Delicious Synergy:

Using DMPs to Build Library Engagement With Data-Intensive Student Programs

***Greg Janée**, University of California Santa Barbara, gjanee@ucsb.edu; **Renata G. Curty**, University of California Santa Barbara, rcurty@ucsb.edu; and **Julien Brun**, University of California Santa Barbara, jb160@ucsb.edu*

NUTRITION INFORMATION

Libraries at research universities want to engage with faculty and departments to address the curation of research data and the advancement of reproducible science practices and FAIR data, but finding opportunities to do so can be challenging. This recipe offers a structured approach to developing a deep and lasting engagement with a department. The essential prerequisite for this approach is that the department in question have a data-intensive student program such as a final project or capstone. By adding a data management plan (DMP) requirement to the program and then building on that, the library can form a delicious synergy with the department.

LEARNING OUTCOMES

Readers of this recipe will be able to form an engagement between the library and a campus department that provides structured opportunities for the library to consult, instruct, and otherwise support faculty and students on data management and open science topics.

NUMBER SERVED

Our experience is that more than one research data librarian is required to handle a cohort of even thirty students, but two to three librarians can handle up to a hundred students.

INGREDIENTS AND EQUIPMENT

- An academic department that requires students to complete a final project or capstone that involves acquiring, analyzing, generating, and possibly archiving diverse datasets.
- Library staff who have expertise in research data management and in open, reproducible science principles and practices.

PREPARATION

The first step is to add a data management plan (DMP) requirement to the department's curriculum. (See the Additional Resources section below for a complete example of curricular DMP requirement text.) It goes without saying that this involves the consent and even active participation of the department. Explain the benefits of adding a DMP requirement, which include:

- Students gain experience in data management planning, which will be required if they ever apply for grant funding in their future careers.
- Faculty advisors, even if they are familiar with research data management recommendations, typically only have time to focus on the topics and substance of student projects, but adding a DMP requirement ensures that students will receive critical data management instruction from the library.
- The students' work will be better organized; their data will be more reusable and their work will be more reproducible.
- A DMP requirement will facilitate the construction and maintenance of an archive of student projects and associated data.

Once a DMP requirement is in place, the library is poised to offer additional services and resources. For example, DMPs provide an opportunity for the library to deliver student workshops on open science principles and data management practices. The library can also consult with students on their specific projects and assist them in curating and publishing their data. Through these efforts, the department will gain familiarity with library staff and services, which can lead to further opportunities for collaboration on topics like data repository solutions and open access publication opportunities. The library can also play a role in assessing student DMPs, just as grant application DMPs are assessed by funders. Assessment of DMPs also provides the opportunity to educate faculty on data management requirements and best practices. (See table 1 for a sample academic calendar that highlights potential library interaction points.)

Table 1. *Timeline of a Two-year Capstone Project. Interactions with the library team are highlighted in underlined text with a triangle bullet*

Semester	Student activity
One	○ Develop work plan in consultation with faculty advisors
	▶ Attend library workshop on data management planning
	○ Incorporate feedback from faculty advisors
	▶ Incorporate feedback from library DMP assessment with optional library consultation
	○ Host work plan review meeting with faculty advisors, clients if any, and external advisors
	○ Submit final work plan
Two	○ Host review meeting with faculty advisors, clients if any, and external advisors
	▶ Attend library workshop on data management
	○ Submit outline of final report
Three	○ Draft background, methods, and preliminary results sections of final report
	○ Receive faculty review of project
	○ Complete draft of final report
Four	○ Draft executive summary
	▶ Consult with the library on data/metadata and code deposit requirements and procedures
	○ Deliver final presentation
	▶ Submit data, metadata, and code to archive and go through the curation process
	○ Submit self/peer evaluation to faculty advisors and program coordinator
	○ Submit faculty advisor evaluations to program coordinator

COOKING METHOD

As students create their DMPs and then work to implement them, the library can offer workshops on research data management, both planning and implementation, and on general open science and reproducibility principles.

We suggest offering two workshops:

- The first workshop is designed to help students develop their data management plan as part of their project proposal. Agenda topics include:
  - open science and its importance
  - data lifecycle
  - introduction to DMPs
  - developing your DMP
  - FAIR principles as a framework for planning data management
  - CARE principles for sensitive data management
  - demonstration of the DMPTool (https://dmptool.org/)
  - how to use your DMP as a living document
- The second workshop, which should take place after students have begun to work on their projects in earnest, focuses on data management and preservation. Agenda topics include:
  - reproducibility and analytical workflows and how scripts can help
  - data management tips
    - tracking data sources
    - keeping a copy of the raw data
    - making work portable (project folder structure and naming conventions)
  - how to document
    - README files
    - code commenting
    - metadata
  - keeping data safe
    - backup strategies
    - file permissions
  - data preservation
    - what to preserve?
    - preparing files
    - preparing README files
    - submission and curation process

If the library offers an institutional repository, as students near the archiving phase of their project, they can reach out to library data curators to consult on how to best develop the necessary documentation and how to select which datasets to reference/cite and which to include in the archive. This early consultation can reduce back-and-forth communication during the curation step and provides its own teaching opportunity.

As a complement to the workshops, consider baking all the necessary information into a website that students can reference as they work on their projects. We recommend organizing the materials according to the phases of the research data life cycle—planning, man-

aging, and preserving. (For an example, see UC Santa Barbara's Project Data Management website, https://ucsb-library-research-data-services.github.io/project-data-management/.)

CHEF'S NOTES

Program-specific DMP requirements and templates can be crafted and added to the DMPTool (https://dmptool.org/) to streamline the DMP writing process while connecting students to campus-specific guidance and local resources. This will also simplify the feedback process since students can use the tool to request help and incorporate feedback.

Library assistance with the assessment of DMPs will very likely be coincident with department evaluation of student project proposals, and will thus require that the library have the resources to perform such work at a specific point and by a certain deadline in the academic year. The same is true if the library assists with deposit of student data: the library must be able to complete curation of data deposits before students graduate.

ADDITIONAL RESOURCES

The following is an example of DMP requirement text that can be inserted into a department curriculum. This example happens to include a data deposit requirement. Notice how the text includes the form the DMP should take, the rationale for its requirement, and advice and examples.

Data, metadata, and associated code emerging from the project will be archived for public use. The only exception is for data restricted by a non-disclosure agreement or a license. For both public and restricted data, the group's work plan must include a Data Management Plan (DMP). The DMP (one to two pages) describes how research data will be managed during the project and, if appropriate, made available to others after completion. There are six major topics to discuss in the DMP:

1. **Describing the research data.** What data are needed? Are such data available? When and how will the data be acquired? Provide a description of the data the group will collect or reuse, including the file types, data set size, the number of expected files or sets, content, and source of the data (creator and method of collection).
2. **Data standards.** Are there any standard formats in the specific research field for managing or disseminating the data sets that have been identified? Who in the group will have responsibility for ensuring that data standards are properly applied and data are properly formatted?
3. **Metadata** is documentation that helps make data sets reusable. Think about what details (metadata) someone would need in order to be able to understand and use your data. For example, perhaps a README.txt file is necessary to explain variables, the structure of the files, etc. In addition, metadata should follow disciplinary standards, including ontologies and vocabularies. If applicable, describe how the group's model construction, scripts, and/or workflows will be documented.
4. **Data sharing and access.** The data may have significant value for other researchers beyond this project, and sharing your data is part of your responsibility as members of the scientific community. Specify the extent to which data can be reused, including any access limitations. List any proprietary software that might be needed to read the files. If there is data that is not appropriate for sharing due to confidentiality, NDA, or disclosure risk, then describe that here.
5. **Intellectual property and reuse.** If data were collected from a client organization or other external distributor or source, does the group have the right to redistribute it? If so, are there any restrictions on redistribution? If the group created its data files, will it assign a Creative Commons license to its data?
6. **Data archiving and preservation.** Throughout the project, the group may produce a large number of files. At the end of the project, groups must submit data used in the project (except data protected by non-disclosure agreements) and associated metadata for archiving. Not all data needs to be saved. If another researcher wanted to replicate the group's work or reuse the group's data, what data and documentation would be required for them to do so? Where will the data and metadata be stored after the project is completed? Is there a subject-specific and/or open-access repository that is appropriate for the data? If students need assistance in evaluating repositories, they should contact the library for consultation.