

# Guidelines for Organizing Your CyVerse Curated Data



For full instructions on requesting a DOI, please see <https://cyverse-doi-request-quickstart.readthedocs-hosted.com/en/latest/>.

There are several steps to properly organizing your dataset. These include determining what data to include, how many identifiers to request, how to organize the data into folders, and creating the ReadMe file and data inventory.

## Step 2.1. Determine what to include

A data collection may be composed of multiple files and different datasets. In preparing your data for publication:

1. Identify the data and other materials that you consider useful for validation and reuse of your research:
  - Data associated to a research project may include multiple files with different roles.
  - If there are components of your dataset that belong in a public repository such as NCBI (e.g., fastq files), submit them to the repository, rather than to CyVerse Curated Data.
2. Beyond data, you will include the **ReadMe file** (see Step 4), and you may include scripts or links to scripts to run your analysis.

## Step 2.2. Determine how many permanent identifiers to request

To determine how many DOIs to request for a given data collection, consider the following:

- Think about its size and components.
- How many studies or publications does it represent?
- Is your data collection formed by different datasets and are those likely to be used separately?
- Do you want to create a data collection with one DOI for the entire project and additional related DOIs for distinct datasets so that they are cited individually?

If you are uncertain about how many DOIs to request, contact us at [doi@cyverse.org](mailto:doi@cyverse.org).

## Step 2.3. Organize your data into folder(s)

1. Organize your data so that there is **one folder for each DOI** (see [CyVerse Curated Data folder-naming guidelines](#) for naming conventions).
2. Within a folder, include all files in your data package plus the ReadMe file and the inventory.
  - You may have subfolders within a data package.
  - You may include compressed files in a package, as described on the [Permanent Identifier FAQs](#), but do not compress the entire folder/package.

## Step 2.4. Name your top level folder according to the guideline

The folder containing your dataset should be named using the **\$Creator\_ \$subject\_ \$date** form at.

For more details on folder naming, see the [CyVerse Curated Data Folder-Naming Guidelines](#).

## Step 2.5. Create a ReadMe file

Create a text file labeled "readMe" with the following information:

- How you obtained, organized, and labeled your dataset.
- How to reuse the data, such as which apps can analyze the data.
- The inventory (see step 2.5) may be included as part of the readme file.
- Examples of good readme files:
  - [http://datacommons.cyverse.org/browse/iplant/home/shared/commons\\_repo/curated/Carolyn\\_Lawrence\\_Dill\\_G2F\\_Mar\\_2017/\\_readme.txt](http://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Carolyn_Lawrence_Dill_G2F_Mar_2017/_readme.txt)

- [http://datacommons.cyverse.org/browse/iplant/home/shared/commons\\_repo/curated/Liang\\_Schnable\\_UNLPlantVision\\_2017/readMe.txt](http://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Liang_Schnable_UNLPlantVision_2017/readMe.txt)

## Step 2.6. Create an inventory

You must create a plain text document that includes an inventory of the contents of the organized dataset (at a minimum, your dataset will contain one data file and one ReadMe file).

The inventory may be part of the ReadMe file or a separate file.

If separate, the inventory should include the ReadMe file and any other additional materials you add to your dataset.

If your dataset contains folders with many files (e.g., large collections of images), you do not need to list each file in the inventory. Simply describe the folder and what it contains.

## 2.7 Supporting documents on data management and organization

Here is a useful guide to data organization: [Research Data Management: File Organization \(PDF\)](#).

Helpful Links
<ul style="list-style-type: none"><li>• <a href="#">Requesting a Permanent Identifier in the Data Commons</a></li><li>• <a href="#">CyVerse Curated Data folder-naming guidelines</a></li><li>• <a href="#">Permanent Identifier FAQs</a></li><li>• <a href="#">Research Data Management: File Organization</a></li></ul>