Title

Expanding research data management to UC Berkeley researchers: a targeted approach to outreach and instruction

Authors

Sackmann, Anna
Neeser, , Amy
Teplitzky, Samantha
et al.

Publication Date

2022-01-03

Peer reviewed

Title

Expanding research data management to UC Berkeley researchers: a targeted approach to outreach and instruction

Authors

Sackmann, Anna
Neeser, Amy
Teplitzky, Samantha
et al.

Publication Date

2022-01-03

DOI

10.25350/B5SG6S

10

EXPANDING RESEARCH DATA MANAGEMENT TO UC BERKELEY RESEARCHERS

A Targeted Approach to Outreach and Instruction

Anna Sackmann, Amy Neeser, Samantha Teplitzky, Ann Glusker, and Elliott Smith[1]
University of California, Berkeley

Data-intensive research has markedly increased over the last ten years, creating a new urgency for the careful curation, storage, documentation, and reuse of data outputs. The importance of properly managing research data and other outputs, such as code, continues to grow in importance, especially as publishers and funders institute mandates and as research activity turns to greater use of computation and data.[2] Data-intensive science, or the fourth paradigm of research, requires that institutions, governments, and other agencies provide a new set of tools and workflows to address the rapid advancement of data generation and reuse.[3] With the increase in computing power and availability, researchers across all disciplines, especially in the life, health, and physical sciences, are utilizing large datasets, code, and modeling in their findings. The various technological components of data and computing add to the complexity of long-term data use and storage due to technological obsolescence, bit decay, and general loss

DOI: 10.25350/B5X592

due to poor management.[4] Increasingly, librarians are providing outreach and instruction to researchers to help them properly manage their research outputs and develop workflows that increase experimental rigor and reproducibility, often through open and transparent workflows. This can be in the form of instruction, consultation, or outreach for services such as cloud storage, data management planning, data curation, and data discoverability and reuse.

RESEARCH DATA MANAGEMENT AT THE UNIVERSITY OF CALIFORNIA, BERKELEY

At the University of California, Berkeley, consultants in the Research Data Management (RDM) Program and librarians teach broad-reaching RDM concepts with a shared goal: to increase rigor and reproducibility. After reviewing our technology services and benchmarking them against several peer institutions in 2013, UC Berkeley's Research Information Technologies (Research IT) group and the library identified research data management support as a growing need and one in which both campus organizations hold expertise.[5] The development of the university's RDM Program emerged out of this collaboration to provide consulting and computing technologies to support the rapidly growing use of data and computation within all disciplines. The program's collaborative partnership leverages the technological infrastructure and expertise within IT with the research support, discovery, curation, and preservation expertise of librarians.[6] At the core of these services, librarians both advise and teach new and seasoned researchers how to best manage their research data outputs for the purposes of personal organization, increased rigor and reproducibility, and increased transparency of the research process.

The RDM Program is composed of consultants from Research IT and librarians who hold research data expertise across subject areas. They are part of a larger team of consultants who, in addition to data topics, consult about research computing. This diverse data and computation consulting team also includes staff from the Lawrence Berkeley National Laboratory, graduate student domain consultants, and undergraduate data peer consultants. They attend shared meetings, office hours, and cross-training and have contributed to a shared consulting handbook. The RDM Program also frequently collaborates with other campus experts and librarians outside of its immediate consultant network, such as the open science librarian and the emerging technologies and bioinformatics librarian at UC Berkeley Library,

to provide specialized instruction. Building and articulating data expertise within the library to better respond to the increasing educational and research needs of students, faculty, and scholars is one of the main goals of the UC Berkeley Library as is articulated in its Data Initiatives Plan.[7] This chapter will address how librarians can develop outreach and instruction strategies that are both targeted and scalable to provide RDM instruction at a large academic research institution.

PROPER DATA MANAGEMENT FOR INCREASED RIGOR AND REPRODUCIBILITY

Science, and research more broadly, is a communal effort built over time through continual assessment and communication of methods and results.[8] Constructing this communal effort to test hypotheses requires the ability to easily access, understand, and reuse data, code, and other research outputs—or proper research data management. Reproducibility, specifically the ability for a different team or individual in a different location to achieve the same measurement or result, has always been a cornerstone of science.[9] The metaphor "standing on the shoulders of giants" is often used to illustrate how research cumulatively builds, most notably, through the tradition of writing literature reviews.[10] In addition to published papers, findable, accessible, and documented data and code are increasingly crucial to show how research develops. Librarians are well-positioned to demonstrate the connection between research data management and reproducibility by approaching research from a transdisciplinary perspective and appraising research data without the connection to traditionally maligned academic incentives. For example, the development of an academic record through publishing may pressure researchers, especially those early in their careers, to inflate their research findings, either consciously or unconsciously.[11]

Proper data management requires detail and attention to documented workflows such that data and code are reproducible to not only indicate a finding but also accurately communicate that finding.[12] UC Berkeley's RDM program advises researchers on developing and implementing workflows to address data and other outputs before, during, and after the research process. The program's consultation services and instruction initiatives aim to help researchers develop and document usable conventions for managing, storing, and versioning data. Left without guidance, researchers may develop their

own bespoke workflows, which require time and may result in lost data or analysis. When researchers take a haphazard approach to data management, their work is often left irreproducible by others and their future selves.[13]

In the following case studies, we show how librarians at UC Berkeley have leveraged RDM instruction to provide targeted, disciplinary expertise to increase rigor and reproducibility and to create open workflows. Open science is not only a commitment to transparency and accessibility throughout the research process that encompasses tools, workflows, and documentation but also to transparent and accessible knowledge that is shared and developed through collaborative networks.[14] The case studies illustrate outreach and instruction structures that are transferable to other disciplines and cohorts of researchers and that are scalable at large, decentralized research institutions.

Librarians and RDM consultants have shifted away from general, one-shot, drop-in workshops to strategically planned departmental or cohort and funder-driven outreach and instruction. After several iterations of offering general data management workshops at the beginning of each semester with low attendance and minimal follow-up, we pivoted our approach to offer targeted education that can translate and scale to larger groups. The case studies highlight current data instruction and outreach initiatives: targeted outreach efforts to new and interested faculty, cohort training for the National Institute of Health's (NIH) required Responsible Conduct of Research Program, and the Open Science workshop series.[15]

CASE 1

Targeted Outreach to Faculty and Graduate Student Researchers in the Engineering and Physical Sciences

UC Berkeley's RDM Program is closely affiliated with Research IT's Berkeley Research Computing (BRC), which supports computing and data analysis needs across campus. Over the course of the last three years, the RDM Program service lead worked closely across the Research IT organization to merge the RDM and BRC consultation services to be jointly managed on a single platform with shared workflows and opportunities for co-consultations. The two services frequently serve the same researchers, and by merging the two consulting services, researchers experience a "one-stop shop" instead of submitting requests for the two separate services. At a larger institutional level, UC

Berkeley continues to invest in data-intensive transdisciplinary programs, such as the NIH-supported Biomedical Big Data Program and the Computational Social Science Training Program, as well as grow the Division of Computing, Data Science, and Society. Now unified, these two campus services have an opportunity to collaborate on comprehensive research services for students and researchers.[16] With the newly combined consultation services, Research IT piloted a dual outreach effort that was targeted, scalable, and provided an introduction to services for computing and data management.

Since its inception in 2015, the RDM Program has taken several approaches to outreach with the goal of increasing researcher engagement and education regarding data management from the beginning of a research project to the final archival and sharing stages. In the beginning years of the RDM Program, consultants set up information tables in key campus locations at the beginning of the academic year and presented at orientations for new graduate students. We offered several general drop-in RDM workshops at the beginning of each semester aimed at introducing researchers to discipline agnostic tools, such as the DMPTool, and general data repositories. Additionally, we scheduled themed programming during the annual, international Love Data Week event that takes place across research institutions in February of each year. However, these general efforts of tabling and drop-in workshops resulted in low engagement and low attendance from graduate students and faculty.

Therefore, we refocused our attention to a targeted outreach pilot, which consisted of identifying and individually emailing new faculty and faculty who had previously used Research IT services (either RDM or BRC) in the engineering and physical sciences departments to request a thirty-minute meeting to talk about data and computing workflows. Librarians collaborated with Research IT to meet with selected researchers to address research computing, data security, and data publication and preservation. This enabled both parties to learn how researchers approach data and information creation, perceive research output value, and envision their research contribution within and beyond their disciplines. These meetings provided key insight into department and lab culture regarding open scholarship and how (or if) researchers are increasing transparency through data management and computing practices.

The initial work of identifying interested faculty—pulled from a previous Research IT user engagement survey and a list of new faculty members from 2017 to 2019—was a time-intensive activity. In the first iteration of this targeted outreach model, we contacted fifty-one new faculty, had further

correspondence with nine of these faculty, and finally met with six faculty and their research groups. We carefully documented the following workflow to make replication at a future date much simpler and faster:

1. Identify department(s) for target faculty outreach.
2. Assemble a small team of consultants to include a subject liaison, BRC consultant, and RDM library consultant.
3. Identify researchers to contact new faculty members and Research IT annual survey respondents.
4. Compose individualized messages requesting a meeting to discuss data workflows and computing needs.
5. Prepare for meetings by reviewing research group membership, research topic areas, and previous journal/data publications.
6. Document workflow and computing/data needs during the meeting.
7. Send a follow-up email within two business days with next steps and linked information.
8. Follow up with main contact within one year to maintain the connection.

Rather than advertising our organizations and services, we used a topical approach and created one-page handouts on topics such as data publication and preservation, secure data, and research computing. We have found this approach to be more relevant to researchers as they do not seem to care about who is helping them, only that they get the help they need. This also allowed us to highlight and promote services specific to UC Berkeley, including institutional cloud storage, computing options, and Dryad, the University of California's data repository. Additionally, Research IT is leading an effort to improve services for researchers working with sensitive and restricted data; so, we are able to promote that work and find potential users to inform the construction of these new services and infrastructure.

Due to our success in the physical science and engineering departments, we have extended the outreach campaign to target researchers in the School of Public Health. Our consulting program has seen an influx of requests from researchers in Public Health, many of whom are working with sensitive data that involves human subjects and/or requires additional data security protections. In addition to the subject librarian and representative from Research IT, a graduate student domain consultant who is pursuing a PhD in public health and the research data management service lead are taking part in the outreach conversations. Combining a holistic approach of combining data

and computation and a topical approach of introducing topics rather than services to support research has been essential in the library and Research IT's collaboration to develop a model of seamless support that leverages the expertise of both organizations. Researchers at UC Berkeley are better able to utilize computing and storage infrastructure because of the high-touch consultations that take into account the variability of research needs as is shown by an increased number of co-consultations between RDM and BRC. By bringing together RDM and research computing, we help researchers make their data open and available when possible and secure when needed.

CASE 2

Responsible Conduct of Research (RCR) Training for Researchers Funded by the NIH

UC Berkeley has developed a program to support the Responsible Conduct of Research (RCR) ethics requirements as first stipulated by the National Institutes of Health (NIH). The federal funding organization has long recognized the need for enhanced rigor and reproducibility in research. RCR is defined by the NIH as "intellectual honesty in the formulation, conduct, and reporting of scientific research."[17] At UC Berkeley, the RCR Program provides training to address ethical research conduct and aims to enhance research reproducibility. More recently, the National Science Foundation (NSF) and the National Institute of Food and Agriculture (NIFA) have added RCR training requirements for certain grant recipients. The NIH specifically stipulates that all "...trainees, fellows, participants, and scholars receiving support through any NIH training, career development award (individual or institutional), research education grant, and dissertation research grant must receive instruction in responsible conduct of research."[18] Moreover, the NIH stipulates that online training is insufficient and must take place in a face-to-face environment. Due to the COVID-19 pandemic, all sessions were successfully transitioned to synchronous training via Zoom.

RCR training sessions are embedded in a variety of research programs at UC Berkeley, including molecular and cell biology (MCB) 293, a course that all incoming graduate students in bioengineering, chemical biology, and molecular and cell biology are required to take. The curriculum for MCB 293 includes research ethics, reproducibility, and statistical validity.[19] After learning of the rigor and reproducibility component of the course, the subject liaison

and librarians within the RDM Program contacted the course instructor to offer a guest lecture on data management and organization, which includes overviews of important campus tools, general practices, and concepts of rigor and reproducibility. Following the first lecture for MCB 293, we have been invited to teach the module to additional graduate cohorts, including a research group in the university's neuroscience institute, a graduate cohort in the Computational Social Science Training Program, and a cohort of graduate researchers in the School of Public Health, where the public health librarian assisted with teaching. In addition to the ongoing involvement with MCB 293, we now regularly teach the module to a multidisciplinary, graduate-level research ethics course that is taken by students in a variety of fields who receive research funding through a number of different organizations. Since the beginning of our collaboration with the RCR program, librarians have taught ten sessions reaching over three hundred graduate students. For each iteration of the class, we are able to utilize the subject expertise of the liaison librarian and the functional expertise of RDM librarians.

Because many of the graduate students in MCB 293 are at the beginning of the research process, we focus on building content around supporting data instead of data management. The ninety-minute module is divided into three sections:

1. Part 1: Introduction to data support and why RCR training matters
2. Part 2: Planning, documenting, and describing data and other outputs during the research process
3. Part 3: Data storage at UC Berkeley, archiving, and sharing

During the introduction component of the workshop, students are introduced to data management within four contexts: the personal benefits, research ethics, funding organization data management policies, and publisher data sharing policies. For graduate students who are at the beginning of their research careers, the importance of data management for purposes of rigor and reproducibility requires a metaphorical carrot and stick explanation. Graduate students in MCB 293 conduct individual research projects during three nine-week laboratory rotations during which they are collecting and analyzing data. During the rotations, they are not yet publishing or writing grant proposals with data management plans. Since they are in the early stages of building their research and data management habits, the daily workflow practices and tips introduced in this session are immediately applicable to this

group. For graduate students who are further along in their research careers, they may already understand the personal benefits to data management because of previous instances of data loss or time lost to poorly organized and documented data. The session content that addresses publisher, institutional, and funder policies is more immediately applicable to this group of researchers. The ethical argument of data management is more difficult to communicate and requires an environment in which departments and advisors communicate the value of open and reproducible workflows, which varies significantly between disciplines.

Following the introduction, students view "Data Sharing and Management SNAFU in 3 Short Acts," created by the NYU Health Sciences Library, in which animated panda bears illustrate the perils of sharing poorly documented data.[20] The short video is relatable for the researchers and injects humor into the process of proper data management. Students then read a data management case study and are asked to discuss questions regarding file types, storage, and organization with a partner or in small groups. Utilizing a video and a case study with discussion questions provides the students with an opportunity to understand the motivation for data management. The librarian then guides the students through best practices for documentation and description, which includes preferred data file formats, file naming conventions, metadata standards, and the crucial components of a descriptive readme.txt file.

Finally, students are introduced to data storage and backup options and tools at UC Berkeley. They are provided with three data-loss scenarios in which data becomes lost or inaccessible due to issues with hardware, encryption, or lack of digitization. They are then asked to share possible solutions with the group. The session closes with a section on data sharing and publishing in which students are provided with a workflow and examples of successful data publishing at UC Berkeley.

Librarians' involvement in RCR instruction on increasing rigor and reproducibility is not a novel concept; rather, it is a natural extension of our current work. Librarianship requires that information sources, data and beyond, be appraised and organized for maximum discoverability and usability. Moreover, the combination of subject and functional expertise grounds the key concepts of RCR training within a domain context while providing UC Berkeley-specific solutions to handling data from the inception of a research project through the final stages of sharing and communicating outputs. With minor adjustments, the material is adaptable across domains.

CASE 3

A Cohort-Focused Open Science Workshop Series for New Graduate Students in the Earth and Planetary Science Department

The Engineering & Physical Sciences Division of the UC Berkeley Library partners with researchers to support the entire research lifecycle, with a focus on opening up key components of that cycle. We aim to understand how workflows have changed within scientific disciplines, identify new opportunities for librarians, and extend support to open science workflows emerging within the departments we support. Throughout the fall semester of 2019, librarians and staff piloted a series of open science workshops in the Department of Earth and Planetary Sciences that focused on trending methods and workflow tools relevant to new graduate students as they launch their research careers. The sessions reached a cohort of twelve students and showed them how to manage both their research and their data as well as encouraged them to adopt open and reproducible workflows.

The pilot had four major objectives:

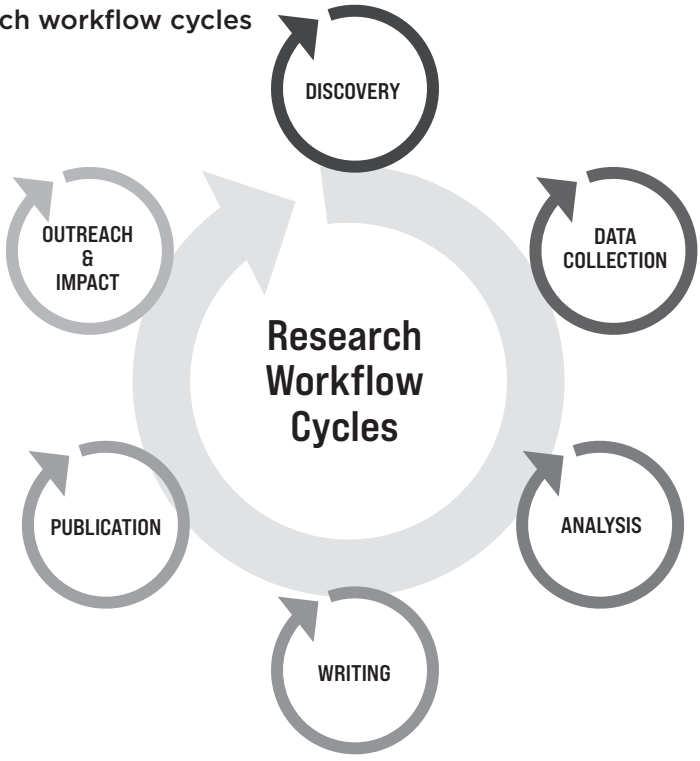
- Incoming graduate students will develop an ongoing community of practice around open science workflows.
- Series participants will be able to apply and explore practices and tools in daily research workflows that support openness, integrity, and reproducibility.
- At least one series participant will take on a role as department ambassador or future workshop instructor.
- The library will be established as a key collaborator and resource for open research workflows.

The series consisted of six workshops within one semester that addressed open science topics as they arise at different points in a typical science research workflow. Research workflows were represented as a larger cycle with smaller, iterative cycles nested within (figure 10.1).

Students were asked to consider the larger cycle as their dissertation or thesis project and the smaller cycles as necessary steps within that project. These steps include the following:

- Discovery (topic selection and initial background research)
- Data Collection (data collected through experiments, fieldwork, or model runs, for example)

FIGURE 10.1
Research workflow cycles



- Analysis (methods and processes used to analyze and interpret data)
- Writing (using open or collaborative tools to write up results)
- Publication (selecting a venue, submitting, and revising the finished product)
- Outreach/Impact (presenting work at conferences, assessing its impact, and determining next steps)

As the series targeted first-year-graduate students, it was not necessary to address each part of their eventual research workflow but rather to start them with good habits and practices that could be applied to later stages as they reach them.

The six sessions included the following:

- Introduction to Research Workflows and Literature Searching
- Citation Management

- Collaborative Writing in LaTeX with Overleaf
- Jupyter Notebooks Best Practices
- Dynamic Data Management
- Publishing and Citation
- Approaches to GitHub and Reproducible Code

Sessions were taught by a mix of librarians and library staff, departmental faculty, and graduate students. Curriculum and materials can be found in the related GitHub repository.[21]

The decentralized culture of large academic campuses presents challenges for developing lasting connections between liaison librarians and graduate students. These workshops drew together expertise from the library and an academic department, offering students a scaffolded introduction to locally supported resources. Unlike a boot camp or one-time workshop model, students learned material over the course of several months, giving them the opportunity to attempt, adapt, and develop new methods. The workshops also filled a need expressed by earlier graduate cohorts for domain-specific sessions in their physical location, making this an effective approach for reaching students to introduce foundational open research methods. Trainers benefitted by encountering a cohesive audience and creating content that could be applied to other disciplines. Pilots such as this build communities of practice, test methodologies, and provide an opportunity to scale training to a broader audience.

The scaffolded, open-science workflows series described here was built on a graduate cohort and was aimed to welcome students into their larger community of practice, department, university, and academic science more broadly. The series follows a similar approach to the Open Science Communities that have taken hold in the Netherlands. These bottom-up networks "create a learning environment where scholars can acquire the skills to conduct open, transparent, and reliable research."[22] Such a setting brings peers together who may have no experience with open science research but are willing to learn, share, and serve as nodes to their lab groups and larger disciplines. This grassroots approach puts new scholars on level footing with more established researchers, dispels the mysteries of scientific workflows, and creates a more inclusive atmosphere within the departmental community.

DISCUSSION

Teaching RDM concepts and tools to researchers in UC Berkeley's large, dispersed organization requires a uniquely scalable yet targeted approach. The three case studies highlighted in this chapter utilize scalable structures for outreach and instruction that can be replicated with minor changes to serve additional departments and researcher cohorts. In order for the education efforts highlighted in each case study to be successful in teaching researchers RDM, we identified specific attributes that make the effort both specific and scalable (table 10.1).

The case studies differ in their approach and execution, yet the structures transition to larger groups of researchers and have the potential to be applied

TABLE 10.1
Comparison of case studies/approaches to RDM outreach

	CASE STUDIES		
THEME	**Case 1 RDM Outreach**	**Case 2 Responsible Conduct of Research (RCR)**	**Case 3 Open Science Series**
Audience	New faculty	NIH-funded researchers	Graduate students
Discipline	Engineering, Earth Sciences, Public Health	Molecular & Cell Biology; Life & Health Sciences; Engineering; Social Sciences	Initial focus on one department (Earth & Planetary Science) with room to scale
Scale	Department, College or School; research groups	Local iteration of federally mandated training	Community; Department
Incentive for participants	Orientation; guidance	Mandated by the NIH and other funders	Informal—bonding, connection, camaraderie
Inclusion & Power Balances	Early career researchers & previous Research IT users	All graduate students and postdoctoral researchers funded by NIH	Graduate students: scaffolded cooperation; mutual sharing and learning
Future directions/ Sustainability	Scalable model that can be extended to any department or program	Established and embedded as part of Graduate Program Training	Get approval to offer for course credit; offer as recurring series beyond department

to different disciplines and institutions. Quantitative metrics that may be traditionally used to measure success, such as the number of sessions taught and the number of session attendees, tell unique stories for each case. For example, in the targeted outreach program, fewer responses and requests for consultations (fifty-one requests for meetings were met with nine email responses and six meetings) may enable librarians and consultants to more deeply engage with the research group and assist with specific computing and data workflows. However, the RCR training consistently reaches roughly sixty researchers each semester, which the librarians and RDM Program consider successful. As discussed in the Open Science series case, a major component of success stems from the opportunity for emerging scholars to engage with established researchers in their field involved in the program series. The cohort consisted of twelve first-year-graduate students who attended sessions led by either librarians, faculty, or other graduate students.

The work of Woodley and Pratt is particularly helpful in understanding how power balances and modes of communication within the UC Berkeley data community influence the interactions within these sessions. [23] Applying their terminology (convey, contribute, collaborate), we can better anticipate how expertise will be shared or communicated beyond the individual sessions. In the RCR training sessions, the organization acts as convenor or expert, informing participants about federal rules and standards. The RCR sessions typically mark the beginning of an interaction between researchers and librarians, after which researchers contact librarians for additional guidance to address data workflows that are unique to their research. RDM Outreach more closely aligns with the "contribute" mode in that researchers share input about their research practices and RDM consultants craft individualized responses. The success of these interactions hinges on the consultant empowering the researcher with tools and guidance to develop new solutions. The Open Science series strikes a balance between "contribute" and "collaborate," distributing power to participants so that they may continue conversations and connections beyond the series. This may result in ongoing changes to their personal research workflows or participants taking new methods of organization and practices back to their research groups.

Librarians and RDM consultants serve as a crucial locus for community-driven approaches to learning and transforming data management and workflow practices. Each case illustrates the need to consider and move beyond flat, quantitative metrics to implement targeted approaches to providing

instruction opportunities for data management and open science. Open science tools, storage options that enable collaboration, and data repositories are only utilized to the fullest extent possible to make data and research outputs findable and reusable if researchers understand why and how they should integrate them into their workflows. Instruction and outreach that address RDM do not need to reach as many researchers as possible in order to see enhanced rigor in research outputs. By identifying ways in which programs may be sustained, scaled, and translated prior to implementation, we prioritize growth and inclusion to connect with emerging and established researchers from a number of different disciplines.

Notes

1. Contributions: Anna Sackmann, Amy Neeser, and Samantha Teplitzky wrote the manuscript. Ann Glusker and Elliott Smith provided review, commentary, and revision. All authors conceptualized the strategies highlighted by the case studies.
2. Carol Tenopir, Ben Birch, and Suzie Allard, A*cademic Libraries and Research Data Services: Current Practices and Plans for the Future*, an ACRL White Paper (Chicago: Association of College and Research Libraries, 2012).
3. Tony Hey, Stewart Tansley, and Kristin Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Redmond, WA: Microsoft Corporation, 2009), www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf.
4. Andrew M. Cox and Stephen Pinfield, "Research Data Management and Libraries: Current Activities and Future Priorities," *Journal of Librarianship and Information Science* 46, no. 4 (2014): 299–316, https://doi.org/10.1177/0961000613492542.
5. UC Berkeley, "RAE Services Peer Benchmarking," 2016, https://rtl.berkeley.edu/rae-services-peer-benchmarking.
6. Jamie Wittenberg and Mary Elings, "Building a Research Data Management Service at the University of California, Berkeley: A Tale of Collaboration," *IFLA Journal* 43, no. 1 (March 23, 2017): 89–97, https://doi.org/10.1177/0340035216686982.
7. Josh Quan and Amy Neeser, "UC Berkeley Library Data Initiatives Plan," UC Berkeley, January 2019, www.lib.berkeley.edu/Staff/Library Data Initiatives Plan.pdf.
8. National Academies of Sciences, Engineering, and Medicine, "Understanding Reproducibility and Replicability," in *Reproducibility and Replicability in Science* (Washington, DC: The National Academies Press, 2019), 39–54, https://doi.org/https://doi.org/10.17226/25303.
9. Melissa Rethlefsen, "Librarians and Reproducibility: It's Time," in Librarians Building Momentum for Reproducibility (Virtual Conference, 2020); Association for Computing Machinery, "Artifact Review and Badging – Version 1.0," Association for

Computing Machinery, 2020, www.acm.org/publications/policies/artifact-review-badging.

10. Jan vom Brocke, Alexander Simons, Kai Riemer, Bjoern Niehaves, Ralf Plattfaut, and Anne Cleven, "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," *Communications of the Association for Information Systems* 37 (2015), https://doi.org/10.17705/1CAIS.03709.
11. National Academies, "Understanding Reproducibility and Replicability."
12. Sara Stoudt, Váleri N. Vásquez, and Ciera C. Martinez, "Principles for Data Analysis Workflows," *PLoS Computational Biology* 17, no. 3 (2021): e1008770, https://doi.org/10.1371/journal.pcbi.1008770.
13. Julia S. Stewart Lowndes, Benjamin D. Best, Courtney Scarborough, Jamie C. Afflerbach, Melanie R. Frazier, Casey C. O'Hara, Ning Jiang, and Benjamin S. Halpern, "Our Path to Better Science in Less Time Using Open Data Science Tools," *Nature Ecology and Evolution* 1, no. 6 (2017), https://doi.org/10.1038/s41559-017-0160.
14. Sonja Bezjak et al., *Open Science Training Handbook* (Zenodo, 2018).
15. National Institutes of Health, "Responsible Conduct of Research Training," 2020, https://oir.nih.gov/sourcebook/ethical-conduct/responsible-conduct-research-training.
16. Amy Neeser and Anna Sackmann, "Holistic Research Support at UC Berkeley: A Library and Research IT Collaboration," *Research IT News*, 2020, https://research-it.berkeley.edu/blog/20/07/22/holistic-research-support-uc-berkeley-library-and-research-it-collaboration.
17. National Institutes of Health, "Update on the Requirement for Instruction in the Responsible Conduct of Research," 2009, https://grants.nih.gov/grants/guide/notice-files/NOT-OD-10-019.html.
18. National Institutes of Health.
19. Elliott Smith, Ann Glusker, and Anna Sackmann, "'Rigor and Reproducibility' Data Support Instruction," in *Librarians Building Momentum for Reproducibility* (Virtual Conference, 2020), https://vickysteeves.gitlab.io/librarians-reproducibility.
20. K Hanson, A Surkis, and K Yacobucci, "Data Sharing and Management Snafu in 3 Short Acts," YouTube video, 2012, www.youtube.com/watch?v=N2zK3sAtr-4.
21. Samantha Teplitzky, A Sackmann, and E Newcome, "Sackmanna/Open_Science_Workshops: Open Science Workshops (Version v1.0)," Zenodo, 2020, https://doi.org/http://doi.org/10.5281/zenodo.4359318.
22. Kristijan Armeni et al., "Towards Wide-Scale Adoption of Open Science Practices: The Role of Open Science Communities," MetaArXiv, 2020, https://doi.org/https://doi.org/10.31222/osf.io/7gct9.
23. Lou Woodley and Katie Pratt, "A Framework to Describe Member Engagement and Information Flow in STEM Communities," Zenodo, 2020, https://doi.org/10.5281/zenodo.3997802.