"Data Management is a 'Sideshow' Pangs" or Data Management and Sharing Plans?

U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics
National Transportation Library

Transportation Research Board 103rd Annual Meeting
Washington, D.C., January 7-11, 2024
Poster: P24-20528
For AJE35 RIIM: Solicited Research Management and Innovation

The Abstract!

Does the phrase **"Data Management"** scare you? Are you feeling DMPs are a distraction rather than a vital part of the research process main event? Does DMP evaluation feel like a rigged carnival game, with no way to win? Does it feel like you have eaten too much cotton candy?

Have no Fear! The National Transportation Library Data Services Team is here!

Many people are still afraid of writing and evaluating Data Management Plans (DMPs). While others may treat DMPs as "just another form," we urge readers to see Data Management Plans as a vital part of Research Project Knowledge Management. DMPs can be utilized to communicate vital information to research team members, and serve as an important succession tool as the team changes. Properly implemented, a DMP can guide a team as they think through research data collection. This in turn, can lead to more accurate research budget requests, as data file size, repository costs, and data management time are estimated during the DMP writing process. Writing robust DMPs help give potential funders confidence in a researcher's ability to shepherd research resources, which can lead to more awards in the future.

The NTL team will introduce researchers and research managers to: the concepts of data management; a simple and useful template for DMP sections; the intersection of Data Management and Knowledge Management; the benefits of regular DMP review and updates; and, evaluating DMPs with confidence.

Don't just 'Hang Around'! Enter, Enter! Thrills a Minute!
We can teach to manage data **without** a net!
1. Data Management Plans Demystified!
2. See the Six Sections to Success!
3. Explore the Hidden Template!
4. Know Your Varietals: Different DMPs for Different Needs!
5. The Secrets of Data Repositories Revealed!
6. Meet the "Change Log"!
Cue the calliope!

1. Data Management Plans Demystified!

I see a Question from the Audience: What is a Data Management Plan? And should I be scared?
Answer: A **data management plan (DMP)** is simply a narrative document created during research proposal writing and planning to capture and record implicit team knowledge into an explicit document. In other words, a DMP is a **Knowledge Management** document! So there is nothing to be frightened about.

DMPs explicitly record team member roles and responsibilities that are vital for succession planning as teams evolve over the course of a research project. For example, a key researcher may win the lottery and suddenly retire (we should all be so lucky!). What were they doing on the project and who best can fill their roles? A well written DMP can help you answer the first part of that question and help to guide you as you answer the second part.

A robust DMP describes:

1. The team's plan for handling the raw and final dataset(s) generated during research;
2. The variables of interest, how they will be captured, and the file formats they are stored in; and,
3. How the research outputs will be stored, preserved, and shared.

A DMP can be any length need to capture the information that needs to be conveyed to the target audience (see "4. Know Your Varietals" for more). If the audience is the research team itself, the DMP should be quite detailed and as long as necessary. If the audience is a funding agency, the DMP can be shorter, as long as it gives funders the confidence that your team has a plan to preserve and share the data.

A good DMP, along with other documentation such as a data dictionary, helps future users understand the data. And since **the most likely future user of a dataset will be you** or the your organization, writing a DMP is being good to yourself.

For more guidance on DMPs see the NTL Research Data Management LibGuide:
https://transportation.libguides.com/researchdatamanagement/dmp

And we have a follow-up question: DMPs are just extra government paperwork and a nuisance!
Answer: Not at all my Friends. A DMP is a living knowledge management document created to help your team, your research program, and your research funder plan for the long life of your research data. A DMP should be reviewed on a regular basis, say at quarterly project meeting. Remember Friends, you and your organization are the most likely reuser of data you collect. So be good to yourself and remember to care for and feed your DMP as needed!

A question from the back row: I have heard them called "Data Management and Sharing Plans" or "DMSPs" recently? Are these different? What is going on?
Answers: Among U.S. federal research funding agencies, the National Institutes of Health (NIH) changed the name to "Data Management and Sharing Plans" with their **2023 Data Management and Share (DMS) Policy** https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan#after. The addition of the word "Sharing" helps to emphasize NIH's commitment to sharing research data as widely as possible, while also account legal, ethical, and technical restraints on sharing.

Other federal research funding agencies have already adopted the **DMSP** acronym, while others have not yet. But as federal agencies are trying to harmonize their requirements to reduce researcher burden, you can expect most others to follow suit. Just remember, **DMP = DMSP**, and use them interchangeably.

2. The Six Sections to Success!

Whether you are writing a DMP or are requiring others to write DMPs that you will review, you will find it beneficial to use a standard format. This will ensure that your DMSP meets the needs of nearly any funder. And if you specify these section in DMPs you will have to evaluate, you can make evaluation more efficient, as you will find the same types of information in the same places. This is especially useful if you are comparing proposed DMPs from competing research applicants. You can also create evaluation tools that help you generate quantitative scores for DMPs if you use and specify a consistent DMP layout.

Between 2013 and 2016, as we began to develop **Public Access** plans, many U.S. federal agencies adopted the DMP format recommended by the National Science Foundation (NSF). The U.S. DOT was among them. While DMPs may vary from funder to funder, you can expect to need to supply information for at least the following six sections.

- **Section 1. Dataset and Contact Information:** Plan to provide the name and contact information for the author of the DMSP, and/or the lead researcher for the project. Information such as contract number and contracting officer are useful if applicable. Don't forget to include each researcher's ORCID!
- **Section 2. Data Description:** You will need to write a narrative description of the expected final research data to be produced. This can include: the purpose of the research; the methods of data creation and data collection; scale and type(s) of data, such as numerical, image, video, source code, etc.; and/or, the reuse of data from other sources.
- **Section 3. Standards Employed:** Provide an explanation of standards used for data and metadata collection, format, and content. This can include: a discussion of whether open or proprietary file formats will be generated; a listing of the data documentation that will be included (think **Data Package**, and see https://transportation.libguides.com/researchdatamanagement/datapackages),
- **Section 4. Access Policies:** Record the policies for access and sharing the final research data, including provisions for appropriate protection of Personal privacy; Business confidentiality; National or State security; Intellectual property; and, Other rights or requirements. Also, describe how you will deidentify any personal or sensitive data, if applicable, in order to create a public use dataset.
- **Section 5. Reuse, Redistribution, and Derivative Products Policies:** Here you will name policies and provisions for reuse, redistribution, and the production of derivatives. Be sure to name the type of license assigned to the data, such as Public Domain or Creative Commons Attribution (CC-BY) [https://creativecommons.org/]. And remember, U.S. Copyright Laws do **not** allow the copyrighting of data, as they are a collection of facts.
- **Section 6. Archiving and Preservation Plans:** State your plans for archiving of, and long-term access to, final research data and publications. Things to include: The name and URL of the specific repository or repositories you are using for the research outputs; backup and disaster recovery plans; the data retention period; and whether the repository employs persistent identifiers (PIDs) for research outputs, people, and/or organizations.

For Section 6 bonus points, discuss how your chosen data repository meets the criteria outlined either on the **Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan** page http://ntl.bts.gov/publicaccess/evaluatingrepositories.html, or with the federal **Desirable Characteristics of Data Repositories for Federally Funded Research** at https://doi.org/10.5479/10088/113528.

Final Tip: Again, while not all DMPs have exactly these sections, some more, some less, and some DMPs have different names for the section, the information needed is much the same. You should provide enough information to give a funder confidence that you are planning to collect, preserve, and share research data ethically and within the guidelines of your contract or grant and the funder's data sharing policy.

3. Explore the Hidden Templates!

Friends! Do your **Data Management Plan** seem more like a **Didn't Manage It Plan?**

Are you worried about "funding loss" and what your colleagues might say?

Do you worry about **bitrot**?

Well Friends, the National Transportation Library (NTL) Data Services team is here to help!

That's right Friends, for the low, low price of free, we have created DMP writing guidance just for you.

Let me first direct your attention to the NTL **Research Data Management LibGuide** paged dedicated to DMPs located at: https://transportation.libguides.com/researchdatamanagement/dmp. Here we reveal excellent tips for writing you DMSP, and we include a **DMP Sufficiency Checklist** tool to see how your DMP scores.

Next I must point out the **US DOT Public Access Guidance** page **Creating Data Management Plans** at: https://doi.org/10.21949/1520562. This page includes detailed prompts to help you think about your data in ways you may never have thought of before.

You are asking "All of this for free?" Yes, Friends, But that is not all!

We also provide you access to example DMPs from your peers in the transportation research community through our Repository & Open Science Access Portal (ROSA P) https://rosap.ntl.bts.gov/

Just navigate to the **US DOT Public Access Data Management Plan Collection** at https://rosap.ntl.bts.gov/collection_pa_dmp and be astounded at the great work your peers have done, and from which you can borrow, learn, and adopt. Among polite company we call that "Technology Transfer" or "Knowledge Sharing." I know you will call that a good deal and time well spent.

4. Know Your Varietals: Different DMSPs for Different Needs!

Come and see; Come and see! Unique DMP varieties for your various needs!

Here we show you how to craft the data management plan your audience needs, depending on the knowledge you are trying to manage and share. We describe 6 types of DMPs for specific parts of the research project lifecycle, the intended audience, and the information most appropriate to that audience. And we provide a few examples!

Type of DMP	Audience	Description
Initial or Planning DMP	The Research Team.	This robust varietal DMP is the fundamental **Knowledge Management (KM)** document for your project data. Think through all of the data you may even think about capturing as you envision the research project. Include detailed descriptions team members and their roles (even if you don't know who will fill each role specifically); document data storage file paths and data access restrictions; etc.
Funding Proposal DMSP	The Funder.	This DMP can edited down to the information requested by the funder in their template or DMP guidance. Or it maybe quite a bit longer, because of the complexity of the data to be collected. The goal of this DMP is to convince the funder that your team will be good caretakers of funded data and have a plan for sharing the data.
In-process or Updated DMP	The Research Team.	This seldom-seen variety should reflect research realities now that research is underway. It may record that fact that there are now more data types or larger file sizes; or that you decided on a different more appropriate repository; etc. This DMP is an "in-process KM document" which is necessary because research conditions & teams change.
Final DMSP	The Funder, the Public, the Repository Manager, the Data Reuser.	Your most public and shared DMSP variety, it comes at the end of project. This DMP will reflects size and scope of final, publicly sharable dataset. It may describe how you anonymized certain variables to reduce the risk of reidentification or disclosure. It will certainly contain the link to the chosen repository and describe how the dataset be managed for long-term. This is a KN document for future data users or reusers. And remember, the future user may be you or member of your team!
Program-level DMP or Organizational Central DMP	Researchers and PIs.	A research organization can author a central DMP that describes the data management practices that all its researchers must follow. It may even direct researchers to use a specific repository for research outputs, such as a university's institutional repository. A major benefit of this variety is that when the Program-level DMP is posted on line, PIs can then reference sections of it in their project DMPs and save rewriting boilerplate language about chosen repositories or data back-up infrastructure.
The "No Data" DMP	Funders. .	This one is the rarest variety of all. Many federal public access policies state that research proposals must be contain a separate document labeled "Data Management Plan," even if your funded research will collect no data. The "No Data" DMP must explicitly state that you do not expect this project to generate data, and that is allowed. However, you should expect skeptical funders to ask questions about why no data will be collected. This variety is a real conversation starter

As you can see, there are a wide variety of DMPs to choose from, based on your needs and your audience.

Keep in mind that all DMSPs are living knowledge management documents and need occasional care and updating.

For more examples, see the **US DOT Public Access Data Management Plan Collection** at https://rosap.ntl.bts.gov/collection_pa_dmp.

5. The Secrets of Data Repositories!

Are you ready Friends? In our next astounding section we reveal ALL!

One section of every DMP describes your plans for the preservation of research data. Now you may be thinking "What do I know about long-term data preservation? Can't someone else do that?" Well, you are not expected to know a great deal about long-term data preservation, because there are data repositories designed to — and eager to — relieve you of the preservation burden!

And why is preservation a burden? While **Data Collection** may take months or years, **Data Preservation** may last decades. Data preservation is series of actions and interventions over the long tail of a dataset's lifespan, that may include, but are not limited to:
1. Having a 3-2-1 Data Backup Strategy: This means having 3 copies of the data, in 2 distinct geographic areas, and on at least 1 other media. Are you going travel around the country to check on your data?
2. Maintaining Robust Documentation: Besides a complete data package that contextualizes the data for future reuse, preservation and curation actions must also be tracked and added to the data package. These might include any of the actions below and others. This documentation can include metadata, READMEs, data dictionaries, code books, and other written instructions.
3. Bit Fixity Checking: Actively monitoring for storage media degradation, and using backups to restore corrupted files.
4. File Format Migration: Software formats change over time, and data files may need migration to the latest compatible file type for future use.
5. Data Disposition Decision-making: This means monitoring data re-use and usefulness to stakeholders over time. This can also mean the conscious, documented, decision to delete data that has outlived its usefulness to make room for new data.

Now, you as researchers don't likely have the time or the training for all of those action. The best practice in that case it to use a repository that actively preserves and curates data over its lifecycle. This means having a conversation with the repository, reading through their documentation, and possible paying for curation services. Remember: this is an allowable expense, but you must include it in your proposal.

"How do I choose a repository" you are asking. Look for a repository that: 1. allows for easy patron **access**; 2. has explicit **retention policies**; 3. has long-term **organizational sustainability**; 4. uses and records **persistent identifiers** for people, outputs, and organizations; 5. has robust and rich **metadata** describing each dataset; and, 6. has explicit **data security** policies and practices.

To help satisfy federal funder requirements refer to the **Desirable Characteristics of Data Repositories for Federally Funded Research** at https://doi.org/10.5479/10088/113528. You can also reference the NTL list of criteria at https://doi.org/10.21949/1520563 or the list of DOT conformant repositories at https://doi.org/10.21949/1520566. NTL Data Services team members are happy to consult with you as well. Just ask!

There are a few types of repositories you can choose from, and each has its benefits and shortcomings:

Generalist Research Data Repository	
Benefits	Shortcomings
. Breadth . Low Cost	. May be hard to find a single dataset among the many; . little to no curation; . outside of your institutional control.

Institutional Data Repository	
Benefits	Shortcomings
. Breadth . Local . Institutional cost sharing	. Curation may be limited by budget and staff skills

Domain-Specific Data Repository	
Benefits	Shortcomings
. Depth . Curation	. Extra cost of curation services; . outside your institutional control

Good Practices: Any repository is better than none. Check with your home institution first. Then look to research partners. Your best option may be to contract for repository services rather than own your own.

Pro tip: As federal agencies update their Public Access policies, they will no longer be able to allow funded-researchers to self-disseminate research data: Repositories will be required. Get to know one today!

6. Meet the Change Log!

We are coming to the ends Friends! But before we go, we want to introduce you to a vital, yet overlooked part of every DMP!

As we have mentioned a number of times, a Data Management & Sharing Plan is a living knowledge management document. But just as with datasets, the versioning and version control of documents can be tricky and confusion.

We have all been embarrassed at one time or another as we have opened a presentation or document that was not the most up to date version. What can we do? How do we prevent research team and funder confusion when we are expected to update our DMSPs?

Have no fear, the **Change Log** is here!

This simple practice can save you many headaches. At the end of each DMP, just add a small header with the words "Change Log". Then on each line below, summarize any changes you made to the DMP, and include the date on which those changes were made. Simple!

Want to see a Change Log in the wild? Just take a look at pages 14 and 15 of the **Plan to Increase Public Access to the Results of Federally-Funded Scientific Research Version 1.1**, also know as the **U.S. DOT Public Access Plan**, published in 2015. In section 9.1 Version, you can see our Change Log table:

9.1 Version

Revision Number	Revision Date	Revision Notes	Owner
1.0	24 Nov 2015		Charles Ducker
1.1	16 Dec 2015	• formatting changes • added cover page and header • corrected Section 7.4.2 to eliminate requirement for TRB to maintain searchable DMPs • corrected references to USDOT	Charles Ducker

As you can see, we also include a column for the name for the author of the changes. This can aid your organizational workflows tremendously.

But if you don't want to create a table, you can use the simpler version from a DMP we previously referenced from the 1977 TIUS dataset:

Change log:
2021-03-24: Initial DMP written

That's it! But that one little addition to your DMPs can save you a lot of bother as you update your DMPs, which we know you will.

This Way to the Egress!

You made it Friends!

Today we learned more than we ever thought we needed to know about Data Management Plans, their uses, their varieties, and some of their pieces and parts.

We hope we have alleviated your fears and given you the tools you need to write Data Management Plans without a net.

The Players

Leighton L Christiansen
https://orcid.org/0000-0002-0543-4268
Data Curator, National Transportation Library
leighton.christiansen@dot.gov

Jesse Ann Long
https://orcid.org/0000-0002-4962-1380
Data Management and Data Curation Fellow,
National Transportation Library
Jesse.long.ctr@dot.gov

Peyton Tvrdy
https://orcid.org/0000-0002-9720-4726
Data Management and Data Curation Fellow,
National Transportation Library
Peyton.Tvrdy.ctr@dot.gov

Tell Your Friends!

Christiansen, Leighton L; Tvrdy, Peyton; and Long, Jesse A. "Data Management is a 'Sideshow Pangs'" or Data Management and Sharing Plans. [2024]. [Presented at Transportation Research Board 103th Annual Meeting.] National Transportation Library, U.S. Department of Transportation. Washington, D.C., USA. https://doi.org/10.21949/1529900

Scan the QR code above to download and save this poster for future reference.