RESEARCH DATA MANAGEMENT

This brief guide presents a set of good data management practices that researchers

can adopt, regardless of their data management skills and levels of expertise.

# Save your raw data in original format

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

# Backup your data

* Use the 3-2-1 rule: Save three copies of your data, on two different storage mediums, and one copy off site.
* Do not backup or store sensitive data on a commercial cloud (Dropbox, Google Drive, etc.).

# Describe your data

* Machine Friendly: Describe your dataset with a metadata standard for discovery.
* Human Friendly: Describe your variables, so your colleagues will understand what you meant. Data without good metadata is useless. Give your variables clear names.
* Do not leave cells blank - use numeric values clearly out of range to define missing (e.g. '99999') or not applicable (e.g. '88888') data and describe these in your data dictionary.
* Convert your data to open, non-proprietary formats
* Name your files well with basic meta-data in the file names

# Process your data

* Make each column a variable.
* Make each row an observation.
* Store units (e.g. kg or cm) as metadata (in their own column).
* Document each step processing your data in a README file.

# Archive and preserve your data

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



[Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)](https://creativecommons.org/licenses/by-nc/4.0/)

**Portage Network |** [**portage@carl-abrc.ca**](mailto:portage@carl-abrc.ca) **|** [**portagenetwork.ca**](https://portagenetwork.ca/)