Title

Open Science and Data Management: Introducing Graduate Students to Research Workflows in a Local Context

ISBN

978-0-8389-3847-8

Authors

Teplitzky, Samantha

Newcome, Erica

Powell, Susan

et al.

Publication Date

2021

Peer reviewed

Open Science and Data Management:
Introducing Graduate Students to Research Workflows in a Local Context

Samantha Teplitzky, Open Science Librarian, UC Berkeley, steplitz@berkeley.edu
Erica Newcome, Technical Processing & Collection Projects Assistant, UC Berkeley, erican@berkeley.edu
Susan Powell, GIS & Map Librarian, UC Berkeley, smpowell@berkeley.edu
Anna Sackmann, Science Data & Engineering Librarian, UC Berkeley, asackmann@berkeley.edu

NUTRITION INFORMATION

The Engineering & Physical Sciences Division of the UC Berkeley Library partners with researchers to support the entire research life cycle, 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 that focused on trending methods and workflow tools relevant to new graduate students as they launch their research careers. The sessions showed students how to manage both their research and their data and encouraged them to adopt open and reproducible workflows. The decentralized culture of large academic campuses presents challenges for developing lasting connections between liaison librarians and graduate students. The type of scaffolded open science workflows series described here builds on existing communities of practice to help overcome those challenges. This recipe provides an outline for implementing an open science workflows series aimed at new graduate students. The recipe can be doubled or tripled to scale and expand across departments.

LEARNING OUTCOMES

- Incoming graduate students will develop an on-going 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.

NUMBER SERVED

- Target the new graduate student cohort from one department; 15–20 participants is ideal for the first "test bake."
- Following the first series, analyze assessment data and identify if the next iteration of the series will target the same department. If expanding, coordinate with the graduate advisor from the new or additional department with scheduling and locations.

COOKING TIME

The preparation time can be considered in three parts: overall program development, individual session development, and session length.

- Program development. Allow one semester or summer session ahead of the series for preparation time. For the first version of the series, much of the program development involves outreach and relationship building. This might require emails to target department staff and faculty, logistical planning to secure instruction space, and possibly outreach to vendors to set up trials or new subscriptions. Time: approximately 2–3 hours/week.
- Session development. Varies by session. Sessions that had been taught before required 1–2 hours of prep time; new sessions required 6–8 hours of prep time.

- Session length. Approximately six sessions were offered in the series. Each was one hour long and ideally could be offered weekly or biweekly at a repeating, agreed-upon time slot.

DIETARY GUIDELINES

As research generates increasing amounts of data, libraries have adapted to support management of that data and tools that enable transparency and reproducibility. These workshops draw together expertise from the library and a discipline-specific department, offering students a scaffolded introduction to locally supported resources. Unlike a boot camp or one-time workshop model, students learn material over the course of several months, giving them the opportunity to attempt, adapt, and develop new methods. The workshops also fill a need expressed by earlier graduate cohorts for domain-specific sessions in their physical location, making this an effective approach for reaching students in order to introduce foundational open research methods. Trainers benefit by encountering a cohesive audience and creating content that can 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.

INGREDIENTS & EQUIPMENT

- Classroom space
- Computers (either library or personal computers)
- Projector/display screen
- Access for each participant to software relevant to each session

PREPARATION

Organizing a workshop series with the continued participation of a specific group of attendees requires different levels of coordination among the target departments, student participants, and contributing instructors.

- First, contact and secure buy-in from the target department at least one semester ahead of instruction. This might involve meeting with administrative staff, student advisors, faculty, or department chairs.
- Confirm that students do not receive similar instruction in other orientation, pedagogy, or department-led courses.
- Then, contact members of the incoming cohort if emails are available, or work with the department to choose and block a time open for all.
- Begin to recruit instructors once the schedule is secure while leaving some flexibility according to their availability. In addition to library colleagues with particular resource expertise, target graduate students and faculty as potential instructors. It is important to include several non-library instructors to make sure domain-specific knowledge and practices are imparted.
- Adapt and update past library workshops if possible and then create new instructional material as needed.

COOKING METHOD

Create a series of approximately six workshops within one semester that address open science topics at different points in a typical science research workflow. Research workflows can be represented as a larger cycle with smaller, iterative cycles nested within (figure 1). Students might 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:

1. Discovery (topic selection and initial background research)
2. Data collection (data collected through experiments, fieldwork, model runs, for example)
3. Analysis (methods and processes used to analyze and interpret data)
4. Writing (using open or collaborative tools to write up results)

Figure 1. Representation of research workflow cycles

5. Publication (selecting a venue, submitting, and revising the finished product)
6. Outreach/impact (presenting work at conferences, assessing its impact and determining next steps)

In a series targeting first-year graduate students, it might not be appropriate to address each part of their eventual research workflow but rather to start them with good habits and practices that can be applied to later stages as they reach them. In a pilot series, focus on the relevance to your local audience as well as department strengths to determine the topics of your sessions.

The example sessions below primarily addressed the following aspects of the research lifecycle: discovery, writing and publication, data, and analysis. Session materials are located in the Github repository in Additional Resources.

Discovery: Introduction to Research Workflows and Literature Searching

- Objective: Introduce participants to the concept of research workflows and approaches to literature search and discovery within their subject domain. Differentiate between library-licensed or subscribed resources and open sources.
- Time: 50-minute session
- Instructional materials: slides and two hands-on exercises
- Brief outline: This session introduces students to the concept of research workflows, laying the groundwork for the series as a whole. They are then led through two exercises. In the first, they analyze and compare two literature databases within their subject domain. In the second, they learn to set up alerts to track current and emerging research on their proposed topics.

Writing: Citation Management with Zotero

- Objective: Participants will understand the purpose and value of citation management software in their research. They will be able to use basic functions in Zotero for saving, organizing, and exporting citations.
- Time: 50-minute session
- Instructional materials: Print handout, including basic contextual information (what is a citation manager, why use one, etc.) and simple exercises to introduce participants to the software interface and managing citations.
- Brief outline:
  - What is citation management software?
  - Why use Zotero?
  - Installing Zotero and browser plugin
  - Adding items from a variety of sources, including subject-specific databases
  - Syncing and grouping your citation library
  - Inserting citations and creating bibliographies

Writing/Publication: Collaborative Writing in LaTeX with Overleaf

- Objective: Students will be able to write in LaTeX syntax and collaborate on documents using the Overleaf platform.
- Time: 50-minute session
- Instructional materials
  - Prepopulated Overleaf template that is relevant to the discipline
  - Exercises that give students hands-on experience with LaTeX syntax
  - Students bring their own laptops or librarians provide computers
  - Overleaf account
- Brief outline:
  - What is LaTeX? Define LaTeX typesetting program for attendees.
  - Introduce Overleaf software as a platform for collaborative writing in LaTeX.
  - Explain how documents are structured in LaTeX.
  - Define basic commands; introduce the first set of hands-on exercises using these commands.
  - Explain how tables are formatted and displayed in LaTeX; introduce the second set of hands-on exercises.
  - Discuss citation management and the creation of bibliographies in LaTeX. Introduce exercise for participants to create or sync a bibliography (or .bib) file and experiment with adding in-text citations.

Data: Dynamic Data Management, Publishing, and Citation

- Objectives:
  - Workshop attendees will be able to describe their research design.
  - Attendees will be able to evaluate their storage needs and identify an appropriate solution.
  - Attendees will understand why they need to share and preserve their data.

- Time: 50-minute session
- Materials:
  - Case study with accompanying worksheet for students to work through a data management scenario
- Brief outline:
  - Reasons for proper data management
    - Personal benefits: instituting good practices circumvents future issues, including data loss
    - Funder requirements: many require data management plans as part of the grant application process
    - Publisher requirements: some may require that data and accompanying outputs be made available upon publication
    - Ethical science: proper data management creates a system that fosters research transparency, accountability, and integrity
  - Develop a usable data management plan in three steps:
    - Before beginning research: describe the research design, including the use of computers (shared or individual), data that will be generated or collected, and the anticipated formats. The DMPTool provides guidance for this process.
    - During research: follow the 3-2-1 storage rule (three copies of the data on two different devices with at least one offsite). Investigate local institutional solutions, such as Box, Google Drive, Dropbox, or other physical server infrastructure.
    - Create a file naming system that is comprehensive and understandable. Document the file naming system in a readme.txt file.
    - After research: revisit the funder mandates and check on publisher requirements. Deposit data and other research outputs in a reputable repository for long-term archiving, like Dryad or Zenodo. Reference re3data.org to learn about additional repositories.

Data/Analysis: Guest Speaker Sessions for Jupyter Notebooks Best Practices, and Approaches to Github and Reproducible Code

- Objective: Guest speaker sessions bring in expertise beyond the library and help students connect with resources on their campus as well as local experts within their departments.
- Time: These sessions are also 50 minutes, but exceptions can be made to accommodate guest schedules
- Materials: Hosted on Github
- Brief outline:
  - Jupyter Notebooks best practices: This session was led by a peer graduate student and introduced students to Jupyter notebooks, with domain-specific examples and packages.
  - Approaches to Github and reproducible code: This session was led by a faculty member and addressed the process for using Github and Zenodo to manage reproducible code.

CLEAN-UP

Consider informal feedback collected through the series as well as formal assessment tools. Follow up with student(s) interested in leading future sessions.

CHEF'S NOTES

- Regardless of the amount of planning and schedule, it is a dynamic process. There will be components of the program that move out of your control, like attrition due to outside forces and schedule changes.
- Reuse/repurpose materials from previously taught workshops.
- Construct workshop materials so that activities are domain-specific (plug-and-play).

ADDITIONAL RESOURCES

References

Bosman, J., & Kramer, B. (2016, June 30). Of shapes and style: Visualising innovations in scholarly communication. figshare. Retrieved March 11, 2020, from https://figshare.com/articles/Of_Shapes_and_Style_visualising_innovations_in_scholarly_communication/3468641/1

GitHub. GitHub repository. Retrieved from https://github.com/EPS-Libraries-Berkeley/Open_Science_Workshops

Teplitzky, S. (2019, October). Hanging out your open science shingle: Launching an open science program by supporting a new graduate cohort. Zenodo. Retrieved from http://doi.org/10.5281/zenodo.3510306