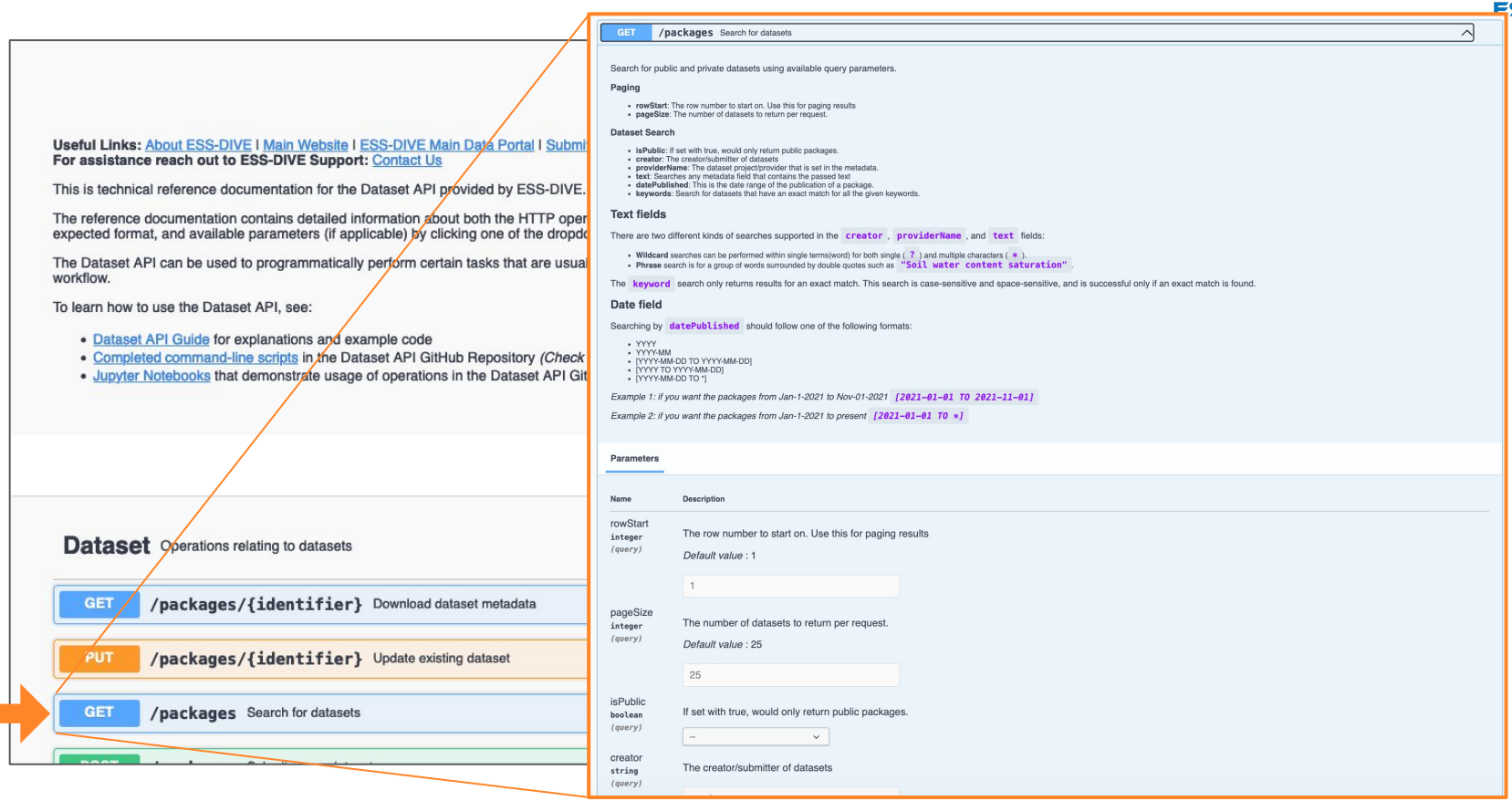
Image Credit: geeksforgeeks.org, What is an API

Data Search: Dataset API

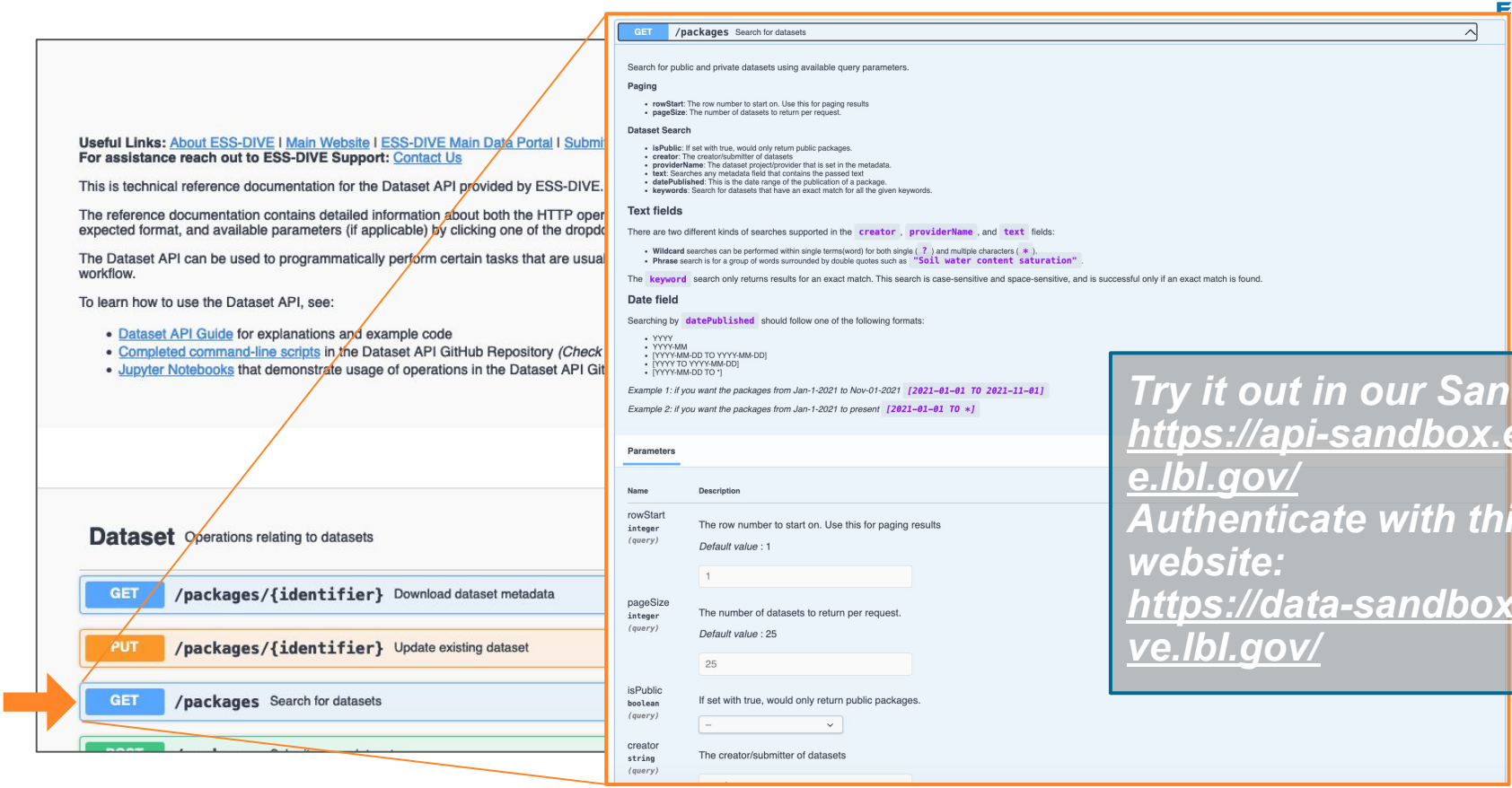


https://api.ess-dive.lbl.gov/#/Dataset/listDatasets

Data Search: Dataset API



Data Search: Dataset API



Try it out in our Sandbox: https://api-sandbox.ess-dive.lbl.gov/
Authenticate with this website: https://data-sandbox.ess-dive.lbl.gov/

Metadata fields have different search capabilities

Metadata Field	Data Search Webpage	Dataset API
General text search	✓	✓
Keywords	—	✓
Project	✓	✓
Identifier	✓	—
Region description	✓	—
All people	✓	—
Creator	✓	✓
Data coverage	✓	—
Dataset Published	✓	✓
Access - Private only	✓	✓
Access - Public only	—	✓
Map-based locations	✓	—

Dataset API Demo

https://github.com/ess-dive/essdive-tutorials/blob/main/search_data/ESS%20PI%20Meeting%202025%20Using%20Data%20-%20Python.ipynb < exact notebook
https://github.com/ess-dive/essdive-tutorials/tree/main < landing page for Colab

Data Search Webpage

Search dataset metadata

Fullest set of metadata fields
Visual map-based search
Portals (collections)

Manual inspection

TOOL: Dataset API

DETAILS:

- Search dataset metadata
- **Subset** of metadata fields
- Supports wildcards
- **Programmatic**, bulk inspection

What did we find? 14 potential datasets and the basic information about each one.

Questions?

Data Discovery Workflow

1. **Search for Data**
    a. Data Search Webpage
    b. Dataset API Service (Jupyter Notebook)
2. **Exploring datasets** (Jupyter Notebook)
    a. Use API Tools to look inside datasets and files
    b. Select and visualize data
    c. Introduction to DeepDive API tools - Extra

From the datasets we found - let's explore inside and identify relevant files.

What are components of well-organized data in ESS-DIVE?

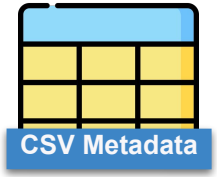
Reporting Formats

Reporting formats are instructions, templates, and tools for consistently formatting data within a discipline.

These data standards were developed by ESS-DIVE and **ESS-DIVE Partner Projects** to standardize metadata and data files of data types commonly collected by DOE ESS projects.

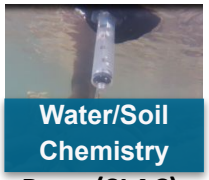
No. of Reporting Formats available:

- **High-level, wide-range of data types**: **6**
- **Data type specific**: **6**

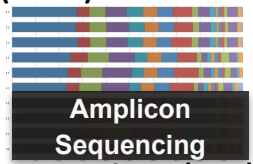


Velliquette, Heinz, Devarakonda (ORNL)

Serbin, Ely (BNL)



Boye (SLAC)



Weisenhorn (ANL)

Goldman (PNNL)

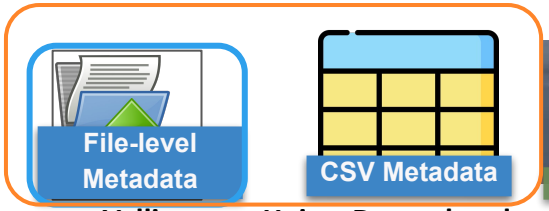
Reporting Formats

Reporting formats are instructions, templates, and tools for consistently formatting data within a discipline.

These data standards were developed by ESS-DIVE and **ESS-DIVE Partner Projects** to standardize metadata and data files of data types commonly collected by DOE ESS projects.

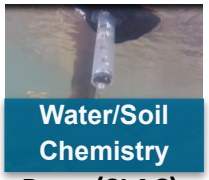
No. of Reporting Formats available:

- **High-level, wide-range of data types**: **6**
- **Data type specific**: **6**

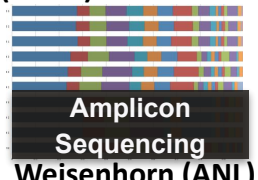


Velliquette, Heinz, Devarakonda (ORNL)

Serbin, Ely (BNL)



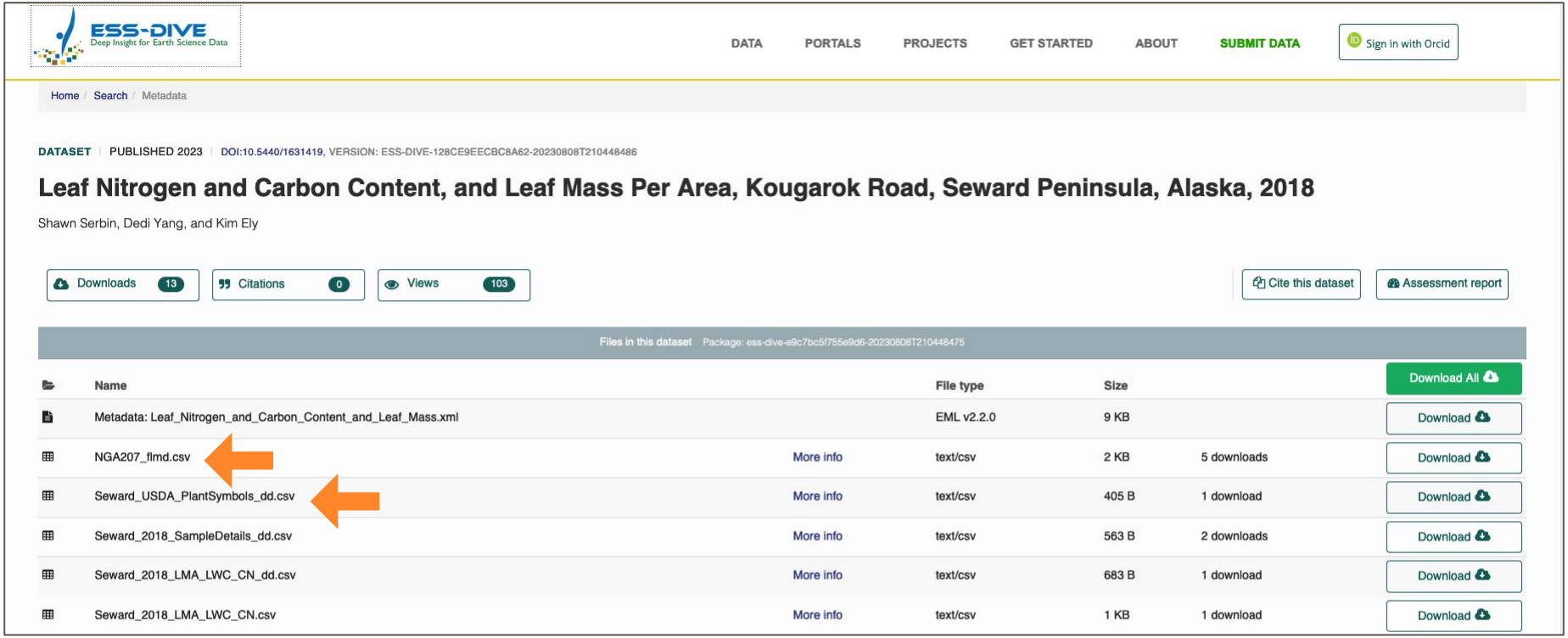
Boye (SLAC)



Weisenhorn (ANL)

Goldman (PNNL)

Datasets with File Level Metadata (FLMD) and Data Dictionaries (*_dd.csv)



FLMD: An index of all files in a dataset (CSV)

Required | **FLMD Example** | Optional

File_Name	File_Description	Standard	UTC_Offset	File_Version	Contact	Start_Date	End_Date	Northwest_Latitude_Coordinate	Northwest_Longitude_Coordinate	Southeast_Latitude_Coordinate	Southeast_Longitude_Coordinate	Latitude	Longitude	Missing_Value_Codes	Notes	Field_Name_Orientation
Seward_2018_SampleDetails.csv	Sample location information, plot, species and photo file name. SampleID links to related leaf spectra and canopy spectra data. Sample photos are in spectra data package.	csv	-8:00	v1.0	Shawn Serbin, sserbin@bnl.gov	7/25/18	7/25/18	65.171006	-164.83776	65.153612	-164.80195	-9999	-9999	-9999	22 records	horizontal
Seward_2018_SampleDetails_dd.csv	data description file	flmd	NA	v1.0	Shawn Serbin, sserbin@bnl.gov	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	horizontal
Seward_2018_LMA_LWC_CN.csv	Leaf carbon and nitrogen content. Leaf mass per area. Leaf water content.	csv	-8:00	v1.0	Shawn Serbin, sserbin@bnl.gov	7/25/18	7/25/18	65.171006	-164.83776	65.153612	-164.80195	-9999	-9999	-9999	22 records	horizontal
Seward_2018_LMA_LWC_CN_dd.csv	data description file	flmd	NA	v1.0	Shawn Serbin, sserbin@bnl.gov	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	-9999	horizontal

Follows the ESS-DIVE File Level Metadata Reporting Format:
https://github.com/ess-dive-workspace/essdive-file-level-metadata

DD: Describes the columns in data files

	A	B	C	D	E	F	G	H	I	J	K	L
1	BURN	CAMPAIGN	BURN_GRI	PLOT_NUM	DATE	BTZ_START	BTZ_END	AT_START	BT_START	BT_END	AT_END	COMP
2	1	1	1	1	3/5/2018	19:09:00	20:34:53	13:54:00	14:09:00	15:34:53	15:34:53	
3	2	1	2	2	3/6/2018	16:27:00	16:46:45	11:12:00	11:27:00	11:46:45	11:46:45	
4	3	1	2	3	3/6/2018	18:27:05	18:55:54	13:12:05	13:27:05	13:55:54	13:55:54	
5	4	2	3	4	3/17/2018	14:53:21	15:17:53	10:38:21	10:53:21	11:17:53	11:17:53	
6	5	2	3	5	3/17/2018	16:42:50	16:57:25	12:27:50	12:42:50	12:57:25	12:57:25	
7	6	2	3	6	3/17/2018	18:44:25	19:11:53	14:29:25	14:44:25	15:11:53	15:11:53	

DD File: SERDP_10x10_BurnSummary**_dd.csv**

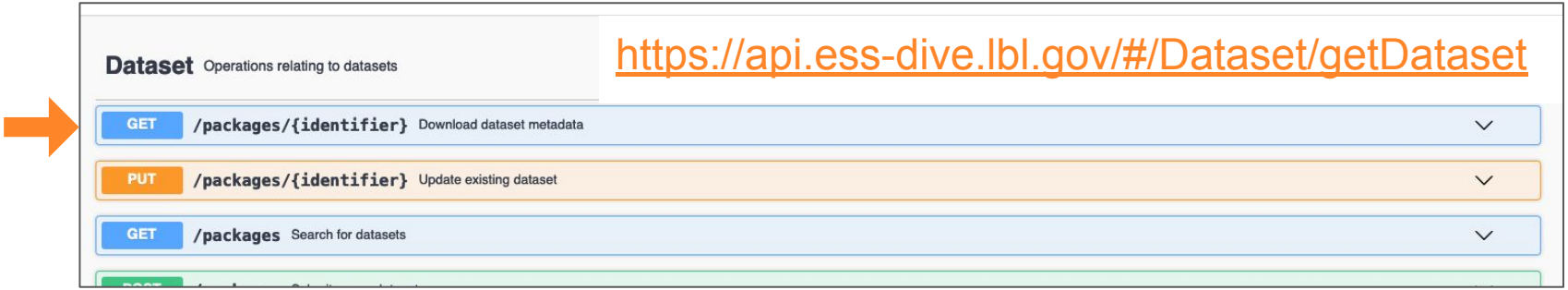
Data File: SERDP_10x10_BurnSummary.xlsx

	A	B	C	D	E
1	Column_or_Row_Name	Unit	Definition	Column_or_Row_Long_Name	Data_Type
2	BURN		numeric value identifying an individual bu	Burn	integer
3	CAMPAIGN		numeric value identifying a series of burr	Field campaign	integer
4	BURN_GRP		numeric value identifying an individual bu	Burn group	integer
5	PLOT_NUM		Originally, a new plot was established for	Plot numeric value identifying a uni	integer
6	DATE	yyyy-mm-dd		Burn date	date
7	BTZ_START	hh:mm:ss	Zulu time when IR sensor first detected 3	Zulu (Coordinated Universal Time (	datetime
8	BTZ_END	hh:mm:ss	Zulu time when IR sensor last detected 3	Zulu (Coordinated Universal Time (	datetime
9	AT_START	hh:mm:ss	The suggested start time of the data arch	Archive start time in local (EST) tim	datetime

Exploring Data within Files

Investigate with detailed metadata from File Distribution

Dataset API:



Inspecting Dataset Contents Demo

https://github.com/ess-dive/essdive-tutorials/blob/main/search_data/ESS%20PI%20Meeting%202025%20Using%20Data%20-%20Python.ipynb < exact notebook
https://github.com/ess-dive/essdive-tutorials/tree/main < landing page for Colab

TOOL: Dataset API

Dataset Details

- Lists all dataset files

Potential challenges:

- No descriptions
- No previews
- Long lists

Programmatic

With Reporting Formats

- Lists dataset files
- Descriptions of files
- Enables file preview / summary

Potential challenges:

- Does not search by field/variable

Programmatic

What did we find? Individual file information that allowed us to select datasets and specific files that we want.

TOOL: Dataset Webpage

Lists all dataset files

Potential challenges:

- Zipped files
- No descriptions
- No previews

Manual inspection

TOOL: Dataset API

Dataset Details

- Lists all dataset files

Potential challenges:

- No descriptions
- No previews
- Long lists

Programmatic

With Reporting Formats

- Lists dataset files
- Descriptions of files
- Enables file preview / summary

Potential challenges:

- Does not search by field/variable

Programmatic

What did we find? Individual file information that allowed us to select datasets and specific files that we want.

Remember to cite your data!

Citations available right in the File Download Log!

essdive_downloaded_files_log.csv

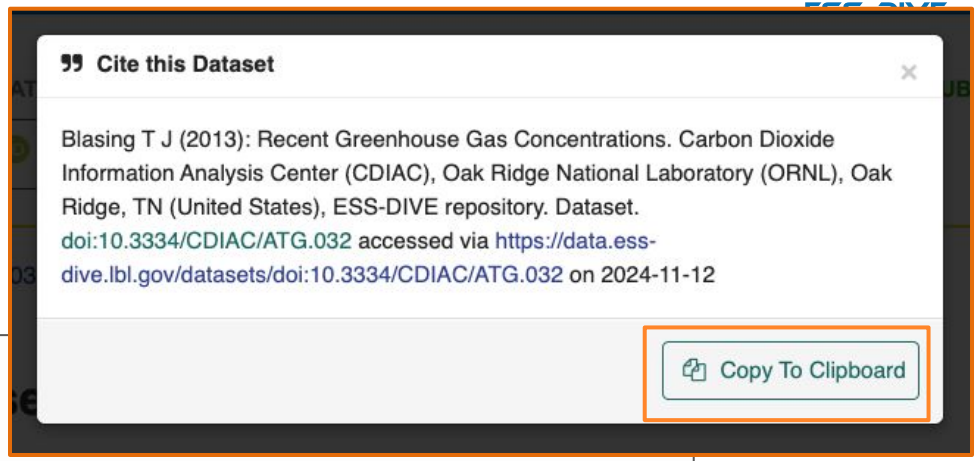
1 entry Filter

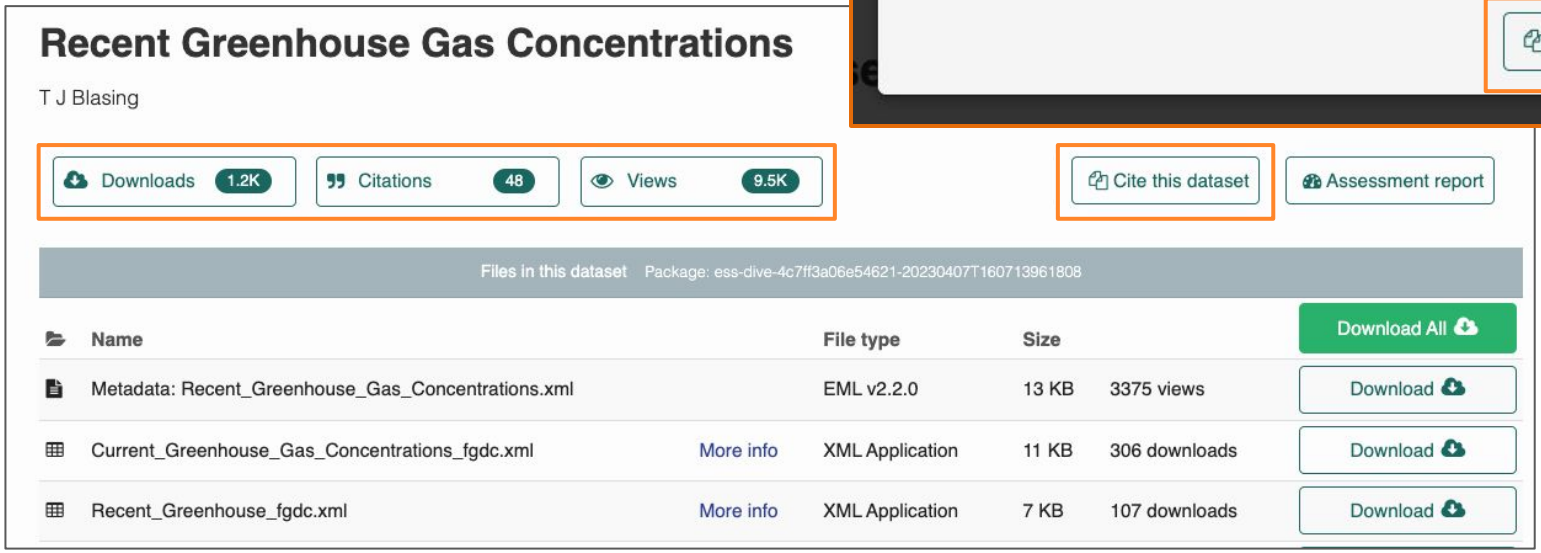
dataset_id	file_name	access_datetime	access_url	dataset_name	citation
doi:10.15485/1923689	v4_CM_SSS_Data_Package.zip	2024-11-15T13:11:02.666573	https://data.ess-dive.lbl.gov/catalog/d1/mn/v2/object/ess-dive-3b2f50c1d4252dc-20240617T222401481	WHONDRS River Corridor Dissolved Oxygen, Temperature, Sediment Aerobic Respiration, Grain Size, and Water Chemistry from Machine-Learning-Informed Sites across the Contiguous United States (v4)	Forbes B; Barnes M; Boehnke B T; Chen X; Cornwell K; Delgado D; Fulton S G; Garayburu-Caruso V A; Gary S; Goldman A E; Gonzalez B I; Grieger S; Hammond G E; Jiang P; Kaufman M H; Laan M; Li B; Li Z; McKever S A; Mudunuru M K; Muller K A; Myers-Pigg A; Otenburg O; Pelly A; Peta K; Powers-McCormack B; Regier P; Renteria L; Roebuck A; Scheibe T D; Son K; Torgeson J M; Stegen J C; Consortium T W (2023): WHONDRS River Corridor Dissolved Oxygen, Temperature, Sediment Aerobic Respiration, Grain Size, and Water Chemistry from Machine-Learning-Informed Sites across the Contiguous United States (v4). River Corridor and Watershed Biogeochemistry SFA. doi:10.15485/1923689

Cite Datasets

Cite Data in Your Papers!

- Data availability and references section



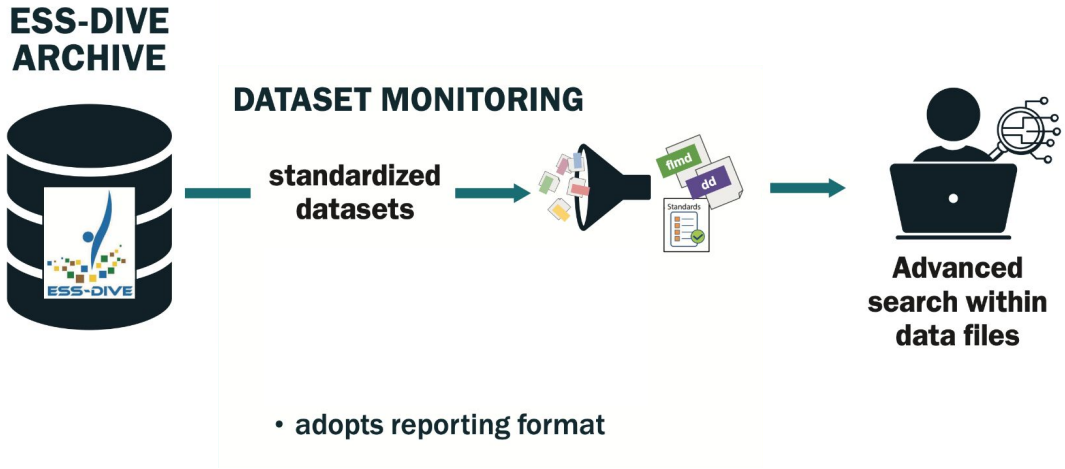


Questions?

What is the Deep Dive API?

Deep Dive API

The **Deep Dive API** is a new search capability that allows you to look for data **within dataset files** on ESS-DIVE



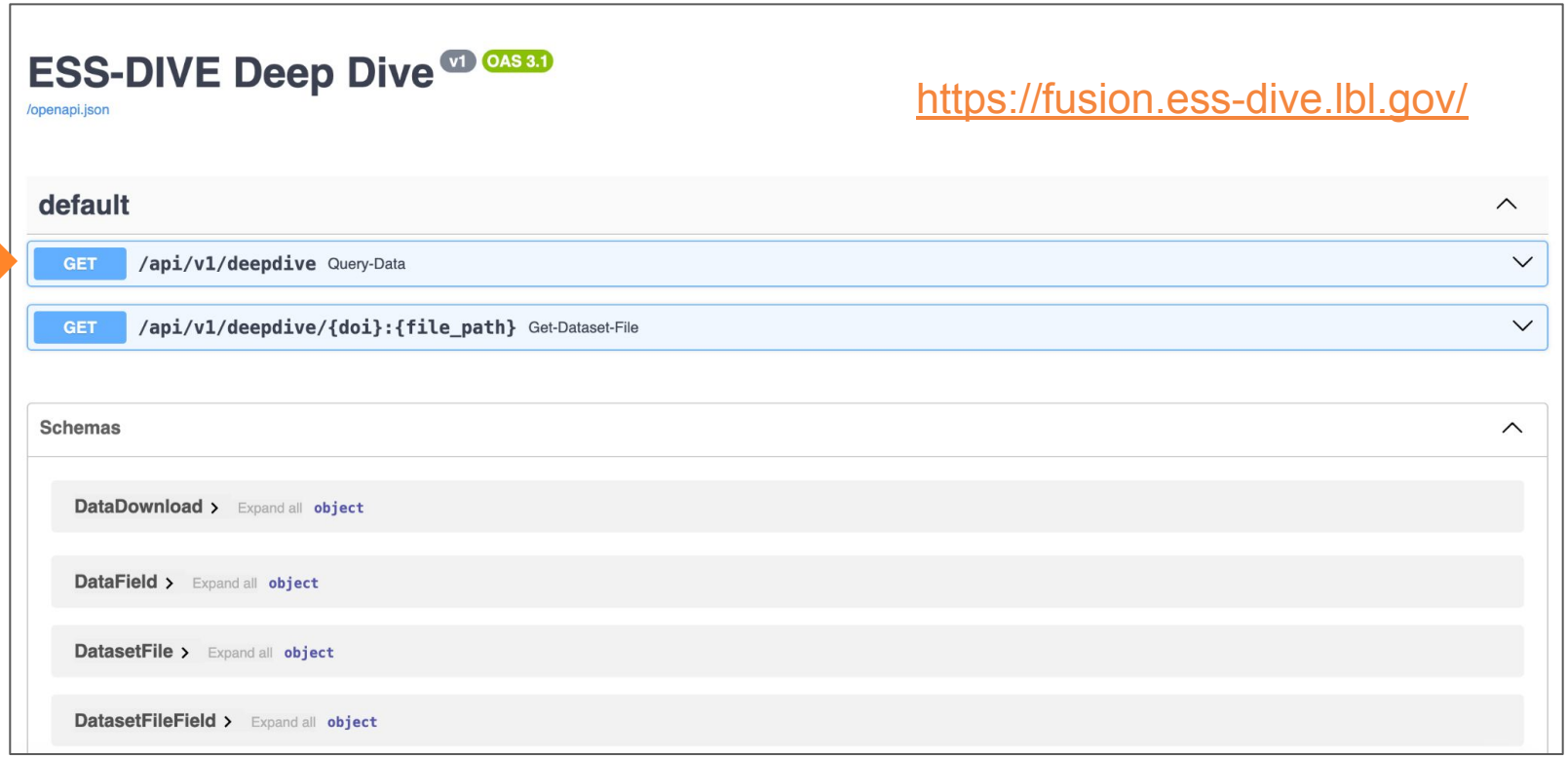
Deep Dive locates and extracts data from published datasets that have **completed File Level Metadata**

Programmatic tool

Very powerful, more narrow in scope

Documentation: https://go.lbl.gov/search-with-DeepDiveAPI

Deep Dive API: Detailed Metadata with File Distribution

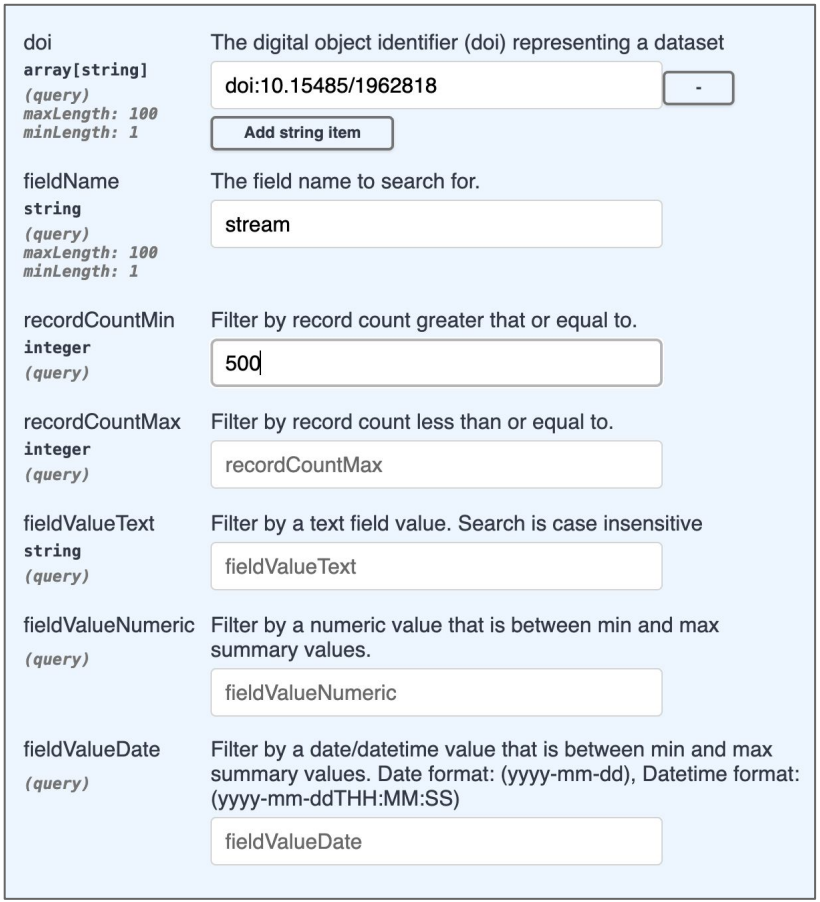


Search Parameters

https://fusion.ess-dive.lbl.gov/

Find datasets relevant to your scientific research

- DOIs
- Field Name
- Record count
- Field Value - text, numeric, date(time)



Comparing Search Tools

Data Search Webpage	Dataset API	Deep-Dive API
Search dataset metadata	Search dataset metadata	Search **within** dataset
Fullest set of metadata fields Visual map-based search Portals (collections)	**Subset** of metadata fields Supports wildcards	Search csv **file contents**
Manual inspection	**Programmatic**, bulk inspection	Only a subset of all of the datasets available
		Online and **programmatic** inspection

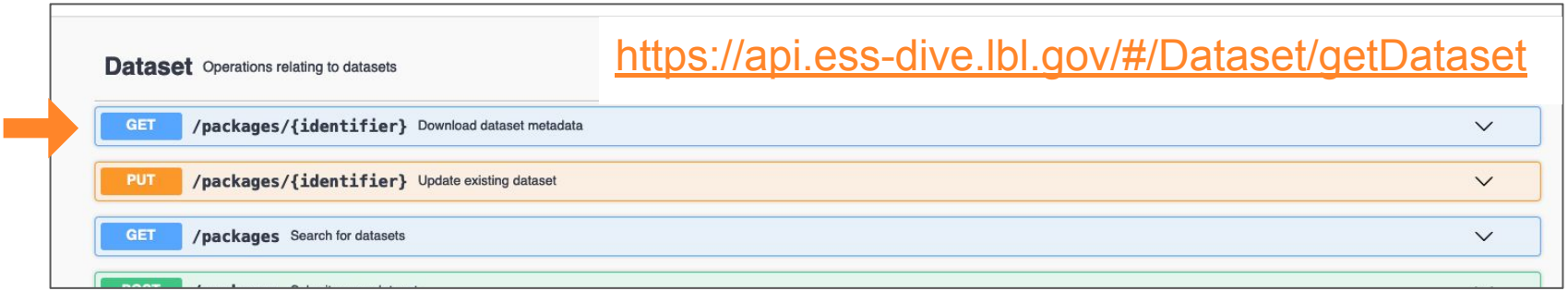
Investigate within Files

Deep Dive API:

https://fusion.ess-dive.lbl.gov/

Dataset API:

https://api.ess-dive.lbl.gov/#/Dataset/getDataset



Ways to explore inside ESS-DIVE Datasets

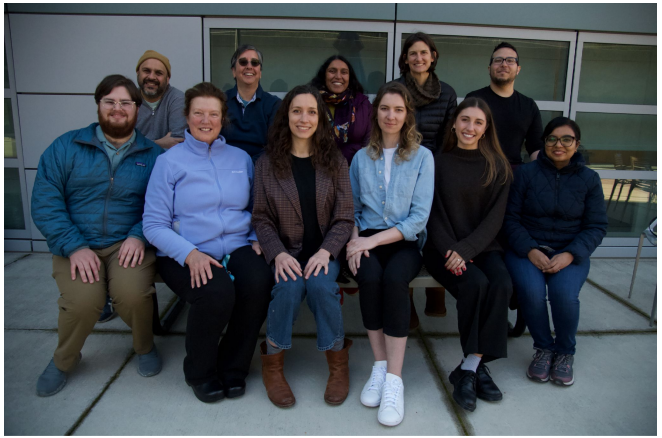
Dataset Webpage	Dataset API		Deep-Dive API
Lists all dataset files Potential challenges: ● Zipped files ● No descriptions ● No previews Manual inspection	**Dataset Details** Lists all dataset files Potential challenges: ● No descriptions ● No previews ● Long lists Programmatic	**With Structured Data (Reporting Formats)** Lists dataset files Descriptions of files Enables file preview Can summarize files Potential challenges: ● Does not search by field/variable Programmatic	**With Structured Data (Reporting Formats)** Searching the data **within** dataset files Potential challenges: ● Limited datasets Programmatic

Questions?

Outcomes

1. **Search for Data**
    a. Broadly we identified potential datasets that fit our interests
    b. We found 14 datasets via the Dataset API
2. **Exploring inside datasets**
    a. We quickly summarize or search through metadata fields of multiple datasets
    b. We downloaded data files that do contain our variable of interest and began visualizing

Thank you! Let's Stay Connected.



Contact us if you are interested in any hands on activities: **ess-dive-support@lbl.gov**

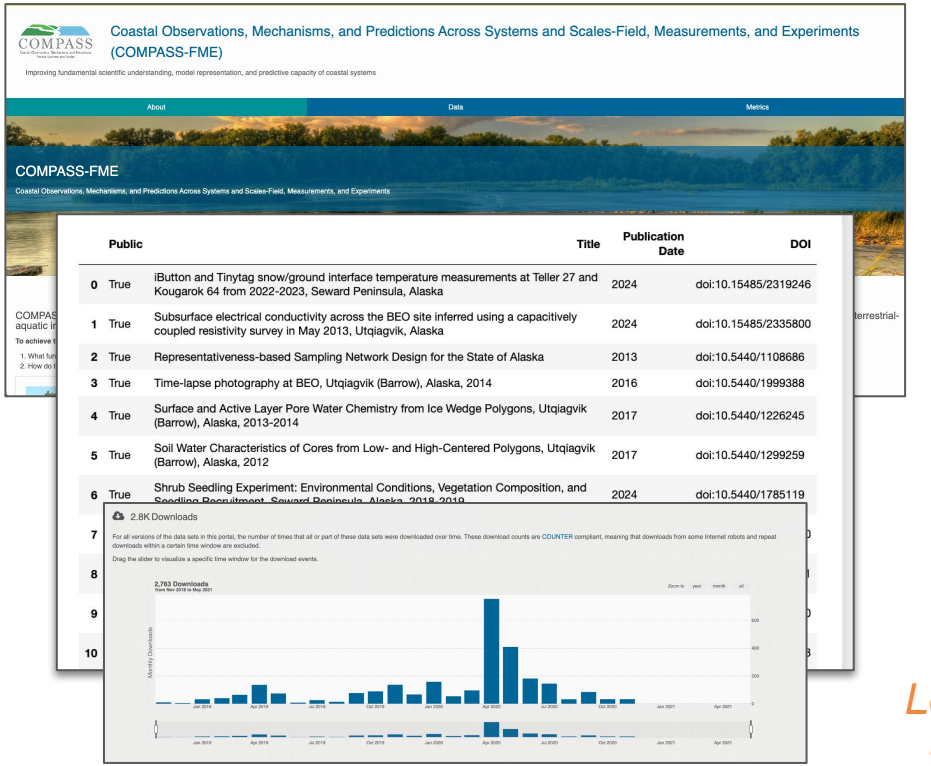
ESS-DIVE Documentation

docs.ess-dive.lbl.gov

Join our mailing list!

https://go.lbl.gov/essdiveMailingList

Don’t miss our lunch tutorial: ESS-DIVE Tutorial for PIs and Data Managers



Lunchtime Session I

Day 2 | Wed April 16

12:30pm - 1:30pm EDT

Grand Ballroom A

Madison Burrus & Joan Damerow (LBNL)

- Checklist for PIs
- Project Management Setup
- Data Portals and Reporting
- Key Features

Learn how to maximize ESS-DIVE features to manage project data

Madison Burrus
(LBNL)