**How to Publish a Manuscript Associated Data Package**

River Corridor SFA Data Management Resource

**Questions?** Contact [Amy.Goldman@pnnl.gov](mailto:Amy.Goldman@pnnl.gov) and [Brieanne.Forbes@pnnl.gov](mailto:Brieanne.Forbes@pnnl.gov)

**Document purpose:** To provide simple, clear guidance on how to put together a data package associated with a specific manuscript. This document will not cover every possible combination of outcomes or optional components. The goal is to make this whole process less overwhelming.

# Contact RC SFA ESS-DIVE Expert

Contact [Amy.Goldman@pnnl.gov](mailto:Amy.Goldman@pnnl.gov) and [Brieanne.Forbes@pnnl.gov](mailto:Brieanne.Forbes@pnnl.gov) to tell them you are preparing a manuscript data package. They will orient you to the process and the expectations and identify places they can help. Amy and Brie will review your data package prior to submission to ESS-DIVE, so when you contact them, be prepared to communicate the date you are aiming to have your data package submitted. The date should be a month or more into the future.

# Identify what will go into your data package

## Data

Include any data used in the manuscript analyses or interpretations. If some of the data are public, you may choose to reference the data rather than include a copy in your data package. The data you include may be data model inputs/outputs, sensor data, sample data, or other data. Please note that there are certain formatting requirements for some data types. It is important to discuss these as part of Step #1. All tabular data should be saved as .csv. Remember that column headers cannot have spaces or special characters that are not underscores. Remember that the same column header cannot be used twice but have two different meanings (i.e., you cannot have “CD” and “cd” columns for count of dogs and count of dolphins in the same data package even if they are in different files). It is also good practice to have the same column name when multiple columns contain the same information (i.e., CD for count of dogs in two files. Not CD in one file and CoDo in another file).

## Code

You can choose to include code that is important to the manuscript generation. This might be model code, analysis code, or figure generation code. You can also reference code on GitHub, but we recommend including code in the data package so the user has the static version used in the paper along with all the other manuscript pieces. If you choose to reference the code on GitHub, we recommend doing a GitHub commit hash so that there is a static release you are pointing to. [If that last sentence doesn’t fully make sense, it’s because Amy doesn’t really know what that means, but she’s hoping you do.] When including code in a data package, remember this code will be opened by future data users. It should be commented and clear. The top of the code should state what it is and preferably have a date and contact information.

## Figures

You can choose to include figures important to the manuscript generation.

# Make changes to your files to make them usable by a future data user

Look at all your files. Consider if they would actually make sense to a future data user with no previous knowledge. Consider if the column headers comply with the rules lists in [Step 2.1](#_Data). Consider if the file names would make sense to a future data user. Make changes accordingly. Please be sure to reflect any changes to data files in the scripts so that they are able to run with the data included in the data package.

# Prepare four **required** metadata files

**ALL** data packages – model, sensor, sample, etc. - require four specific metadata files.

## ESS-DIVE data package metadata

This is the metadata displayed on the ESS-DIVE website landing page when you access a data package. It includes things like abstract, authors, keywords, data license, and more. There is an [RC SFA template](https://pnnl.sharepoint.com/:w:/r/teams/SubsurfaceBiogeochemicalResearchSFA/_layouts/15/Doc.aspx?sourcedoc=%7B3ECA4949-F2A5-44A6-9B54-AA86268F1C35%7D&file=ESSDIVE_Metadata_Template.docx&action=default&mobileredirect=true) for you to fill out. Using the template is important.

## readme

This is a readme that will go into the data package as a pdf. It will provide some high-level information about your data package, including abstract, data package navigation, contact, change history, and more. There is an [RC SFA template](https://pnnl.sharepoint.com/:w:/r/teams/SubsurfaceBiogeochemicalResearchSFA/_layouts/15/Doc.aspx?sourcedoc=%7B620670A6-20D9-45D5-BA49-3BDA9999B396%7D&file=readme_template.docx&action=default&mobileredirect=true) for you to fill out. Using the template is important.

*\* The readme and ESS-DIVE data package metadata contain some similar information but serve different purposes. Both are very necessary pieces.*

## File-level metadata (flmd)

This is a csv file that lists all of the files in your data package with an accompanying description of each file. Creating an flmd is part of complying with the ESS-DIVE Reporting Formats. You can generate this by [using a template](https://pnnl.sharepoint.com/:x:/r/teams/SubsurfaceBiogeochemicalResearchSFA/Shared%20Documents/General/SFA%20Data%20and%20Software%20Management/Data-Publishing/template_flmd.csv?d=weaba923acd574dde81346457cee651ee&csf=1&web=1&e=X08oGh) or by running an R script (contact [Brieanne.Forbes@pnnl.gov](mailto:Brieanne.Forbes@pnnl.gov) to run the script) and then filling out the descriptions for the resulting file. See a filled-out example [here](https://pnnl.sharepoint.com/:x:/r/teams/SubsurfaceBiogeochemicalResearchSFA/Shared%20Documents/General/SFA%20Data%20and%20Software%20Management/Files-to-Comply-with-ESSDIVE-Reporting-Formats/example_flmd.csv?d=w8dfbd747f9ba42b6ba34bdc9da7a1010&csf=1&web=1&e=DxNMs4) before you get started.

## Data dictionary (dd)

This is a csv file that defines any column or row headers in csv files contained in your data package. You may decide to define other terms here also. Creating a dd is part of complying with the ESS-DIVE Reporting Formats. You can generate this by [using a template](https://pnnl.sharepoint.com/:x:/r/teams/SubsurfaceBiogeochemicalResearchSFA/Shared%20Documents/General/SFA%20Data%20and%20Software%20Management/Data-Publishing/template_dd.csv?d=wf6417d9f6b5048c6af69936c621ea39b&csf=1&web=1&e=gxgUR5) or by running an R script (contact [Brieanne.Forbes@pnnl.gov](mailto:Brieanne.Forbes@pnnl.gov) to run the script) and then filling out the descriptions for the resulting file. See a filled-out example [here](https://pnnl.sharepoint.com/:x:/r/teams/SubsurfaceBiogeochemicalResearchSFA/Shared%20Documents/General/SFA%20Data%20and%20Software%20Management/Files-to-Comply-with-ESSDIVE-Reporting-Formats/example_dd.csv?d=wd1296a9e8c284394b8b32a89ed7d3133&csf=1&web=1&e=agovjg) before you get started. *Please note that if you think you do not have any csv files in your data package, you are incorrect, because at a minimum, the flmd file is a csv, and you need a dd for that file.*

# Ask RC SFA ESS-DIVE Expert to review data package

Data packages for the RC SFA submitted to ESS-DIVE should be reviewed before submission by Brieanne Forbes or Amy Goldman. They will provide feedback on items to revise. Contact [Amy.Goldman@pnnl.gov](mailto:Amy.Goldman@pnnl.gov) and [Brieanne.Forbes@pnnl.gov](mailto:Brieanne.Forbes@pnnl.gov) to tell them you are ready for them to review your data package. Expect the review to take at least two weeks.

# Make corrections to data package

Edit files according to the feedback provided in Step #5 and communicate with Brieanne Forbes or Amy Goldman to make sure that the edits address all of the feedback.

# Upload data package or ask ESS-DIVE Expert to upload data package

You can choose to upload your data package to ESS-DIVE on your own or Brieanne Forbes or Amy Goldman can upload it for you. Discuss with them. They have some pieces of this process automated.

# Frequently Asked Questions

1. **I am confused. Who can help explain this more?**

Amy Goldman and Brieanne (Brie) Forbes

1. **Can I ask for Amy and Brie to review my files earlier in the process?**

Yes. We are happy to do interim reviews and answer questions.

1. **Why are we doing all of this? It seems like a lot.**

Our funding agency requires that we use ESS-DIVE. The RC SFA agreed to comply with ESS-DIVE Reporting formats. These steps are necessary for that.

1. **Can I upload my data to Zenodo or somewhere else to avoid ESS-DIVE?**

No. You will need to put it on ESS-DIVE.

1. **I am a modeler. Can I just upload my scripts and normal files and Git readme and ignore everything else?**

No. The RC SFA agreed to comply with ESS-DIVE Reporting formats.

1. **I found an error in something I published on ESS-DIVE. Can I fix it?**

Yes, ESS-DIVE data packages can be revised after they have been published. Amy and Brie can help you do this correctly. We record changes in a change history.

1. **What are some helpful things I can keep in mind?**
   * When you write a description with an acronym, define the acronym.
   * When you write units in a description, write the words out fully.
   * Capitalize sentences.
   * Put periods at the ends of sentences.
   * Do not put spaces in your column headers.
   * Aim for consistency and clarity.
   * Avoid using commas in csv files. Use semicolons when needed.