



## The Research Data Management Workbook: Building a Collection of Data Management Exercises to Bridge Data Information Literacy and Data Management Implementation

Kristin Briney, California Institute of Technology, Pasadena, CA, USA, [briney@caltech.edu](mailto:briney@caltech.edu) 

---

### Abstract

**Objective:** There are limited opportunities and resources for data information literacy at small universities, requiring instructors to make the most of the time they have in the classroom. This article describes the creation of a collection of data management exercises, collectively called *The Research Data Management Workbook*, which supplement one-shot instruction and help students implement specific data management tasks.

**Methods:** Exercises were developed using backward design and authentic assessment, with the goal of scaffolding data management implementation yet allowing for customization to research workflows. Exercises cover activities across data lifecycle and take the form of worksheets, checklists, and procedures. The exercises were collectively formatted as a book using the tool bookdown.

**Results:** For a one-hour library session, students can work through one or two exercises during class and the instructor can refer to specific exercises for follow up on various data management topics. The exercises have also proved useful for consultation, as a researcher can develop an understanding of a way to address the data problem ahead of a more in-depth consultation.

**Received:** April 28, 2024 **Accepted:** May 1, 2025 **Published:** July 9, 2025

**Keywords:** data information literacy, research data management, publishing

**Citation:** Briney, Kristin. 2025. "The Research Data Management Workbook: Building a Collection of Data Management Exercises to Bridge Data Information Literacy and Data Management Implementation." *Journal of eScience Librarianship* 14 (1): e937. <https://doi.org/10.7191/jeslib.937>.

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0>.

OPEN ACCESS

## Abstract Continued

**Conclusions:** The workbook has been a useful supplement to limited data management instruction time at a small university. Further work needs to be done to quantify the efficacy of this form of data information literacy.

---

## Introduction

An unfortunate amount of information literacy, which includes data information literacy, occurs via the “one-shot” workshop (Pagowsky 2021). This means that, while librarians may be experts at teaching research data management (RDM), there can be limited instructional opportunities to help students and researchers transition from learning best practices to actually implementing data management in the research process. There are several options that go beyond the one-shot model, including: for-credit courses on RDM (Nelson and Kong 2020; Griffin 2021); books on RDM for scientists and social scientists (Briney 2015; Corti et al. 2019); online RDM modules (Lamar Soutter Library - University of Massachusetts Medical School 2023; The University of Edinburgh 2022); and RDM toolkits (LEARN Project 2017; Read and Surkis 2017; University of Michigan et al. 2015; “RDMkit” 2021). All of these are solid educational options, but not every institution has the resources or opportunity to provide a full curriculum on data management.

This article describes a different model for supporting data management implementation at a smaller university: a set of reproducible, structured data management worksheets used to complement one-shot workshops. The worksheets guide students and researchers through implementing various RDM tasks and can be used independently by researchers or during library instruction, either as a single worksheet or a set. The advantages of creating a set of worksheets are that worksheets were easier to develop and implement at a small university and they can be used selectively to supplement limited RDM instructional opportunities.

This article describes the process of creating a comprehensive set of RDM exercises, formatting those exercises as a book using the tool “bookdown,” and using the resultant workbook in teaching and consultations. The resultant workbook, *The Research Data Management Workbook*, is freely available and licensed with a Creative Commons Attribution NonCommercial license (Briney 2023).

## Writing the Workbook Exercises

The idea for the workbook originated from using handouts to augment one-shot data management instruction at the California Institute of Technology (Caltech), an institution with 2,000 students. The university has a very high research activity in science and technology – with research occurring at all levels from undergraduates to faculty – meaning that data management support and education is necessary. The author had already created a handful of worksheets/handouts for local RDM instruction, and these were edited then supplemented with new exercises to create the workbook.

The goal for each exercise was to create a scaffolded activity for students and researchers to work through to implement discrete, realistic data management tasks. Since so much of data management balances best principles with customization to specific research workflows, exercises needed to provide enough structure and guidance for students to follow yet allow for myriad implementation options.

Exercise development followed the principles of backward design and authentic assessment (Wiggins and McTighe 2005). The process started by identifying: the real data management task that the student should be able to implement; the assessment to reach that objective; and, finally, the structure for the student to do that task. Taking the lab notebook exercise as an example, the learning objective was to have students take better notes in their notebook. The assessment for reaching this objective was for the student to self-identify areas needing improvement in their current note taking. This was done through a worksheet where students evaluate an old entry in their lab notebook using the following prompts:

1. Summarize the entry
2. Determine ease of understanding the work done
3. Judge whether documentation was sufficient for reproducibility
4. Identify strengths and weaknesses of the note taking
5. State one improvement they would make in notetaking going forward

Exercises were structured so that students wouldn't need to know data management best practices, though having some foundation in RDM is helpful. To aid with this, most exercises included realistic examples to provide context for how to do the exercise, especially as exercises may not always be done in class with a data management expert on hand to answer questions.

The first edition of the workbook included the following 15 exercises:

- Evaluate a laboratory notebook
- Write a project-level README.txt
- Create a data dictionary
- Set up a file organization system
- Create a file naming convention
- Pick storage and backup systems
- Test your backup
- Write a living data management plan
- Determine data stewardship

- Pick a data repository
- Share data
- Prepare data for future use
- Convert data files types
- Create an archive folder
- Separate from the institution

To keep exercises discrete and focused, several exercises refer to or require the completion of other exercises to complete the task. For example, the “Share data” exercise prompts the student to work through other exercises: pick a data repository, convert to open file formats, and to create a data dictionary.

The style for each exercise was determined by what best fit the learning objective. Most (nine) of the exercises were formatted as worksheets, which listed key questions and provided space for students and researchers to write in answers. Another four of the exercises were checklists; these corresponded to actions done during data sharing and project wrap up. The final two exercises were procedures; one was a card-sorting exercise to brainstorm a file organization scheme and the other a procedure for test recovering a file from backup.

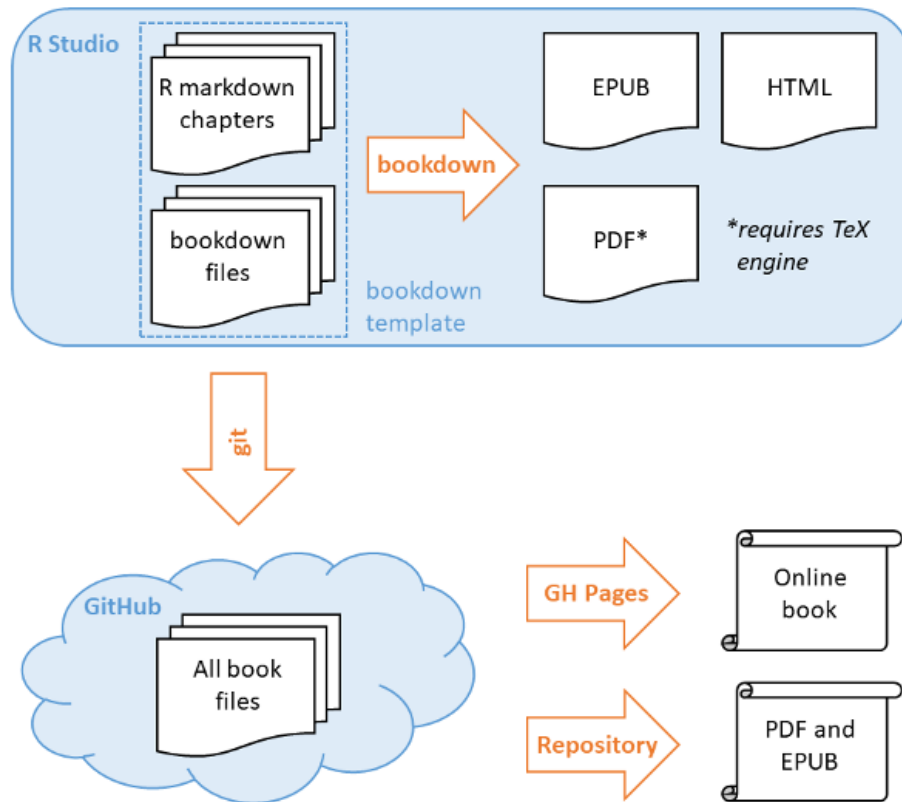
Exercises were organized into six chapters, containing two-to-four exercise each, loosely themed by the data lifecycle: documentation, file organization and naming, data storage, data management, data sharing, and project wrap-up. A very short introduction chapter was also added to the workbook, which concisely explained general concepts of data management and referred readers to further educational materials on the topic. The workbook was intended as a supplement to existing resources and not a stand-alone book.

## Building the Workbook

*The Research Data Management Workbook* used a number of software tools to turn the collection of exercises in to a formatted book available on the internet. A visual representation of this process is provided in Figure 1 and summarized in this section.

The principle tool used to create the book was a software package called “bookdown” (version 0.34) (Xie 2023), which turned individual R Markdown files – one Markdown file per chapter – into a book-shaped output. The software assembled the chapters in the proper order, added a table of contents, handled references, and took care of all of the other formatting. By default, bookdown produced an HTML website, an EPUB eBook, and a PDF file versions of the book.

The book was written within the software R Studio (“Ghost Orchid” release, R version 4.3.1) starting with a bookdown template; the template was a built-in option for new projects created in R Studio. The template provided example chapter files which explained how to format R Markdown text – this text was replaced



**Figure 1:** A flow diagram of the process of creating the workbook: the process started with a bookdown template in R Studio; bookdown output the assembled book as EPUB, HTML, and PDF files; git was used to back up all files onto GitHub; and different variants of the book were hosted in the institutional repository and on GitHub Pages.

with the actual book content – as well as all of the necessary supporting files to assemble a bookdown book. Converting the R Markdown files into a PDF book required a TeX engine – in this case, the package TinyTeX (version 0.45) was used. Once chapters were written, creating the book was as easy as clicking the “Build Book” button in R Studio to render the book as HTML, EPUB, and PDF.

git and GitHub were used to back up and version control the book throughout the writing process. Toward the end of the writing process, the workbook’s repository was transferred from a personal GitHub account to the Library’s GitHub account for long-term management.

Once the book was complete, the PDF and EPUB versions were deposited into the Library’s institutional repository and received a DOI and ISBNs. GitHub Pages (GitHub 2023) was used to make the HTML version of the book available on the internet.

The benefit of using this suite of tools is that it's fairly easy to update the book. The HTML version of the workbook can be changed within minutes and the repository files can be versioned as different editions of the book. This allows for continual improvements to the workbook in the future.

## Teaching with the Workbook

*The Research Data Management Workbook* has predominantly been used when teaching one-shot lectures and workshops to undergraduates, grad students, and post docs. There is not enough time in a one-hour session to cover every RDM topic, let alone work through all of the workbook exercises. Select exercises are therefore added to the standard lecture material to provide an interactive individual activity, followed by volunteers sharing out their results so that students can learn from one another. For most one-hour sessions, the instructor lectures on a range of topics but only guides the students through one or two exercises – most often the file naming conventions exercise which is universally applicable to research – and refers to other exercises for follow up after class if students would like to implement a particular RDM task in their own research. There has not yet been a formal assessment to determine if students actually work through other workbook exercises after class; this is a good direction for future work.

Some classes merit receiving print copies of the whole workbook as they are larger or are comprised of students at a critical time in their development as researchers. Having print copies of the workbook allows for more interactive instruction sessions, as students can physically write in worksheets, and exposes students to the range of RDM tasks as they can take the workbook home after class. For other classes, students are referred to one of the online versions of the book.

Surprisingly, the workbook has also been helpful for RDM consultations. Upon receiving a question from a researcher, the librarian can share specific workbook exercises and other information on the relevant RDM topic. This lays a solid foundation for the subsequent in-person consultation, as the researcher has been given a preliminary framework for addressing the problem. During the consultation, the librarian can either work through the exercise with the patron or use the exercise as a touchstone for more in-depth problem solving. Using the workbook for consultation has been an unexpected benefit of this resource but one that has been very helpful.

## Conclusions

This article focused on the development of *The Research Data Management Workbook* and future work needs to be done to evaluate the effectiveness of this method of education. Future improvements to the workbook are also planned, including adding new exercises and adjusting existing exercises based on feedback from teaching with them. Thankfully, the technology used to create the workbook allows for such continual updates, which would not be possible with a traditionally published book.

The workbook has been helpful at a small university with few opportunities for RDM instruction. These sessions are often limited to an hour or less, making it necessary to prioritize what topics to cover and in what depth. The workbook provides options for interactive RDM exercises to work through during this precious time and, where topics cannot be covered comprehensively, exercises are made available to help students bridge from a brief overview of an RDM task to actually implementing it. It is challenging to make the most of one-shot library instruction, but the workbook provides a pathway beyond the classroom to integrate RDM into research.

---

### Competing Interests

The author declares that they have no competing interests.

### Data Availability

No data was generated by this project.

---

### References

- Briney, Kristin. 2015. *Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success*. Exeter, UK: Pelagic Publishing.
- Briney, Kristin. 2023. *The Research Data Management Workbook*. 1.0. Pasadena, CA: Caltech Library. <https://caltechlibrary.github.io/RDMworkbook>.
- Corti, Louise, Veerle Van den Eynden, Libby Bishop, and Matthew Woollard. 2019. *Managing and Sharing Research Data: A Guide to Good Practice*. Second edition. Thousand Oaks, CA: SAGE Publications Ltd.
- GitHub. 2023. "GitHub Pages." GitHub Pages. 2023. <https://pages.github.com>.
- Griffin, Tina M. 2021. "Knowledge and Practice Changes Following a Student Data-Focused Data Management Education Program." *Journal of Librarianship and Scholarly Communication* 9 (1). <https://doi.org/10.31274/jlsc.12906>.
- Lamar Soutter Library - University of Massachusetts Medical School. 2023. "New England Collaborative Data Management Curriculum." 2023. <https://library.umassmed.edu/resources/necdmc/index>.
- LEARN Project. 2017. "Research Data Management Toolkit: Now Available." April 4, 2017. <https://learn-rdm.eu/en/research-data-management-toolkit-now-available>.
- Nelson, Megan Sapp, and Ningning Nicole Kong. 2020. "Capturing Their 'First' Dataset: A Graduate Course to Walk PhD Students through the Curation of Their Dissertation Data." *IASSIST Quarterly* 44 (3). <https://doi.org/10.29173/iq971>.
- Pagowsky, Nicole. 2021. "The Contested One-Shot: Deconstructing Power Structures to Imagine New Futures." *College & Research Libraries* 82 (3). <https://doi.org/10.5860/crl.82.3.300>.

"RDMkit." 2021. June 23, 2021. <https://rdmkit.elixir-europe.org>.

Read, Kevin, and Alisa Surkis. 2017. "Research Data Management Teaching Toolkit." figshare. <https://doi.org/10.6084/m9.figshare.5042998.v3>.

The University of Edinburgh. 2022. "Research Data MANTRA." 2022. <https://mantra.ed.ac.uk>.

University of Michigan, Jake Carlson, Megan Sapp Nelson, Marianne Bracke, and Sarah Wright. 2015. "The Data Information Literacy Toolkit." Purdue University. <https://doi.org/10.5703/1288284315510>.

Wiggins, Grant, and Jay McTighe. 2005. *Understanding By Design*. 2nd Expanded edition. Alexandria, VA: Assn. for Supervision & Curriculum Development.

Xie, Yihui. 2023. "Bookdown." <https://github.com/rstudio/bookdown>.