Data Management Plan Implementation, Assessments, and Evaluations: Implications and Recommendations

COLLECTION:
DATA MANAGEMENT PLANNING ACROSS DISCIPLINES AND INFRASTRUCTURES

ESSAY

BRADLEY WADE BISHOP
PETER NEISH
JI HYUN KIM
RAPHAËLLE BATS
A. J. MILLION
JAKE CARLSON
HEATHER MOULAISON-SANDY
MINH T. PHAM

*Author affiliations can be found in the back matter of this article

ABSTRACT

Data management plans (DMPs) have become nearly a worldwide requirement for research funding. To meet these new funding agency expectations, information professionals across domains and the world have worked to create resources and services to successfully implement and sometimes assess DMPs. This essay presents a series of case studies from different institutions across the globe to highlight current practices and share recommendations for future work. A summary of various projects related to DMP implementation, assessment, and evaluation in different contexts provides a useful overview of current practices. The essay concludes with recommendations for practical oversight and scoring to improve DMPs' utility in enabling the sharing of data.

CORRESPONDING AUTHOR:
Bradley Wade Bishop
University of Tennessee, US
wade.bishop@utk.edu

KEYWORDS:
research data management; data management plan; machine-actionable DMP; maDMP; assessment; evaluation

TO CITE THIS ARTICLE:
Bishop, BW, Neish, P, Kim, JH, Bats, R, Million, AJ, Carlson, J, Moulaison-Sandy, H and Pham, MT. 2023. Data Management Plan Implementation, Assessments, and Evaluations: Implications and Recommendations. *Data Science Journal*, 22: 27, pp. 1–8. DOI: https://doi.org/10.5334/dsj-2023-027

INTRODUCTION

The inclusion of data management plans (DMPs) for research proposals has become a near-global expectation for research funding. Well-executed DMPs add to research transparency, accountability, and reproducibility; however, more work needs to be done to validate these assumptions. Regardless of the efficacy of any single DMP, success of data sharing and reuse requires interoperability of data across domains. Beyond this need for interoperability, the replicability and reproducibility crisis in science demands that provenance, ownership, and any use limitations to data be captured across systems to increase trust and authenticity in research as well as control proper access (Tullis & Karr 2021).

Nearly every country with a research enterprise through their affiliated funding agencies and institutions have promulgated a wide variety of DMP expectations with varying levels of guidance, support, and oversight. Despite the differing approaches and start times for each country's DMP implementation journey, similar themes emerge when discussing DMP implementation, assessment, and evaluation. This essay presents the background, considerations, and recommendations to inform future study of DMP implementation, assessment, and evaluation. At SciDataCon in 2022 during International Data Week, a panel of DMP experts from across the globe gathered to share their recent DMP efforts. The panelists were recruited through personal contact because of their reputations for advancing work on DMP implementation, assessment, and evaluation. Every effort was made to have representation from as many continents as possible, but ultimately only four countries were represented from the six projects presented. These pioneers provide some concrete, measurable advances in DMPs to further scientific research.

BACKGROUND

As data-intensive approaches become the dominant method across research domains, the cyberinfrastructure, support services, and research data management (RDM) tools for all research grows in importance. The efforts referred to often as RDM encompass many tasks that most succinctly enable data to align with the FAIR Data Principles (to be as findable, accessible, interoperable, and reusable as possible) (Wilkinson et al. 2016). DMPs are neither a new concept nor a new practice. DMPs are a structured, formal document describing the roles, responsibilities, and activities for managing data during and after research (Bishop & Hank 2020). Historically, DMPs captured details at data collection to prevent any information entropy or loss when it came time to deposit data into a repository (Michener et al. 1997). Today, DMPs serve purposes beyond data deposit as they capture the details that inform data reuse (Smale et al. 2018). Gaps exist across all regions when it comes to creating accountability of DMP implementation despite their proliferation as a stipulation to receive research dollars.

Successful DMPs may be one of the linchpins to deliver on the promises of open science. As more funding agencies and academic journals require researchers to make public the data and digital outputs associated with a publication, DMPs value rises (The Royal Society 2017; PLOS n.d.). Supporting sustainable, widespread utilization of DMPs for their initial and tertiary purposes requires an understanding of what happens beyond the grant submission and before data deposit into a repository. More systematic work needs to be done on if and how researchers adhere to DMPs throughout the research lifecycle. Dietrich et al. (2012) reviewed funding organizations and found lacking coverage in data management policies concerning storage, licensing, metadata, and sharing. In other words, funders require DMPs but provide little prescriptive guidance on their data element expectations. Research questions abound, but for this essay and panelists focused on these overarching inquiries:

- What are the lessons learned from DMP implementation?
- What are the lessons from DMP assessment and evaluations?

More empirical research is needed to address these questions. This essay's authors represent four continents, countries, and several institutions where DMPs are a commonplace requirement of publicly funded research and other proposal applications. The intention of having several global perspectives on the panel, along with the subsequent essay, allows for everyone to learn from various approaches to DMP implementation and assessment. As a result, DMP implementation, assessment, and evaluation approaches are reported here from various global perspectives, and the themes that emerge may assist others charged with these tasks. Each panelists' context shines some light on the ways to best support DMP success despite individual, governmental, institutional, technological, cultural, and economic influences and restrictions. Certainly, sharing experiences in this essay is not generalizable to the entire spectrum of DMP adoption, implementation, assessment, and evaluation, but we hope these case studies benefit others.

DMP IMPLEMENTATION ACROSS COUNTRIES

Implementing DMPs may cover many phases, from writing a plan to other actions such as selecting metadata standards or conducting quality control and assurance. The first phase for most is the practice of writing a DMP. The writing phase is primarily driven by research funders' requests to have a document to help plan, steer, and evaluate the data aspects of projects. The actual DMP writing falls to the researchers and their institutions, and this precedent to delegate all responsibility to the individual researchers (e.g., principal investigators) impacts all later phases of DMP implementation, assessment, and evaluation. The information contained in a written DMP has the potential to be mobilized by many other stakeholders beyond the investigators, especially if a DMP can be represented in a machine-actionable form. These first three case studies from Australia, Korea, and France focus on the implementation of DMPs in their respective institutions to (1) create machine-actionable DMPs, (2) standardize ontologies, (3) define research data, and (4) build human capacity to support the new RDM resources and services.

One recent lesson learned and shared by the Research Data Alliance (RDA) DMP Common Standards Working Group is a standard application profile that allows a DMP to be represented as atomized structural data (i.e., a machine-actionable DMP (maDMP)) (Miksa et al. 2019; https://www.rd-alliance.org/madmps). This standard enables information to be shared across systems, to be reused, and to trigger actions adding value to additional stakeholders in the research lifecycle such as researchers, data stewards, repository operators, and others. For example, a repository could set up a backup strategy or preservation plan in response to a data steward choosing a particular repository for data deposit. This RDA standard application profile is being adopted by DMP systems, repositories, and other related services across the world (Miksa et al. 2021). The standard has also been developed into an ontology, the DMP Common Standard Ontology (DSCO), which enhances the standard to explicitly link to related data models and ontologies, to standardize controlled vocabularies and to enhance extensibility (Cardoso et al. 2022). The DSCO allows business rules or heuristics to be developed that can be applied to DMPs allowing much of the information to be evaluated computationally rather than by a person. Any entity pursuing DMP implementation likely will benefit from these efforts to create maDMPs.

Several countries are at the start of their DMP implementation journeys and may utilize the advances made by RDA and other data organizations such as Committee on Data (CODATA) of the International Science Council (https://codata.org/) and the World Data System (https://worlddatasystem.org/about/). Recently in the Republic of Korea, researchers must create DMPs for publicly funded research. The Ministry of Science and ICT (MSIT) proposed principles for RDM and data sharing and enactment of the National R&D Innovation Act moved DMP-related provisions to administrative rule. In 2019, the Regulation of Management required mandatory DMP submission, and by 2022, DMPs could be considered in the evaluation process of research proposals. First, it was necessary to establish a legal basis for managing and sharing research data prior to DMP implementation. Each country has its own considerations based on existing intellectual property law and information policy. In Korea, research data is defined to be factual records produced from various experiments, observations, investigations, analyses, and so forth from research and development projects that are necessary to validate research findings. Many other countries may benefit from pursuing a more formal definition of what qualifies as research data. Korean researchers also assessed the preparedness of government-funded research institutes (GRIs) to implement DMPs. In 2021, 3 of 25 government-funded research institutes (GRIs) implemented DMPs, but all others were preparing for implementation (Park & Kim 2021).

A clear advantage of countries whose research enterprise is more centralized is that each funding agency creates and manages repositories as opposed to having universities, organizations, and individual researchers coordinating those responsibilities and roles. Yet, even with this centralization of data repositories, many of the related RDM tools and services still fall to individual institutions and researchers. Oversight for assessment and evaluation of the DMPs at least remains the task of the Ministry rather than individuals. Still, the challenge of staffing exists. At any Korean institution, more dedicated to data curation is now needed to implement DMP requirements. In one study, university librarians were interviewed, and they discussed some obstacles that included a low level of RDM services demand from researchers, insufficient support at their libraries, and overall under-staffing and lack of resources (Park & Kim 2021). The adoption of DMPs and RDM in the Republic of Korea is still in the early stages, but all countries and institutions newly adopting DMPs stand to benefit from the past work of others, and more mature DMP infrastructures may learn from the innovation of later adopters as well. Like any new requirement, researchers have many questions for such a change to their workflow, and institutions have quickly responded by developing RDM services and resources.

In France, DMP implementation has begun as a response to a series of political injunctions, from 2019 to 2021 with the obligation to provide DMPs for all projects funded by the National Research Agency (ANR), in concert with the launch of the National Plans for Open Science, that demanded the attribution of the label “Atelier de la donnée” to the universities engaged in research data policies. Between these national level mandates and the DMP implementation at universities and other research institutions, several obstacles appeared (Le Béchec et al. 2022). First, each location needed a local roadmap on their unique approach to open science and associated DMPs. The inertia and pace of change of universities often leads to a lag of several years between the commencement of a new approach and the determination of key actions because of the reluctance of researchers (time, academic freedom, fear of open access, and so forth). A culture change of this magnitude to any research enterprise is a challenge, but like other countries, France’s more centralization approach has the same advantages to alter the overall behavior of any scholar working in the country using public funds using the *power of the purse*.

The most vital work at one institution in France was raising awareness of these cultural changes to RDM and not strictly showing researchers how to write DMPs. In all countries and agencies, assistance in writing DMPs is a marginal request compared to the institutional commitment in general to implementing, assessing, and evaluating DMPs. DMP models at each institution help researchers understand the issues of repositories, metadata, legal issues, and so forth. Given the newness of these mandates in France, institutions are facing the difficulty of developing a DMP implementation strategy that would not be too disciplinary-specific and work across different disciplines with unique RDM needs and practices. As with any workflow change, the implementation relies on actors who have in common a lack of knowledge about new things and a need for additional training for RDM.

Across the globe, a central actor has emerged from this implementation process at many research universities and institutions, namely the librarian. Librarians have long been pioneers in open science and scholarly communication. Librarians play a local lobbying role to promote the implementation of new national recommendations such as DMPs. The skills of librarians in terms of metadata and repositories also make them well-informed players in DMP implementation. Much like the adoption of maDMPs, librarians face adoption obstacles. Often, the library lacks power in the overall research enterprise that reduces their ability to establish or enforce policies. In addition to the lack of visibility as an obvious source for DMP help for researchers, many academic librarians lack RDM knowledge and require new training themselves. DMP implementation in any setting will benefit from the creation of information professional roles similar to Data Services Librarians (DSLs) (Bishop et al. 2022).

DMP ASSESSMENT AND EVALUATION

DMP assessment and evaluation occurs prior to, during, and after DMP implementation. Like any aspect of the research enterprise in the age of accountability, DMPs also benefit from review. Since 2011, the US National Science Foundation (NSF) required DMPs and many other funding agencies within the US have followed suit. Like DMPs anywhere, the intent is for the document to communicate vital information between the researchers developing data and the data curators who may take on the responsibility for making the data reusable by others. In practice, DMPs may or may not be accomplishing their initial objectives to enable reuse. Given the current decentralized structure of the US research enterprise, inaccessibility to DMPs themselves, and the anecdotal known uneven quality of DMPs limits their utility as a tool to support data curation. More formal assessments and evaluations of DMPs have occurred in the following case studies, and their findings could help inform all others. Those in a centralized role at research universities and institutions, like librarians and other research support services, may be able provide a solution. These next three case studies include two from the US and one with a multinational DMP requirement. Each scenario presented provides different insight into how to (1) assist researchers at the start and throughout projects to improve data quality for data deposit, (2) insert program officers into a role of DMP assessment and evaluation that increases compliance with open data practices, and (3) consider digital curation procedures for more sensitive qualitative data and the range of other digital objects needed to access and use data.

One DMP assessment and evaluation study conducted at the University of Michigan (UM) gathered DMPs from awarded grants over the period of one year (March 2020 to February 2021). A large research university, UM had 744 awarded grants where UM was the lead institution, and of those awarded grants, 461 had a DMP and were generating data. Findings indicate that DMPs did not directly describe data types, and only 21% had a statement on metadata. Fortunately, two-thirds of the DMPs indicated their intention to share data through a repository, and 63% included statements relating to intellectual property, copyright, or other restrictions on data. Although librarians may assist with many aspects of DMP implementation, other entities at universities and institutions will need to be involved in addressing more specialized areas, such as intellectual property concerns. One concern for any research enterprise is the cost of data management. Only 25% of DMPs in that study included information about the duration with most defaulting to 'in-perpetuity'. Institutions of all kinds committing to data preservation at a growing scale without clear policies and additional cost considerations should be addressed by funding agencies.

The assessment of these DMPs provided insight to the institutional repository to better plan for the projects that would generate data that needed to be ingested and consider any intellectual property issues well before data deposit time. With this insight, their institution may help researchers with data curation and have a system to check, understand, and request any missing information before augmenting, transforming, evaluating, and documenting for deposit. Other places may benefit from this service model where consultation exists throughout the data lifecycle.

In another US context, DMPs are regulatory compliance artifacts used by the US Department of Justice (DOJ), an entity that deals with crime and criminal justice research (Office of Justice Programs 2021). To facilitate data sharing and open scientific practices, like many other funders, DOJ requires DMPs in all proposals with the key caveat that funding is contingent on DMP compliance by requiring grant recipients to write and submit a DMP not only at the start of a research project, but also through semi-annual updates (Bureau of Justice Statistics 2022; National Institute of Justice 2022). The funding agency takes on the role of DMP assessment and evaluation to protect its financial investment in scientific research. Program officers evaluate DMPs before and during a research study until data are archived. This improves DMP implementation because, at the discretion of program officers, if DMP deliverables are not met, they withhold funding. As a result, research data are consistently shared with the broader academic research community for analysis and reuse.

As research methods in criminology have changed from largely quantitative to mixed methods, downstream challenges for data curation have emerged (Million 2023). The DMPs researchers create provide roadmaps for long-term data preservation, but the time and cost constraints associated with protecting study participants in qualitative data means not everything can be curated (A.J. Million, personal communication, March 15, 2023). Current practice for US DOJ suggests DMPs are highly effective in facilitating open scientific practices and allow funders to evaluate and correct data practices with compliance checks. However, they do not address long-term data preservation issues, but that notwithstanding, other funding agencies could provide incentives for long-term sharing and reuse with similar DMP assessments and evaluations to

DOJ. This may be a best practice because unlike librarians, program officers retain the ability to incentivize DMP implementation by withholding funds.

Finally, another DMP study investigated DMPs from the Belmont Forum, an international funding agency. This study analyzed 21 DMPs from funded projects, focusing on their discussion of digital objects. Digital objects are a spectrum of resources created as part of the research endeavor that are worthy of attention; as such, they are likewise worth the efforts of planning for retention and access. Examples of important digital objects in research can include such unexpected resources as'Objects on the Web, such as YouTube videos, Facebook profiles, Flickr images etc. that are composed of data and formalized by schemes or ontologies that one can generalize as metadata' (Hui 2012).The Belmont Forum has data and digital object (DO) management planning as the requirements for their DMPs (https://www.belmontforum.org/archives/resources/data-and-digital-outputs-management-plan-ddomp).

The study evaluated overall compliance with these elements as they pertain to digital objects and found that almost all the DMPs (20/21, 95.2%) mentioned digital objects in some way (or files that could be assumed to be digital objects). The analysis also indicated that information about the data management personnel was consistently included in the DMPs, implying that experts would be on hand to guide the collection and treatment of digital objects by the research teams. Still, similar challenges fell to these DMPs on the tougher-to-predict elements of data size, retention, and timeframe for access (Pham et al. 2023). With more transdisciplinary research that involves a variety of data and digital objects, the work of DMP implementation also becomes more complex. Ultimately, all funding agencies would benefit from detailed expectations for DMP elements to encourage compliance and indeed streamline DMP assessment and evaluation for a range of data files and digital objects beyond traditional datasets.

CONCLUSION

To summarize the lessons learned from these global perspectives on DMP implementation, assessment, and evaluation, the essay concludes with a set of recommendations to support data stewardship of DMPs. Across the research enterprise of many countries and organizations, DMP tools and services have been created to buttress the writing and implementation of DMPs. DMPs serve as the central document to address the FAIR, TRUST, and CARE Data Principles and the resulting culture change to move all research data to be more machine-actionable, trusted, and just. Although many legal and other administrative considerations will be necessarily local in context and country/domain specific, each entity embarking on DMPs stands to benefit from:

1) utilizing librarians and other information professionals as embedded data assistants;
2) expecting all funders to provide more detailed DMPs guidance for researchers, which likely encourages a move towards maDMPs; and
3) selecting some entity, whether at the funding agency or institutional level, to serve in a role to continually evaluate DMPs for compliance.

The first two recommendations build on established work of many, with several success stories outlined in this essay's case studies. The final recommendation of enforcing compliance through DMP assessment and evaluation will require the most work. Research compliance administrators and other subject-area data experts likely require additional DMP training. Ultimately, relying on individual researchers and teams to do all DMP-related tasks is problematic and may lead to data fabrication and falsification, whether intended or not, and the saddest thing of all–data loss.

FUNDING INFORMATION

Dr. Million's contribution was supported by DJO-BJS-17-RO-0253 "Data Archiving Services," an inter-agency agreement between the University of Michigan and the U.S. Office of Justice Programs.

Dr. Kim's work was supported by Ewha Womans University Research Grant of 2022.

Dr. Bishop's work was partially funded by the University of Tennessee-Knoxville's Faculty Development Leave.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Bradley Wade Bishop orcid.org/0000-0002-5022-2707
University of Tennessee, US

Peter Neish orcid.org/0000-0002-7171-2334
University of Melbourne, AU

Ji Hyun Kim orcid.org/0000-0003-0209-4365
Ewha Womans University, South korea

Raphaëlle Bats orcid.org/0000-0003-2165-8453
University of Bordeaux, FR

A. J. Million orcid.org/0000-0002-8909-153X
Director of the National Archive of Criminal Justice Data, US Department of Justice, University of Michigan, US

Jake Carlson orcid.org/0000-0003-2733-0969
University of Buffalo, US

Heather Moulaison-Sandy orcid.org/0000-0001-7783-7069
University of Missouri, US

Minh T. Pham orcid.org/0000-0002-9776-3981
University of Florida, US

REFERENCES

Bishop, BW and **Hank, C.** 2020. Curation, Digital. In: Kobayashi, A (ed.), *International Encyclopedia of Human Geography*, 2e. Amsterdam, Netherlands: Elsevier. DOI: https://doi.org/10.1016/B978-0-08-102295-5.10531-1

Bishop, BW, et al. 2022. Data Services Librarians' Responsibilities and Perspectives on Research Data Management. *Journal of eScience Librarianship*, 11(1): e1226. DOI: https://doi.org/10.7191/jeslib.2022.1226

Bureau of Justice Statistics. 2022. *Data Management Plan for the Bureau of Justice Statistics*. Retrieved from https://bjs.ojp.gov/sites/g/files/xyckuh236/files/media/document/BJS%20Data%20Management%20Plan%20Model%20Template%20August%202022.pdf.

Cardoso, J, et al. 2022. DCSO: towards an ontology for machine-actionable data management plans. *Journal of Biomedical Semantics*, 13(1): 21. DOI: https://doi.org/10.1186/s13326-022-00274-4

Dietrich, D, Adamus, T, Miner, A and **Steinhart, G.** 2012. De-mystifying the data management requirements of research funders. *Issues in Science and Technology Librarianship*, 70(1). DOI: https://doi.org/10.5062/F44M92G2

Hui, Y. 2012. What is a digital object? *Metaphilosophy*, 43(4): 380–395. DOI: https://doi.org/10.1111/j.1467-9973.2012.01761.x

Le Béchec, M, Bouchard, A, Charrier, P, Denecker, C, Gallezot, G, et al. 2022. Pratiques et usages des outils numériques dans les communautés scientifiques en France. *[Rapport de recherche] Comité pour la science ouverte*, 112: ⟨hal-03545512⟩.

Michener, WK, Brunt, JW, Helly, JJ, Kirchner, TB and **Stafford, SG.** 1997. Nongeospatial metadata for the ecological sciences. *Ecological Applications*, 7(1): 330–342. DOI: https://doi.org/10.2307/2269427

Miksa, T, Walk, P and **Neish, P.** 2019. RDA DMP Common Standard for Machine-actionable Data Management Plans. DOI: https://doi.org/10.1109/BigData.2018.8622618

Miksa, T, et al. 2021 Application Profile for Machine-Actionable Data Management Plans. *Data Science Journal*, 20(1): 32. DOI: https://doi.org/10.5334/dsj-2021-032

Million, AJ. 2023. Successes, pain points, and lessons learned when curating data at scale: The NACJD experience. In: *IASSIST 2023*, Philadelphia, PA on 30 May–3 June 2023.

National Institute of Justice. 2022. *Data Archiving*. Retrieved from https://nij.ojp.gov/funding/data-archiving#data-archiving-plan.

Office of Justice Programs. 2021. *FY 2022 Budget Request Overview*. Retrieved from https://www.ojp.gov/sites/g/files/xyckuh241/files/media/document/dojojpfy2022budget.pdf.

Park, HJ and **Kim, JH.** 2021. A Study on the Considerations for Data Management Plan Services in University Libraries. *Journal of the Korean Biblia Society for Library and Information Science*, 32(3): 187–215.

Park, Y and **Kim, J.** 2021. A study on research data creation and management by researchers in mechanical engineering. *Journal of Korean Society of Archives and Records Management*, 21(4): 137–162.

Pham, MN, et al. 2023. Data Management Plans: Implications for automated analyses. *Data Science Journal*. DOI: https://doi.org/10.5334/dsj-2023-002

PLOS. n.d. Data availability. http://journals.plos.org/plosone/s/data-availability.

Smale, N, Unsworth, K, Denyer, G and **Barr, D.** 2018. The history, advocacy and efficacy of data management plans. *bioRxiv* (pre-print). DOI: https://doi.org/10.1101/443499

The Royal Society. 2017. Data sharing and mining. https://royalsociety.org/journals/ethics-policies/data-sharing-mining/.

Tullis, J and **Kar, B.** 2021. Where Is the Provenance? Ethical Replicability and Reproducibility in GIScience and Its Critical Applications. *Annals of the American Association of Geographers*, 111(5): 1318–1328. DOI: https://doi.org/10.1080/24694452.2020.1806029

Wilkinson, MD, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(160018). DOI: https://doi.org/10.1038/sdata.2016.18

TO CITE THIS ARTICLE:
Bishop, BW, Neish, P, Kim, JH, Bats, R, Million, AJ, Carlson, J, Moulaison-Sandy, H and Pham, MT. 2023. Data Management Plan Implementation, Assessments, and Evaluations: Implications and Recommendations. *Data Science Journal*, 22: 27, pp. 1–8. DOI: https://doi.org/10.5334/dsj-2023-027

Submitted: 16 December 2022
Accepted: 25 May 2023
Published: 25 August 2022

COPYRIGHT:
© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.