

Arcus Research Data Management
File Naming Activity Worksheet
arcus-support@email.chop.edu

This worksheet will help you think through possible file naming conventions for your project and identify appropriate conventions to use that meet your needs.

This activity involves two major steps:

1. Walk through parts A and B outlined in this paper to develop file naming conventions.
2. Fill out the file naming README (or add this to your current documentation README) to make the conventions clear and accessible to your group.

Part A.

Identify your group's file naming needs

1. Consider how you currently name and organize your files. Write out some of the file names you use now.
2. Reflect on these names. What do they tell you about the files? How do they help you identify the files you need? What do you like or not like about these names / naming conventions?
3. Who will you be designing these file naming conventions for? Yourself? Your PI? Other researchers who need to access this data? List out the names or roles of people and groups who will be using these files. If you know of any preferences for file naming, note those here as well.

Part B.

Create a file-naming schema

1. Create your project file inventory

Consider the files you will receive and generate in the course of your research project. It may be difficult or daunting to think about ALL the files, so you can start tackling this in phases as well: prep-to-research, data collection, analysis, results and publication, etc.

Use the table below to capture the information on all your different data file types.

Column A: What different **types of data** are you using/creating? (e.g., MRI, REDCap data in csv, whole exome sequence, figures, protocols, r scripts, pipelines, etc.)

Column B: What are the **file formats**? (e.g., .csv, .fq)

Column C: What are the **meaningful unique characteristics** of this data file that are important to know about it (e.g., date created, study participant, experimental conditions)? Are there standard abbreviations for any of these characteristics? You may have more than one unique characteristic.

Column D: Will you expect **multiple versions** of the same file? If so, will there be multiple versions in a given day? How many?

Column E: Keep blank; we'll tackle that in the next section.

Arcus Research Data Management
 File Naming Activity Worksheet
 arcus-support@email.chop.edu

A. Data Type	B. File Format	C. Unique characteristics	D. Multiple versions?	E. Draft file name
<i>Ex: MRI</i>	<i>Ex: DICOM</i>	<i>Ex: Date collected Study participant Pre or Post treatment Anatomical Region</i>	<i>Ex: N BUT multiple MRIs per participant over course of study</i>	<i>Ex: YYYYMMDD_par001_pre.DC M</i>

2. Draft your file names

Things to include: Include unique characteristics in column C can be used to create a file name. Not all unique characteristics will be necessary to include in the file name; consider items that will be useful in helping users select the appropriate file for their needs. Other unique characteristics that are *not* helpful for this purpose can still be recorded elsewhere in project documentation, but consider leaving them out of the file names.

If you will have multiple versions of a file (indicated in Column D), make sure to add version information into your file name. If you anticipate having more than one version in a day, you can't rely on the date alone and should add a version number to the file name. For example, YYYYMMDD_par001_v01.csv

Ordering: Put the unique characteristics in the order that you want your data files to sort by. For example, if you prefer to sort by study participant and then date, the participant ID should be first followed by the experiment number and so on, eg: par001_YYYYMMDD_v01.csv.

Tip: Format dates according to ISO-8601 date standard (four digit year, two digit month, two digit day) so they always sort in chronological order. For more tips and best practices, check out the file naming best practices one-sheet developed by MIT.

Arcus Research Data Management
File Naming Activity Worksheet
arcus-support@email.chop.edu

README.

Record and document your new file naming conventions

Now that you have developed meaningful file naming conventions, record them in a space that's easily accessible to all who need to reference them. We suggest including a readme file in each folder describing the naming convention(s) you have chosen. If you are using the Arcus Project Template to organize your project files, these README should already exist. If not, we recommend you create them!

Below is a suggested format for documenting naming conventions in a README.

```
File type:  
Filename schema:  
Schema key:  
Example filename:
```

Example:

```
File type: MRI DICOM  
Filename schema: [date]_[subjectID]_[preOrPost]  
Schema key: date: date of imaging in format YYYYMMDD  
             subjectID: coded study subject ID  
             preOrPost: indicates if imaging was pre or post  
                       treatment.  
Example filename: 20200119_par1355_post.dcm
```

Arcus Research Data Management
File Naming Activity Worksheet
arcus-support@email.chop.edu

Citation: "Worksheet: Naming and Organizing your Files and Folders." Handout. MIT Libraries Data Management Service. Cambridge, MA. 2020-02-03. Digital.