

Research Data Management Workbook

Kristin Briney

2023-08-11

Contents

About this Book	5
Description	5
License	5
The Author	5
1 Introduction	7
2 Documentation	9
2.1 Checking Laboratory Notebooks	9
2.2 README.txt	9
2.3 Data Dictionary	9
3 File Organization and Naming	11
3.1 File Organization	11
3.2 File Naming	11
4 Data Storage	17
4.1 Backups	17
5 Data Management Planning	19
5.1 Living Data Management Plan	19
5.2 Data Governance	19
6 Data Sharing	21
6.1 Picking a Repository	21
6.2 Sharing Data	21

7 Project Wrap Up	23
7.1 Prepare Files for Future Use	23
7.2 The Archive Folder	24
7.3 Separation From The Institution	25

About this Book

Description

The Research Data Management Workbook is made up of a collection of exercises for researchers to improve their data management. The workbook contains exercises across the data lifecycle, though the range of activities is not comprehensive. Instead, exercises focus on discrete practices within data management that are structured and can be reproduced by any researcher. The book is divided into chapters by phases of the data lifecycle, with one or more exercises in each chapter. Every exercise comes with a description of its importance, notes on how to do the exercise, and the exercise instructions.

The workbook is intended as a supplement to existing data management education. If you would like to learn more about the principles of data management, please see the article by Briney, Coates, and Goben Briney et al. [2020] or read the book “Data Management for Researchers” Briney [2015].

License

This book is available under a Creative Commons Attribution-NonCommercial (CC BY-NC) 4.0 International license

The Author

Kristin Briney is the Biology & Biological Engineering Librarian at the California Institute of Technology and author of the books “Data Management for Researchers” (Pelagic Publishing, 2015) and, with Becky Yoose, “Managing Data for Patron Privacy” (ALA Editions, 2022). She has a PhD in chemistry and an MLIS, both from the University of Wisconsin-Madison. Her research focuses on research data management, institutional data policy, and patron privacy vis-a-vis library data handling. Kristin is an advocate for the adoption of

the international date standard ISO 8601 (YYYY-MM-DD) and likes to spend her free time making data visualizations out of yarn and fabric.

Chapter 1

Introduction

Chapter 2

Documentation

2.1 Checking Laboratory Notebooks

Description: **

Instructions: **

2.2 README.txt

Description: **

Instructions: **

2.3 Data Dictionary

Description: **

Instructions: **

Chapter 3

File Organization and Naming

3.1 File Organization

Description: **

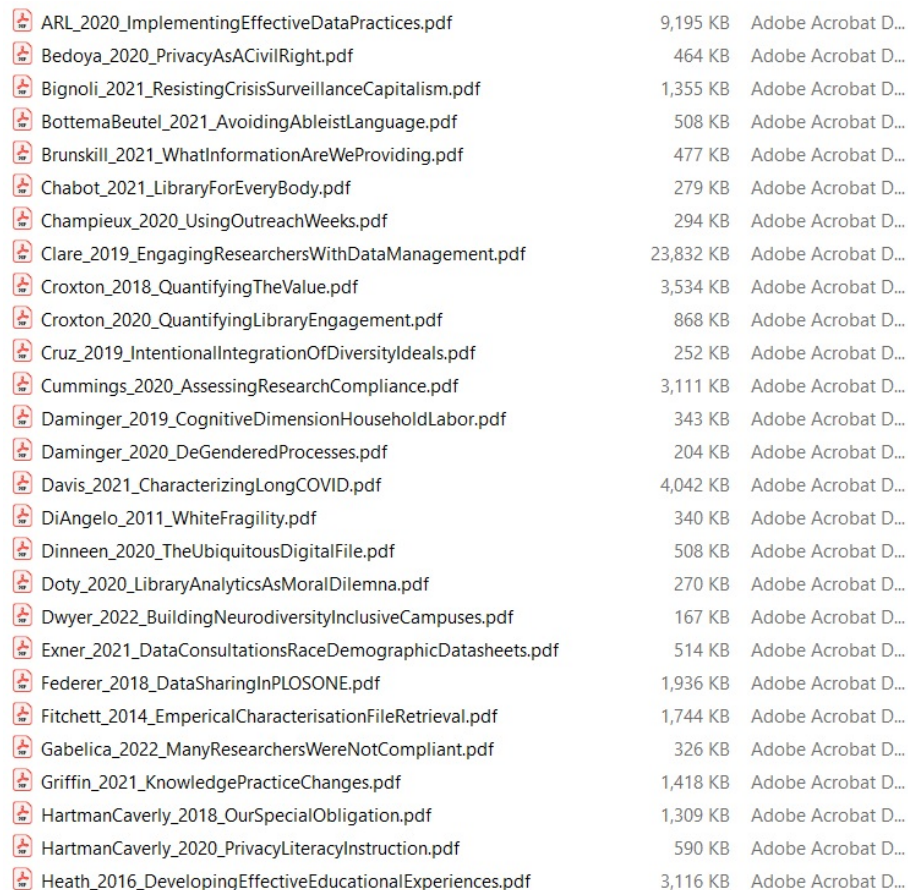
Instructions: **

3.2 File Naming

Description: *File naming conventions are a simple way to add order to your files and help to find them later. Rich and descriptive file names make it easier to search for files, understand at a glance what they contain, and tell related files apart.*

Instructions: *This exercise guides researchers through the process of creating a file naming convention for a group of related files. Fill in each section for a group of related files following the instructions; an example for microscopy files is provided. This exercise may be redone as needed, as different groups of files require different naming conventions.*

This exercise is based on the “File Naming Convention Worksheet” Briney [2020a].



ARL_2020_ImplementingEffectiveDataPractices.pdf	9,195 KB	Adobe Acrobat D...
Bedoya_2020_PrivacyAsACivilRight.pdf	464 KB	Adobe Acrobat D...
Bignoli_2021_ResistingCrisisSurveillanceCapitalism.pdf	1,355 KB	Adobe Acrobat D...
BottemaBeutel_2021_AvoidingAbleistLanguage.pdf	508 KB	Adobe Acrobat D...
Brunskill_2021_WhatInformationAreWeProviding.pdf	477 KB	Adobe Acrobat D...
Chabot_2021_LibraryForEveryBody.pdf	279 KB	Adobe Acrobat D...
Champieux_2020_UsingOutreachWeeks.pdf	294 KB	Adobe Acrobat D...
Clare_2019_EngagingResearchersWithDataManagement.pdf	23,832 KB	Adobe Acrobat D...
Croxton_2018_QuantifyingTheValue.pdf	3,534 KB	Adobe Acrobat D...
Croxton_2020_QuantifyingLibraryEngagement.pdf	868 KB	Adobe Acrobat D...
Cruz_2019_IntentionalIntegrationOfDiversityIdeals.pdf	252 KB	Adobe Acrobat D...
Cummings_2020_AssessingResearchCompliance.pdf	3,111 KB	Adobe Acrobat D...
Daminger_2019_CognitiveDimensionHouseholdLabor.pdf	343 KB	Adobe Acrobat D...
Daminger_2020_DeGenderedProcesses.pdf	204 KB	Adobe Acrobat D...
Davis_2021_CharacterizingLongCOVID.pdf	4,042 KB	Adobe Acrobat D...
DiAngelo_2011_WhiteFragility.pdf	340 KB	Adobe Acrobat D...
Dinneen_2020_TheUbiquitousDigitalFile.pdf	508 KB	Adobe Acrobat D...
Doty_2020_LibraryAnalyticsAsMoralDilemma.pdf	270 KB	Adobe Acrobat D...
Dwyer_2022_BuildingNeurodiversityInclusiveCampuses.pdf	167 KB	Adobe Acrobat D...
Exner_2021_DataConsultationsRaceDemographicDatasheets.pdf	514 KB	Adobe Acrobat D...
Federer_2018_DataSharingInPLOSONE.pdf	1,936 KB	Adobe Acrobat D...
Fitchett_2014_EmpiricalCharacterisationFileRetrieval.pdf	1,744 KB	Adobe Acrobat D...
Gabelica_2022_ManyResearchersWereNotCompliant.pdf	326 KB	Adobe Acrobat D...
Griffin_2021_KnowledgePracticeChanges.pdf	1,418 KB	Adobe Acrobat D...
HartmanCaverly_2018_OurSpecialObligation.pdf	1,309 KB	Adobe Acrobat D...
HartmanCaverly_2020_PrivacyLiteracyInstruction.pdf	590 KB	Adobe Acrobat D...
Heath_2016_DevelopingEffectiveEducationalExperiences.pdf	3,116 KB	Adobe Acrobat D...

Figure 3.1: Screenshot of pdf's with consistent file names using the convention FirstAuthorLastName_YEAR_ShortTitle.pdf

1. What group of files will this naming convention cover?

You can use different conventions for different file sets.

Example: This convention will apply to all of my microscopy files, from raw image through processed image.

2. What information (metadata) is important about these files and makes each file distinct?

Ideally, pick three pieces of metadata; use no more than five. This metadata should be enough for you to visually scan the file names and easily understand what's in each one.

Example: For my images, I want to know date, sample ID, and image number for that sample on that date.

- 1.
- 2.
- 3.
- 4.
- 5.

3. Do you need to abbreviate any of the metadata or encode it?

If any of the metadata from step 2 is described by lots of text, decide what shortened information to keep. If any of the metadata from step 2 has regular categories, standardize the categories and/or replace them with 2- or 3-letter codes; be sure to document these codes.

Example: Sample ID will use a code made up of: a 2-letter project abbreviation (project 1 = P1, project 2 = P2); a 3-letter species abbreviation (mouse = "MUS", fruit fly = "DRS"); and 3-digit sample ID (assigned in my notebook).

4. What is the order for the metadata in the file name?

Think about how you want to sort and search for your files to decide what metadata should appear at the beginning of the file name. If date is important, use ISO 8601-formatted dates (YYYYMMDD or YYYY-MM-DD) at the beginning of the file names so dates sort chronologically.

Example: My sample ID is most important so I will list it first, followed by date, then image number.

- 1.
- 2.
- 3.
- 4.
- 5.

5. What characters will you use to separate each piece of metadata in the file name?

Many computer systems cannot handle spaces in file names. To make file names both computer- and human-readable, use dashes (-), underscores (_), and/or capitalize the first letter of each word in the file names.

Example: I will use underscores to separate metadata and dashes between parts of my sample ID.

6. Will you need to track different versions of each file?

You can track versions of a file by appending version information to end of the file name. Consider using a version number (e.g. “v01”) or the version date (use ISO 8601 format: YYYYMMDD or YYYY-MM-DD).

Example: As each image goes through my analysis workflow, I will append the version type to the end of the file name (e.g. “_raw”, “_processed”, and “_composite”).

7. Write down your naming convention pattern.

Make sure the convention only uses alphanumeric characters, dashes, and underscores. Ideally, file names will be 32 characters or less.

**Example: My file naming convention is “SA-MPL-EID_YYYYMMDD_###_status.tif”*
Examples are “P1-MUS-023_20200229_051_raw.tif” and “P2-DRS-285_20191031_062_composite.tif”.*

8. Document this convention in a README.txt (or save this worksheet) and keep it with your files.

Chapter 4

Data Storage

4.1 Backups

Description: **

Instructions: **

Chapter 5

Data Management Planning

5.1 Living Data Management Plan

Description: **

Instructions: **

5.2 Data Governance

Description: **

Instructions: **

Determine with the project PI/your collaborators:

- Who is allowed to reuse this data later?
- Who will store the master copy of the data and for how long?
- Will this data be shared publicly?
- Are there security or intellectual property restrictions on the data?
- Who keeps any physical research notebooks?

Chapter 6

Data Sharing

6.1 Picking a Repository

Description: **

Instructions: **

6.2 Sharing Data

Description: **

Instructions: **

Chapter 7

Project Wrap Up

7.1 Prepare Files for Future Use

Description: **

Instructions: **

This exercise has been modified from the “Project Close Out Checklist” Briney [2020b].

Prepare Data

Prepare data:

Convert data to more open/common file formats (e.g. .CSV or .TXT)

Record General Project Information in a README.txt File:

Project title

Project description

Dates

Personnel

Where files are stored

File organization and naming conventions

7.2 The Archive Folder

Description: To save your future-self time spent digging through all of your research files, set aside the most important files into a separate “Archive” folder. Do this at the end of the project while you still remember which files are important and where they are located. The Archive folder should only contain a small subset of the most important documents that are likely to be reused; you may still need to go through all of your files but, in the majority of instances, you will save time by easily finding what you need in the Archive folder.

Instructions: This exercise consists of a checklist of the key documents that are likely to be most useful in a research project archive. Create a separate folder within the larger project folder (or in a highly visible place within the storage system) labelled “Archive”. Copy – do not move – the files on this checklist into the Archive folder. Add copies other important research documents, as needed. Remember, the Archive folder does not need to be comprehensive, so focus on the subset of files that are most likely to be reused or referenced in the future.

This exercise has been modified from the “Project Close Out Checklist” Briney [2020b].

Project Documentation

- README file of project information

Data Snapshots

- Raw data

- Key data analyses

- Final data

Code

- Analysis code

- Record software version, as appropriate

Other Research Documents

- Protocols

- Survey instruments

Research Notes

- Scan of research notebook

- Digital notes

Images

Flat files of figures (e.g. .JPG or .TIFF)

Editable image files (e.g. Photoshop)

Publications

Published article in .PDF format

Final version of the article in editable document format (e.g. .DOCX)

Posters

Administrative Documents

Grant proposals

Grant progress reports and final report

7.3 Separation From The Institution

Description: **

Instructions: **

Bibliography

Kristin Briney, Heather Coates, and Abigail Goblen. Foundational Practices of Research Data Management. *Research Ideas and Outcomes*, 6:e56508, July 2020. ISSN 2367-7163. doi: 10.3897/rio.6.e56508. URL <https://riojournal.com/article/56508/>.

Kristin A. Briney. *Data Management for Researchers : Organize, Maintain and Share Your Data for Research Success*. Pelagic Publishing, 2015. URL <https://pelagicpublishing.com/products/data-management-for-researchers-briney>.

Kristin A. Briney. File Naming Convention Worksheet, June 2020a. URL <https://doi.org/10.7907/894q-zr22>.

Kristin A. Briney. Project Close-Out Checklist for Research Data, May 2020b. URL <https://resolver.caltech.edu/CaltechAUTHORS:20200519-142758925>.