



# Using ESS-DIVE Data

# 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



[DATA](#)
[PORTALS](#)
[PROJECTS](#)
[GET STARTED](#)
[ABOUT](#)
[SUBMIT DATA](#)

Search 

**Filter by:**

- Project
- Identifier
- Region description
- Creator
- Year
- Access

**DATASETS 1 TO 25 OF 1,123**

1	2	3	...	45	Next
---	---	---	-----	----	------

Sort by

Grieger S ; Aronstein P ; Bailey J ; Barnes M ; Barton R ; Bladon K D ; Chu R ; Forbes B ; Garayburu-Caruso V A ; Graham E B ; Goldman A E ; Homolka K ; Kew W ; Lipton A S ; McKeever S A ; Munson K M ; Myers C R ; Nieto-Pereira N ; O'Day P ; Otenburg O ; Regier T ; Renteria L ; Roebuck A ; Scheibe T D ; Torgeson J M ; Toyoda J G ; Wagner S ; Winston I ; Young R P ; Myers-Pigg A (2022): **Organic matter concentration and composition of experimentally burned open air and muffle furnace vegetation chars across differing burn severity and feedstock types from Pacific Northwest, USA (v4)**. River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1894135

Rodrigues J ; Solander K ; Cropper S ; Collins A ; Newman B ; Warren J ; Negron-Juarez R ; Gimenez B ; Spanner G ; Menezes V ; Rios-Villamizar E ; Ferreira S ; Higuchi N (2024): **Soil Water Percolation Chemistry, April 2017 to March 2019, BR-Ma2, Manaus**. Next-Generation Ecosystem Experiments (NGEE) Tropics, ESS-DIVE repository. Dataset. doi:10.15486/NGT/1995493

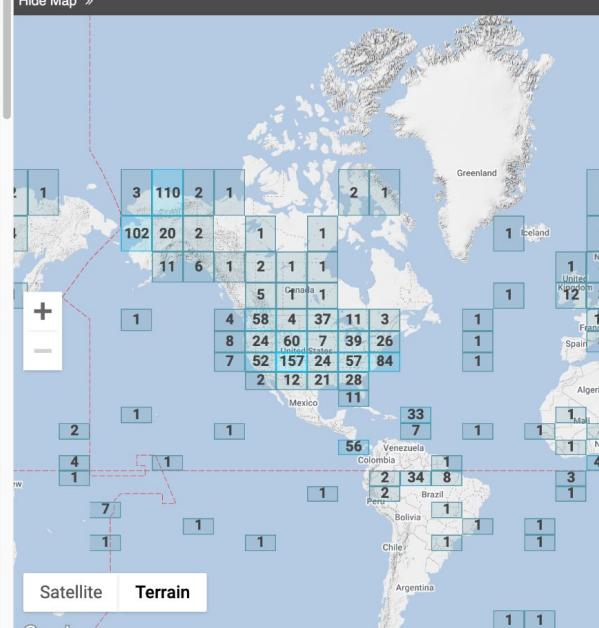
Davidson K ; Serbin S ; Ely K ; Rogers A (2022): **Stomatal response curves of hybrid poplar, New York, USA, 2020**. Next-Generation Ecosystem Experiments (NGEE) Tropics, ESS-DIVE repository. Dataset. doi:10.15485/2007075

Gimenez B ; Bomfim B ; Camelo S ; Oliveira R ; Chambers J ; Higuchi N ; Lima A (2024): **Comparative analysis of nutrient concentrations in generalist and specialist tree species and soils, Manaus, Brazil**. Next-Generation Ecosystem Experiments (NGEE) Tropics, ESS-DIVE repository. Dataset. doi:10.15485/2007075

[Hide Map >](#)



Satellite  Terrain

Google

Keyboard shortcuts | Map data ©2024 | 1000 km | Terms

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

9

# Data Search: Main Search Webpage

[DATA](#)
[PORTALS](#)
[PROJECTS](#)

[GET STARTED](#)
[ABOUT](#)
[SUBMIT DATA](#)
[Sign in with Orcid](#)

[Clear all filters](#)

**Search** ([?](#))

**My Search**

- [Specific conductance](#) ([X](#))
- [Dissolved oxygen](#) ([X](#))
- [Yakima](#) ([X](#))
- [Data coverage: 2020 to 2024](#) ([X](#))

**Filter by:**

- [Project](#)
- [Identifier](#)
- [Region description](#)
- [Creator](#)
- [Year](#)

**Year**

**Data coverage**

- [Publish year](#)

**ACCESS**

- [Open](#)
- [Protected](#)
- [Closed](#)

**DATASETS 1 TO 11 OF 11**

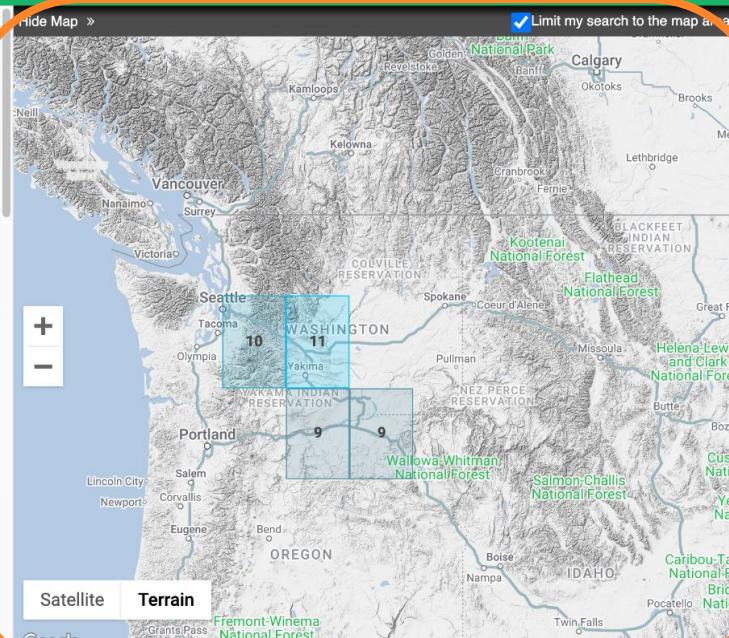
Sort by [Most recent](#)

[Delgado D ; Barnes M ; Boehnke B T ; Chen X ; Chen Y ; Cornwell K ; Forbes B ; Fulton S G ; Garayburu-Caruso V A ; Goldman A E ; Gonzalez B I ; Grieger S ; Hammond G E ; Jiang P ; Kaufman M H ; Laan M ; Li Z ; Lin X ; 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 ; Zheng J ; Stegen J C \(2023\): Spatial Study 2022: Surface Water Samples, Cotton Strip Degradation, and Hydrologic Sensor Data across the Yakima River Basin, Washington, USA \(v3\). River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1969566](#)

[Lin X ; Fulton S G ; Barnes M ; Borton M A ; Chen X ; Farris Y ; Forbes B ; Garayburu-Caruso V A ; Goldman A E ; Grieger S ; Kaufman M H ; McKever S A ; Myers-Pigg A ; Otenburg O ; Pelly A ; Ren H ; Renteria L ; Scheibe T D ; Son K ; Torgeson J M ; Stegen J C \(2022\): Temporal Study 2021-2022: Sensor-Based Time Series of Surface Water Temperature, Specific Conductance, Total Dissolved Solids, Turbidity, pH, and Dissolved Oxygen from across Multiple Watersheds in the Yakima River Basin in Washington, USA. River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1892054](#)

[Otenburg O ; Barnes M ; Borton M A ; Chen X ; Chu R ; Farris Y ; Forbes B ; Fulton S G ; Garayburu-Caruso V A ; Goldman A E ; Gonzalez B I ; Grieger S ; Kaufman M H ; McKever S A ; Myers-Pigg A ; Pelly A ; Renteria L ; Scheibe T D ; Son K ; Torgeson J M ; Toyoda J G ; Stegen J C \(2022\): Temporal Study 2021-2022: Sample-Based Surface Water Chemistry and Organic Matter Characterization across Watersheds in the Yakima River Basin, Washington, USA \(v2\). River Corridor and Watershed Biogeochemistry SFA, ESS-DIVE repository. Dataset. doi:10.15485/1898912](#)

Limit my search to the map area



Satellite [Terrain](#)

Google

Keyboard shortcuts Map data ©2024 Google, INEGI 100 km

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

10



# Dataset Landing Page

 **ESS-DIVE**  
Deep Insight for Earth Science Data

DATA PORTALS PROJECTS GET STARTED ABOUT **SUBMIT DATA** 

Home / Search / Metadata

**DATASET** | doi:10.15485/1969566, version: ess-dive-3531f1661cd538c-20241024T211306613

## Spatial Study 2022: Surface Water Samples, Cotton Strip Degradation, and Hydrologic Sensor Data across the Yakima River Basin, Washington, USA (v3)

Dillman Delgado, Morgan Barnes, Brandon T Boehnke, Xingyuan Chen, Yunxiang Chen, Kali Cornwell, Brieanne Forbes, Stephanie G Fulton, Vanessa A Garayburu-Caruso, Amy E Goldman, Brianna I Gonzalez, Samantha Grieger, Glenn E Hammond, Peishi Jiang, Matthew H Kaufman, Maggi Laan, Bing Li, Zhi Li, Xinming Lin, ... and James C Stegen  
[+ SHOW 15 MORE AUTHORS](#)

 Downloads 489     Citations 1     Views 4.4K     Cite this dataset     Assessment report

Files in this dataset Package: ess-dive-1e7f68d52b84322-20241024T211306599

Name	File type	Size	Download All
Metadata: Spatial_Study_2022_Surface_Water_Samples_Cotton.xml	EML v2.2.0	71 KB	
v3_SSS_Data_Package.zip	More info	ZIP file	
SedimentQuadratPhotos_Part1.zip	More info	ZIP file	
CottonStripPhotos.zip	More info	ZIP file	

[Show 2 more items in this data set](#)

## 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?

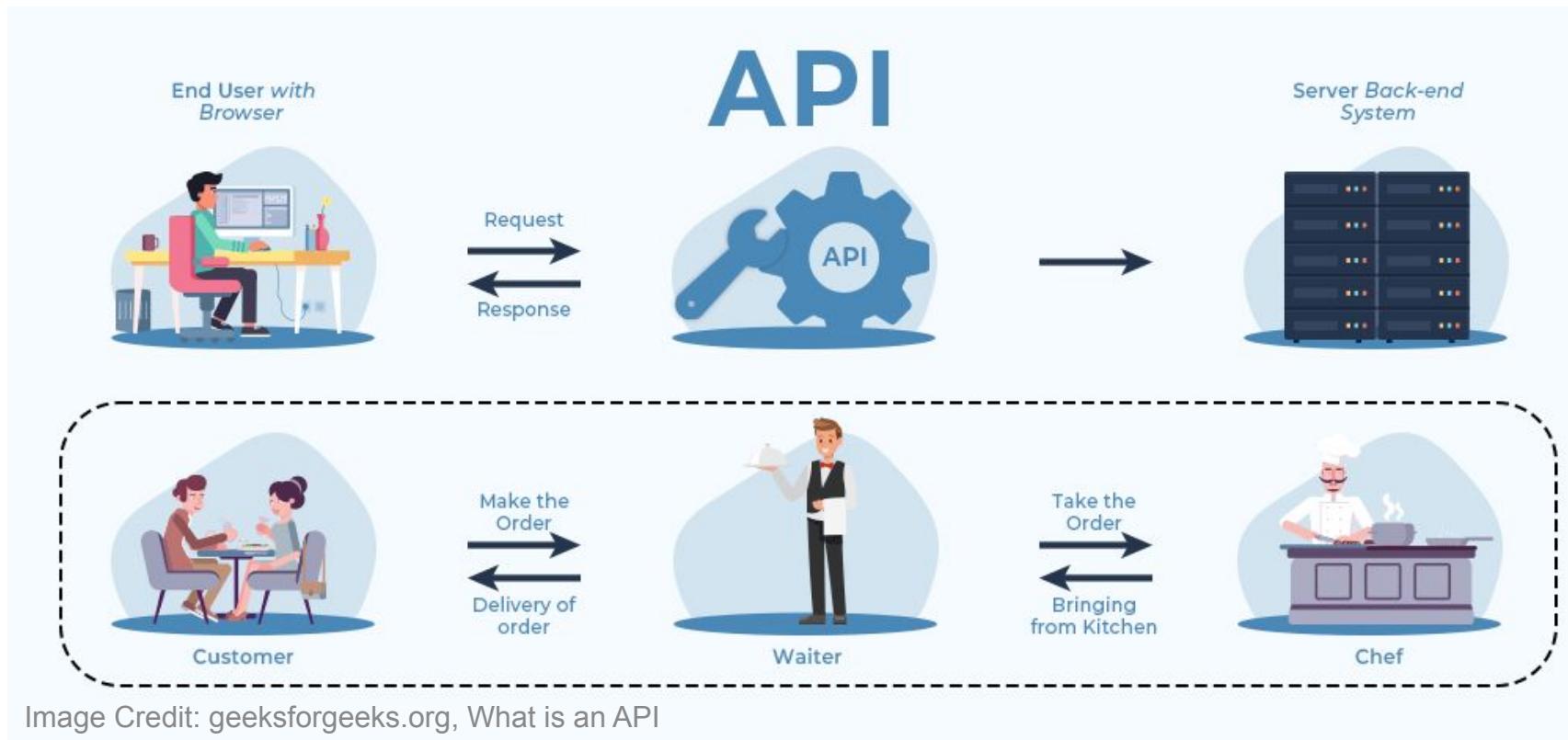


Image Credit: geeksforgeeks.org, What is an API

# Data Search: Dataset API

The ESS-DIVE logo, featuring a stylized blue and green graphic element.

**ESS-DIVE**  
Deep Insight for Earth Science Data

Useful Links: [About ESS-DIVE](#) | [Main Website](#) | [ESS-DIVE Main Data Portal](#) | [Submit Data](#)  
For assistance reach out to ESS-DIVE Support: [Contact Us](#)

This is technical reference documentation for the Dataset API provided by ESS-DIVE.

The reference documentation contains detailed information about both the HTTP operations available for use and the various schemas that are used by the Dataset API. You can review what each request or schema does, its expected format, and available parameters (if applicable) by clicking one of the dropdowns and reading the description. The reference documentation assumes you have an understanding of any key concepts.

The Dataset API can be used to programmatically perform certain tasks that are usually done through ESS-DIVE's web interface at [data.ess-dive.lbl.gov](https://data.ess-dive.lbl.gov). Consider using the API to automate aspects of your data publication workflow.

To learn how to use the Dataset API, see:

- [Dataset API Guide](#) for explanations and example code
- [Completed command-line scripts](#) in the Dataset API GitHub Repository (*Check out the README first for instructions*)
- [Jupyter Notebooks](#) that demonstrate usage of operations in the Dataset API GitHub Repository (*Only available for certain operations in certain coding languages*)

**Dataset** Operations relating to datasets ^

<b>GET</b>	<a href="#"><code>/packages/{identifier}</code></a>	Download dataset metadata	<span style="color: blue;">▼</span>
<b>PUT</b>	<a href="#"><code>/packages/{identifier}</code></a>	Update existing dataset	<span style="color: orange;">▼</span>
<b>GET</b>	<a href="#"><code>/packages</code></a>	Search for datasets	<span style="color: lightblue;">▼</span>
<hr/>			

A large orange arrow pointing to the right, indicating the direction of the link below.

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

16

# Data Search: Dataset API

**Useful Links:** [About ESS-DIVE](#) | [Main Website](#) | [ESS-DIVE Main Data Portal](#) | [Submit](#)

For assistance reach out to ESS-DIVE Support: [Contact Us](#)

This is technical reference documentation for the Dataset API provided by ESS-DIVE.

The reference documentation contains detailed information about both the HTTP operation expected format, and available parameters (if applicable) by clicking one of the dropdowns.

The Dataset API can be used to programmatically perform certain tasks that are usually part of the workflow.

To learn how to use the Dataset API, see:

- [Dataset API Guide](#) for explanations and example code
- [Completed command-line scripts](#) in the Dataset API GitHub Repository (Check out the [examples](#) folder)
- [Jupyter Notebooks](#) that demonstrate usage of operations in the Dataset API GitHub Repository (Check out the [notebooks](#) folder)

**Dataset** Operations relating to datasets

<b>GET</b>	<b>/packages/{identifier}</b> Download dataset metadata
<b>PUT</b>	<b>/packages/{identifier}</b> Update existing dataset
<b>GET</b>	<b>/packages</b> Search for datasets

# Data Search: Dataset API

**Useful Links:** [About ESS-DIVE](#) | [Main Website](#) | [ESS-DIVE Main Data Portal](#) | [Submit](#)  
**For assistance reach out to ESS-DIVE Support:** [Contact Us](#)

This is technical reference documentation for the Dataset API provided by ESS-DIVE.

The reference documentation contains detailed information about both the HTTP operation expected format, and available parameters (if applicable) by clicking one of the dropdowns.

The Dataset API can be used to programmatically perform certain tasks that are usually part of the workflow.

To learn how to use the Dataset API, see:

- [Dataset API Guide](#) for explanations and example code
- [Completed command-line scripts](#) in the Dataset API GitHub Repository (Check out the [README](#))
- [Jupyter Notebooks](#) that demonstrate usage of operations in the Dataset API GitHub Repository

**Dataset** Operations relating to datasets

<b>GET</b>	<a href="#">/packages/{identifier}</a>	Download dataset metadata
<b>PUT</b>	<a href="#">/packages/{identifier}</a>	Update existing dataset
<b>GET</b>	<a href="#">/packages</a>	Search for datasets

**GET /packages** Search for datasets

Search for public and private datasets using available query parameters.

**Paging**

- `rowStart`: The row number to start on. Use this for paging results
- `pageSize`: The number of datasets to return per request.

**Dataset Search**

- `isPublic`: If set with true, would only return public packages.
- `creator`: The creator/submitter of datasets
- `provider`: The data provider/provider that is set in the metadata.
- `text`: Searches any metadata field that contains the passed text
- `datePublished`: This is the date range of the publication of a package.
- `keywords`: Search for datasets that have an exact match for all the given keywords.

**Text fields**

There are two different kinds of searches supported in the `creator`, `providerName`, and `text` fields:

- Wildcard searches can be performed within single terms(word) for both single (?) and multiple characters (\*).
- Phrase search is for a group of words surrounded by double quotes such as `"Soil water content saturation"`.

The `keyword` search only returns results for an exact match. This search is case-sensitive and space-sensitive, and is successful only if an exact match is found.

**Date field**

Searching by `datePublished` should follow one of the following formats:

- YYYY
- YYYY-MM
- (YYYY-MM-DD TO YYYY-MM-DD)
- (YYYY TO YYYY-MM-DD)
- (YYYY-MM-DD TO )

Example 1: If you want the packages from Jan-1-2021 to Nov-01-2021 `[2021-01-01 TO 2021-11-01]`

Example 2: If you want the packages from Jan-1-2021 to present `[2021-01-01 TO *]`

**Parameters**

Name	Description
<code>rowStart</code> <code>integer (query)</code>	The row number to start on. Use this for paging results  Default value : 1
<code>pageSize</code> <code>integer (query)</code>	The number of datasets to return per request.  Default value : 25
<code>isPublic</code> <code>boolean (query)</code>	If set with true, would only return public packages.
<code>creator</code> <code>string (query)</code>	The creator/submitter of datasets

**Try it out in our Sandbox:**  
<https://api-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/Using\\_Data\\_with\\_Dee](https://github.com/ess-dive/essdive-tutorials/blob/main/search_data/Using_Data_with_Dee)  
pDiveAPI\_Python.ipynb < exact notebook

<https://github.com/ess-dive/essdive-tutorials/tree/main> < landing page for Colab

## TOOL: Dataset API

### DETAILS:

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

### Data Search Webpage

Search dataset metadata

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

### Manual 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



File-level  
Metadata



CSV Metadata



UAS

Velliquette, Heinz, Devarakonda  
(ORNL)



Water/Soil  
Chemistry  
Boye (SLAC)



Amplicon  
Sequencing  
Weisenhorn (ANL)



Hydrologic  
Monitoring  
Goldman (PNNL)

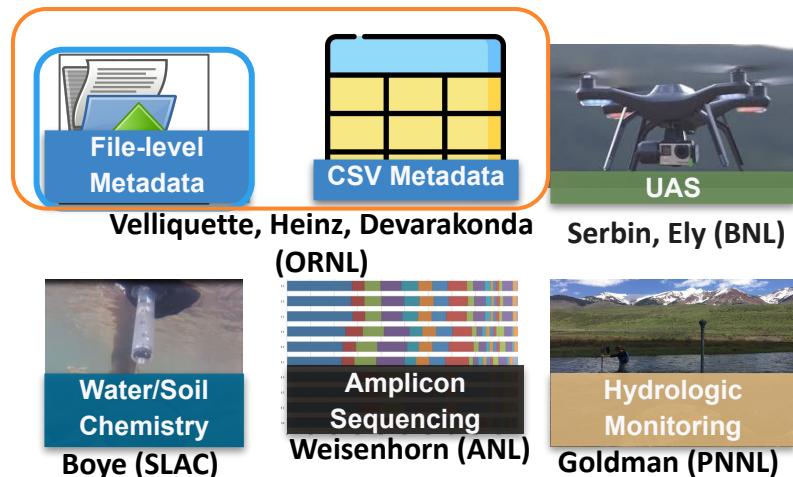
# Reporting Formats

**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



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



ESS-DIVE Deep Insight for Earth Science Data

DATA PORTALS PROJECTS GET STARTED ABOUT SUBMIT DATA Sign in with Orcid

Home / Search / Metadata

DATASET | PUBLISHED 2023 | DOI:10.5440/1631419, VERSION: ESS-DIVE-128CE9EECBC8A62-20230808T210448486

## Leaf Nitrogen and Carbon Content, and Leaf Mass Per Area, Kougarok Road, Seward Peninsula, Alaska, 2018

Shawn Serbin, Dedi Yang, and Kim Ely

Downloads 13 Citations 0 Views 103 Cite this dataset Assessment report

Files in this dataset Package: ess-dive-e9c7bc5f755e9d6-20230808T210448476					
Name	File type	Size	Actions		
Metadata: Leaf_Nitrogen_and_Carbon_Content_and_Leaf_Mass.xml	EML v2.2.0	9 KB	Download All		
NGA207_flmd.csv	text/csv	2 KB	5 downloads	Download	Cloud
Seward_USDA_PlantSymbols_dd.csv	text/csv	405 B	1 download	Download	Cloud
Seward_2018_SampleDetails_dd.csv	text/csv	563 B	2 downloads	Download	Cloud
Seward_2018_LMA_LWC_CN_dd.csv	text/csv	683 B	1 download	Download	Cloud
Seward_2018_LMA_LWC_CN.csv	text/csv	1 KB	1 download	Download	Cloud

# 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_GRP	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 burn	Burn	integer
3	CAMPAIGN		numeric value identifying a series of burns	Field campaign	integer
4	BURN_GRP		numeric value identifying an individual burn group	Burn group	integer
5	PLOT_NUM		Originally, a new plot was established for Plot	numeric value identifying a unique plot	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 archive	Archive start time in local (EST) time	datetime



# Exploring Data **within** Files

# Investigate with detailed metadata from File Distribution

## Dataset API:

**Dataset** Operations relating to datasets

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



<b>GET</b>	/packages/{identifier}	Download dataset metadata	▼
<b>PUT</b>	/packages/{identifier}	Update existing dataset	▼
<b>GET</b>	/packages	Search for datasets	▼
<b>POST</b>			▼

# Inspecting Dataset Contents Demo

[https://github.com/ess-dive/essdive-tutorials/blob/main/search\\_data/Using\\_Data\\_with\\_Dee\\_pDiveAPI\\_Python.ipynb](https://github.com/ess-dive/essdive-tutorials/blob/main/search_data/Using_Data_with_Dee_pDiveAPI_Python.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	<a href="https://data.ess-dive.lbl.gov/catalog/d1/mn/v2/object/ess-dive-3b2f50c1d4252dc-20240617T222401481">https://data.ess-dive.lbl.gov/catalog/d1/mn/v2/object/ess-dive-3b2f50c1d4252dc-20240617T222401481</a>	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

### Recent Greenhouse Gas Concentrations

T J Blasing

Downloads 1.2K Citations 48 Views 9.5K

Files in this dataset Package: ess-dive-4c7ff3a06e54621-20230407T160713961808					
Name	File type	Size		Download All	
Metadata: Recent_Greenhouse_Gas_Concentrations.xml	EML v2.2.0	13 KB	3375 views		
Current_Greenhouse_Gas_Concentrations_fgdc.xml	<a href="#">More info</a>	XML Application	11 KB	306 downloads	
Recent_Greenhouse_fgdc.xml	<a href="#">More info</a>	XML Application	7 KB	107 downloads	

„ Cite this Dataset

Blasing T J (2013): Recent Greenhouse Gas Concentrations. Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), ESS-DIVE repository. Dataset.  
doi:10.3334/CDIAC/ATG.032 accessed via <https://data.ess-dive.lbl.gov/datasets/doi:10.3334/CDIAC/ATG.032> on 2024-11-12

Copy To Clipboard



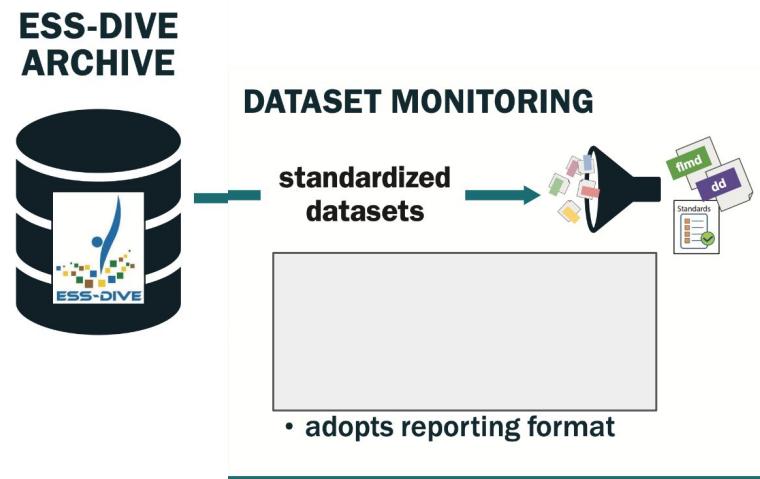
# Questions?

# 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

**ESS-DIVE Deep Dive** v1 OAS 3.1

</openapi.json>

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

**default**

GET /api/v1/deepdive Query-Data

GET /api/v1/deepdive/{doi}:{file\_path} Get-Dataset-File

**Schemas**

DataDownload > Expand all object

DataField > Expand all object

DatasetFile > Expand all object

DatasetFileField > Expand all object



# 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)

<b>doi</b> <b>array</b> [string] <i>(query)</i> <b>maxLength:</b> 100 <b>minLength:</b> 1	The digital object identifier (doi) representing a dataset <input type="text" value="doi:10.15485/1962818"/> <span style="border: 1px solid #ccc; padding: 2px;">-</span> <input type="button" value="Add string item"/>
<b>fieldName</b> <b>string</b> <i>(query)</i> <b>maxLength:</b> 100 <b>minLength:</b> 1	The field name to search for. <input type="text" value="stream"/>
<b>recordCountMin</b> <b>integer</b> <i>(query)</i>	Filter by record count greater than or equal to. <input type="text" value="500"/>
<b>recordCountMax</b> <b>integer</b> <i>(query)</i>	Filter by record count less than or equal to. <input type="text" value="recordCountMax"/>
<b>fieldValueText</b> <b>string</b> <i>(query)</i>	Filter by a text field value. Search is case insensitive <input type="text" value="fieldValueText"/>
<b>fieldValueNumeric</b> <i>(query)</i>	Filter by a numeric value that is between min and max summary values. <input type="text" value="fieldValueNumeric"/>
<b>fieldValueDate</b> <i>(query)</i>	Filter by a date/datetime value that is between min and max summary values. Date format: (yyyy-mm-dd), Datetime format: (yyyy-mm-ddTHH:MM:SS) <input type="text" value="fieldValueDate"/>

# Comparing Search Tools

## Data Search Webpage

Search dataset metadata

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

Manual inspection

## Dataset API

Search dataset metadata

**Subset** of metadata fields  
Supports wildcards

**Programmatic**, bulk inspection

## Deep-Dive API

Search **within** dataset

Search csv **file contents**

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/>

default	
GET	/api/v1/deepdive Query-Data
GET	/api/v1/deepdive/{doi}:{file_path} Get-Dataset-File



## Dataset API:

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

Dataset Operations relating to datasets	
GET	/packages/{identifier} Download dataset metadata
PUT	/packages/{identifier} Update existing dataset
GET	/packages Search for datasets



# Ways to explore inside ESS-DIVE Datasets

Dataset Webpage	Dataset API	Deep-Dive API
<p>Lists all dataset files</p> <p>Potential challenges:</p> <ul style="list-style-type: none"> <li>• Zipped files</li> <li>• No descriptions</li> <li>• No previews</li> </ul> <p>Manual inspection</p>	<p><b>Dataset Details</b></p> <p>Lists all dataset files</p> <p>Potential challenges:</p> <ul style="list-style-type: none"> <li>• No descriptions</li> <li>• No previews</li> <li>• Long lists</li> </ul> <p>Programmatic</p>	<p><b>With Structured Data (Reporting Formats)</b></p> <p>Lists dataset files</p> <p>Descriptions of files</p> <p>Enables file preview</p> <p>Can summarize files</p> <p>Potential challenges:</p> <ul style="list-style-type: none"> <li>• Does not search by field/variable</li> </ul> <p>Programmatic</p>



# Questions?

# Outcomes

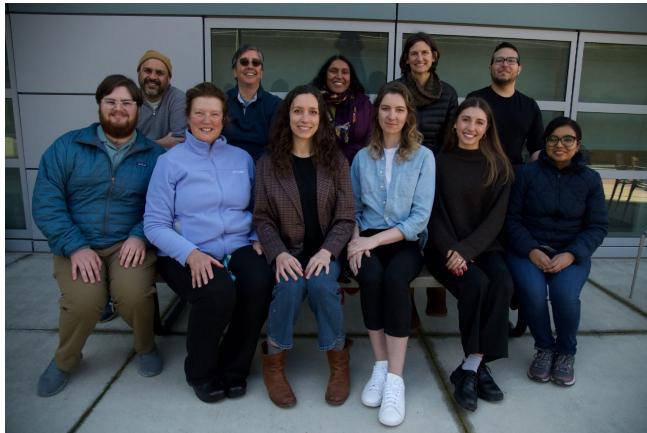
## 1. Search for Data

- 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](mailto:ess-dive-support@lbl.gov)

**ESS-DIVE Documentation**

[docs.ess-dive.lbl.gov](https://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



Coastal Observations, Mechanisms, and Predictions Across Systems and Scales-Field, Measurements, and Experiments  
(COMPASS-FME)

Improving fundamental scientific understanding, model representation, and predictive capacity of coastal systems

About      Data      Metrics

### COMPASS-FME

Coastal Observations, Mechanisms, and Predictions Across Systems and Scales-Field, Measurements, and Experiments

**Public**

	Title	Publication Date	DOI
0	iButton and Tinytag snow/ground interface temperature measurements at Teller 27 and Kougarok 64 from 2022-2023, Seward Peninsula, Alaska	2024	doi:10.15485/2319246
1	Subsurface electrical conductivity across the BEO site inferred using a capacitively coupled resistivity survey in May 2013, Utqiaġvik, Alaska	2024	doi:10.15485/2335800
2	Representativeness-based Sampling Network Design for the State of Alaska	2013	doi:10.5440/1108686
3	Time-lapse photography at BEO, Utqiaġvik (Barrow), Alaska, 2014	2016	doi:10.5440/1999388
4	Surface and Active Layer Pore Water Chemistry from Ice Wedge Polygons, Utqiaġvik (Barrow), Alaska, 2013-2014	2017	doi:10.5440/1226245
5	Soil Water Characteristics of Cores from Low- and High-Centered Polygons, Utqiaġvik (Barrow), Alaska, 2012	2017	doi:10.5440/1299259
6	Shrub Seedling Experiment: Environmental Conditions, Vegetation Composition, and Seedling Recruitment, Seward Peninsula, Alaska, 2018-2019	2024	doi:10.5440/1785119

2.8K Downloads

For all versions of the data sets in this portal, the number of times that all or part of these data sets were downloaded over time. These download counts are COUNTER compliant, meaning that downloads from some Internet robots and repeat downloads within a certain time window are excluded.

Drag the slider to visualize a specific time window for the download events.

2,763 Downloads  
From Nov 2018 to May 2021

Monthly Downloads

## 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 publish project data*



Madison Burrus  
(LBNL)