

Data Management Best Practices Evaluation Checklist

This checklist is designed to help you evaluate your data management activities within your research projects by providing a list of common data management best practices. Best practices enable the proper organization, documentation, and preservation of data files that will result in more easily discoverable and reusable data, addressing funding agency requirements for transparency and reproducibility of research methods.

File Formats

Retain the original, unedited outputs from software and hardware to preserve source data.

- Do not edit or alter the raw data file. Keep it in its native format and create a copy for editing or further manipulation.

Ensure future access to your data files by using standard, stable, commonly-used file formats.

- Non-proprietary formats are preferred (particularly for final versions).
- Be aware of what software is required to view and process data files, and be wary of software lifespans.

File Organization

A logical and organized folder structure can make it easier to keep track of project information.

- Avoid complex directory hierarchies and consider that folder names will sort alphabetically.
- Avoid keeping duplicate working copies of files (backup copies are not considered duplicates in this context).

Develop a file and folder naming convention and document it so all team members can follow it.

Good practices in choosing file and folder names:

- Uniquely name each file.
- Be consistent and include similar information in all file names of the same file type.
- Consider sorting order (usually lexicographic) and logical hierarchies in file directories.
- Avoid ambiguous and confusing names, such as 'MyData' or 'sample'
- Derivatives and versions should have similar (but differentiated) names to keep them co-located but still uniquely identified.
- Names should reflect the contents of the file and/or the stage of development.
 - When using dates, if you want the files to sort chronologically, put the year first and use numerical two-digit months and days (YYYY-MM-DD). (Example: March 7, 2004 would be written '2004-03-07'.)
- Use only alphanumeric characters but use dashes (-) or underscores (_) instead of spaces; avoid special characters such as colons (:) and slashes (/).
- Avoid using case differences to distinguish between files: 'Record', 'record', and 'RECORD' may be three different file names or the same file name, depending on the operating system.

Documentation

If possible, document data characteristics and workflows in a digital format at the time that data files are created or altered.

- Create readme files and data dictionaries to provide digital documentation of data characteristics, workflows, progress, results, software, etc.
- Document any database data field (variable/column) characteristics for later interpretation. Possible elements are field name, field description, and permitted range of values.
- Document data file and collection characteristics that are relevant for later interpretation. Possible elements are file name and path, relevant dates, creation method, and status.
- Digitize (scan) relevant paper laboratory or field notes so that they can be more easily shared along with the data.

Data files are more easily and accurately interpreted when they are documented using a formal, standardized metadata format.

- Consider using discipline-specific metadata specifications and schemas. (see a list of schemas at <http://rd-alliance.github.io/metadata-directory/standards/>)
- Develop strategies for streamlining the metadata data entry process, such as through the use of templates to input information that is consistent across all project metadata.
- Store the metadata file close to the data (or embedded if possible) to ensure discovery.

Storage and Security

Ensure data redundancy and replication, and avoid single points of failure.

- Never rely on a single copy of data. Have at least two backup copies (with at least one in a remote location) in addition to the working copy.
- Document your data storage system and data backup policy.
- Back up data regularly. Backups are particularly important if using portable media, such as laptops and flash drives.
- Use managed, networked storage whenever possible (Example: departmental network drive with system administrator).

Access and Use Restrictions

Data files may be protected by ownership rights or licenses.

- Ensure that you have proper permissions to use and share data, considering any license agreements or ownership issues.
- Document any access or use restrictions in the metadata, readme file, or data dictionary.

Protect sensitive and confidential information.

- Datasets that include confidential information should have that information de-identified or suppressed before being shared.

NOTE:

This checklist does not include every best practice, but instead focuses on those that are the most critical or the most applicable to a wide variety of research data types. Not all of these best practices will apply to, or be appropriate for, every project or data type; this document is intended only as a guide to help you identify possible practices that could be improved.

Sources:

- Hook, Les A., Suresh K. Santhana Vannan, Tammy W. Beaty, Robert B. Cook, and Bruce E. Wilson. (2010). *Best Practices for Preparing Environmental Data Sets to Share and Archive*. Retrieved from <http://daac.ornl.gov/PI/BestPractices-2010.pdf> (doi:10.3334/ORNLDAAAC/BestPractices-2010)
- Inter-university Consortium for Political and Social Research (ICPSR). (2012). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (5th ed.). Ann Arbor, MI. Retrieved from <http://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- Michigan State University Libraries. (2013). *Research Data Management Fundamentals*. Retrieved from <http://img.lib.msu.edu/rdmg/RDMFundamentalsBooklet.pdf>
- Strasser, C. (2015). *Research Data Management*. Baltimore, MD: National Information Standards Organization. Retrieved from http://www.niso.org/apps/group_public/download.php/15375/PrimerRDM-2015-0727.pdf
- Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). *Primer on Data Management: What you always wanted to know*. A DataONE publication. Retrieved from https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf (doi:10.5060/D2251G48)