

---

AI translation · View original & related papers at  
[chinaxiv.org/items/chinaxiv-202308.00440](https://chinaxiv.org/items/chinaxiv-202308.00440)

---

## Application of Data Management Plans in Library Scientific Data Management Services: Postprint

**Authors:** Wang Dandan

**Date:** 2023-08-26T00:00:00+00:00

### Abstract

[Purpose/Significance] Aimed at the current application of data management plans in library user research and scientific data management services, this study synthesizes and summarizes the application value of scientific data management plans to guide libraries in better leveraging data management plans to enhance their scientific data management services.

[Method/Process] By reviewing existing research findings on data management plans, this paper summarizes three application values of data management plans for libraries in advancing scientific data management services. Taking the content analysis practice of the scientific data management plan at Nanyang Technological University, Singapore as a case study, it further validates the proposed three application values and considerations for application.

[Results/Conclusion] The value of data management plans for libraries in advancing scientific data management services is primarily manifested in three aspects: as an information resource that facilitates in-depth understanding of researchers' data management practices and behavioral characteristics within the institution; as a feedback channel that provides insights for improving current services and developing potential services; and as a practice-rooted training platform that helps librarians effectively enhance data service capabilities and build stakeholder collaboration networks. However, analyzing data management plans has certain inherent limitations and should be combined with other methods to better guide institutions in implementing and improving scientific data management services. Nevertheless, the importance of data management plans is unquestionable, and relevant domestic institutions should prioritize requiring data management plans as soon as possible.

**Full Text**

**Preamble**

**Vol. 62 No. 1, January 2018**  
**ChinaXiv Partner Journal**

**Application of Data Management Plans in Library Research Data Management Services**

**Wang Dandan**

School of Management, Henan University of Science and Technology, Luoyang 471023

**Abstract**

**[Purpose/Significance]** This paper summarizes the application value of data management plans (DMPs) in library user research and scientific data management services, guiding libraries to better utilize DMPs to enhance their research data management (RDM) services. **[Method/Process]** Existing research on DMPs was reviewed, and three types of application value for libraries advancing RDM services were identified. The DMP content analysis practice at Nanyang Technological University (NTU) in Singapore was used as a case study to further verify these three application values and associated implementation considerations. **[Result/Conclusion]** The value of DMPs for libraries in advancing RDM services manifests in three aspects: as an information resource to help understand researchers' data management practices and behavioral characteristics; as a feedback channel to inspire improvements to current services and development of potential services; and as a practice-based training platform to help librarians effectively enhance data service capabilities and build stakeholder cooperation networks. However, analyzing DMPs has inherent limitations and should be combined with other methods to better guide institutions in implementing and improving RDM services. Nevertheless, the importance of DMPs is undeniable, and domestic institutions should prioritize implementing DMP requirements as soon as possible.

**Keywords:** data management plan; libraries; content analysis

**1. Research Review on Using DMPs to Optimize and Improve Institutional RDM Services**

As awareness of the importance of scientific data sharing continues to grow, an increasing number of countries, international funding agencies, and research institutions are requiring researchers to submit data management plans (DMPs) that document their plans and measures for describing, storing, protecting, sharing, and preserving scientific data. A DMP is essentially a written document describing the existing data to be acquired or new data to be generated during

a research project, explaining how the data will be managed, analyzed, and stored, and how it will be shared and preserved upon project completion [1]. While DMPs are documents prepared by researchers to meet policy requirements of funding agencies or institutions, they represent an extremely valuable resource for libraries advancing RDM services. This paper reviews research and practices on how libraries can use DMPs to optimize and improve institutional RDM services, and combines this with DMP analysis practices at NTU to summarize the application value of DMPs in library RDM services, providing references for domestic institutions in formulating data management policies and implementing RDM services.

Although libraries currently offer various forms of RDM services, services centered on DMPs are the most common and universal [2]. DMP-related RDM services include: (1) summaries of funding agency DMP requirements; (2) DMP writing guidelines and best practices; (3) embedded online tools such as DMP-Tool to support DMP creation; (4) institution-specific DMP templates; (5) DMP FAQs and workshops; and (6) DMP review services [3]. For example, Stanford University provides data management training, DMP writing consultation, and support for selecting metadata and data repositories.

The application of DMPs in library RDM services is primarily reflected in three aspects:

### **1.1 Using DMPs to Understand Researchers' Data Management Practices**

Syracuse University Library first conducted an online survey of 966 National Science Foundation (NSF)-funded researchers from multiple countries to understand their data management attitudes and practices, then analyzed 68 voluntarily submitted DMPs. The analysis revealed that researchers primarily used informal methods (such as email requests) to share data, rarely used metadata to describe data, and did not consider data reuse issues [4]. As part of a pilot program at the University of Michigan to provide RDM services to NSF grant applicants, N. H. Nicholls et al. obtained 104 DMPs from successfully funded engineering faculty and analyzed the extent to which these DMPs met NSF guidelines, finding that most faculty had unclear assignment of data management responsibilities [5]. Analysis of 182 DMPs from researchers at the University of Minnesota revealed significant differences in data sharing and preservation behaviors across different fields [6]. Georgia Tech's analysis of 181 DMPs found that even when using institutional repositories for data sharing, different disciplines used inconsistent terminology, suggesting that different academic departments had their own language preferences and recommending unified terminology and targeted efforts to raise awareness of institutional repository usage [7]. The University of Michigan Library used DART Rubric (DMP Assessment Rubric) [8] to evaluate 29 DMPs from engineering researchers, finding that most researchers had unclear understanding of data management roles, inadequate responsibility allocation, and lacked intellectual property awareness

[9].

These studies demonstrate that DMPs, as documents written by researchers themselves, can reflect their understanding of the potential value of their data, the environment in which data is created and prepared, and their willingness and ability to ensure data can be utilized by others now and in the future. While researchers generally understand funding agency requirements, there remains a gap from best practices. For instance, they often fail to designate specific personnel responsible for data management, describe data storage duration after project completion, include policies for overseeing data reuse or dissemination, mention specific metadata standards or data description methods, and rely on informal, unreliable, and unsustainable methods such as email requests or personal/project websites for sharing. These issues reflect that researchers lack thorough consideration of the entire data lifecycle, do not understand metadata and its role in making data discoverable and reusable, and are unaware of the limitations of informal data sharing methods. Despite writing DMPs, their data may not be easily discoverable or reusable in the long term. Therefore, libraries have significant room to improve RDM services.

## 1.2 Providing a Basis for Developing Institutional RDM Services

Since NSF introduced DMP requirements in 2011, libraries have actively explored their potential roles in the DMP landscape. One approach to supporting researchers is providing DMP templates. Participants in focus group studies at Colorado State University noted they were aware of and had used the library's DMP templates, finding them very useful [10]. The University of Illinois Library provides specific DMP templates for different NSF disciplinary divisions [11]. The University of Houston Library offers online forms for researchers [12]. Johns Hopkins University provides data management and curation services covering the entire data lifecycle, from DMP consultation to using institutional repositories for data storage [13]. Purdue University developed an institutional repository primarily for disseminating and curating scientific data [14], collaborating with faculty to address their diverse data needs [15-16]. Other libraries provide data information literacy education [17-18].

Structured review of DMPs can reveal researchers' deficiencies and weaknesses in understanding and applying data management concepts and practices, identifying barriers to achieving optimal data management. DMP content analysis results can directly inform improvements to DMP consultation services and provide valuable information for developing potential new services. For example, J. E. V. Loon et al. analyzed differences between DMPs from engineering and arts researchers, finding that compared to arts researchers, engineering researchers less frequently mentioned data sharing through journal article appendices, indicating a need to raise awareness among engineering researchers about this data sharing method. Arts researchers, compared to their engineering counterparts, less frequently specified data preservation duration, described reuse and redistribution policies, or addressed protection of sensitive data and intellectual

property issues, suggesting they lacked understanding of the data lifecycle's importance and specific measures for ensuring data security. Libraries can use these findings to improve data education content [19]. S. W. Parham et al. evaluated 500 DMPs from five universities using DART Rubric and found that in addition to training on data management topics (such as metadata and its applications, appropriate formats for shareable data, and data reuse documentation), researchers clearly needed guidance on data licensing and intellectual property policies. This expertise is dispersed across multiple departments within institutions, making it essential to build alliances and cooperation networks among libraries, IT centers, funding administrators, and other stakeholders for successful training programs. This is a priority for building data management capacity and meeting local needs [20].

### 1.3 Building a Cooperative Network of RDM Service Stakeholders

To effectively deliver RDM services, educating and training librarians is essential foundational work. K. Antell et al. found that while librarians considered “experience gained from helping write DMPs” very useful, only 2% believed they possessed these skills or were actively acquiring them [21]. A. Cox et al. also found that lack of domain-specific expertise and limited research experience hindered librarians' ability to play key supporting roles, recommending that librarians learn through real-world practice for optimal results [22]. Currently, different libraries implement various training forms for librarians based on institutional culture, available resources, and researcher needs. The widely recognized New England Collaborative Data Management Curriculum is considered effective, though users note it requires substantial time and effort [23-25].

This makes DMP review services a relatively time-efficient, low-effort, and practice-based training platform. Conducting DMP review services first creates opportunities for librarians to become more familiar with DMPs and funding agency requirements. While librarians have a basic understanding of what DMPs should contain, requirements vary by discipline. Librarians must familiarize themselves with different requirements to provide targeted, practical feedback, thereby demonstrating the value of librarians and library services and establishing librarians as important partners in the research process. Second, it provides librarians with opportunities to better understand how researchers plan and consider data management, helping them identify directions for providing assistance. Using DMP review services as a platform and training ground enables librarians to continuously improve data service awareness, acquire knowledge, and develop professional skills.

Wayne State University Library found that conducting DMP quality evaluation was a valuable experience for its RDM service team, providing opportunities to enhance understanding of funding applications and DMP processes, strengthen relationships between the data services team and university administrators and other research support staff, and provide valuable information for designing DMP workshops for other librarians [19]. Currently, North Carolina State Uni-

versity Library has implemented team-based DMP review services and explored the feasibility of using DMP review for staff skills training. The library integrates librarians with different skills and expertise from various departments to quickly and effectively support researchers in writing DMPs that meet funding agency requirements. Through shared responsibility, staff rotation, and inviting subject librarians to participate in DMP review processes, participants have opportunities to share best practices and learn from each other, forming a broader campus-wide research support network. Additionally, reviewing actual DMPs enables librarians to quickly gain firsthand experience, identify gaps in their knowledge and skills, and build networks with administrators from other schools and departments [26].

## 2. DMP Analysis Practice at Nanyang Technological University

Like other research universities worldwide, Nanyang Technological University (NTU) recognizes that effective RDM is an important component of research integrity. In 2014, NTU Library established a “data librarian” position and formed an RDM working group, beginning to plan RDM services as a new business direction. Unlike the United States and United Kingdom, Singapore has no national-level open data policy, and research funding agencies have not issued data management requirements. Consequently, the NTU RDM team made promoting the development of university-level scientific data policy its primary task. In April 2016, NTU enacted its research data policy, becoming the first research institution in Singapore to require researchers to submit DMPs.

All principal investigators (PIs) of funded research projects must create and submit a DMP in the Research Information Management System (RIMS) at the project’s outset; otherwise, research funds cannot be accessed. PIs are responsible for managing data generated during the project and for submitting DMPs in RIMS, updating them when substantive changes occur. The policy stipulates that unless specially permitted by the head of the research management department, all scientific data must be stored in facilities, equipment, or virtual spaces accessible and controlled by the university. All scientific data related to the project must be retained for at least 10 years after publication or project completion. Final data used to establish and verify research results must be deposited in the open data repository DR-NTU (Data) or other recognized open data repositories. If the latter, URLs and dataset access methods must be provided [27].

Based on these policy requirements, the NTU RDM team made DMP support services and the development and promotion of DR-NTU (Data) the entry point and main direction of service advancement. On one hand, the library designed DMP templates [28] and RDM guidelines [29] and began offering DMP writing workshops in May 2016. On the other hand, based on Harvard University’s Dataverse software, the library developed DR-NTU (Data), which was launched in August 2017. As of August 2017, over 700 DMPs had been submitted in the

RIMS system, with 298 people trained through workshops. To further optimize and improve NTU's RDM services, the RDM working group decided to analyze DMPs exported from the RIMS system. The analysis framework is shown in Figure 1 [Figure 1: see original paper].

**Part 1: NTU DMP Quality Evaluation.** First, by analyzing DMPs that could not be used for quality evaluation, the team found that NTU indeed had cases where scientific data would not be generated during project implementation and DMPs were unnecessary. Second, through comprehensive quality analysis of DMPs, the team found room for improvement in overall NTU DMP quality, with significant differences not only between different colleges but also between different departments within the same college. Third, through analysis of DMP content elements, the team found that NTU researchers could adequately describe data types, file extensions, software and tools for reading files, conditions for data access and sharing, data documentation, storage and backup, and long-term preservation. However, descriptions of data formats, collection methods, data usage, file organization and description, and data management responsibilities and authority were incomplete and unclear. Moreover, significant variations existed between colleges and between departments within the same college regarding these content elements.

Based on these findings, the NTU RDM team planned to: (1) consider cases where DMPs are unnecessary in the RIMS system; (2) implement DMP review services, referencing practices at North Carolina State University; and (3) conduct in-depth interviews with specific colleges and departments based on differences in DMP quality and content elements to better understand the reasons for these differences, assess the applicability of the NTU DMP template to different colleges and disciplines, identify departments for deeper collaboration, and design differentiated RDM services.

**Part 2: Analysis of NTU Researchers' RDM Practices.** First, by analyzing descriptions of long-term preservation plans for final data in DMPs, the team found that most research projects did not yet have appropriate open data repositories for storing final scientific data and planned to use DR-NTU (Data) for long-term preservation. A small group of researchers already had good data sharing awareness and practices, using existing open data repositories such as OSF (<https://osf.io>), GitHub (<https://github.com/>), and Dryad (<http://datadryad.org/>) for long-term preservation and sharing. However, some researchers misunderstood what constitutes an open data repository, listing platforms that were not true open data repositories as their planned preservation locations. Second, by analyzing descriptions of preservation plans for data generated during the project, the team found that some researchers could not accurately distinguish between final scientific data and process-generated data, and the platforms they mentioned for long-term preservation of process data were primarily traditional, such as various servers, hard drives, portable storage devices (DVDs, USB drives), and personal computers. Third, some researchers could not effectively differentiate between storing data and preserving data, as

evidenced by identical or similar descriptions for these two components in their DMPs.

Based on these findings, the NTU RDM team planned to: (1) further promote DR-NTU (Data) to increase awareness and usage among NTU researchers; (2) investigate external open data repositories already used by NTU researchers and study how to integrate DR-NTU (Data) with these platforms—the NTU RDM team has already written guidelines to help researchers integrate OSF and DR-NTU (Data); and (3) reference DMP themes proposed by the UK Digital Curation Centre (DCC) [30] and, combined with NTU's specific circumstances, further standardize terminology in the DMP template and refine its questions.

As descriptions of researchers' own data management practices, DMPs contain researchers' awareness, knowledge, and capabilities for data management, making them extremely valuable resources. Meanwhile, standardized tools for supporting best-practice DMP development, such as DMPOnline and DMPTool, and widely used and continuously improved rubrics for cross-institutional DMP quality evaluation, provide strong support for mining this resource's value. Although funding agencies' DMP requirements vary significantly and different research groups have different data practices—meaning DMPs may contain complex information written for domain experts that is difficult to fully understand without relevant disciplinary background, and researchers may write DMPs to meet institutional requirements rather than accurately describe their actual practices and intentions—the value of DMPs and the advantages of DMP content analysis cannot be denied.

Based on current research findings, the value of DMPs is primarily manifested through three main application directions: (1) as an information resource, enabling in-depth understanding of researchers' data management practices and behavioral characteristics through structured DMP content analysis, serving as the first step in user research for institutional scientific data services, providing background material for further in-depth interviews, and laying the foundation for service planning and design; (2) as a feedback channel, where deficiencies and weaknesses in researchers' understanding and application of data management concepts and practices, as well as barriers to achieving optimal data management revealed in DMPs, provide ideas for improving current services and designing data information literacy education content, and inspire the development of potential new services; and (3) as a practice-based training platform, where team-based DMP review services not only promote library RDM services and build service influence but also facilitate mutual learning and knowledge sharing among different personnel, strengthen connections and collaboration with relevant departments within the institution, build a service network covering the entire institution, and effectively enhance librarians' data service knowledge and skills.

NTU Library fully recognizes the importance of DMPs. In the absence of national-level policies and funding agency requirements in Singapore, the university formulated its own policy and collects DMPs written by researchers through

the RIMS system. NTU's DMP content analysis fully demonstrates the value of DMPs for libraries implementing RDM services as summarized in the first part of this paper. Using DMPs as an entry point not only helps the NTU RDM team understand researchers' data management practices and identify RDM service needs, providing a basis for advancing RDM services, but also serves as a connection point to help build a cooperation network between the library RDM team and internal university stakeholders such as research management, research integrity, legal, IT departments, and researchers, ensuring effective RDM service advancement. In view of this, the NTU RDM team plans to make DMP analysis a regular part of its routine work. However, the team also recognizes inherent limitations in DMP analysis: it can answer "what" questions (e.g., what is the overall quality of institutional DMPs, whether differences exist between academic departments and to what extent) but cannot answer "why" questions (e.g., what factors cause these differences). Therefore, to better guide institutions in implementing and improving RDM services, libraries should consider DMP analysis as a first step, followed by targeted in-depth interviews using existing tools such as Purdue University's DCPT (data curation profiles toolkit) [31], University of Virginia's DMVitals [32], and University of Illinois at Urbana-Champaign's risk assessment tool [33] to understand the reasons for differences and provide a basis for designing differentiated RDM services.

## References

- [1] STANFORD UNIVERSITY LIBRARIES. About data management plan [EB/OL]. [2017-03-06]. <https://library.stanford.edu/research/data-management-services/data-management-plans>.
- [2] RICE R, SOUTHALL J. The data librarian's handbook [M]. London: Facet Publishing, 2016.
- [3] STANFORD UNIVERSITY LIBRARIES. Data management plan [EB/OL]. [2017-03-31]. <http://library.stanford.edu/research/data-management-services/data-management-plans>.
- [4] CURTY R, KIM Y, QIN J. What have scientists planned for data sharing and reuse? a content analysis of NSF awardees' data management plans [EB/OL]. [2017-02-28]. <http://surface.syr.edu/istudents/2>.
- [5] NICHOLLS NH, SAMUELS M, LALWANI LN, et al. Resources to support faculty writing data management plans: lessons learned from an engineering pilot [J]. *International journal of digital curation*, 2014, 9(1): 242-252.
- [6] BISCHOFF C, JOHNSTON L. Approaches to data sharing: an analysis of NSF data management plans from a large research university [EB/OL]. [2017-11-25]. <https://jisc-pub.org/articles/abstract/10.7710/2162-3309.1231/>.
- [7] PARHAM SW, DOTY C. NSF DMP content analysis: what are researchers saying? [J]. *Bulletin of the American Society for Information Science and Technology*, 2012, 39(1): 37-38.

[8] ROLAND O, CARLSON J, HSWE P, et al. Data management plans as a research tool [J]. *Bulletin of the American Society for Information Science and Technology*, 2015, 41(5): 43-45.

[9] SAMUELS M, GROCHOWSKI PF, LALWANI LN, et al. Analyzing data management plans: where librarians can make a difference [EB/OL]. [2017-02-28]. <https://www.asee.org/public/conferences/56/papers/12072/view>.

[10] MCLURE M, LEVEL A, CRANSTON CL, et al. Data curation: a study of researcher practices and needs [J]. *Portal libraries and the academy*, 2014, 14(2): 139-164.

[11] MISCHO WH, SCHLEMBACH MC, O'DONNELL MN. An analysis of data management plans in University of Illinois National Science Foundation Grant Proposals [EB/OL]. [2017-11-25]. [http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1045&context=refinst\\_](http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1045&context=refinst_)

[12] PETERS C, DRYDEN AR. Assessing the academic library's role in campus-wide research data management: a first step at the University of Houston [J]. *Science & technology libraries*, 2011, 30(4): 387-403.

[13] SHEN Y, VARVEL VE. Developing data management services at the Johns Hopkins University [J]. *The journal of academic librarianship*, 2013, 39(6): 552-557.

[14] WITT M. Co-designing, co-developing, and co-implementing an institutional data repository service [J]. *Journal of library administration*, 2012, 52(2): 172-188.

[15] GARRITANO JR, CARLSON J. A subject librarian's guide to collaborating on e-science projects [EB/OL]. [2017-02-28]. <http://www.istl.org/09-spring/refereed2.html>.

[16] CARLSON J, STOWELL-BRACKE M. Data management and sharing from the perspective of graduate students: an examination of the culture and practice at the Water Quality Field Station [J]. *Portal libraries and the academy*, 2013, 13(4): 343-361.

[17] CARLSON J, FOSMIRE M, MILLER CC, et al. Determining data information literacy needs: a study of students and research faculty [J]. *Portal libraries and the academy*, 2011, 11(2): 629-657.

[18] CARLSON J, JOHNSTON L. *Data information literacy: librarians, data and the education of a new generation of researchers* [M]. West Lafayette: Purdue University Press, 2015.

[19] LOON JEV, KATHERINE G, HUDSON AC, et al. Quality evaluation of data management plans at a research university [J]. *International federation of library associations and institutions*, 2017(1): 1-7.

[20] PARHAM SW, CARLSON J, HSWE P, et al. Using data management plans to explore variability in research data management practices across domains [J]. *The international journal of digital curation*, 2016, 11(1): 53-67.

- [21] ANTELL K, FOOTE JB, TURNER J, et al. Dealing with data: science librarians' participation in data management at Association of Research Libraries Institutions [J]. College & research libraries, 2014, 75(4): 557-574.
- [22] COX A, PINFIELD S. Research data management and libraries: current activities and future priorities [J]. Journal of librarianship and information science, 2014, 46(4): 299-316.
- [23] KAFEL D, CREAMER A, MARTINE R. Building the New England collaborative data management curriculum [EB/OL]. [2017-11-25]. <http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1064&context=jeslib>.
- [24] ISHIDA M. The new England collaborative data management curriculum: a Canadian experience pilot at the University of Manitoba [EB/OL]. [2017-11-25]. <https://mspace.lib.umanitoba.ca/handle/1993/30222>.
- [25] PETERS C, VAUGHN P. Initiating data management instruction to graduate students at the University of Houston using the New England collaborative data management curriculum [EB/OL]. [2017-11-25]. <http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1066&context=jeslib>.
- [26] DAVIS HM, CROSS WM. Using a data management plan review service as a training ground for librarians [EB/OL]. [2017-11-25]. <https://jlscepub.org/articles/abstract/10.7710/2162-3309.1243/>.
- [27] NANYANG TECHNOLOGICAL UNIVERSITY. NTU Research data policy [EB/OL]. [2017-09-01]. <http://research.ntu.edu.sg/rieo/RI/Pages/Research-Data-Policies.aspx>.
- [28] NANYANG TECHNOLOGICAL UNIVERSITY. DMP template [EB/OL]. [2017-09-01]. <http://blogs.ntu.edu.sg/lib-data-management/dmp/>.
- [29] NANYANG TECHNOLOGICAL UNIVERSITY. Research data management guide [EB/OL]. [2017-09-01]. <http://blogs.ntu.edu.sg/lib-data-management/>.
- [30] DIGITAL CURATION CENTRE. DMP themes [EB/OL]. [2017-11-25]. <http://www.dcc.ac.uk/sites/default/files/documents/publications/DMP-themes.pdf>.
- [31] CARLSON J. The data curation profiles toolkit: the profile template [EB/OL]. [2017-02-06]. <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1003&context=dcptoolkit>.
- [32] SALLANS A, LAKE S. DMVitals: a data Management assessment tool [EB/OL]. [2017-06-13]. <https://journals.lib.virginia.edu/catalog/libra-oa:469>.
- [33] UNIVERSITY of ILLINOIS at URBANA-CHAMPAIGN. Making data management manageable: a risk assessment activity for managing research data [EB/OL]. [2017-06-13]. <http://hdl.handle.net/2142/95768>.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv — Machine translation. Verify with original.*