Tutorial:
ESS-DIVE Model Data Guidelines

Madison Burrus
Data Management Assistant

ESS-DIVE BERKELEY LAB NERSC NCEAS DataONE U.S. DEPARTMENT OF ENERGY Office of Science

Tutorial Overview

- About Model Data Archiving Guidelines
- Components of a Model Data data package
- How to start your Model Data data package

About these Guidelines

These guidelines are the culmination of user-feedback forms and interviews with land modelers. Designed, conducted and synthesized by **Maegen Simmonds** at ESS-DIVE.

- Based on a **review of existing model archiving practices** and surveys of land modelers.
- Help modelers decide **how to organize and archive data** from their land model simulations.

Takeaways from this Tutorial

- Easy to follow, **3-step guide** to archiving model data
- Guidelines will help you fulfill **journal and funding requirements**
- Guidelines were developed **by modelers and are for modelers**

Archiving model data is complex, but the ESS-DIVE Team has resources to make the process simpler!

Components of a Model Data data package

Major Components

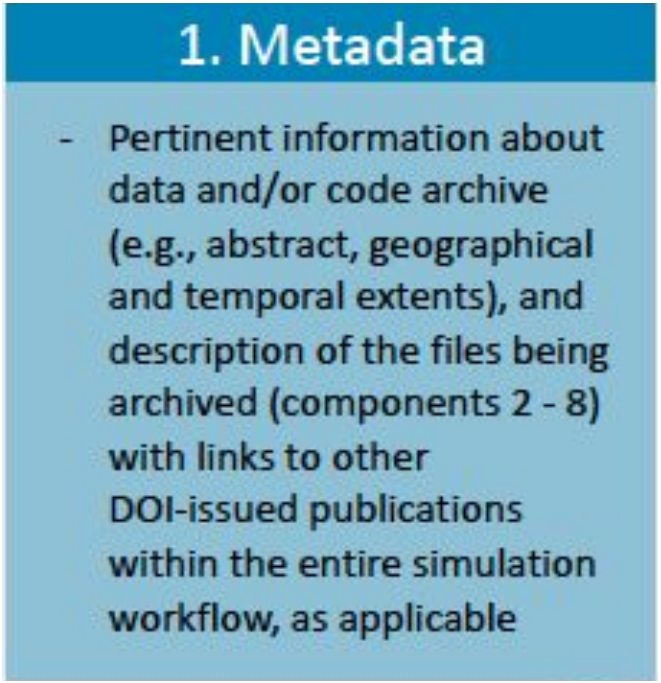
Every model data package will look different, but we recommend including these major components:

Metadata

Data Files

User Guide

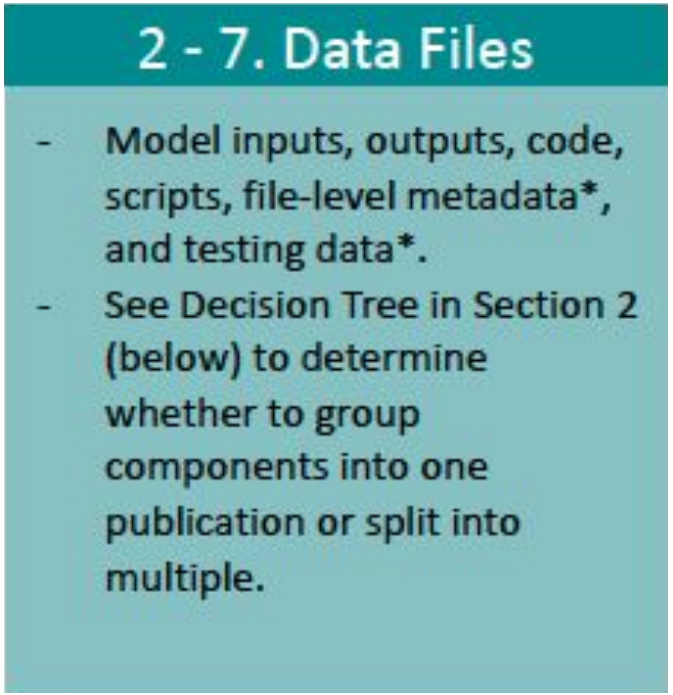
Major Components



Metadata

- Data Package Metadata
- Description of attached data files
- References to other DOI-issued publications within the workflow

Major Components



Data Files

- Model code and pre/post processing scripts
- Input/output files
- Testing data
- File-level Metadata (FLMD)

We are giving you a lot of options of files to include in your data package, but you decide what is needed to reuse your model

Major Components

8. User Guide

- Information required to operate the model(s)
- A workflow description of inputs, outputs, models, simulations, file naming;[2]
- *Description of the post-processing scripts;
- *Linking workflow to table and figure identification numbers

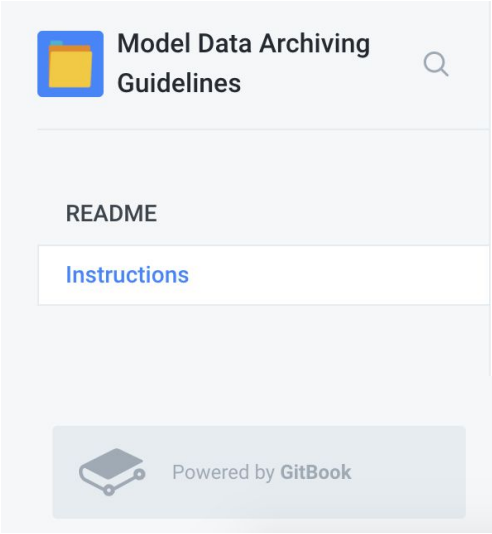
User Guide

- How to operate the model(s)
- Describe model workflow
- Describe model outputs

How to start your Model Data data package

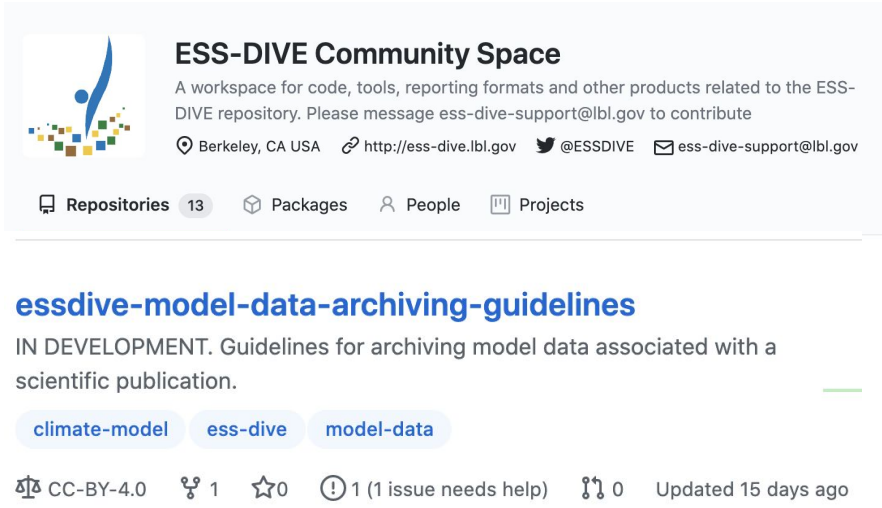
Open instructions in GitHub or Git Books

GitBook



https://ess-dive.gitbook.io/model-data-archiving-guidelines/instructions

GitHub



https://github.com/ess-dive-community/essdive-model-data-archiving-guidelines/blob/main/instructions.md

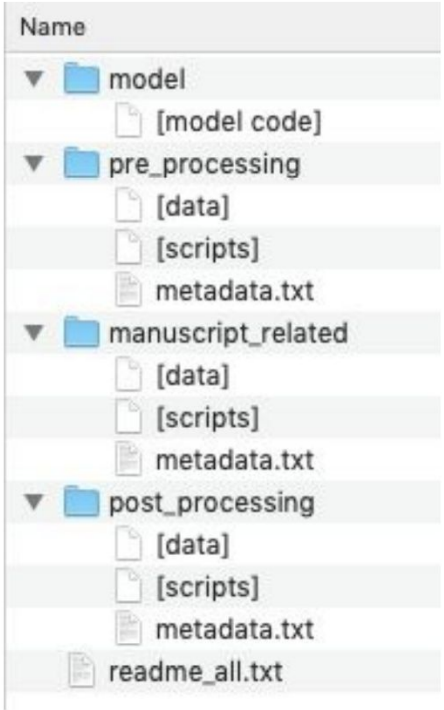
Step One:
Gather model data and associated files

This step helps determine:

- **Which** files to include in a data package
- **What** details each file will need
- **Recommendations** for file naming and formatting

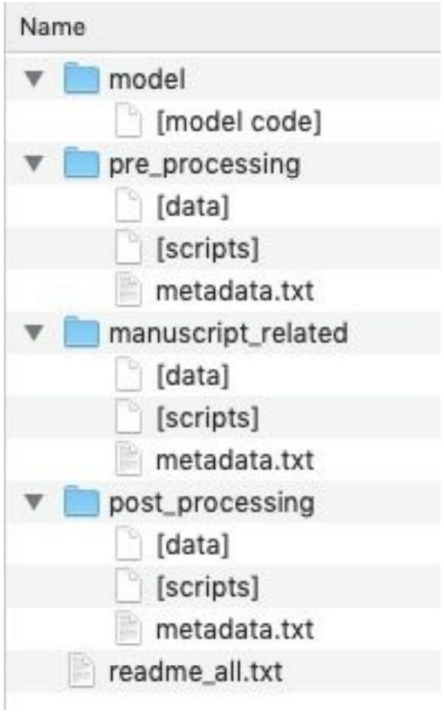
For the modelers in the room...

What are examples of data files that you may need to include in order to reuse your model code?



Data Files to Include

- Model Inputs
- Model Outputs
- Model Code
- Scripts
- Files Level Metadata
- Testing Data



Addressing Model Data Archiving Needs for the Department of Energy's Environmental Systems Science Community

PUB

Simmonds et al., 2020

Guidelines for publicly archiving terrestrial model data to enhance usability, intercomparison, and synthesis

Simmonds et al., 2021

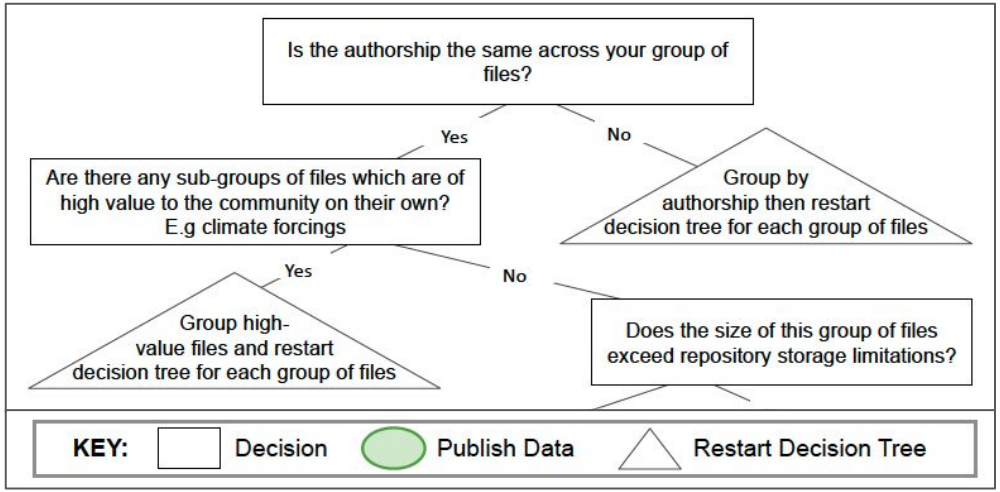
Step Two:
Decide how to bundle data files

As you are gathering your data files you may observe that:

- **Authorship** is not uniform across all data products
- The cumulative files size exceeds ESS-DIVE's **upload or storage capacity**
- Model output **downstream value** is not equivalent for all outputs

Decision Tree

These are some of the **most common reasons** why you may need to bundle data files into more than one data package.

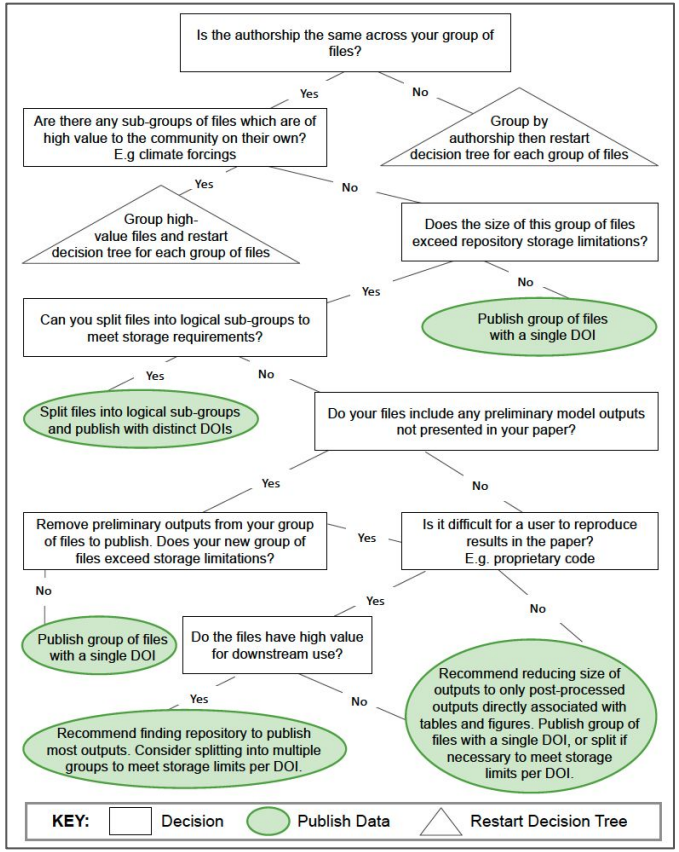


https://github.com/ess-dive-community/essdive-model-data-archiving-guidelines/blob/main/instructions.md#2-deciding-how-to-bundle-files

Decision Tree

These are some of the **most common reasons** why you may need to bundle data files into more than one data package.

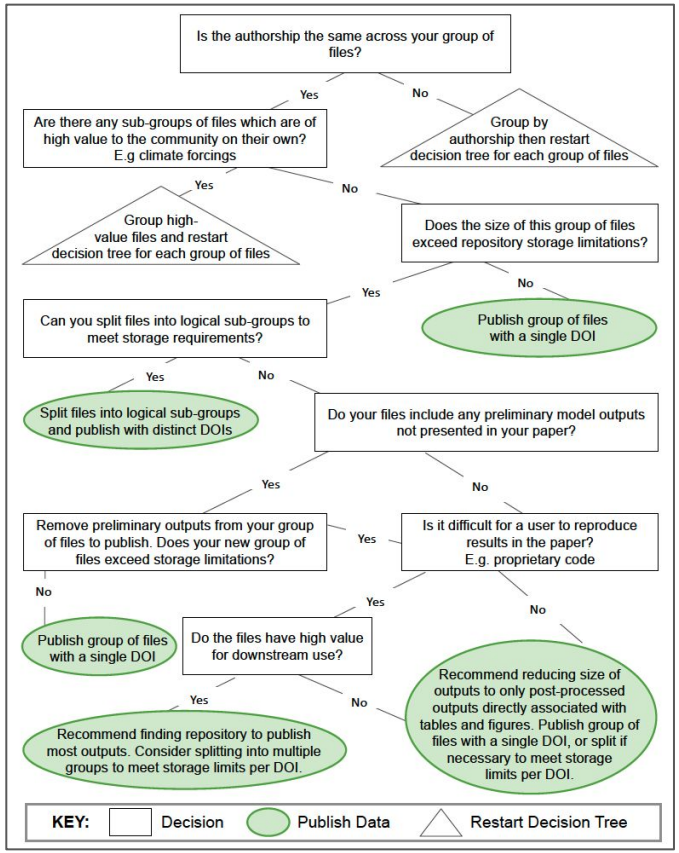
Use this decision tree to **learn when** to group or separate data files based on those three considerations.



https://github.com/ess-dive-community/essdive-model-data-archiving-guidelines/blob/main/instructions.md#2-deciding-how-to-bundle-files

Any other considerations when bundling model data?

- Authorship
- File size
- Downstream value
- ?



https://github.com/ess-dive-community/essdive-model-data-archiving-guidelines/blob/main/instructions.md#2-deciding-how-to-bundle-files

Step Three:
Include File-Level Metadata in data package

In the opening talk, **Terri Velliquette** walked us through FLMD reporting format…
why bring it up again?

File-level metadata (FLMD)[1]

FLMD is an optional component of files to include in data publications, which include two types of files. To ensure machine-readability of files, adhere to CSV reporting format[2].

FLMD is a component that appears across **all** the reporting formats!

https://github.com/ess-dive-community/essdive-model-data-archiving-guidelines/blob/main/instructions.md#3-file-level-metadata

Step Three:
Include File-Level Metadata in data package

File Level Metadata (FLMD) is a versatile **ESS-DIVE reporting format**

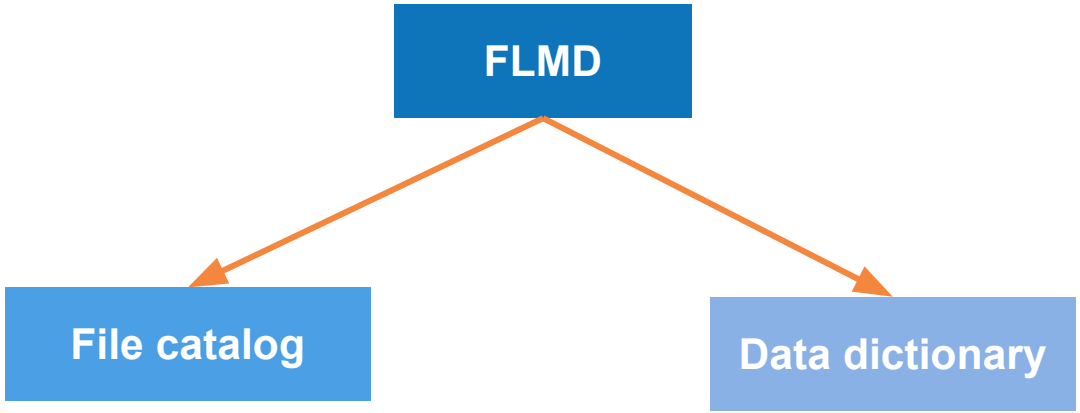
While it may not be necessary to run the model, FLMD will help data users **reuse** your model

File-level metadata (FLMD)[1]

FLMD is an optional component of files to include in data publications, which include two types of files. To ensure machine-readability of files, adhere to CSV reporting format[2].

https://github.com/ess-dive-community/essdive-model-data-archiving-guidelines/blob/main/instructions.md#3-file-level-metadata

Data Dictionary and File Catalog



- One catalog **per data package**
- Describes all data files in the data publication

- One dictionary **per unique csv file**
- Provides metadata for each column within a csv file

Summary of Model Data data package

Archiving Model Data just got easier!

1. Model data packages have **3 major components**
2. You might need **more than one data package** for a single journal publication
3. Attaching **file level metadata** increases the reusability of models

Thanks for stopping by!

Questions?