

Data Management Planning for an Eight-Institution, Multi-Year Research Project

Kristin A. Briney
California Institute of Technology

Abigail Goben
University of Illinois Chicago

Kyle M. L. Jones
Indiana University-Indianapolis (IUPUI)

Abstract

While data management planning for grant applications has become commonplace alongside articles providing guidance for such plans, examples of data [management](#) plans as they have been created, implemented, and used for specific projects are only beginning to appear in the scholarly record. This article describes data management planning for an eight-institution, multi-year research project. The project leveraged four data management plans (DMP) in total, one for the funding application and one for each of the three distinct project phases. By understanding researcher roles, the development and content of each DMP, team internal and external challenges, and the overall benefits of creating and using the plans, these DMPs provide a demonstration of the utility of this project management tool.

Submitted 14 February 2022 ~ Revision received 9 May 2022 ~ Accepted 10 May 2022

Correspondence should be addressed to Kristin Briney, 1200 E California Blvd, Pasadena, CA 91125, USA. Email: briney@caltech.edu

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Introduction

Data management plans (DMPs) have become a commonplace feature of grant applications for federal and private granting agencies in the United States, the United Kingdom, and the European Union. Beyond grant applications, however, these plans have the potential to be integrated and preserved alongside other protocols, data dictionaries, standards, and metadata requirements in order to facilitate many aspects of research documentation both during data collection and also in preparation for data archiving or sharing. Further, while researchers may have written a DMP one time for a grant or have relied upon department or institutionally provided boilerplate, they need examples of the variety and depth to which a DMP can be written and how it can guide team workflows.

The Data Doubles project was a United States-based, multi-year, grant-funded research project that studied student perceptions of privacy in academic library learning analytics. Research was conducted by a nationally distributed team of eight researchers at eight different institutions, with additional partner institutions for various portions of the project. The project was made up of three one-year phases, consisting of: student interviews, a large student survey, and student focus groups. Each phase collected a different type of data and studied different research themes. Because of the large collaborative nature of the project and the high variety and volume of data collected across time, multiple DMPs were instrumental to the project's success.

In total, the project had four data management plans, which are available in OSF (Briney et al., 2022). The first DMP was submitted with the funding application and the other three DMPs were internal documents, each covering one phase of the three-phase project. These three DMPs demonstrated the benefit of using DMPs as “living documents,” having multiple DMPs, and standardizing data handling across all research sites. This article reviews the process by which the DMPs were developed and their important role within the larger scope of research work.

Literature Review

Many researchers consider DMPs in the scope of funder requirements, as these requirements have become normalized in the last decade. However, DMPs can also be living documents that serve as touchpoints during the research process and grow and adapt as the project does.

As funding agencies have increased their data management plan requirements, a number of articles have undertaken to address the content of what should go into the plan — whether aimed at supporting the entire research project or solely as a piece of the grant. Several general articles have provided overarching recommendations on creating and navigating the DMP process (Briney et al., 2020; Fadlilmola et al., 2021; Michener, 2015; Schiermeier, 2018; Wright, 2016) and identified the challenges for researchers and institutions as the work in the underlying processes (Fadlilmola et al., 2021; Lefebvre et al., 2020). Further, more detailed guidance has included recommendations related to dates (Briney, 2018); spreadsheet management (Broman & Woo, 2018); guidance for reviewers (Fearon et al., 2018); calls for more formal data management instruction integrated into science education (Tenopir et al., 2016). One limitation to the generation of these plans, however, is that the officers at the federal agencies charged with reviewing DMPs have received little training on assessing these (Bishop et al., 2021).

Due to the plethora of data management plans generated, there has also been interest in understanding what researchers have chosen to include and value in their grant DMPs and their perceptions of DMPs. Bishoff and Johnston, after reviewing three years of the United States' National Science Foundation DMPs, found a focus on the topic of data sharing with large

inconsistencies in mechanism, duration, and amount (2015). Van Tuyl and Whitmire also found an emphasis on sharing but when compared to actual practices found that researchers did not follow through with their stated plans (2016). Poole and Garwood found a lack of structure in digital humanities DMPs, leading to highly disparate contents and projects that did not include appropriate consideration for staffing and data sharing and reuse, as well as creating challenges for funders interested in assessing compliance (2020). In 2021, Mushi surveyed researchers about available services and found over half of the respondents described themselves as having poor knowledge of a variety of types of data management resource support available (2021). Spichtinger reviewed data management planning for Horizon 2020 projects in the European Union and found that while researchers rely on templates and peers to create DMPs, they also wanted more assistance from funders such as in form of a Research Data Helpdesk (2022).

Recently, researchers have begun to recognize the value of DMPs as their own independent scholarly object, which can add to our understanding of research and increase reproducibility. In one example, Lagring et al., describe implementing a DMP as part of protecting four decades of historical marine and continental shelf data (2016); in another, Yang et al., describe a data management plan for an oyster sperm repository (2021). Cauchick-Miguel et al., have reviewed the guidance provided to researchers and the published recommendations for DMP content (2020). The journal *RIO*¹ provides the opportunity for researchers to publish their data management plan independently as a scholarly object. Excellent DMPs are also getting noticed and shared through competitions, such as that hosted by the Qualitative Data Repository in 2021 (Qualitative Data Repository, 2021).

Universities are also beginning to acknowledge the value of DMPs outside of funding requirements. A handful of universities, primarily in the United Kingdom, now have policy that requires or recommends researchers create a data management plan for all research projects that process data regardless of the project's funding source (Digital Curation Centre, 2016; University of Bath, 2019; University College London, 2020). There has additionally been recognition of the need for DMPs for graduate student final projects (Raszewski et al., 2021).

There continue to be emerging next stages for DMPs. Current trends focus on making data and DMPs findable, accessible, interoperable, and reusable (FAIR) (Wilkinson et al., 2016). For DMPs, this can entail making plans readable for machine learning or allowing for programmatic system connections through DMPs (Cardoso et al., 2020, 2021; Miksa et al., 2019, 2021). However, it is unclear whether ensuring the documents themselves are machine-accessible will improve the content of most DMPs or will significantly change data practices.

Discussion

Project Roles

The project benefited from having several research team members with expertise in data management, including two librarians leading research data management services at their respective universities (herein referred to as “research data management librarians”) and one iSchool instructor who teaches data curation courses. Other team members brought extensive research experience to the project, with various levels of data management experience developed therein. One of the research data management librarians was designated as the data manager for the overall project and was in charge of DMP development, ensuring team members followed the DMPs, and generally keeping project files organized.

Development and Contents of the DMPs

Development of the DMPs and their contents was different for the formal grant application DMP and the three informal project-phase DMPs. For the DMP portion of the grant

¹ <https://riojournal.com/>

application (available at Briney et al., 2022), the project's Principal Investigator (PI) worked with the two research data management librarians to create the DMP, with the PI writing and the research data management librarians editing the document. The structure of the DMP was determined by the funding agency, U.S. Institute of Museum and Library Services (IMLS), with DMP contents giving a broad outline of the expected data types and data issues the project would encounter. This DMP was the only one that covered how project data would be shared outside of the research team, as the funder mandates data sharing.

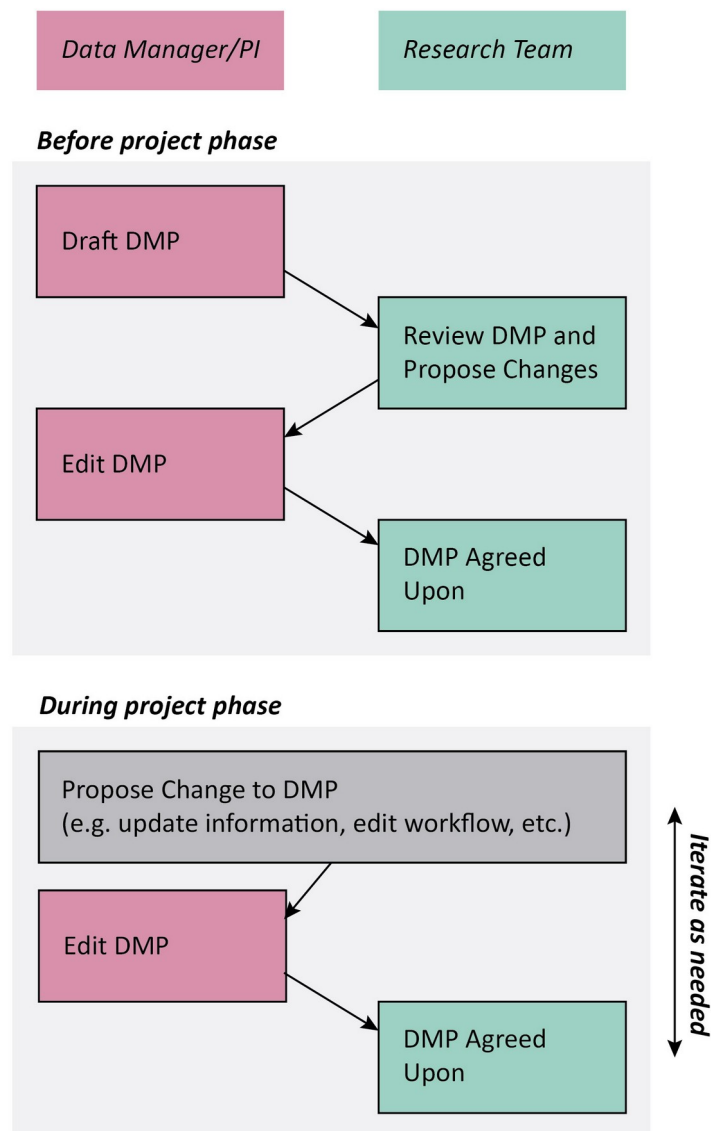


Figure 1. This flow chart outlines the process and roles for developing a phase-based DMP (top box) and editing that DMP, as needed, during the research phase (bottom box). The data manager and PI were responsible for drafting and editing the DMP, while the research team reviewed the DMP, proposed changes, and collectively agreed to follow the DMP during the project phase.

The development of the three phase-based DMPs (available at Briney et al., 2022) was led by the project's data manager at the start of each one-year project phase. Figure 1 outlines the development and editing process for a phase-based DMP. The data manager drafted an initial

DMP with help from the PI or another team member. The DMP was then reviewed by the entire team to ensure that the proposed management system worked for everyone. Minor adjustments to a DMP were made after the team review stage. DMPs were then put into use guiding the team's organization of that phase's project data. As these DMPs were living documents, updates to the DMP were sometimes made throughout the phase; for one phase's DMP, this included adding a new section for a data analysis procedure developed mid-phase.

The team used Google Drive to organize project files (sensitive data was stored separately in Box, with file permissions and folder organization noted in phase-based DMPs as relevant) and each project phase had a separate folder hierarchy for project files. Phase-based DMPs were stored at the highest file level or in a "data management" subfolder of the main phase folder so that they could be easily found by team members.

The contents of the phase-based DMP focused only on the data collected and analyzed during that phase (for reference, the phases were: individual interviews, a large survey, and small online focus groups). Each DMP provided information on how files should be organized, consistently named, and saved in the proper file formats. The team heavily leveraged codes in file naming, particularly for site (i.e. the university or college that each student participant was from) and research theme (with themes varying across project phases). Codes were alphabetic to be as human readable as possible, with site codes consisting of two-letter abbreviations and theme codes consisting of three-letter abbreviations. Dates were always written in ISO 8601 format (Briney, 2018).

DMPs also provided structure for how to move files through the analysis process, as appropriate to that phase's research plan. For example, the DMP for the student interview phase required format information to be appended onto file names (e.g. "_Audio.mp3", "_CaseSummary.pdf", "_OriginalTranscript.pdf", etc.) to keep track of analysis processing for individual interviews. The DMP for the survey phase outlined the timelines by which each site needed to distribute the survey, remind participants to complete the survey, and close the survey, etc., as all sites shared the same survey platform but timelines for survey distribution at each site varied. The DMP for the focus group phase described how files should move back and forth between Google Drive and the coding software, which was hosted on a separate online platform.

The interview- and focus group-phase DMPs also leveraged one other data management tool to deal with the massive number of files created during these two phases: an index. The interview phase generated data for almost 120 interviews across eight different sites covering five different research themes. Each interview was recorded, summarized, transcribed, and coded, which meant that this phase generated multiple file variants for each of the 120 interviews. This meant that several hundred files needed to be organized and tracked.

The focus group phase generated data for 21 focus groups across seven different sites covering three different research themes; each focus group was recorded, transcribed, and coded. While DMPs described the process by which files moved through the analysis workflow, index spreadsheets were critical to tracking the current state of any one interview or focus group in the analysis pipeline. Team members were responsible for continually updating indexes to reflect the current analysis status of interviews and focus groups conducted at their site; the PI kept the overall analysis process on the project's timeline.

External Factors and Impact

In addition to the challenges in developing data management plans over a large project and variety of data types with a distributed team, there were unanticipated external factors that both created challenges for implementation of and were mitigated by the DMP. The biggest overall challenge was the natural evolution of the project as it was conducted. As a living document, the DMP continued to be refined as researchers identified best practices and worked in concert with the project advisory board to improve research methodology.

Unanticipated challenges, however, were numerous. One of these included a mandatory platform change, when the lead institution ceased using Box for cloud storage and forcibly

migrated data researchers to an IRB-approved Google Drive with minimal advance notice. Here, having established file naming conventions mitigated the fact that folder structure was compromised in the migration. The team also had changes in personnel over the course of the project, with researchers both leaving and joining the team. The research group chose not to include the administrative aspects of personnel changes in the data management plan and to document it elsewhere. A final unanticipated challenge was the COVID-19 pandemic, which came in the midst of the project's second year and forced the delay of some work in the project's third year as institutions pivoted online. Having a flexible data management plan allowed researchers to continue work and document changes as they were navigated.

Reflections from the PI on the Success of Using the DMPs

The complexity of this research agenda combined with the number of collaborators required a sustained focus on composing purposeful, useful documentation. That was especially true given the largely asynchronous nature of the team's communications and unparallel alignment in work schedules. Collaborators needed to reference instructional documents and, effectively, policies for creating, storing, and using key research documents (e.g., data, software tool guides, research protocols, meeting agendas and notes, etc.). As the PI, it was easy for me to reflect back on the project's successes and state that they would have been less so had the team not been educated on data management practices and committed to their requirements. These practices created a detailed audit trail of our discussions and joint agreements on key research practices. We typically knew where we were headed with the project at a granular level because we could trace our work back through the path of documentation that we had created and organized. While the success of the project was itself a desirable goal with regard to wide dissemination via publications and presentations, other outcomes arose out of the team's commitment to resolved data management. Data management practices useful for ourselves also enable other library and information science (LIS) practitioners and researchers to replicate our work efficiently and correctly by accessing a vast majority of the documents we created for ourselves and deposited in a publicly accessible research repository.

Conclusion

DMPs are sometimes thought of as just another document that needs to be included in a funding application, but this project demonstrated how useful DMPs can be outside of this limited model. There were three key features of how this project used DMPs that are worth broader adoption by other researchers. First, the team used DMPs as living documents that were regularly refined to meet the project's and researchers' needs; if something in the DMP wasn't working, it was changed. Second, the project leveraged multiple DMPs; each project phase had distinct data collection and analysis methods, making individual DMPs beneficial. Finally, the project's DMPs were vital to coordinating work done by multiple people in different locations; the large and distributed project required consistent data handling and the DMPs facilitated this. The project's four DMPs augmented the funding application, streamlined data collection, documented analysis workflows, and made data handling uniform, allowing researchers to focus on research instead of being mired in the data problems that can easily arise in a multi-site, multi-year research project.

Acknowledgements

This project was made possible in part by the Institute of Museum and Library Services (LG-96-18-0044-18). The views, findings, conclusions, or recommendations in this article do not

necessarily represent those of the Institution of Museum and Library Services. The research team expresses its profound thanks to our funder for supporting this important research. The authors also thank the other members of the Data Doubles research team – Andrew Asher, Michael Perry, Mariana Regalado, Dorothea Salo, and Maura Smale – for their feedback on project DMPs, for thoroughly following the DMPs during each research phase, and for providing comments on a draft of this article.

References

- Bishoff, C., & Johnston, L. (2015). Approaches to data sharing: An analysis of NSF data management plans from a large research university. *Journal of Librarianship & Scholarly Communication*, 3(2), 1–27. doi:10.7710/2162-3309.1231
- Bishop, B. W., Nobles, R., & Collier, H. (2021). Research Integrity Officers' responsibilities and perspectives on data management plan compliance and evaluation. *Journal of Research Administration*, 52(1), 76–101.
- Briney, K. (2018). The problem with dates: Applying ISO 8601 to research data management. *Journal of EScience Librarianship*, 7(2). doi:10.7191/jeslib.2018.1147
- Briney, K., Coates, H., & Goben, A. (2020). Foundational practices of research data management. *Research Ideas and Outcomes*, 6, e56508. doi:10.3897/rio.6.e56508
- Briney, K., Jones, K. M. L., Asher, A., Goben, A. H., Perry, M., Regalado, M., Salo, D., & Smale, M. (2022). *Data Doubles Data Management Plans*. doi:10.17605/OSF.IO/JE7QP
- Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. *The American Statistician*, 72(1), 2–10. doi:10.1080/00031305.2017.1375989
- Cardoso, J., Castro, L. J., & Miksa, T. (2021). Interconnecting systems using machine-actionable data management plans – hackathon report. *Data Science Journal*, 20(1), 35. doi:10.5334/dsj-2021-035
- Cardoso, J., Proença, D., & Borbinha, J. (2020). Machine-actionable data management plans: A knowledge retrieval approach to automate the assessment of funders' requirements. *Advances in Information Retrieval*, 12036, 118–125. doi:10.1007/978-3-030-45442-5_15
- Cauchick-Miguel, P. A., Moro, S. R., Rivera, R., & Amorim, M. (2020). Data management plan in research: Characteristics and development. In R. Mugnaini (Ed.), *Data and Information in Online Environments* (pp. 3–14). Springer International Publishing. doi:10.1007/978-3-030-50072-6_1
- Digital Curation Centre. (2016). *UK Institutional data policies*. Digital Curation Centre. <https://www.dcc.ac.uk/guidance/policy/institutional-data-policies>
- Fadlelmola, F. M., Zass, L., Chaouch, M., Samtal, C., Ras, V., Kumuthini, J., Panji, S., & Mulder, N. (2021). Data management plans in the genomics research revolution of Africa: Challenges and recommendations. *Journal of Biomedical Informatics*, 122, 103900. doi:10.1016/j.jbi.2021.103900

- Fearon, D., Boehm, R., & Chiu, C. (2018). *Grant reviewer's guide for data management plans*. doi:10.17605/OSF.IO/SNYFB
- Lagring, R., Adam, M., Goffin, A., Stojanov, Y., Tyberghein, L., & Vandenberghe, T. (2016). The implementation of a data management plan to uplift historical data: Long-term change detection in the Belgian Continental Shelf. *Bollettino Di Geofisica Teorica Ed Applicata*, 57, 127–127.
- Lefebvre, A., Bakhtiari, B., & Spruit, M. (2020). Exploring research data management planning challenges in practice. *It - Information Technology*, 62(1), 29–37. doi:10.1515/itit-2019-0029
- Michener, W. K. (2015). Ten simple rules for creating a good data management plan. *PLOS Computational Biology*, 11(10), e1004525. doi:10.1371/journal.pcbi.1004525
- Miksa, T., Oblasser, S., & Rauber, A. (2021). Automating research data management using [journal article] machine-actionable data management plans. *ACM Transactions on Management Information Systems*, 13(2), 18:1-18:22. doi:10.1145/3490396
- Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLOS Computational Biology*, 15(3), e1006750. doi:10.1371/journal.pcbi.1006750
- Mushi, G. (2021). Research data management and services: Resources for different data practitioners. *IASSIST Quarterly*, 45(3–4), Article 3–4. doi:10.29173/iq995
- Poole, A. H., & Garwood, D. A. (2020). Digging into data management in public-funded, international research in digital humanities. *Journal of the Association for Information Science and Technology*, 71(1), 84–97. doi:10.1002/asi.24213
- Qualitative Data Repository. (2021, May 19). *DMP competition winners: DMPs so good they go to 11*. Qualitative Data Repository. <https://qdr.syr.edu/qdr-blog/dmp-competition-winners-dmps-so-good-they-go-11>
- Raszewski, R., Goben, A. H., Bergren, M. D., Jones, K., Ryan, C., Steffen, A. D., & Vonderheid, S. C. (2021). A survey of current practices in data management education in nursing doctoral programs. *Journal of Professional Nursing*, 37(1), 155–162. doi:10.1016/j.profnurs.2020.06.003
- Schiermeier, Q. (2018). Data management made simple. *Nature*, 555(7696), 403–405. doi:10.1038/d41586-018-03071-1
- Spichtinger, D. (2022). Data Management Plans in Horizon 2020: What beneficiaries think and what we can learn from their experience. *Open Research Europe*, 1(42). doi:10.12688/openreseurope.13342.2
- Tenopir, C., Allard, S., Sinha, P., Pollock, D., Newman, J., Dalton, E., Frame, M., & Baird, L. (2016). Data management education from the perspective of science educators. *International Journal of Digital Curation*, 11(1), 232–251. doi:10.2218/ijdc.v11i1.389
- University College London. (2020). *UCL research data policy*. Retrieved from https://www.ucl.ac.uk/isd/sites/isd/files/ucl_research_data_policy_v6.pdf

- University of Bath. (2019). *Research data policy guidance*. Retrieved from <https://www.bath.ac.uk/guides/research-data-policy-guidance/>
- Van Tuyl, S., & Whitmire, A. L. (2016). Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. *PLOS ONE*, *11*(2), e0147942. [doi:10.1371/journal.pone.0147942](https://doi.org/10.1371/journal.pone.0147942)
- [Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. [doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)
- Wright, A. (2016). Electronic resources for developing data management skills and data management plans. *Journal of Electronic Resources in Medical Libraries*, *13*(1), 43–48. [doi:10.1080/15424065.2016.1146640](https://doi.org/10.1080/15424065.2016.1146640)
- Yang, H., Huo, Y., Yee, J. C., Rikard, S., Walton, W. C., & Saillant, E. (2021). Sperm repository for a breeding program of the Eastern Oyster *Crassostrea virginica*: Sample collection, processing, cryopreservation, and data management plan. *Animals*, *11*(10), 2836. [doi:10.3390/ani11102836](https://doi.org/10.3390/ani11102836)