



# Expand the Discoverability of your ESS-DIVE Data through External Linking

**Joan Damerow**

Community Engagement Lead Scientist



February 2022 Webinar

# What are your external linking needs?

I have published one or more dataset(s) elsewhere (e.g. project data archive) that need to be mirrored on ESS-DIVE. **A**

I need to link my dataset to specific files that are published elsewhere (e.g. a data analysis platform). **B**

I would like to link my dataset to related methods, samples, data publications, journal publications. **C**



# Why are you interested in linking ESS-DIVE datasets to external sources?

To comply with funding requirements

My project is jointly funded by DOE ESS and another funder that requires publishing elsewhere

To make my ESS dataset more discoverable across repositories

I have a data type that fits scientifically with another repository

I have processed or analyzed data on another platform that would be scientifically useful to link to

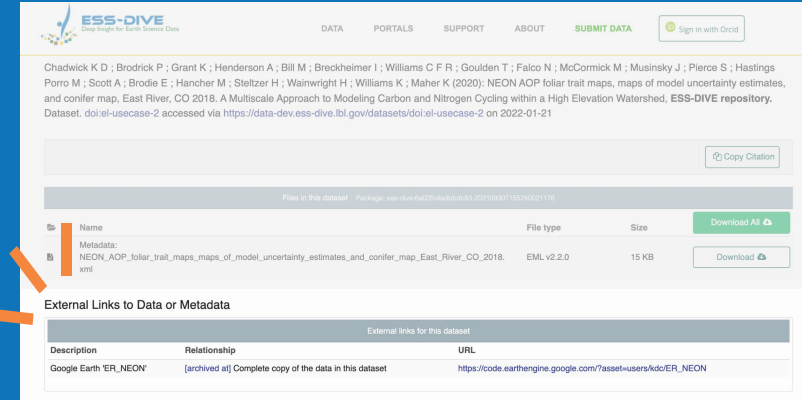


# Objectives

- Review current use cases for external linking:
  - Link to the 1) **original publication of a dataset** (e.g. your project's data archive) where metadata and data can be found
  - Link to 2) **complete copies of your data** or 3) **individual data files** stored elsewhere
- How to Request Linking External Data to ESS-DIVE Dataset
- **Feedback on additional needs** for linking to external data, methods, samples, and publications.

*Learn about ESS-DIVE's new approach and capability to link your ESS-DIVE data to other online repositories and data systems!*

# \*NEW\* External Linking



ESS-DIVE  
Deep Insights for Earth Science Data

DATA PORTALS SUPPORT ABOUT **SUBMIT DATA** Sign in with Orcid

Chadwick K D ; Brodrick P ; Grant K ; Henderson A ; Bill M ; Breckheimer I ; Williams C F R ; Goulden T ; Falco N ; McCormick M ; Musinsky J ; Pierce S ; Hastings Porro M ; Scott A ; Brodie E ; Hancher M ; Steltzer H ; Wainwright H ; Williams K ; Maher K (2020): NEON AOP foliar trait maps, maps of model uncertainty estimates, and conifer map, East River, CO 2018. A Multiscale Approach to Modeling Carbon and Nitrogen Cycling within a High Elevation Watershed, **ESS-DIVE repository**. Dataset. doi:el-usecase-2 accessed via <https://data-dev.ess-dive.lbl.gov/datasets/doi:el-usecase-2> on 2022-01-21

[Copy Citation](#)

Files in this dataset Package: ess-dive-6d020e4e8b0a0530210b071020d0031176

| Name  | File type  | Size  | Download All             |
|---|------------|-------|--------------------------|
| Metadata:   |            |       |                          |
| NEON_AOP_foliar_trait_maps_maps_of_model_uncertainty_estimates_and_conifer_map_East_River_CO_2018.xml | EML v2.2.0 | 15 KB | <a href="#">Download</a> |

External Links to Data or Metadata

| External links for this dataset |   |   |
|---------------------------------|---|---|
| Description                     | Relationship  | URL   |
| Google Earth 'ER_NEON'          | [archived at] Complete copy of the data in this dataset | <a href="https://code.earthengine.google.com/?asset-users/do/ER_NEON">https://code.earthengine.google.com/?asset-users/do/ER_NEON</a> |

# What is “External Linking”?

Data stored on other repositories are directly linked to dataset metadata on ESS-DIVE and **some/all data do not have to be on ESS-DIVE**

Copy Citation

Files in this dataset Package: ess-dive-6a122b4adbbdc83-20210930T155240021176

| Name   | File type  | Size  | Download All |
|--|------------|-------|--------------|
| Metadata:<br>NEON_AOP_foliar_trait_maps_maps_of_model_uncertainty_estimates_and_conifer_map_East_River_CO_2018.xml | EML v2.2.0 | 15 KB | Download     |

External Links to Data or Metadata

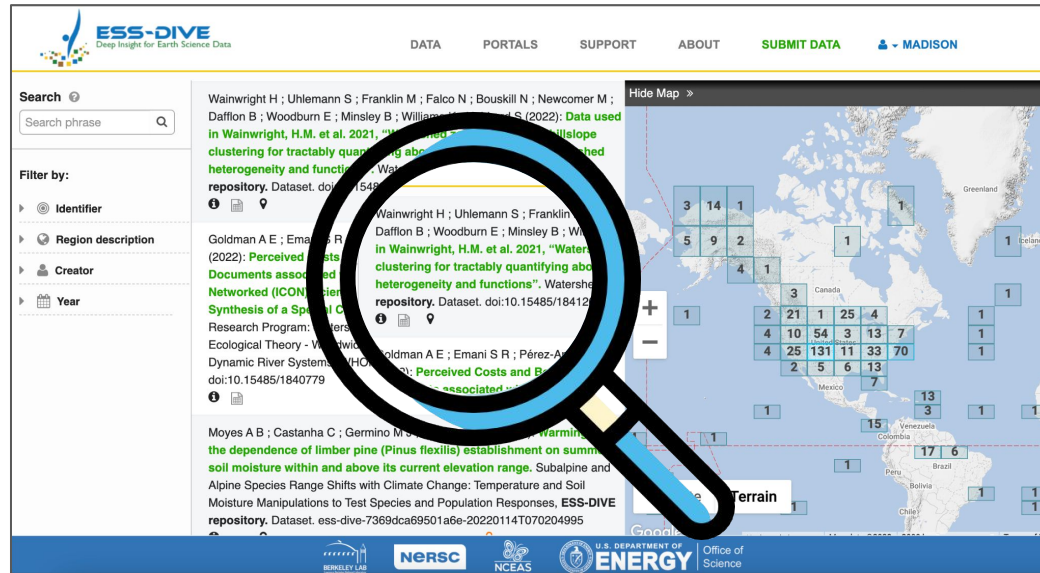
External links for this dataset

| Description            | Relationship  | URL   |
|------------------------|---|---|
| Google Earth 'ER_NEON' | [archived at] Complete copy of the data in this dataset | <a href="https://code.earthengine.google.com/?asset=users/kdc/ER_NEON">https://code.earthengine.google.com/?asset=users/kdc/ER_NEON</a> |

# What is “External Linking”?

- Store your data where it makes the most sense scientifically or for practical reasons
- **Metadata** will be on ESS-DIVE,
- Original dataset will be **searchable** on ESS-DIVE using metadata,
- Data can only be externally linked if stored on an approved repository
- External dataset metadata can be bundled into a **Data Portal**, and

# Meet ESS Publication Policy Requirements



- The **ESS publication policy requires** mirroring data on ESS-DIVE even if the data are located in another repository (e.g. project data archive)
- Makes data more findable and accessible long-term



# Use Cases: Why Externally Link Datasets?



# Relationships between ESS-DIVE datasets and externally linked data



| Relationship | Definition   |
|--------------|--|
| same as      | Original publication of this dataset (where the data+metadata can be found). The DOI url of this dataset, starting with ' <a href="https://doi.org/">https://doi.org/</a> ', redirects to original source. |
| archived at  | Complete copy of the data files in the dataset resides in external source.   |
| has part     | One or more files that are part of the dataset held outside of the repository. This could be a link to an individual file or a directory.  |

## Schema.org - EML, JSON-LD

### EML annotations - semantic triple

([subject] [predicate] [object])

- subject: The dataset
- predicate: 'has part', 'same as' or 'archived at'
- object: The externally linked resource

[DataONE Community Github issue](#)

**Handled by ESS-DIVE Team**

**\*Important to enable provenance tracking, and updating relevant information when needed**

# Use Cases to Store your Data Externally

Scientific, practical, or other funding compliance reasons

## 1) Original Dataset Publication

same as

## 2) Complete Copy of Data Files

archived at

- Large Project Data Archives
  - SPRUCE
  - Watershed Function SFA
  - NGEE Tropics
- Approved Data Repositories
  - EDI
  - USGS ScienceBase

## 3) Individual Files

has part

- Data Analysis Platform
  - Github
  - Google Earth Engine
- Large Data
  - File size
  - Number of files

# Use Case 1: Original Publication of Dataset External (e.g. Project Data Archive)

same as



Smith R ; Nelson P ; Jovan S ; Hanson P ; McCune B (2018): SPRUCE Epiphytic Lichen Annual Biomass Growth in Experimental Plots, 2013-2016. Climate Change-Terrestrial Ecosystem Science SFA. doi:el-usecase-3

Copy Citation

Edit

Publish

Files in this dataset Package: ess-dive-51d23d68ecd5642-20210930T144927611602

| Name  | File type  | Size  | Download All |
|---|------------|-------|--------------|
| Metadata:<br>SPRUCE_Epiphytic_Lichen_Annual_Biomass_Growth_in_Experimental_Plots_2013_2016_.xml | EML v2.2.0 | 10 KB | Download     |

## External Links to Data or Metadata

| External links for this dataset |   |   |   |
|---------------------------------|---|---|---|
| Label                           | URL   | Metadata Scheme   | Relationship  |
| doi:10.25581/spruce.048/1425889 | <a href="https://doi.org/10.25581/spruce.048/1425889">https://doi.org/10.25581/spruce.048/1425889</a> | <a href="https://schema.org/sameAs">https://schema.org/sameAs</a> | Original publication of this dataset (where the data+metadata can be found) |

DOI url (e.g. <https://doi.org/10.3334/CDIAC/SPRUCE.042>) will direct to original publication

# Use Case 2: Complete Copy of the Data Files External

archived at



Copy Citation

Files in this dataset Package: ess-dive-6af22b4adbbdc83-20210930T155240021176

| Name   | File type  | Size  | Download All |
|--|------------|-------|--------------|
| Metadata:<br>NEON_AOP_foliar_trait_maps_maps_of_model_uncertainty_estimates_and_conifer_map_East_River_CO_2018.xml | EML v2.2.0 | 15 KB | Download     |

External Links to Data or Metadata

| Description            | Relationship  | URL   |
|------------------------|---|---|
| Google Earth 'ER_NEON' | [archived at] Complete copy of the data in this dataset | <a href="https://code.earthengine.google.com/?asset=users/kdc/ER_NEON">https://code.earthengine.google.com/?asset=users/kdc/ER_NEON</a> |

- Only metadata in ESS-DIVE - make data more findable
- Avoid the need to update data files in multiple locations

# Use Case 3: Individual Files External (e.g. Analysis Platform)

hasPart



Brodrick P ; Goulden T ; Chadwick K D (2020): Custom NEON AOP reflectance mosaics and maps of shade masks, canopy water content. Watershed Function SFA. doi:el-usecase-1.1

Copy Citation

Edit

Publish

Files in this dataset Package: ess-dive-abaaf5bbdac34e6-20210930T145833351466

| Name  | File type  | Size  | Download All |
|---|------------|-------|--------------|
| Metadata:<br>Custom_NEON_AOP_reflectance_mosaics_and_maps_of_shade_masks_canopy_water_content.xml | EML v2.2.0 | 18 KB | Download     |

## External Links to Data or Metadata

| External links for this dataset          |   |   |                                    |
|--|---|---|------------------------------------|
| Label                                    | URL   | Metadata Scheme   | Relationship                       |
| orthomosaiced estimated reflectance data | <a href="https://portal.nersc.gov/wfsfa/doi-10-15485-16181314/">https://portal.nersc.gov/wfsfa/doi-10-15485-16181314/</a> | <a href="https://schema.org/hasPart">https://schema.org/hasPart</a> | Data that is part of this dataset. |

Some data in ESS-DIVE and one or more data files are stored externally

# Any Questions?



# How to Request Linking External Data to ESS-DIVE Dataset



# How to Create an Externally Linked Dataset

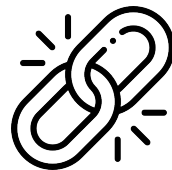
1. **Email** [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov) to start a request
2. **Send any existing DOI(s)** to ESS-DIVE Team
3. If there is an existing DOI, **the ESS-DIVE Team** uses a programmatic tool to transfer the metadata from the original repository (e.g. EDI) to new ESS-DIVE data packages
4. If there is NOT an existing DOI, you will create a new dataset for your data and send the associated ESS-DIVE identifier to the ESS-DIVE Team  
(e.g. `ess-dive-ea25aaddf4b47a3-20211103T152543118`).



# How to Create an Externally Linked Dataset

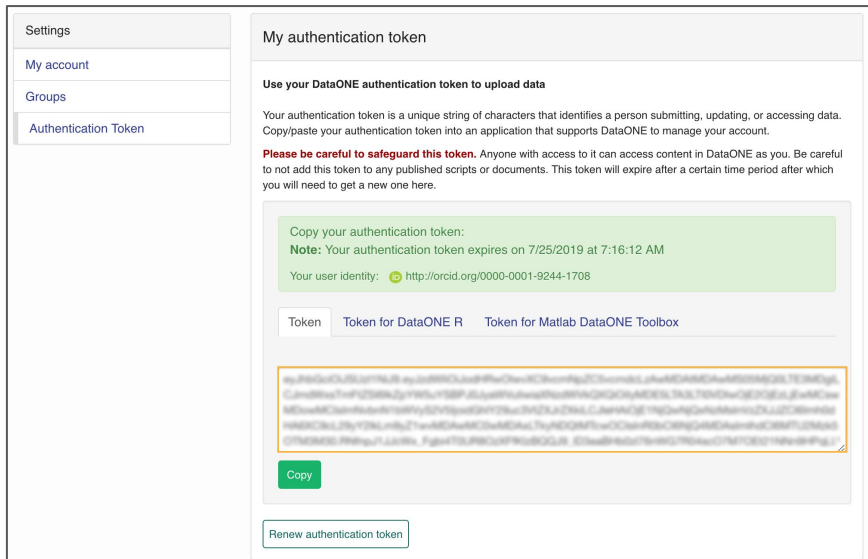


6. **The ESS-DIVE Team** will populate the External Links table in your dataset
7. **The ESS-DIVE Team** will send you the dataset URL and give you ownership
8. **Review the dataset metadata** to ensure that it's complete.
  - a. Programmatic metadata **transfers are not comprehensive** and there will be missing information that you need to complete
9. **Request** to publish the dataset



# Use the Package Service API to Link External Data to ESS-DIVE Metadata

# First, setup your script to create a dataset

A screenshot of the 'My authentication token' page in the DataONE interface. The left sidebar shows a menu with 'Settings', 'My account', 'Groups', and 'Authentication Token' (which is highlighted). The main content area is titled 'My authentication token' and contains instructions on how to use the token. It includes a green box with a 'Copy' button and a 'Renew authentication token' button at the bottom. The token itself is displayed as a long string of characters, with a yellow box highlighting it and a 'Copy' button next to it.

Settings

My account

Groups

Authentication Token

My authentication token


Use your DataONE authentication token to upload data

Your authentication token is a unique string of characters that identifies a person submitting, updating, or accessing data. Copy/paste your authentication token into an application that supports DataONE to manage your account.


**Please be careful to safeguard this token.** Anyone with access to it can access content in DataONE as you. Be careful to not add this token to any published scripts or documents. This token will expire after a certain time period after which you will need to get a new one here.

Copy your authentication token:

**Note:** Your authentication token expires on 7/25/2019 at 7:16:12 AM

Your user identity:  <http://orcid.org/0000-0001-9244-1708>

Token Token for DataONE R Token for Matlab DataONE Toolbox



Copy

Renew authentication token

1. Copy your Authentication Token
  - a. **NOTE:** if you have not registered as a data contributor, your token will not allow you to create or edit datasets with the API.
2. Copy code to setup environment

```
1 import requests
2 import os
3 import json
4
5 token = "<Enter your authorization token here>"
6 base = "https://api-sandbox.ess-dive.lbl.gov/"
7 header_authorization = "bearer {}".format(token)
8 endpoint = "packages"
```

Code is available at [docs.ess-dive.lbl.gov](https://docs.ess-dive.lbl.gov)

# First, setup your script to create a dataset

```
1 json_ld = {  
2   "@context": "http://schema.org/",  
3   "@type": "Dataset",  
4   "@id": "http://dx.doi.org/10.3334/CDIAC/spruce.001",  
5   "name": "SPRUCE S1 Bog Environmental Monitoring Data: 2010-2016",  
6   "keywords": [  
7     "EARTH SCIENCE > BIOSPHERE > VEGETATION",  
8     "Climate Change"  
9   ],  
10  "license": "http://creativecommons.org/licenses/by/4.0/",  
11  "funder": {  
12    "@id": "http://dx.doi.org/10.13039/100006206",  
13    "name": "U.S. DOE > Office of Science > Biological and Environmental Research"  
14  },  
15  "provider": provider_spruce,  
16 }
```

3. Copy the example **JSON-LD** code and replace the example text with your dataset metadata

# Then, add external link relationships to your dataset JSON-LD

|                   |                           |
|-------------------|---------------------------|
| <b>hasPart</b>    | <b>HasPart</b> > {...}    |
| <b>archivedAt</b> | <b>ArchivedAt</b> > {...} |
| <b>identifier</b> | <b>Identifier</b> > {...} |
| <b>sameAs</b>     | <b>SameAs</b> > {...}     |

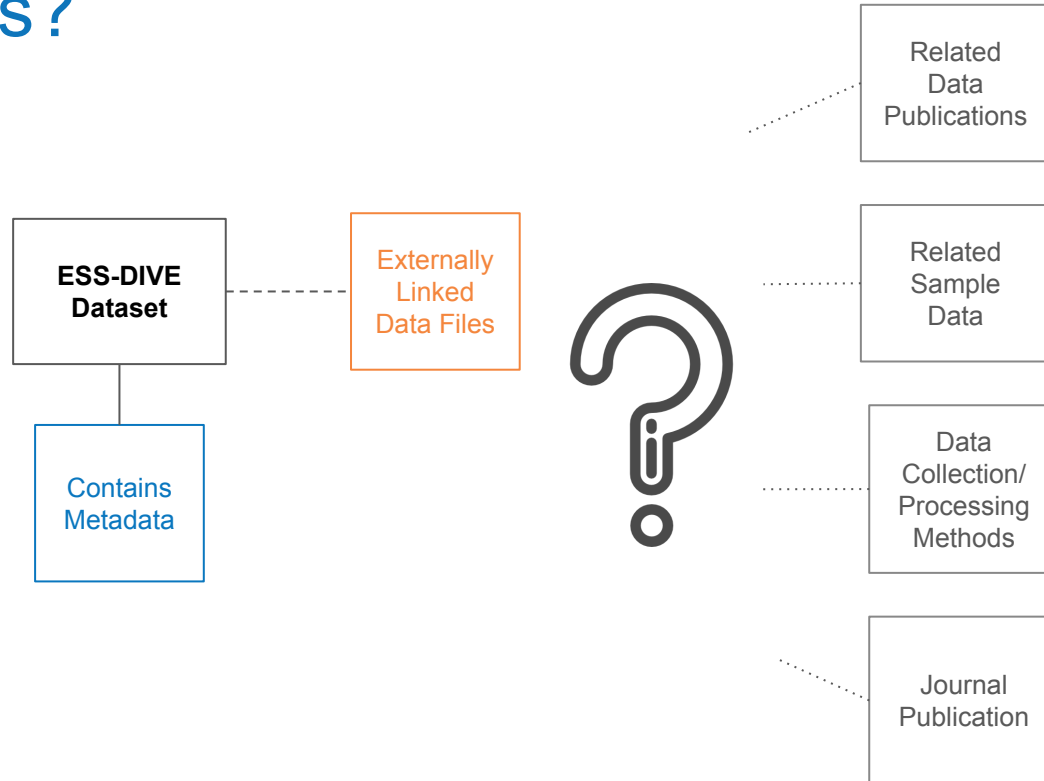
```
1 post_packages_url = "{}{}".format(base,endpoint)
2 post_package_response = requests.post(post_packages_url,
3                                     headers={"Authorization":header_authorization},
4                                     json=json_ld)
5
6 if post_package_response.status_code == 201:
7     # Success
8     response=post_package_response.json()
9     print(f"View URL:{response['viewUrl']}")
10    print(f"Name:{response['dataset']['name']}")
11 else:
12     # There was an error
13     print(post_package_response.text)
```

4. Select which **external linking relationship** matches your use case and add it to your JSON-LD
5. Copy the code that will execute your script and run it!

Code is available at [docs.ess-dive.lbl.gov](https://docs.ess-dive.lbl.gov)

**\*Feedback\*** Tell us what you need!

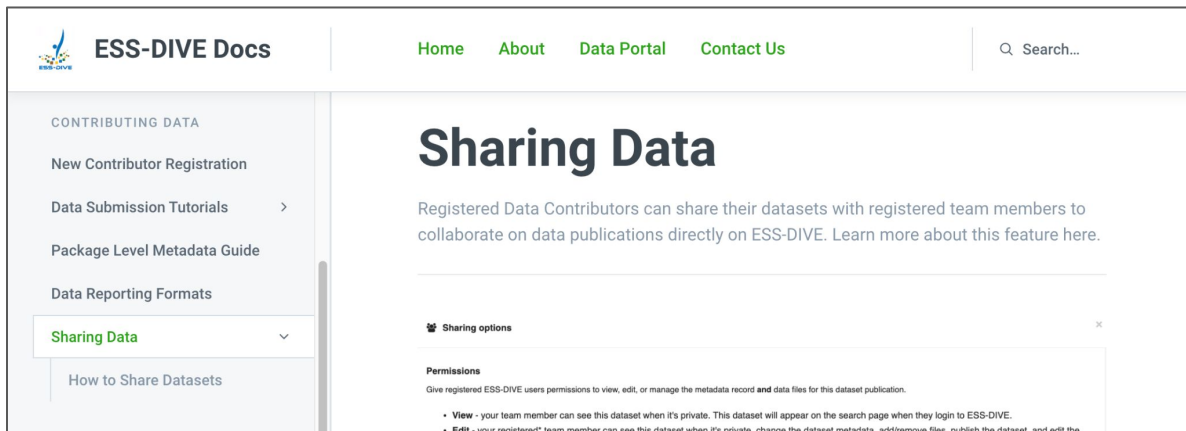
# What about external linking to other related resources?



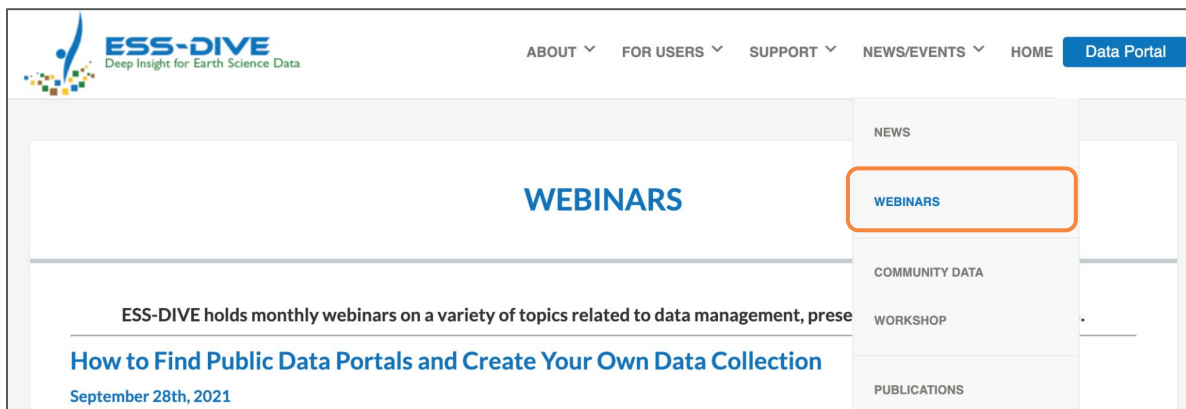


# Linked Data Resources

# You can find all this material on ESS-DIVE's...



**COMING SOON:**  
**External Linking**  
**Documentation page**  
<https://docs.ess-dive.lbl.gov/>



**Webinar page**  
where this slide deck is  
available for download  
<https://ess-dive.lbl.gov/>

# Summary



**ESS-DIVE has a new feature that allows you to mirror ESS project data that is on another repository or data analysis platform**

- Store your data where it makes sense, for scientific or practical purposes
- ESS program **requires** mirroring data on ESS-DIVE
- External Linking **reduces redundant effort**
- Makes data more findable and accessible

*Make a request to link external data today! [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov)*

# Thank You!



**Join ESS-DIVE's Community Mailing List!**

<http://bit.ly/essdiveMailingList>

**Contact us at** [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov)