

Good Enough Research Data Management - A Very Brief Guide

Version 1.1, July 2017

By Eugene Barsky, University of British Columbia, eugene.barsky@ubc.ca

This very brief guide presents a set of good data management practices that researchers can adopt, regardless of their data management skills and levels of expertise. We believe that this practices' synthesis allows better research computing principles and enhance research collaborations.

We organize our recommendation into the following five (5) steps:

- Save your raw data
- Backup your data
- Describe your data
- Process your data
- Archive and preserve your data

Please let us elaborate:

1. Save your raw data in original format

- 1.1. Don't overwrite your original data with a cleaned version
- 1.2. Protect your original data by locking it or making it read-only
- 1.3. Refer to this original data if things go wrong (as they often do)

2. Backup your data

- 2.1. Use the 3-2-1 rule - save three copies of your data, on two different devices, and one copy off site
- 2.2. Do not backup or store sensitive data on a commercial cloud (Box, Google Drive, etc)

3. Describe your data

- 3.1. Machine friendly. Describe your dataset with a metadata standard for discovery (e.g. DataCite, Dublin Core, DDI, etc)
- 3.2. Human friendly. Describe your variables, so your colleagues will understand what you meant. Data without good metadata is useless. Give your variables clear names

- 3.3. Use NA for missing data, since computer programs don't like blank cells
- 3.4. Convert your data to open, non-proprietary formats (see this [guide](#) for more)
- 3.5. Name your files well with basic metadata in file names (see this [guide](#) for more)

4. Process your data

- 4.1. Make each column a variable
- 4.2. Make each row an observation
- 4.3. Store units (e.g. kg or cm) as metadata (in their own column)
- 4.4. Document each step processing your data in a README file (some tools, e.g. Open Refine or Git allow documenting steps automatically)

5. Archive and preserve your data

- 5.1. Submit final data files to a repository assigning a persistent identifier (e.g. handles or DOIs)
- 5.2. Provide good metadata for your study so others could find it (use your discipline's metadata standard, e.g. Darwin Core, DDI, etc)

* This guide was inspired by PLoS's "[Ten Simple Rules](#)" - quick guides for mastering some of the professional challenges research scientists face in their careers and especially: Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, et al. (2017) Good enough practices in scientific computing. PLOS Computational Biology 13(6): e1005510. <https://doi.org/10.1371/journal.pcbi.1005510>

** Please consider engaging your Library and/or Advanced Research Computing for any help along the way. You will be pleasantly surprised.



This work is licensed under a Creative Commons [Attribution 4.0 International](#) License.