



Information Research – Vol. 31 No. iConf (2026)

Revising the research data lifecycle: Evidence from oceanography

Yubing Tian, Thomas Struett, Amelia Acker, and Megan Finn

DOI: <https://doi.org/10.47989/ir31iConf64146>

Abstract

Introduction. This research examines temporal elements of data management for ocean science researchers funded by the United States' National Science Foundation (NSF). By uncovering data practices for sharing data, this paper finds that researchers' experience diverges significantly from traditional models of research data lifecycle.

Method. Corpus creation of 379 Data Management Plans (DMP) from NSF funded oceanography awards between 2011-2021 and 34 semi-structured interviews with principal investigators (PIs) and data managers from purposive and snowball sampling.

Analysis. Qualitative coding of both DMP documents and interview transcripts using MaxQDA software using a codebook drawn from Adam's (1998) timescapes framework and themes developed through inductive qualitative analysis of the interview transcripts.

Results. Research data management to support sharing data lacks an established place in the organization of research practice and often leads to end-loaded project data management. Three ways that researchers are developing new practice-time profiles include creating routines, automating workflows, and anticipating the data practices of reusers.

Conclusion. Data management for sharing is not only constrained but also actively shaped through time. Research data lifecycles need to place more emphasis on the actors within this lifecycle.

Introduction

The United States' National Science Foundation's (NSF) data management and sharing policy, like many other countries and regions' data sharing policies, envisions data sharing as a routine part of research. However, researchers often struggle to make time for the very tasks that would realize this vision. Data management for public sharing is frequently treated as one of the final steps in a research project, something to be completed once the core scientific work is completed. This sequencing of tasks, combined with time pressures and constraints, contributes to the deferral of data sharing work. While traditional research data lifecycle models and data sharing policy presume a progression from data collection to long-term availability, this paper shows that researchers' experiences diverge significantly from that model. The difficulties of 'finding time' for data management is shaped by temporal conflicts that reflect labour constraints and the mismatched assumptions about when, how, and who does data work.

These findings are based on analysis of 379 data management plans (DMPs) submitted to the NSF between 2011 and 2021 for successfully funded oceanography research projects as well as 34 interviews with principal investigators and data managers. DMPs have been a required two-page document with every NSF grant submission since 2011 where scientists lay out their plans for the management of their data and how they plan to share data in compliance with NSF's open data policies (NSF, 2011).

Traditionally, prescriptive models of research data management and sharing presume a linear, standardized trajectory from collection to sharing and tend to obscure invisible data work and labour problems. With the exception of some models, such as the Data Curation Continua (Treloar & Harboe-Ree, 2008) and Stahlman's (2022) researcher-centred data lifecycle model, human actors and the labour required to move data from one stage to the next are absent from these models. Scholarship has established that in practice, researchers struggle to find time for key data management tasks (Borgman et al., 2016; Tenopir et al., 2011, 2015).

To explore why researchers often find data management and sharing difficult to prioritize, we draw on Shove and colleagues' (2009) notion of '*practice-time profiles*' which '*refer to embedded conventions of duration, sequence and timing associated with the competent performance of a practice*' (Shove et al., 2009, p. 25). Shove's concept builds on Zerubavel's seminal observation that practices are governed by socially shared understandings of '*the proper time*,' which emphasizes that timing is collectively shaped rather than individually determined (Zerubavel, 1981, p. 8 in Shove, 2009, p. 25). Based on interviews with oceanographers, this appears to be in part, because researchers conceive of data management for public sharing as work that occurs at the end of their projects, rather than as a distributed component of research. Using our analysis to build on previous scholarship, this paper makes two claims: (1) there is no established practice-time profile for data management and sharing and (2) knowing the '*right time*' to make data sharable is a marker of competence and is necessary for developing accurate models for data management and sharing.

While traditional data lifecycle models presume a linear progression from data collection to sharing, this research shows that data management for sharing data publicly lacks an established place in the temporal organization of research practice. The absence of an established temporal organization for data management results in uncertainty about '*the proper time*' to do that work and contributes to it being concentrated over a narrow window what we term an '*end-loaded*' practice-time profile. To better understand this this pattern, we pair Shove's concept with British sociologist Barbara Adam's (2008) '*timescapes*' theoretical framework, which asserts that time is '*complex and multiplex, involving many different features and dimensions*' (Adam, 2008, p. 1). While timescapes help characterize the broader temporal organization of contemporary research in which data management occurs, such as compressed

project timelines and competing temporal expectations of publications and policies. Practice-time profiles help illustrate how researchers navigate these conditions in their data practices. Together, this analytical pairing allows us to examine both the structural and practice-level temporalities that shape data management and sharing and explain why time pressures materialize unevenly across the research lifecycle.

Here we also report findings related to three ways that researchers are developing new practice-time profiles for data management. With the caveat that there is no universal or ideal practice-time profile, examples from this study include early planning, creating routines, and automating workflows. We conclude by discussing the implications of practice-time profiles for the data management lifecycle.

Literature: the temporalities in data lifecycles

This literature review focuses on research data lifecycle models, beginning with prescriptive research data lifecycle models where we surface the implicit assumptions about time, how data move through time, when they are made available, and the relationship of the model to actual data practices.

Prescriptive research data lifecycle models

Most models of data and its lifecycle are prescriptive models (Mosconi et al., 2019). Prescriptive models of research data are abstract and include pre-defined stages for research data (see Figure 1). Within a given model, the number of stages that research data is thought to pass through, as well as their sequencing, which stage comes before and after, may vary. However, in broad strokes, these research data models share several key similarities. Movement of data is depicted cyclically, and the stages are individually contained and fixed.

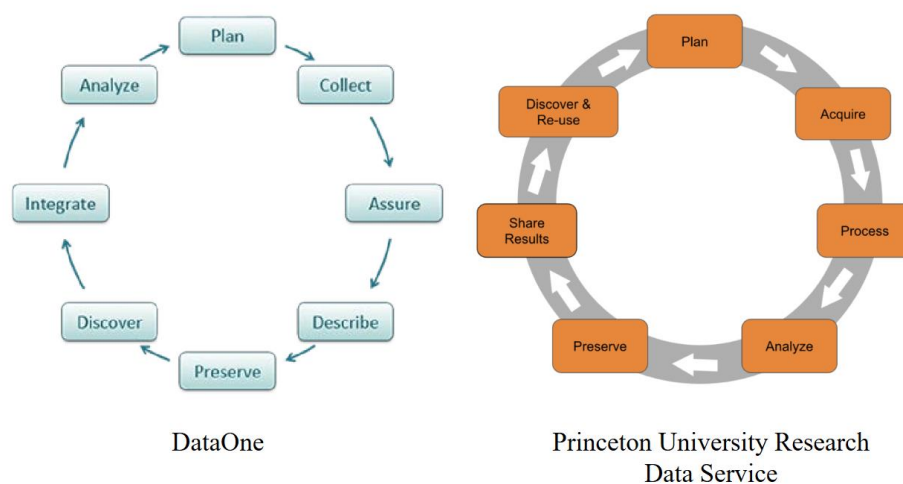


Figure 1. Examples of research data lifecycle models

First, positioned at the top of the circular diagram is the planning stage and this is often depicted as the first phase in the data lifecycle. Second, stages that are typically associated with active phases of a research project are presented, such as collection and analysis. Sharing and preservation are typically the last stages that are associated with research data being made available beyond the research team that produced or collected that data. Data reuse is imagined as directly spawning new avenues of inquiry and further planned research to begin this cycle anew. The exception is the prescriptive research data model from DataOne, which begins with data re-

use or what they term the ‘discover stage,’ which is reflective of their organization, a network of interoperable data repositories. These models are devoid of specific actors that would be involved in enabling research data to move through different stages. Overall, there is a sense that research data inherently possess this lifecycle, with each stage requiring equal attention.

Mosconi et al., argue that these abstract models help illustrate what is being requested by funding bodies, but that they neglect the ‘collaborative infrastructure in which researchers actually engage in the business of storing, managing, and archiving data’ (2019, p. 753). The authors point to Treloar and Harboe-Ree’s (2008) research data model, the Data Curation Continua, as a closer representation of research data as it moves across space and time. The model, as illustrated in Figure 2, represents a more dynamic model of data management.

Notable differences from the data curation continua model and the prescriptive research data lifecycle models are the recognition that labour is required to move data between private, shared, and public domains. The model highlights how PI’s and institutional resources play a critical role in moving data between domains. Whilst this model highlights the data management activities involved in producing accessible and open data, there is an implicit linearity that remains. This model also exclusively concerns the active phases of a research project as there is no treatment of data management planning or data preservation.

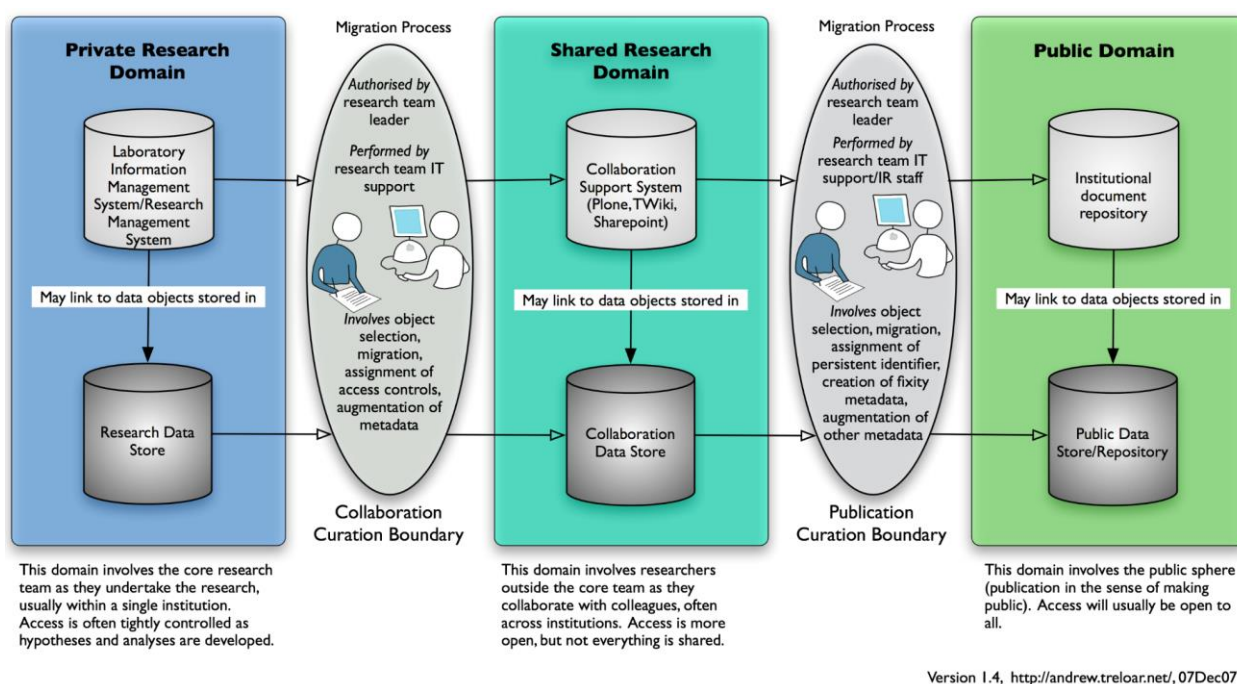


Figure 2. From Treloar & Harboe-Ree (2008) The data curation continua model

Although there is a type of linear time implied within these models, research data is depicted to move unidirectionally from stage to stage, or domain to domain, as the hypothetical research project progresses. Through the example of ocean data repositories, Baker and Chandler suggest that scholars rethink data flows as non-linear as the movement of data is a ‘complex system of frequently ill-defined relationships between local repositories and a larger-scale community web of institutional repositories, discipline-specific centers, and national archives.’ (2008, p. 2134). Mosconi et al., (2019) call ‘pragmatic models of data,’ a bottom-up representation of data curation and management, embedded within researchers’ data practices. We contribute to scholarship by

developing a pragmatic model of data management lifecycle by focusing on researchers' discussion of how time factors into their data management practices.

Methods

Data management plan corpus construction

The first data generation activity consisted of collecting DMPs of active and non-active research projects funded by NSF's Division of OCE between January 2011 and June 2021. In the tradition of constructivist grounded theory, these documents are considered '*extant documents*'; the researcher did not play a role in their construction (Charmaz, 2014; Ralph et al., 2014). While DMPs are not made publicly available by NSF because they can contain personally identifiable information, PIs are free to share them publicly if they wish. So, in order to create a corpus of DMPs, we emailed PIs asking if they would share their DMPs with us. While this research only uses DMPs from the Division of OCE, DMPs from the Division of Biological Infrastructure, Civil, Mechanical, and Manufacturing Innovation, Secure and Trustworthy Cyberspace, and Science and Technology Studies were also collected.

To begin the DMP corpus construction, we first found all NSF OCE-funded projects between 2011 and 2021 in June 2021 using information compiled from NSF's publicly accessible award database using the advanced search interface (<https://www.nsf.gov/awardsearch/advancedSearch.jsp>). Grants focused on teaching, conferences, or from short duration projects were excluded (RAPID, EAGER, CAREER, EDU, RCN, Workshop, Symposium, and RCU grants). Grants under \$100,000 were also excluded from our corpus. Moreover, if a PI had more than one project funded during 2011 - 2021, we selected the oldest grant for that PI. This decision was made to not fatigue participants with more than one DMP solicitation request.

From July to August 2021, 1689 personalized emails were sent to OCE-funded PIs asking if they would share their DMP for our research. 400 OCE DMPs were received, of which 379 qualified for this study. This resulted in a response rate of 23.39%. Figure 3. Shows the total number of OCE DMPs received and included in the final corpus organized by the year that the project was funded.

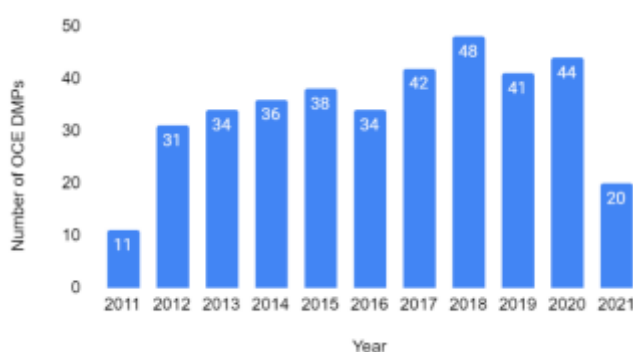


Figure 3. Number of OCE DMPs by year the project was funded

Semi-structured interviews

The second data generation activity was semi-structured interviews with PIs of funded NSF OCE research and their project collaborators. Interviews help to surface perspectives, descriptions of practices, and reflections in participants' own words (Dumit, 2004) which provide

descriptions and reflections on important activities and events that occurred in the past that may have little official documentation (Tracy, 2013). Semi-structured interviews helped to provide context about the research project and the DMP creation which is key to placing texts into context; self-reported data management practices and their perceptions on how, if at all, the DMP mandate has affected their data practices.

The first round of 21 one-hour interviews was conducted between March - June 2022, and the second round was conducted between January - March 2025. Initially, purposive sampling was used to contact PIs who had, in their email replies to our DMP solicitation request, expressed a willingness to discuss their DMP and their research project further with us. Interviews covered researchers' data types and data formats, who performed data-related work during the research project, how collaborators learned about data management best practices, the role of DMPs in their day-to-day work, discipline-related changes in data collection and dissemination methods, and data challenges and open access to data discussions. The second round of interviews included fourteen more scientists who were found using snowball sampling from the first round of interviews. The second phase of interviews focused on data management practices for sharing and preserving data and who was responsible for that work.

Overall, the first phase established context and helped situate the data-related work and decisions within the broader aims of the specific project, and the second phase focused on the ending phases of the project, including how data sharing was approached, the types of research artifacts shared, and who had access to these.

Document & interview analysis

In this study, the DMPs from funded projects provided a starting point for interviews with oceanography PIs to learn more about their projects and the role that DMPs played in their data management and sharing practices. A preliminary cycle of coding OCE DMPs in the qualitative coding software MaxQDA included 'temporality' as a code, which captured references to time in DMPs. This code was subsequently expanded using a second codebook drawing on Adam's (1998) timescapes framework and Steinhardt and Jackson's (2015) 'anticipation work.' This codebook was refined through four rounds of revision where two researchers independently applied the codebook to a portion of the corpus to improve definitional clarity and ensure it was applicable to both DMPs and interview transcripts. This paper focuses on findings associated with the sensitizing concept of 'practice-time profiles,' (Shove, 2009).

In addition, interview transcripts were also analysed through inductive qualitative analysis (Saldaña, 2013). After the second cycle of coding, the lead author wrote analytical memos to gain a deeper insight into emerging themes and examine co-occurring codes, to help establish links between themes.

Findings

The missing practice-time profiles of research data management

When asked how the NSF mandate had impacted their day-to-day data work, interviewees often responded 'not at all' (P20). However, in the same breath, they remarked that the 'only thing is now I have to spend time making the data so somebody else can use it' (P20). While on its face this seems like a contradiction, interviewees understood metadata, documentation, re-formatting, and following standards, as extensions of existing data work, but when researchers find themselves doing data work in preparation for sharing, they find that not to be the case.

As it turns out, data sharing, and managing research outputs to be shareable is far from a trivial task as PIs provided concrete estimates ranging from '10-20 hours of effort', to '30 hours' (P14),

as well as more nebulous ones such as that it just ‘takes forever,’ which nevertheless got across the point that these tasks are demanding on PIs time and effort. Instead, we can see that the requirement of management and sharing of data policies generates new forms of labour and asks researchers to work outside their established temporal norms and rhythms for research data practices. Working outside of temporal norms and rhythms leads to lacking temporal norms for research data management practices. The next section explores the consequences of this absence of missing practice-time profile.

End-loaded practice-time profiles of data management

Existing models of research data management, which present data sharing as occurring at the sunset stages of the research project, match some experiences of interviewee's informal practice-time profiles for data management and sharing. As one PI noted:

So, I think we understand that at the end of our project and our output, publications, thesis that we want to ... we have to share data. It's making sure [that] at this stage we're still in a position where often, when we've done all the work that we're interested in, that, we still then have to do extra time to make sure the data is in the right formats. (P22)

This quote shows that researchers may be aware of what the mandate requires but because projects span multiple years, the task of data management and sharing doesn't get the attention it should be given until the very end. When done at the end, PIs often do not *'realis[e] that data management takes a lot of time, and often dedicated resources and staff'* (P25). Instead of incrementally doing data management through the active phases of the research, when left until the end means that all the time and effort that are required for data management and sharing is experienced over a short duration. Specifically, this practice-time profile can be characterized as end-loaded which can make data sharing vulnerable to delays. While there were no instances where data sharing did not eventually occur, this could be due to sampling procedures and that those who are interested in this subject are more likely to agree to being interviewed.

Key personnel and the dynamics of data availability and delay

As the research data, per the mandate, is expected to be usable beyond the small circle of researchers, additional data management tasks are required for this move to be made successfully. Common data management tasks necessary during this phase of the project include re-formatting data to open-source formats and ensuring the data has sufficient documentation and metadata. These data management tasks are the intermediate steps that are omitted from the final research outputs and the labour that went into their production is rendered invisible. These intermediate steps in researcher's data work are performed by key project personnel, such as students, technicians and even project staff such as laboratory managers. With the exception of a handful of DMPs reported here, key personnel are typically omitted from the documents. Through interviews though it becomes clear how students, technicians, and other non-PI research members, to be referred through this section as personnel, play a significant role in shaping the dynamics of data availability and delay in an end-loaded practice-time profile.

Starting first with graduate and doctoral students, their theses are considered a key research output and one way that PIs write that they anticipate project data being shared. Depending on the graduation timelines and how this lines up with the timing of the project close-out phase, these may further push the availability of data from the policy-mandated time to further down the road. For example, some DMPs write about making their data available according to the mandate but with *'the exceptions will be data related to graduate thesis projects, which will not*

be made available until 12 months following graduation' (DMP 281). One interviewee explained to me how managing both student graduation timelines and data availability timelines is difficult:

It's very often that people, you know, have a grad student write a bunch of code, collect a bunch of data, and then in a hurry when they're close to graduating, they're like, 'oh, we gotta do this obligation' and they do it after the fact. (P11)

This scenario is all too familiar, and even for PIs and graduate students who are proponents of data sharing, it is easy to see how the work required to graduate takes precedence over data management and sharing.

When personnel times are out of sync with the project's lifetime, the timing of data sharing is affected. One interviewee who described their PI's approach to data management as 'hands-off' relied heavily on her predecessor's work which was used as an example for her to follow when producing the documentation required to deposit their data. She shared:

I was responsible for doing all of the intercalibration reports and. All of that. I would say. At least in our lab. [PI] has a pretty hands-off approach in terms of like paperwork and stuff like that. So yeah, I built off of some previous reports from a student who had submitted cobalt data, for I think a different GEOTRACES cruise in the past, which was really helpful. (P28)

The data management and sharing work done by project personnel can shape data availability even after they have moved on. As almost all DMPs note, the PI and co-PIs are ultimately responsible for overseeing and coordinating data sharing. However, interviews illustrate that the departure of personnel can create disruptions in data sharing timelines, particularly when those individuals authored code, maintained documentation or had tacit knowledge of data management and sharing. These disruptions do not necessarily prevent data from being shared, but they can contribute to delays and illustrate how end-loaded practice-time profiles are vulnerable to personnel turnover and making sure that personnel timelines and depend on the alignment between personnel timelines and data sharing requirements.

'We like to first publish and then share'

Many researchers tend to think about data sharing when they are about to publish, instead of during data collection or other active phases of the project. Likewise, many DMPs do not refer to the policy's timeframe for public data availability, *'two (2) years after collection'* and instead *substitute the mandate's timeframe with making their data publicly available 'upon publication.'* This is for many reasons, such as for fear of being scooped, but of interest is the implications of tying data availability so tightly to publication. Importantly, the issue is not the act of sharing data at the point of publication, as many journals increasingly require this, but that the labour required to prepare shareable data is deferred until that moment. This sequencing implicitly positions publication as the "proper time" for data management and therefore concentrating data work that could be undertaken earlier in the active phases of a project. The PI whom we quote in the heading shared:

For our work we typically, as many scientists, we like to first publish it and then share. I think the NSF, they have a 2-year time period where they really want to say, after 2 years after production of the data you really have to share it. So typically, our challenge is to make sure that within those 2 years that we have open publications ready. And of course, sometimes that takes a little while. (P22)

Synchronizing data management and sharing labour around publication introduces delays for the project and creates time pressure for data managers. For example, a data managers

recounted an incident of when researchers left data sharing until after their manuscript was accepted for publication, mimicking a panicked researcher, the interviewee continued:

We've got a paper in the pipeline, 'Help! help!' ... You can't wait until the last minute to say, 'Oh gosh, I need to get this 50 terabytes of data up and published.' Because the paper's being held up because [the data's] not there yet. That's bad. I hate that. I basically have to drop everything and go do what I call a fire drill. (P13)

The term, '*fire drill*,' used by the interviewee is apt because from the researcher's perspective, this is an urgent or emergency requiring an immediate response. And much like a firefighter, to address the situation, the data manager must disrupt their work, rush in, and put out this metaphorical fire.

When asked what they would do differently in their next project, almost all PIs replied with some variation of thinking about and doing data management and data sharing 'earlier' in their projects. One PI shared that *'if we still go through BCO-DMO for future data archiving, I think I would send them data way in advance, before our paper got published. Probably as soon as we submit our manuscript, we will send our data for archiving'* (P23).

Timing the public availability of publications and the data associated with the publications is tricky because additional labour is required that are not mere extensions of existing data work. The synchronization of data availability and publications aligns and meets journal requirements for data sharing. For oceanographers, one of the key publication venues is the American Geophysics Union and its journals which require that publications share their data publicly with a DOI. However, tying data to publications reinforces an end-loaded practice-time profile of data management and sharing. Our point here is not that data should be shared before publication, but that deferring labour, such as documentation, formatting and metadata work, to this narrow moment produces a bottleneck during the project's sunset stages.

Making time at '*right time*' for data management and sharing

Interviewees repeatedly emphasized the importance of doing data management '*in advance*' or '*earlier*.' This suggests that there is a perceived temporal misalignment between when data management is being done and when it ideally *should* be done. From a theoretical perspective, '*timing*' is one of the eight temporal dimensions of a '*timescape*' (Adam, 2008) and became salient as we coded DMP segments and interview transcripts in the later stages analysis. It is not only that researchers lack information or even necessarily training on data management and data sharing, but that there exists a temporal misalignment in researcher's current practice-time profiles (end-loaded) missing in the discussion. Making data sharable requires that practice-time profiles be more integrated throughout the entire active phase of the research.

Although almost all researchers interviewed advised others, or their future selves, to do data management '*ahead of time*,' earlier does not automatically mean better, nor does it guarantee that delays won't occur. The emphasis on doing data management earlier surfaced repeatedly. A data manager at a large research centre expressed that they have seen researchers slowly come to recognize this and it is reflected in their data management practices. They said:

The best time to look at [research data management] is when your project is getting started. I love it when I get called to a meeting. They say, we're gonna do a whole bunch of these [simulation] runs, we're going to look at this aspect of the science, how best can we do this? Ask me at the beginning rather than wait until the project is done. (P13)

Researchers had to learn through their previous, often stressful, experiences that they should prepare their research data for sharing much earlier than when they expect to share that data.

Towards integrated practice-time profiles for data management

We found three ways that researchers are developing practice-time profiles for data management and sharing that suit their data, research aims, and workflows. In other words, the deliberate efforts made towards reconfiguring timing may result in developing integrated practice-time profiles for data management. The three examples differ in their approach, but what they share in common is that they are preventing an end-loading of data management and sharing work by integrating data management into the active phases of their research and specifically into their intermediary research steps, which are usually made invisible in final research outputs such as peer-reviewed publications and even in datasets. This is in line with pragmatic models of research data management, i.e. that there is no universal or ideal practice-time profile as this differs depending on the researcher, the project, and the data in question, but as noted one common factor is an orientation toward timing. That is, they are all ways to do the right data management work at the right time.

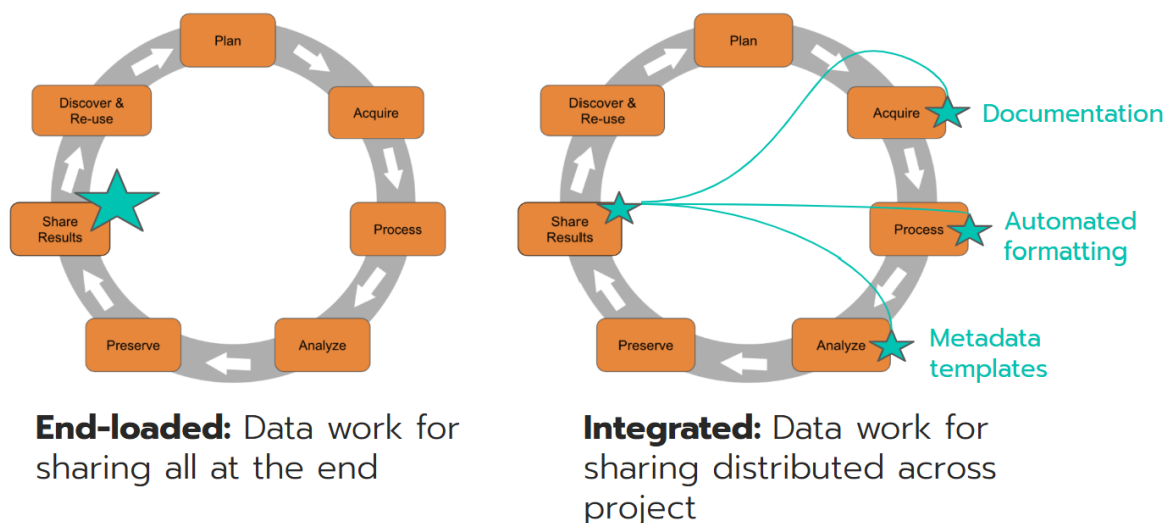


Figure 4. End-loaded and Integrated practice-times profiles for data sharing work with examples

Integrating data management into the intermediary steps of research

In reflecting on what they would do differently for their next project; several interviewees discussed the idea of thinking about the end goal of data sharing consistently throughout their project. As an interviewee explained, for their next project they planned to have the metadata templates, that many repositories provide, on hand as they analyse their data:

I would potentially make sure I had all of the like [metadata] templates and stuff even before I started processing my data, or running samples, if possible. So that I didn't have to reformat things a bunch of times. I just would prepare them as they needed to be for data sharing right out the gate. I think, that would just be a lot easier to keep everything organized, and I could also do more of the methodology reporting as I go. Rather than having to write it all at once. (P28)

In the example above, we see a shift in when data management is done, which suggests that creating metadata and documentation are becoming part of the researcher's workflow, rather than a separate task. In doing so, they redistribute the work and labour of data management and

sharing across the project's key phases, mitigating the time pressure that typically occurs when all this work is done at the end. As preparing to share data becomes habitual and integrated into the research process, activities, like reformatting data or preparing metadata, develops into a 'routine' one that happens during, not after, data collection and analysis (DMP 132).

Redistributing data management work through automation

Physical oceanographers who create ocean models are automating their data management and sharing workflows. One DMP noted that

We already have scripts in place to generate and submit sequence data to the [National Center for Biotechnology Information] NCBI Sequence Read Archive (SRA) in an automated process. Scripts written during the course of developing our bioinformatics pipeline will be made open source and available upon validation. (DMP 207)

For example, one PI, a self-proclaimed MATLAB user, explained that with the help of their doctoral students, they had *'developed some scripts that do exporting from the way we're working internally in MATLAB and exports in [Network Common Data Form] netCDF'* (P14), noting that the netCDF format is what public data repositories required. Another PI described a similar situation and explained to me that *'the data [our project] produces doesn't actually fit very well into that [referring to National Archives]. It feels not very useful [for us] ... so I wrote computer functions that would write those tables'* (P6). In this case, 'those tables' refer to additional information required by the repository's data dictionary. A third PI, also discussed adapting their workflow to meet repository requirements, sharing, *'this data is literally sitting in a data structure in MATLAB so I wrote a script in MATLAB that would make is so that NCDC likes it'* (P9).

Though using technical artifacts to automate work has long promised to save time, scripts take time to produce. A PI explained that *'it takes a lot of time at the beginning, but then the end is so much easier'* (P9). In other words, automation involves fronting labour that will make the onerous aspects of data management and sharing less taxing at the end. These practices redistribute time across the research process, transforming what would have been a final bottleneck into a manageable task and one that can potentially be used across multiple research projects.

Automating data management tasks may be a good solution for some PIs, but it is not a silver bullet. Not only does it take time to code scripts, but it may also take even more time and experience to know which aspect of data management to automate and to what end and many projects, particularly those that don't involve models, cannot be automated.

Anticipating future reuse in the present

Finally, some researchers are developing integrated practice-time profiles by anticipating the needs of future reusers during the active phases of their own research. Researchers are anticipating and integrating reusability into their data practices not only to reduce their own personal workload but also take into consideration the data practices of future reusers. For example, one evolving norm in sharing climate models is for researchers to share scripts and programs to run the research data alongside the data itself. In other words, reusable data is no longer just about data, but includes the scripts, software, or setup needed to run that data. As one oceanographer explained:

if the goal is to make the data easily accessible, you don't want the user to then struggle in writing their own program and code. It's a lot more efficient if you host the data and here's the script you need to run to access the data. (P8)

This example highlights two motivations for changing norms of data sharing in this context, efficiency through avoiding duplication of effort in writing programs and code and equity through making reuse easier for a broader group, not only those who can write their own programs and code. This quote also further shows that some researchers are orienting their present labour toward the needs of future reusers and pre-emptively integrating their imagined needs into their data practices. Integrated practice-time profiles enable researchers to configure their present-day data practices with future publics in mind.

Conclusion

The data management lifecycle needs revision to reflect the practices of scientists. Here we have shown that data management for sharing is not only constrained but actively shaped through time. While data policy and researchers often assume that data sharing will be an extension of existing data work, in practice, researchers' experiences reveal otherwise. Data management for sharing, instead, tends to follow end-loaded practice-time profiles, in which the data work for sharing is deferred until the final stages of a research project. The sunset stages of a project are also oftentimes when key personnel depart or when time is the most limited. In response, researchers are beginning to develop situated ways to manage this by integrating data management tasks into the active phases of their research. There is no ideal, or one-size-fits-all way to integrate data management for sharing into one's data practices, as this is shaped by the researcher, data, aims of the project, and where as well as how they plan to share their data. However, some recurring examples from this research includes, creating routines, automating workflows, and anticipating the needs of future data reusers. These 'integrated' practice-time profiles align better with the desired goal of data sharing as it is temporally distributing the labour and data work required to make shareable data across the research project instead of concentrating it at the end.

This research contributes to our understanding of research data management for sharing as a practice with distinct temporal characteristics. Drawing on Shove's (2009) concept of practice-time profiles, this research evinces a grounded understanding of research data management. By highlighting time and temporality, it argues that time is not just a background constraint but a central and contested factor shaping data management and sharing practices. The findings highlight the importance of considering time and temporality in data policy formation.

Rather than assuming a uniform progression from collection to sharing, data policies and institutional practices should explicitly acknowledge that managing data for public sharing introduces additional forms of labour that are not accounted for in traditional lifecycle models. These tasks are not mere extensions of research work but instead create new work that requires time, experience, and coordination. Strategies to address this could include funding mechanisms that support intermediary curation activities. Data repositories could encourage researchers' early engagement with data sharing requirements by familiarizing researchers with expectations such as file formats and required metadata. Moreover, data policies could clarify that public release denotes the end of data sharing, not the beginning of the work required to prepare data for sharing. Such measures would be a first step towards distributing data management and sharing work across a project's duration, reduce end-loaded practice-time profiles, and reshape the temporal organization of making data publicly available. They would also shift responsibility away from individual researchers exclusively and begin to support conditions in which the labour of making data shareable is supported across the research lifecycle.

Although our findings presented here are grounded in oceanography, the temporal tensions we identify may be transferable to other domains. For example, fields such as ecology, climate science, and geology may also rely on long-term data collection cycles, key personnel

responsible for data management work, and repository mandates that create similar end-loaded practice-time profiles. Other domains may exhibit comparable end-loaded practice-time profiles for different reasons. For example, human-subject data privacy concerns in the social sciences or intellectual property constraints in the materials sciences and related fields. Future work could therefore examine how researchers in different disciplines interpret data management policies and integrate the labour of data management and sharing during the active phases of a project. Such studies would help identify which of our findings are oceanography-specific and which reflect broader temporal organization in contemporary research data management and sharing.

Lastly, our findings also respond to growing calls within information science to pay attention to time and temporality (e.g. Haider et al., 2022). This research contributes to this body of work by showing how data management for sharing practices are temporally structured and distributed. By examining how researchers lack and create practice-time profiles, this research offers an empirically grounded understanding that time is not only background or context but is a dynamic and contested dimension of data management practice itself.

Acknowledgements

Our sincere thanks go to the oceanographers who shared their DMPs and participated in interviews for this research. We appreciate the thoughtful feedback received during the development of this project from Jaime Snyder and Ricardo Gomez, as well as the constructive input of three anonymous reviewers. Finally, we would like to thank Sarika Sharma, who contributed to the development of the initial DMP codebook during first cycle coding. This work was supported by the National Science Foundation grants 2020183 and 2429325.

About the authors

Yubing Tian is an independent researcher affiliated with American University, Washington, D.C., USA. She received her PhD from University of Washington's Information School, and her research interests include research data management and sharing, information infrastructures and the temporalities underlying them. She can be contacted at yubing.tian@gmail.com

Thomas Struett is a communication PhD student at American University. He is also the research director at the Digital Trade and Data Governance Hub located at George Washington University. His research has focused on data governance, folk theories of algorithms, and internet governance.

Amelia Acker is Associate Professor of Library and Information Science at Rutgers, The State University of New Jersey. She is the author of *Archiving Machines: From Punch Cards to Platforms* (MIT, 2025).

Megan Finn is an Associate Professor at the School of Communication at American University where she teaches about Media, Technology, and Democracy. She is also an Affiliate Associate Professor at the Information School at the University of Washington and the author of *Documenting Aftermath: Information Infrastructures in the Wake of Disasters*.

References

- Adam, B. (1998). *Timescapes of modernity: The environment & invisible hazards*. Routledge.
- Adam, B. (2008). OF TIMESCAPES, FUTURESAPES AND TIMEPRINTS.
- Baker, K. S., & Chandler, C. L. (2008). Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics. *Deep Sea Research Part II: Topical Studies in Oceanography*, 55(18–19), 2132–2142.
<https://doi.org/10.1016/j.dsr2.2008.05.009>
- Borgman, C. L., Golshan, M. S., Sands, A. E., Wallis, J. C., Cummings, R. L., Darch, P. T., & Randles, B. M. (2016). Data Management in the Long Tail: Science, Software, and Service. *International Journal of Digital Curation*, 11(1), 128–149.
<https://doi.org/10.2218/ijdc.v11i1.428>
- Charmaz, K. (2014). *Constructing grounded theory* (2nd edition). Sage.
- Dumit, J. (2004). *Picturing personhood: Brain scans and biomedical identity*. Princeton University Press.
- Haider, J., Johansson, V., & Hammarfelt, B. (2022). Time and temporality in library and information science. *Journal of Documentation*, 78(1), 1–17. <https://doi.org/10.1108/JD-09-2021-0171>
- Mosconi, G., Li, Q., Randall, D., Karasti, H., Tolmie, P., Barutzky, J., Korn, M., & Pipek, V. (2019). Three Gaps in Opening Science. *Computer Supported Cooperative Work (CSCW)*, 28(3–4), 749–789. <https://doi.org/10.1007/s10606-019-09354-z>
- NSF. (2011). PAPPG NSF 11-1 Grant Proposal Guide, Chapter II.C.2.j. U.S. National Science Foundation. https://nsf.gov-resources.nsf.gov/pubs/policydocs/pappguide/nsf11001/nsf11_1.pdf (Internet Archive)
- Ralph, N., Birks, M., & Chapman, Y. (2014). Contextual Positioning: Using Documents as Extant Data in Grounded Theory Research. *SAGE Open*, 4(3), 215824401455242.
<https://doi.org/10.1177/2158244014552425>
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed). SAGE.
- Shove, E., Trentmann, F., & Wilk, R. (Eds.). (2009). *Time, consumption and everyday life: Practice, materiality and culture*. Berg Publishers.
- Stahlman, G. R. (2022). From nostalgia to knowledge: Considering the personal dimensions of data lifecycles. *Journal of the Association for Information Science and Technology*, 73(12), 1692–1705. <https://doi.org/10.1002/asi.24687>
- Steinhardt, S. B., & Jackson, S. J. (2015). Anticipation Work: Cultivating Vision in Collective Practice. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 443–453. <https://doi.org/10.1145/2675133.2675298>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE*, 6(6), e21101.
<https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE*, 10(8), e0134826.
<https://doi.org/10.1371/journal.pone.0134826>

- Tracy, S. J. (2013). *Qualitative Research Methods: Collecting evidence, crafting analysis, communicating impact*. Wiley-Blackwell.
- Treloar, A., & Harboe-Ree, C. (2008). Data management and the curation continuum: How the Monash experience is informing repository relationships.

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).