

1. Book Section

Data Management & Sharing

2. Author(s) Name(s) and brief biographies

Hannah C. Gunderman (any pronouns) is a Data, Gaming, and Popular Culture Librarian at Carnegie Mellon University Libraries. She received her Ph.D. in Geography (with a focus on cultural geography) from UT-Knoxville in 2018 and her M.S. in Information Sciences from UT-Knoxville in 2021. She is passionate about bringing her background in cultural geography and popular culture to the data management classroom, using fun metaphors and activities to convey foundational data management concepts to students, staff, and faculty on the CMU campus. ORCID: <https://orcid.org/0000-0002-7710-7055>.

Emma Slayton is the Data Curation, Visualization, and GIS Specialist at Carnegie Mellon University Libraries. She obtained an MPhil from the University of Oxford in 2013 and completed her Ph.D. at the Faculty of Archaeology, Leiden University in 2018. Her current work centers around improving and supporting digital literacy efforts. ORCID: <https://orcid.org/0000-0003-2230-3101>.

3. Title

7 Weeks, 7 DMPs: Iterative Learning around Data Management Plan Creation

4. Nutrition Information

Data management plans (DMP) are an essential part of any healthy research diet. This recipe is used in the *Discovering the Data Universe* course, a 7-week undergraduate course taught by data librarians at Carnegie Mellon University (CMU), which includes teaching students

the proper techniques for cooking up a flavourful and filling data management plan. This recipe is useful for anyone teaching foundational data literacy concepts in an academic library, or others who want to leverage data librarians' expertise to provide data literacy instruction and outreach.

5. Target Audience & Number Served

The target audience is undergraduate students in data literacy instruction settings, with a recommendation of at least one chef per 25 students (if there are more students, bring more chefs to the kitchen!). One chef will suffice if resources are limited.

6. Learning Objectives

Like any good chef, success is not only about how delicious the food is, but also the skills they teach to the sous-chefs (or students). This includes teaching what knife is needed for each kind of cutting technique a recipe requires. Specific knife skills (learning objectives) we want our sous-chefs to come away with include:

- Recognize data: Identify sources of data (both research-based datasets and data from everyday life), and describe how those data fit into different possible narratives and perspectives in stories.
- Communicate data: Interpret meaning from data and explain the ways in which data can be manipulated. Use data to answer key questions.
- Manage data: Plan and perform proper data preparation and storage (including the use of DMPTool and Open Science Framework).
- Critique data stories and themes: Identify common data narrative structures in a variety of examples including their classmates (and their own) work, critically evaluate and convey feedback about a classmate's work.

7. Cooking Time

Allow 7 weeks for marinating, with approximately 1-2 hours each week set aside to check DMPs for temperature, flavour, and adjusting spice levels as needed.

8. Dietary Guidelines

This recipe prepares students for future academic research projects and for cooking after graduation. Learning how to prepare a DMP helps students think through project design and connect that to what information is needed to answer research questions. To properly carry out a DMP, students must recognize what elements of data need to be explained, communicate those aspects effectively, and prepare for storing the data in a way that is accessible. They have to find the right recipe for their discipline, recognize and find the ingredients, and be able to describe those ingredients and how to use them.

Two learning objectives from this assignment that map well to the *ACRL Framework for Information Literacy* are *Communicate data* and *Manage data*. These learning objectives fit most strongly to the *Information Literacy* characteristics of *Information Creation as a Process* and *Authority Is Constructed and Contextual*. As the course progresses and students learn more about data literacy topics, their expertise and credibility in the subject increases and this is reflected in their DMP updates throughout the course.

9. Ingredients

- DMPTool (<https://dmptool.org/>) If you can't access this particular ingredient, any word-processing platform will suffice. However, the DMPTool provides templates, so your recipe will definitely flourish by using the DMPTool.

- Information from at least 3 funding agencies requiring DMPs. We recommend the National Science Foundation, National Institutes of Health, and National Endowment for the Humanities.
- Well-described, open datasets available for download. Our favourite datasets are from the U.S. Census, City of Pittsburgh, and datasets from our institution.

10. Preparation

In preparation for this class, facilitate a scenario much like the cooking show *Chopped*, prepare all the ingredients (or datasets) for students in advance, and ask them to create a meal (DMP) using these ingredients.

To prep the data ingredients, look for different examples of data. These include qualitative and quantitative data and public and private datasets. The open data sets are on a national level (U.S. Census Data), and local level (Pittsburgh Police Incident Reports from the Western PA Regional Data Center). The private data set is from our University Archives (a historical photograph data set). Students are individually assigned one of these datasets to use as their main ingredients in the course.

11. Instructions

The assignments task students with creating and iterating upon a Data Management Plan (DMP) which is based on the assigned dataset they receive in the first week of class. Below are step-by-step instructions for guiding students through their DMP recipe.

Week 1

Topic: Sample DMPs, Introducing DMPTool, and Choosing a Funding Agency Template

Lesson Specifics:

Ask students to imagine they are the head-chefs (creators) of the dataset they are assigned, walk them through the process of data collection, data analysis, and data management. Show several DMPs to students, using examples from the collection of public DMPs from the Digital Curation Centre (https://dmponline.dcc.ac.uk/public_plans). When showing sample DMPs, highlight areas where the ingredients have been used effectively and areas where the recipe could have been improved.

1. Guide students through the DMPTool and highlight appropriate templates that the students can employ to create their DMP. We recommend the National Science Foundation, National Institutes of Health, and National Endowment for the Humanities templates, as they serve as excellent recipes to draw from and cover a wide range of disciplinary areas. Ask students to choose a template that reflects the nature of their assigned dataset (quantitative data, ethnographic data, etc.).
2. Help students navigate to the DMP requirements from the organization on which their template is based and ask them which elements should be included in their DMP (see *Additional Resources* for links to guidance from the three funding agencies).

Student Deliverables: Students turn in a written explanation of which template they selected, and screenshots that demonstrate that they successfully logged in to the DMPTool.

Recipe Critique Guidelines: Verify that students chose a DMP template that makes sense for their data type, and included screenshots verifying that they logged in to the DMPTool.

Week 2

Topic: Data Description and Data Collection Basics

Lesson Specifics: Students begin to work in DMPTool with their chosen DMP template. This week's lecture should focus on basic practices in describing data and data collection

processes. What variable types are present? What do the date ranges appear to be for the data collection? How does the data appear to have been collected? This will help inspire students to apply the same level of inquiry and description to their own datasets.

Student Deliverables: Students turn in their first official DMP created in DMPTool which addresses what their data look like and how they have been collected, framing the DMP as though the students are the original collectors of the data.

DMP Critique Guidelines: Instructors repeat this step weekly throughout the remainder of the course. Review this round of DMPs by highlighting in what areas students successfully addressed the topics covered this week, and what areas of opportunity exist for the next round of revisions.

Week 3

Topic: Data Management Basics and Introduction to Metadata

Lesson Specifics: Students learn about data management basics using the DataONE recommended practices as guidelines (see *Additional Resources*) including: storage backups, file naming, and descriptive documentation, and an introduction to metadata. Through lecture, address questions such as: What is the importance of having a good file naming scheme for my data? How many backups should I have of my data? Should I write down the steps I take to interact with my data throughout a project? What is a README file? How do I choose a metadata standard to apply to my data?

Student Deliverables: Students turn in a DMP which addresses the revisions provided by the instructor for the Week 2 content, and include new content on how the student would employ recommended practices in data management and metadata to their assigned dataset.

Week 4

Topic: Data Analysis Foundations and Storytelling with Data

Lesson Specifics: Students learn about types of data analyses that are suitable for different types of data and research projects (qualitative, quantitative, etc.), such as statistical tests, ethnographic inquiry, and geospatial analysis. Discuss how to choose an analysis type for a dataset, and how to employ that analysis to tell an effective story about the dataset.

Student Deliverables: Students should identify at least one data analysis technique they would like to apply to their dataset, and indicate how this analysis technique may change any of their data management strategies listed in the previous week's DMP.

Week 5

Topic: Formatting and Visualizing Data

Lesson Specifics: Lead a discussion on how to choose appropriate data visualization techniques to display the results of a data analysis, showing students examples of good visualizations and visualizations which could be improved. Discuss several data visualization platforms, including Tableau, ArcGIS, and programming-based visualization packages in R and Python.

Student Deliverables: Students should choose at least one data visualization technique that they would like to apply to the results of their data analysis identified in Week 4, and describe how this visualization technique may change any of their data management strategies listed in the previous week's DMP.

Week 6

Topic: Ethics of Data, Data Licensing, and Data Discovery

Lesson Specifics: In this multi-topic lesson, (1) guide the students through a discussion on why it is important to be ethical when collecting and working with data, (2) describe why it is important to assign datasets a license when sharing them publicly (using the Creative Commons licenses as an example), and (3) guide students through the process of identifying an appropriate data repository for their data.

Student Deliverables: Students should describe the ethical considerations that exist around their data. Do the data refer to real people? Is there any identifying information within the dataset? What would happen if the dataset got into the wrong hands? How might their data management strategies mitigate any ethical risks? Ask students to choose an appropriate license for their data from the Creative Commons, and identify at least one possible data repository in which they would share this data publicly.

Week 7

Topic: Wrapping it All Together: Final DMP Creation

Lesson Specifics: Provide students with an open-working session in class to ask any questions about the DMP revisions they have received during the course, and receive instructor help and feedback as they work on their final DMP iteration.

Student Deliverables: Students turn in their final DMP which should address all revisions and lesson topics from the course.

Critique Guidelines: Review this final round of DMPs by highlighting in what areas students have incorporated course concepts, and how well the students have addressed the feedback received during each week of revisions.

12. Reviews/Assessment Strategy Notes

As noted above, when judging the flavour of the DMPs, pay attention to how the students make use of available ingredients (information on their assigned dataset), cooking techniques (how they incorporate class materials into refining their DMP), and presentation (how well they address all the elements of the DMP).

During the critique/review process, we caution instructors on using negative language or framing a DMP as “bad.” Given the wide range of skillsets, neurodiversity, and positionality that students may bring to this assignment, instructors can craft a more positive learning environment by offering opportunity-based critique which highlights areas where students can improve their recipe.

13. Chef’s Notes

Because students have not collected their own data and they are working with the information provided with their dataset, they must take creative liberties to fill out some sections of the DMP. Some students will lean into this adventurous cooking more than others! In our course we asked students to carry out their chosen data analysis and data visualization techniques on their data, in your cooking environment you may ask the students to describe what they *would* do with the data. It all depends on the cooking experience you want to create!

14. Additional Resources

ArcGIS. “Discover your Power with ArcGIS.” Accessed May 28, 2021.

<https://www.esri.com/en-us/arcgis/about-arcgis/overview>

Carnegie Mellon University. “Carnegie Mellon University Archives.” Accessed May 28, 2021. <https://findingaids.library.cmu.edu/>.

(Contains historical photographs used as a dataset in the course)

Creative Commons. “Share your Work.” Accessed May 28, 2021.

<https://creativecommons.org/share-your-work/>

DataONE. “Education Modules.” Accessed May 28, 2021. <https://old.dataone.org/education-modules>

National Endowment for the Humanities. “Data Management Plans for NEH Office of Digital Humanities Proposals and Awards.” Accessed May 28, 2021.

https://www.neh.gov/sites/default/files/2018-06/data_management_plans_2018.pdf

National Institutes of Health. “NIH Data Sharing Policy and Implementation Guidance.” Accessed September 12, 2021.

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

National Science Foundation. “Dissemination and Sharing of Research Results – NSF Data Management Plan Requirements.” Accessed May 28, 2021.

<https://www.nsf.gov/bfa/dias/policy/dmp.jsp>

U. S. Census Bureau. “American Community Survey.” Commuting (Journey to Work) Data Tables (2015). Accessed May 28, 2021.

<https://www.census.gov/topics/employment/commuting/data/tables.html>.

Western Pennsylvania Regional Data Center. "Crime Incidents Reports." Accessed May 28, 2021. <http://tools.wprdc.org/guides/public-safety/incidents/>.

Tableau. "Meet the World's Leading Analytics Platform." Accessed May 28, 2021. <https://www.tableau.com/>