



Jan 19, 2021

Checklist for organizing and managing data

Chris Berndsen¹¹James Madison University*In Development*

This protocol is published without a DOI.



Chris Berndsen
James Madison University

ABSTRACT

When planning an experiment, making a plan for data organization is a key step. Data are the currency of science and allow for confirmation of results. Moreover, your data may aid others in their work.

This protocol is a recommended checklist for planning (OR organizing data in hindsight) for students in Biochemistry lecture or labs at James Madison University. It was designed to help new scientists make their data accessible.

PROTOCOL CITATION

Chris Berndsen 2021. Checklist for organizing and managing data. **protocols.io**
<https://protocols.io/view/checklist-for-organizing-and-managing-data-bqv6mw9e>

LICENSE

————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Dec 18, 2020

LAST MODIFIED

Jan 19, 2021

PROTOCOL INTEGER ID

45726

BEFORE STARTING

If planning for a new experiment or set of experiments, ensure that you have a clear idea of the types and amount of data that you will be generating.

If organizing data for deposition or in hindsight, know what types of data are in the data set.

Data organization

- 1 Create a separate folder for each type of experiment. Data should be sorted to keep data from the same experiment together.
- 2 Create separate sub-folders for distinct trials or data collection times within each experiment folder.
- 3 Keep analysis files separate from the raw data files

File naming

Files names should be clear as to the experiment and avoid computer added versioning (i.e. the (1) that computers add

4 when saving a file with the same name).

4.1 Experiment titles should contain information including type (NMR, SAXS), initials of scientist (if known), and type of data (intensity, counts, etc.)

5 If known, raw data files should have a collection date or experiment data in YYYYMMDD format OR YYMMDD format.

6 Files that are collected from iterations of the same experiment like multiple reads of the same sample should have leading numbers in 0001, 0002, 0003 so that they sort in order.

7 Avoid punctuation marks in file names (except - (dash) and _ (underscore)) in file names.

Tables/spreadsheets

8 Tables or spreadsheet data should be tidy, specific, and clearly indicate units. An example of a reasonable table is shown below.

<i>maker</i>	<i>efficiency (in mpg)</i>	<i>color</i>	<i>type</i>	<i>passenger doors</i>
Toyota	23	green	sedan	4
Ford	21	green	sedan	4
Ford	14	green	sport utility	4
Dodge	28	red	sport utility	2

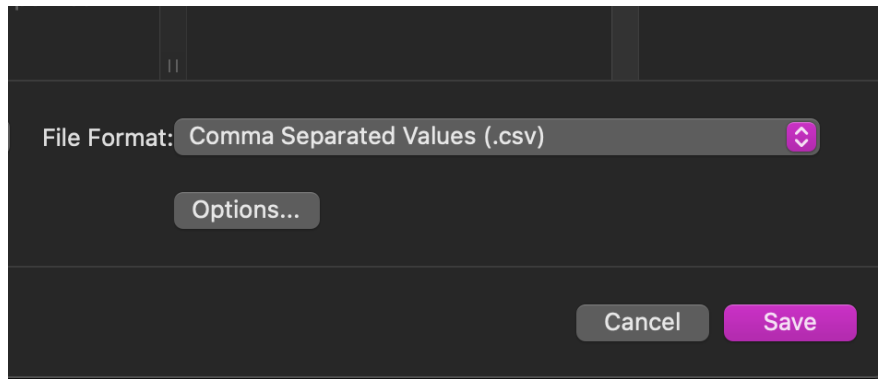
A tidy table with one type of information per column, units on numerical data, and specificity in column titles.

File formatting

9 File extensions should be non-proprietary and accessible. Table below shows suggested formats.

A	B
File type	Suggested File extension
spreadsheet	.csv
biomolecule coordinates/structure	.pdb or mmCIF
sequence file	.FASTA or .FASTQ or .txt
words/text	.txt
image	.tff or .png
video	.mov

9.1 To convert a file, most programs offer a Save As functionality with a File Format option.



README file

- 10 Write a text file or a wiki entitled as README or file guide to describe the contents of each folder. The descriptions should include information on how the files are internally organized and how the data were generated in enough detail that an interested observer can read this file and easily find which files they need, understand how to access them, and have an idea of the contents.

Saving checklist

- 11 For class projects, this completed checklist should be turned in. The easiest way is to export directly to OSF via MORE>Export>PDF>to OSF and then place in your project.