



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





Deep Insight for Earth Science Data

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

Search ⓘ

Search phrase

Filter by:

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

DATASETS 1 TO 25 OF 1,123

1 2 3 ... 45 Next

Sort by Most recent ▾

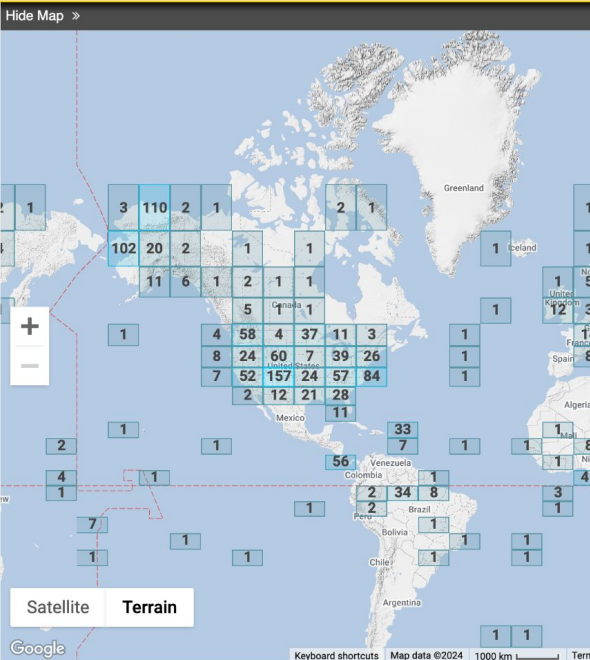
Grieger S ; Aronstein P ; Bailey J ; Barnes M ; Barton R ; Bladon K D ; Chu R ; Forbes B ; Garayburu-Caruso V A ; Graham E B ; Goldman A E ; Homolka K ; Kew W ; Lipton A S ; McKeever S A ; Munson K M ; Myers C R ; Nieto-Pereira N ; O'Day P ; Otenburg O ; Regier T ; Renteria L ; Roebuck A ; Scheibe T 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)

Hide Map »



Satellite Terrain

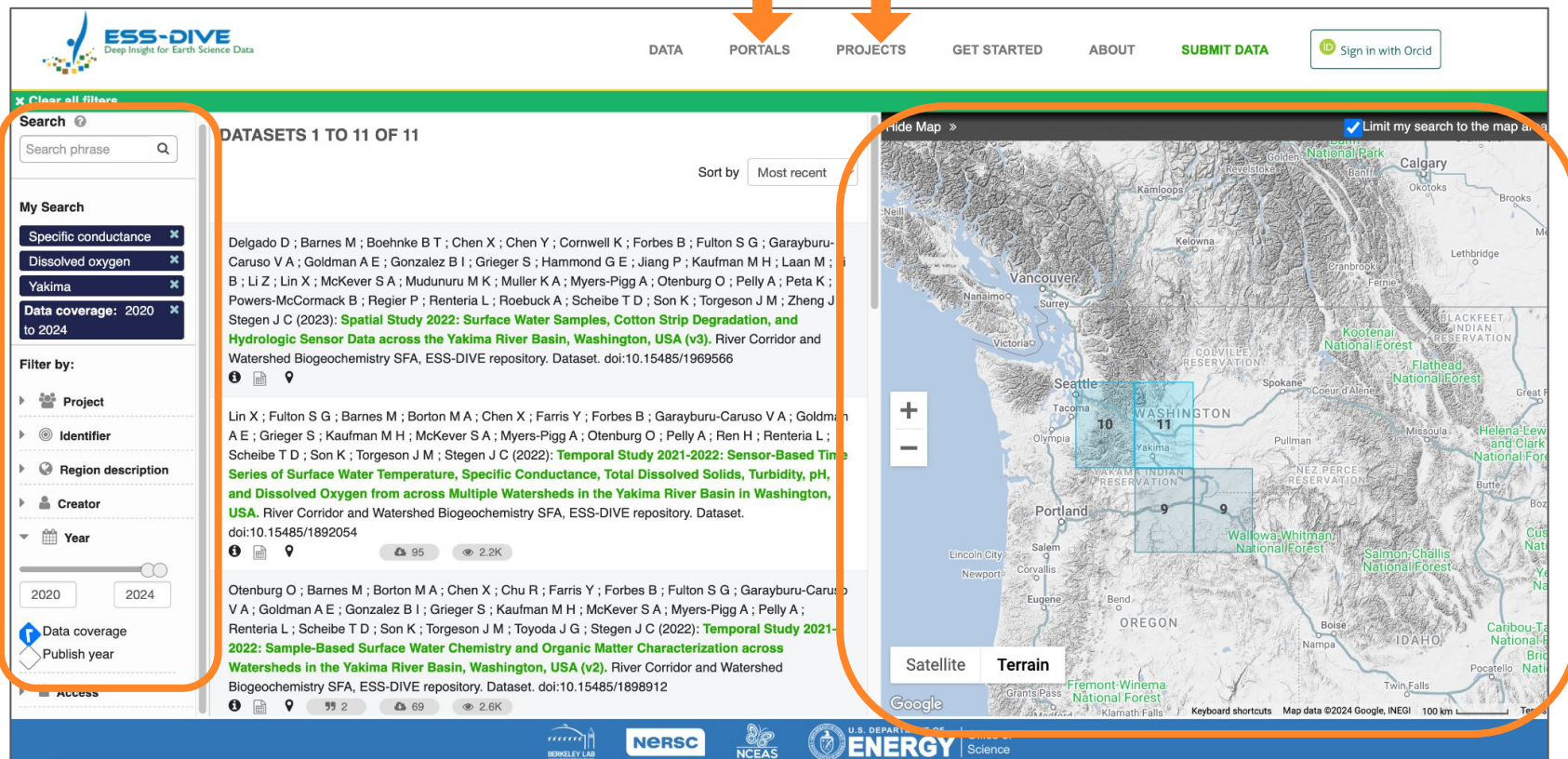
Google

Keyboard shortcuts Map data ©2024 1000 km Terms

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

9

Data Search: Main Search Webpage



The screenshot displays the ESS-DIVE Data Search Main Search Webpage. The top navigation bar includes links for DATA, PORTALS, PROJECTS, GET STARTED, ABOUT, and SUBMIT DATA, along with a 'Sign in with Orcid' button. The main content area is titled 'DATASETS 1 TO 11 OF 11' and features a search bar, a 'Sort by' dropdown (set to 'Most recent'), and a list of datasets. The left sidebar contains a 'Search' section with a 'Search phrase' input field and a 'My Search' section with filters for 'Specific conductance', 'Dissolved oxygen', 'Yakima', and 'Data coverage: 2020 to 2024'. Below these are 'Filter by' options for 'Project', 'Identifier', 'Region description', 'Creator', and 'Year'. The right sidebar shows a map of the Yakima River Basin with a 'Limit my search to the map area' checkbox and a 'Hide Map' button. The map displays the river basin and surrounding areas, with a 'Satellite' and 'Terrain' view toggle. The bottom of the page features a blue banner with logos for BERKELEY LAB, NERSC, NCEAS, and the U.S. DEPARTMENT OF ENERGY Science.

ESS-DIVE
Deep Insight for Earth Science Data

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

Clear all filters

Search Search phrase

My Search

- Specific conductance
- Dissolved oxygen
- Yakima
- Data coverage: 2020 to 2024

Filter by:

- Project
- Identifier
- Region description
- Creator
- Year

2020 2024

Data coverage Publish year

Access

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 ; B ; Li Z ; Lin X ; McKeever 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 ; McKeever 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 ; McKeever 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

Hide Map » 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>

Dataset Landing Page





DATAPORTALSPROJECTSGET STARTEDABOUTSUBMIT DATA

Sign in with Orcid

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, Brienne 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](#)

Downloads489

Citations1

Views4.4K

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	Download
v3_SSS_Data_Package.zip	ZIP file	365 MB	Download
SedimentQuadratPhotos_Part1.zip	ZIP file	5 GB	Download
CottonStripPhotos.zip	ZIP file	8 GB	Download

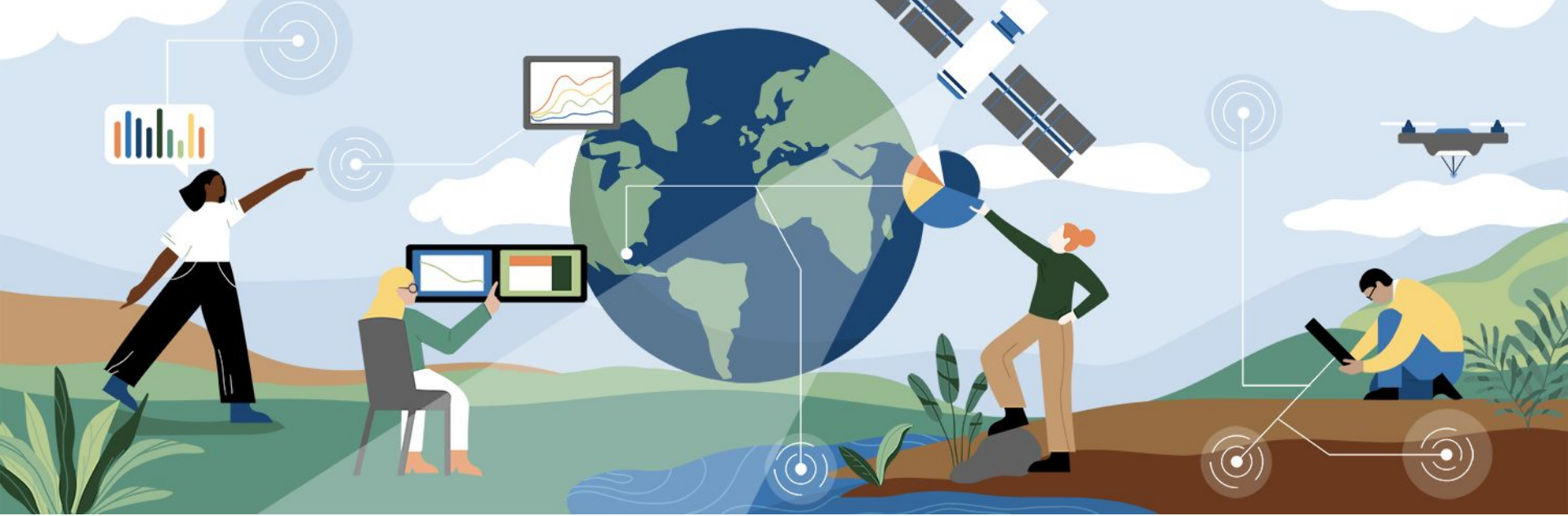
Show 2 more items in this data set

TOOL: Data Search Webpage

DETAILS:

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

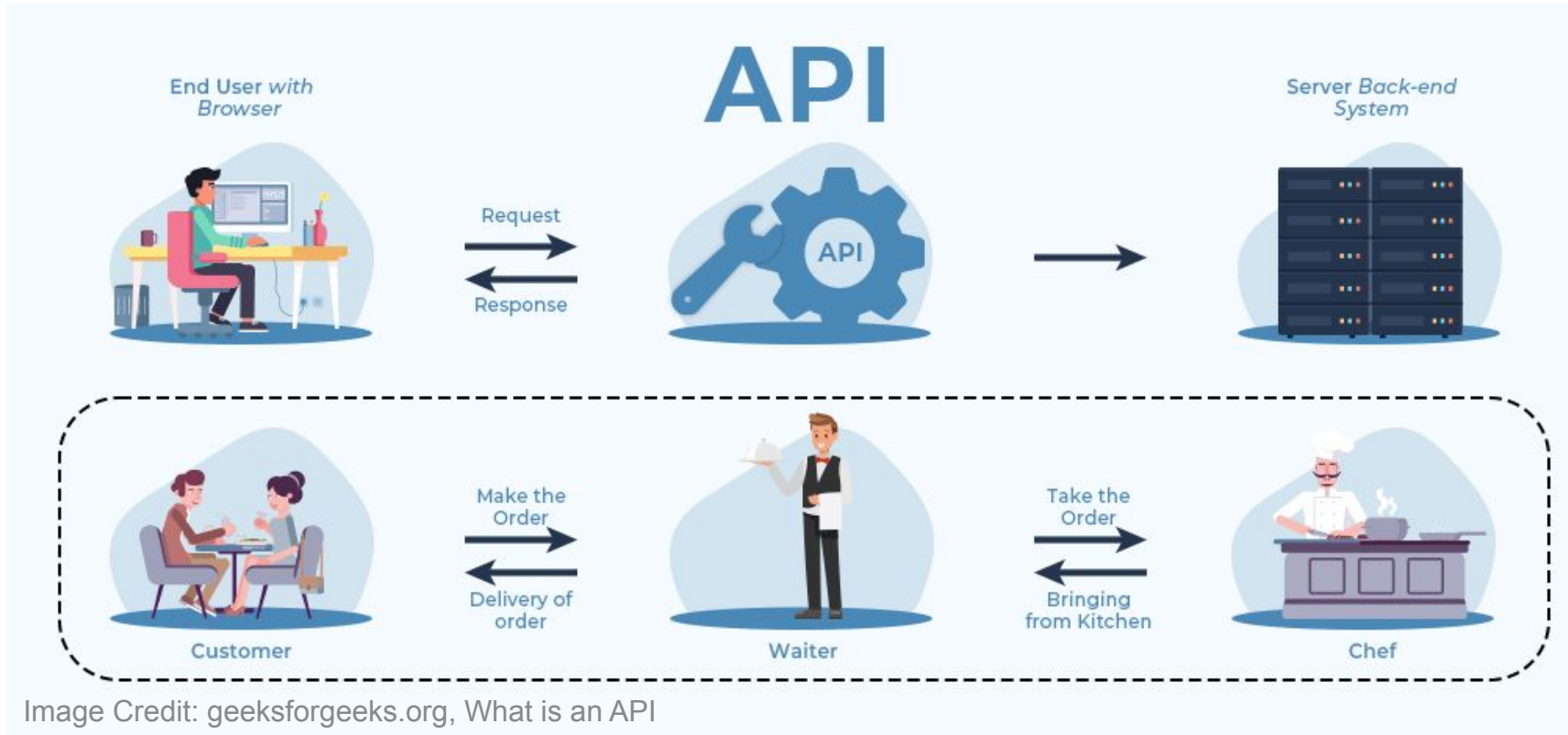
What did we find?: 11 datasets that I need to read through that may or may not have the variables I'm interested in



Questions?

What is the **Dataset API**?

What is an API?



Data Search: Dataset API



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, it's 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. 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

GET /packages/{identifier} Download dataset metadata

PUT /packages/{identifier} Update existing dataset

GET /packages Search for datasets



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

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 and the 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 usual in a data 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](#))
- [Jupyter Notebooks](#) that demonstrate usage of operations in the Dataset API GitHub Repository

Dataset Operations relating to datasets

- GET** `/packages/{identifier}` Download dataset metadata
- PUT** `/packages/{identifier}` Update existing dataset
- GET** `/packages` 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
- `providerName`: The dataset project/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> integer (query)	The row number to start on. Use this for paging results Default value : 1
<code>pageSize</code> integer (query)	The number of datasets to return per request. Default value : 25
<code>isPublic</code> boolean (query)	If set with true, would only return public packages.
<code>creator</code> string (query)	The creator/submitter of datasets

Data Search: Dataset API



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 operation and the 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 usual in a data 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 scripts](#))
- [Jupyter Notebooks](#) that demonstrate usage of operations in the Dataset API GitHub Repository

Dataset Operations relating to datasets

GET `/packages/{identifier}` Download dataset metadata

PUT `/packages/{identifier}` Update existing dataset

GET `/packages` 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
- providerName: The dataset project/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
rowStart	The row number to start on. Use this for paging results
integer (query)	Default value : 1
1	
pageSize	The number of datasets to return per request.
integer (query)	Default value : 25
25	
isPublic	If set with true, would only return public packages.
boolean (query)	-- v
creator	The creator/submitter of datasets
string (query)	

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%202022%20Using%20Data%20-%20Python.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.

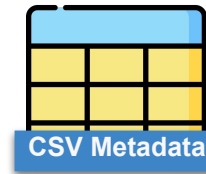
These data standards were developed by ESS-DIVE and **ESS-DIVE Partner Projects** to standardize metadata and data files of data types commonly collected by DOE ESS projects.

No. of Reporting Formats available:

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



Velliquette, Heinz, Devarakonda (ORNL)



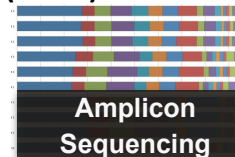
Weisenhorn (ANL)



Serbin, Ely (BNL)



Boye (SLAC)



Weisenhorn (ANL)



Goldman (PNNL)

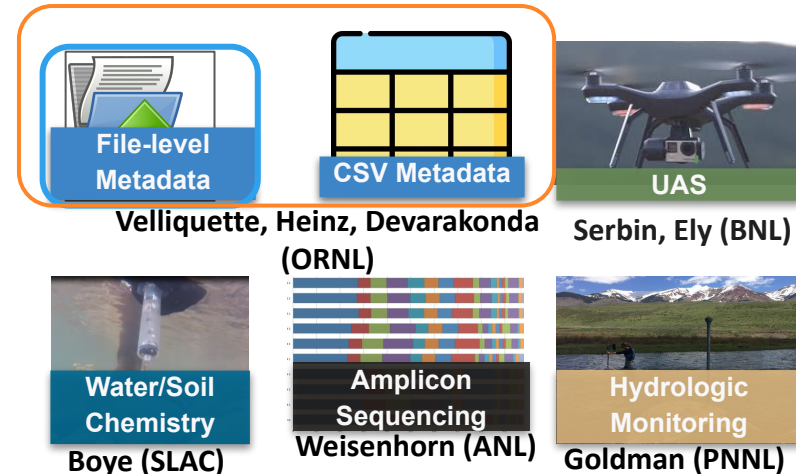
Reporting Formats

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**



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

[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, Kougarak 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-e9c7bc5f755e9d6-20230808T210448475

							Download All
	Name		File type	Size			Download
	Metadata: Leaf_Nitrogen_and_Carbon_Content_and_Leaf_Mass.xml		EML v2.2.0	9 KB			Download
	NGA207_flmd.csv	More info	text/csv	2 KB	5 downloads		Download
	Seward_USDA_PlantSymbols_dd.csv	More info	text/csv	405 B	1 download		Download
	Seward_2018_SampleDetails_dd.csv	More info	text/csv	563 B	2 downloads		Download
	Seward_2018_LMA_LWC_CN_dd.csv	More info	text/csv	683 B	1 download		Download
	Seward_2018_LMA_LWC_CN.csv	More info	text/csv	1 KB	1 download		Download

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_Sample_Details.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_Sample_Details_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	COMPI
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 each burn	Plot numeric value identifying a unit	integer
6	DATE	yyyy-mm-dd		Burn date	date
7	BTZ_START	hh:mm:ss	Zulu time when IR sensor first detected a burn	Zulu (Coordinated Universal Time) time	datetime
8	BTZ_END	hh:mm:ss	Zulu time when IR sensor last detected a burn	Zulu (Coordinated Universal Time) 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>

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

Inspecting Dataset Contents Demo

https://github.com/ess-dive/essdive-tutorials/blob/main/search_data/ESS%20PI%20Meeting%202022%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 X

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; McKeever 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

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](https://data.ess-dive.lbl.gov/datasets/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

Recent Greenhouse Gas Concentrations

T J Blasing

Downloads

1.2K

Citations

48

Views

9.5K

Cite this dataset

Assessment report

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	Download
Current_Greenhouse_Gas_Concentrations_fgdc.xml	XML Application	11 KB 306 downloads	Download
Recent_Greenhouse_fgdc.xml	XML Application	7 KB 107 downloads	Download



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

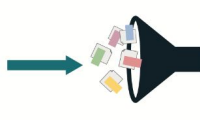
Deep Dive locates and extracts data from published datasets that have **completed File Level Metadata**

ESS-DIVE
ARCHIVE



DATASET MONITORING

standardized
datasets



**Advanced
search within
data files**

- adopts reporting format

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)

doi array[string] <i>(query)</i> <i>maxLength: 100</i> <i>minLength: 1</i>	The digital object identifier (doi) representing a dataset <input type="text" value="doi:10.15485/1962818"/> <input type="button" value="Add string item"/>
fieldName string <i>(query)</i> <i>maxLength: 100</i> <i>minLength: 1</i>	The field name to search for. <input type="text" value="stream"/>
recordCountMin integer <i>(query)</i>	Filter by record count greater than or equal to. <input type="text" value="500"/>
recordCountMax integer <i>(query)</i>	Filter by record count less than or equal to. <input type="text" value="recordCountMax"/>
fieldValueText string <i>(query)</i>	Filter by a text field value. Search is case insensitive <input type="text" value="fieldValueText"/>
fieldValueNumeric <i>(query)</i>	Filter by a numeric value that is between min and max summary values. <input type="text" value="fieldValueNumeric"/>
fieldValueDate <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">• Zipped files• No descriptions• No previews <p>Manual inspection</p>	<p>Dataset Details</p> <p>Lists all dataset files</p> <p>Potential challenges:</p> <ul style="list-style-type: none">• No descriptions• No previews• Long lists <p>Programmatic</p>	<p>With Structured Data (Reporting Formats)</p> <p>Lists dataset files Descriptions of files Enables file preview Can summarize files Potential challenges:</p> <ul style="list-style-type: none">• Does not search by field/variable <p>Programmatic</p>	<p>With Structured Data (Reporting Formats)</p> <p>Searching the data within dataset files</p> <p>Potential challenges:</p> <ul style="list-style-type: none">• Limited datasets <p>Programmatic</p>



Questions?

Outcomes

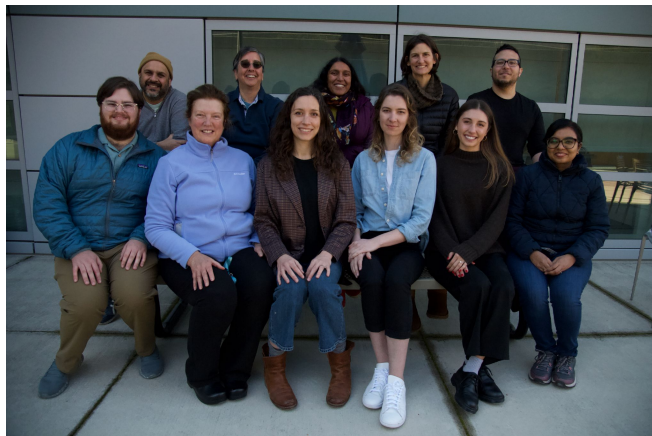
1. Search for Data

- 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

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



COMPASS
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 True	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 True	Subsurface electrical conductivity across the BEO site inferred using a capacitively coupled resistivity survey in May 2013, Utqiagvik, Alaska	2024	doi:10.15485/2335800
2 True	Representativeness-based Sampling Network Design for the State of Alaska	2013	doi:10.5440/1108686
3 True	Time-lapse photography at BEO, Utqiagvik (Barrow), Alaska, 2014	2016	doi:10.5440/1999388
4 True	Surface and Active Layer Pore Water Chemistry from Ice Wedge Polygons, Utqiagvik (Barrow), Alaska, 2013-2014	2017	doi:10.5440/1226245
5 True	Soil Water Characteristics of Cores from Low- and High-Centered Polygons, Utqiagvik (Barrow), Alaska, 2012	2017	doi:10.5440/1299259
6 True	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.

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)