



26 May 2016

*Creating DMP Guidance for Iowa DOT Researchers*

*Leighton Christiansen (National Transportation Library), Vanessa Goetz (Iowa DOT), Brian Worrell (Iowa DOT), Akiema Buchanan (Interim Iowa DOT Librarian), Megan O'Donnell (Iowa State University)*

# Creating DMP Guidance for Iowa DOT Researchers

A Long & Winding Road...

Leighton Christiansen, Former Iowa DOT Librarian  
Vanessa Goetz, Iowa DOT Research Office  
Brian Worrell, Iowa DOT Research Office  
Megan O'Donnell, Iowa State University, Librarian  
Akiema Buchanan, Interim Iowa DOT Librarian

Transportation Librarians Roundtable  
May 26, 2016



# Agenda

- Guidance Creation
- Contract Language
- Our First DMP
- Repository Issues
- Future Plans
- Questions



In case you want to follow along...

- **Iowa Department of Transportation Office of Research & Analytics Guidance on Data Management Plan (DMP) Requirements for Federally Funded Research Projects**
- <http://publications.iowa.gov/id/eprint/21913>
  - Note: persistent identifier, will not change even as we update documents
  - Includes Guidance Document and Template
  - Now on Version 2
- <http://ntl.bts.gov/publicaccess/creatingaDMP.html>



# Needs Identified

- Carefully read through US DOT PAP
- Outreach to Iowa DOT Researchers and Contractors to Explain new Public Access Plan
- Choose a Repository
- Update Research Contract Language
- Create Guidance for Data Management Plans

# Contract Language/Process

- Contact Legal
- Contract Amendment
- Outreach needed
- Evolving process

Iowa Department of Transportation Office of Research & Analytics  
Data Management Plan (DMP) Requirements for Federally Funded Research Projects

Iowa DOT Project No: \_\_\_\_\_  
Name of Federal Funding Agency, Funding Program Name, Agency Code and/or Contract/Grant numbers: \_\_\_\_\_

#### Introduction

The Iowa Department of Transportation (Iowa DOT) Office of Research & Analytics (R&A), in compliance with the United States Department of Transportation (USDOT) Public Access Plan (PAP) "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research" requires researchers and research contractors receiving federal funding to meet the data management requirements of the USDOT PAP version current at the time of contract signing for the each project: see Attachment A. The intent of the US DOT PAP is to improve public access to the reports and final data sets created by federally funded research. The US DOT PAP plan and guidance documents can be read in detail at the US DOT Public Access webpages, available at <http://www.ntl.bts.gov/publicaccess/> as of March 1, 2016.

In order to guide data management efforts, each research project must be accompanied by a carefully considered data management plan (DMP). A complete DMP is required. This document explains the DMP requirements for Iowa DOT R&A research projects.

#### Definitions

"Scientific Research" is defined as activities comprising creative work undertaken on a systemic basis in order to increase the stock of knowledge, including knowledge of man, culture and society. (US DOT)

"Publications" are defined as any final peer-reviewed manuscript accepted for publication, any technical reports, final reports, summaries, and/or project written deliverable research funded, either fully or partially, by federal funds awarded through an Iowa DOT-managed contract, grant, or other agreement.

"Digital Data Sets," "Data," or "Data Sets" are defined as all scientific data collected through research projects funded, either fully or partially, by federal funds awarded through an Iowa DOT contract, grant or other agreement or collected by DOT employees or contractors. Such scientific data are the digitally recorded factual materials resulting from research that is necessary to validate research findings. These may include sensor data and/or video materials. If video data is present, it must be secured to prevent disclosure or personally identifiable information. These do not include laboratory notebooks, partial datasets, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as gels or laboratory specimens.

"Metadata" is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. (NISO)



# Contract Language Special Notes

- **Publication Management:** All final reports, tech transfer summaries, and other textual products of research shall be **deposited in an appropriate digital repository** that makes use of **persistent links** and has a plan for long-term preservation of electronic text materials.
- **Note:** Project PIs are responsible for ensuring all final reports, tech transfer summaries, and other textual products are publically accessible for a period of ten [10] years from the end of the contract period.
- **Note:** Project PIs are responsible for ensuring all final datasets are publically accessible for a period of ten [10] years from the end of the contract period.

# Guidance: Shameless Borrowing

United States Department of Transportation  
 About DOT | Briefing Room | Our Activities

OFFICE OF THE ASSISTANT SECRETARY FOR RESEARCH AND TECHNOLOGY  
 About OST-R | Press Room | Programs | OST-R Publications | Library | Contact Us

National Transportation Library

OST-R > BTS > NTL > Public Access

## Creating Data Management Plans

### Data Management Plans (DMPs) Content Overview

DMPs will describe how the research proposal conforms to DOT policy on the dissemination and sharing of research results in a narrative description covering:

- The final research data to be produced in the course of the project;
- The standards to be used for data and metadata format and content;
- Policies for access and sharing the final research data, including provisions for appropriate protection of privacy, confidentiality, and other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives; and
- Plans for archiving final research data and other research products, and for preservation of access to them.

### Guidance on DMP Sections

**Description of the Data:** Provide a description of the data that you will be gathering in the course of your project. You should describe the scale of the data that will be collected. Describe the characteristics of the data, their relationship to other data, and provide sufficient detail to understand any disclosure risks that may apply. Discuss value of the data over the long-term.

**Standards to be Used:** Your DMP should describe the anticipated formats that your data and related files will use. To the maximum extent possible, use generally accepted practices in your field. Your DMP should address how you will use platform-independent formats to ensure maximum utility of the data in the future. If you are unable to use platform-independent and non-proprietary formats, you should describe the formats that will be used and the rationale for using those standards and formats.

**Policies for Access:** Protecting research participants and guarding against the disclosure of identities and/or confidential business information in scientific research. Your DMP should address these issues and outline the efforts you will take to provide informed consent forms will permit sharing with the research community and whether additional steps, such as an Institutional Review Board, are necessary to protect privacy and confidentiality.

**Policies for Re-Use, Re-Distribution, and Derivative Products:** Describe who will hold the intellectual property rights for the data. Describe whether you will transfer those rights to a data archive, if appropriate. Identify whether any copyrights apply to the data.

<http://ntl.bts.gov/publicaccess/creatingaDMP.html>

Iowa Library Services  
 State Library of Iowa  
 Iowa Publications Online

<http://publications.iowa.gov/id/eprint/21913>

### Data Management Plan

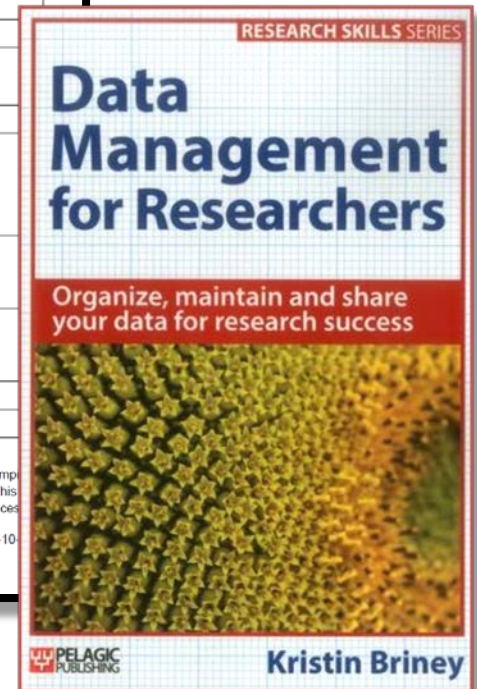
Name of Contractor	
Name of project	
Project Duration	Start date : End:
DMP Version	
Date Amended, if any	
Name of all authors, and ORCID number for each author	
WYDOT Project Number	
Any Digital Object Identifier (DOI), including any CROSSREF number, which has been assigned to any peer reviewed publication or data generated by this project	
Name of all peer reviewed publications which have been generated using data from this project	
URLs for all peer reviewed publications which have been generated using data from this project	
RIP RH Display ID Number	
Dataset URL, if available	

What constitutes such data will be determined by the Principle Investigator, Project Champion and the Research Manager. In general, your plan should address final research data. This includes recorded factual material commonly accepted in the scientific community as essential to the research.

1

Rev. 2-10

Attachment E  
 InterAgency Agreement between Wyoming Department of Transportation  
 And the University of Wyoming





## Data Management Plan (DMP) for Iowa DOT Research Projects

### Section 1: Research Data Management Documentation Table

Name of Principle Investigator(s) or Contractor(s) and ORCID number		
Current Project Title, and all previous project titles		
Iowa DOT Project Manager		
Iowa DOT Project number		
Other contract or grant numbers		
Iowa DOT Research-assigned project Digital Object Identifier (DOI), or researcher acquired DOI		
TRB Research in Progress (RiP) Title, Accession Number, and URL		
Project Duration (projected)	Start Date:	End date:
Do the data management requirements of the US DOT "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research" apply to this project	Yes or No; and if No, why not:	
Name(s) of Federal funder(s), Funding Program Name(s), Agency Code(s) and/or Contract/Grant numbers		
DMP Version		
Date DMP amended, if any		
Name and ORCID number of each author		
Persistent links or identifiers assigned to this project, datasets, reports, or peer reviewed publications generated by this project		
Name and URL of all peer reviewed publications which have been generated from this project		

This table is to be filled out as completely as possible before the beginning of the project, and updated as needed, including at the end of the of the project, and after, as derivative publications are created.

[Note: For ease of copy and paste, a Word .docx template of this table is available at <http://publications.iowa.gov/id/eprint/21913> ]

# Guidance & Template: Section 1 Research Data Management Documentation Table



# Entering DMP Information into RiP

**Research in Progress**

RiP > LiveEdit > LiveEdit Intrans\_IADot

To edit the record below, click on any highlighted field. Tab and Shift+Tab may be used to move between fields. When you are finished editing the field, click Save. \* denotes required fields

Title \*

Translated Title

Subject to Public Access Plan compliance

Subject to Public Access  ←

Plan compliance \*

Save  Cancel

Managing Organizations \*  ←

Project Managers - select -

Performing Organizations \*  ←

Principal Investigators \*  ←

USDOT Program \*  ←

Managing Organizations - select -

Project Managers - select -

Performing Organizations - select -

Principal Investigators - select -

USDOT Program - select -

USDOT Sub-Program - select -

USDOT Program \*

USDOT Sub-Program

Index Terms

Subject Areas \*

- Advanced Materials/Structural Safety
- Advanced Research
- Aeromedical Research
- Air Traffic Control/Technical Operations Human Factors
- Aircraft Catastrophic Failure Prevention Research
- Aircraft Icing/Digital Systems Safety
- Airport Cooperative Research Program

<b>Do the data management requirements of the US DOT "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research" apply to this project</b>	Yes or No; an
<b>Name(s) of Federal funder(s), Funding Program Name(s), Agency Code(s) and/or Contract/Grant numbers</b>	
<b>DMP Version</b>	





# Guidance & Template: Section 2 Description of the Data

**Section 2: Description of the Data:** Provide a description of the data that you will be gathering in the course of your project. You should address the nature, scope, and size of the data that will be collected. Describe the characteristics of the data, their relationship to other data, and provide sufficient detail so that reviewers will understand any disclosure risks that may apply. Discuss value of the data over the long-term.

*Possible questions to answer in this section of your DMP:*

- What type of data will be produced? (Tabular, sensor, video, audio, etc.)
- How will data be collected? In what formats? (.txt, .csv, .tiff, etc.)
- Are there special tools or software needed to create/process/visualize the data?
- How will the data collection be documented?
- What project and data identifiers will be assigned?
- Will the data collected be unique or will the data be reproducible? What would happen if the data got lost or became unusable later?
- How much data will there be, and at what growth rate? (1 GB, 3 TB etc.) How often will it change?
- Will you use pre-existing data? If so, from where?
- How will you store, backup, and protect data from lost during the research project?
- Who will potentially use the data?
- Who is responsible for managing the data?
- What value does the data have over the long-term? (Please consider not only your research team, but third parties as well.)



# Guidance & Template: Section 3 Standards to be Used

**Section 3: Standards to be Used:** Your DMP should describe the anticipated formats that your data and related files will use. To the maximum extent practicable, and in accordance with generally accepted practices in your field, your DMP should address how you will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future. If you are unable to use platform-independent and non-proprietary formats, you should specify the standards and formats that will be used and the rationale for using those standards and formats.

*Possible questions to answer in this section of your DMP:*

- What formats will the data be in? (.txt, .csv, .tiff, etc.)
- Are these formats open or proprietary? If proprietary, what is the rationale for using that format?
- What standards will be used for documentation and metadata?
- What directory and file naming convention will be used?
- What documentation or descriptive metadata will you be creating in order to contextualize the data for future users?



# Guidance & Template: Section 4

## Data Organization and Description

**Section 4: Data Organization and Description:** Your DMP should list how you will organize, document, and describe the data. Descriptive metadata is vital to contextualize the dataset for future data users, including the original data creator. Descriptive metadata should be written following the rules and format of a published metadata schema appropriate to the type of data or to the research discipline.

*Possible questions to answer in this section of your DMP:*

- What are the file naming conventions to be employed?
- How will the data be organized?
- What metadata schemas are appropriate for describing these types of data?
- What metadata schema will be chosen for this data?



# Guidance & Template: Section 5 Policies for Access

**Section 5: Policies for Access:** Protecting research participants and guarding against the disclosure of identities and/or confidential business information is an essential norm in scientific research. Your DMP should address these issues and outline the efforts you will take to provide informed consent statements to participants, the steps you will take to protect privacy and confidentiality prior to archiving your data, and any additional concerns. If necessary, describe any division of responsibilities for stewarding and protecting the data among Principal Investigators.

If you will not be able to deidentify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, you should describe the necessary restrictions on access and use. In general, in matters of human subject research, your DMP should describe how your informed consent forms will permit sharing with the research community and whether additional steps, such as an Institutional Review Board (IRB), may be used to protect privacy and confidentiality.

*Possible questions to answer in this section of your DMP:*

- Does the data contain any personally identifiable information (PII)?
- How will you anonymize or deidentify the data if PII is present?
- What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights?
- Does your data have any access concerns? Describe the process someone would take to access your data.
- Who controls the data (e.g., funder, PI, student, lab, University)?
- Are there any special privacy or security requirements (e.g., personal data, high-security data)?
- Are there any embargo periods to uphold?



# Guidance & Template: Section 6

## Policies for Re-Use, Re-Distribution, and Derivative Products

**Section 6: Policies for Re-Use, Re-Distribution, and Derivative Products:** Describe who will hold the intellectual property rights for the data created by your project. Describe whether you will transfer those rights to a data archive, if appropriate. Identify whether any copyrights apply to the data, as might be the case when using copyrighted instruments. If you will be enforcing terms of use or a requirement for data citation through a license, indicate the license type in your DMP. Describe any other legal requirements that might need to be addressed.

*Possible questions to answer in this section of your DMP:*

- If you allow others to reuse your data, how will the data be discovered and shared?
- Any sharing requirements (e.g., funder data sharing policy)?
- What license type is being used (e.g.: Creative Commons 0, etc.)
- Who might be the audience for data reuse? Who will use the data now? Later?
- When will the data be published and where?
- What special tools and/or software are needed to work with data?



# Guidance & Template: Section 7 Plans for Archiving and Preservation

**Section 7: Plans for Archiving and Preservation:** Describe how you intend to archive your data and why you have chosen that particular option. You may select from a variety of options including, but not limited to:

- Use of an institutional repository;
- Use of an archive or other community-accepted data storage facility;
- Self-dissemination

You must describe the dataset that is being archived with an appropriate amount of metadata that ensures its discoverability. Whatever archive option you choose, that archive must support the capture and provision of the US Federal Government "Project Open Data Metadata Schema" metadata current at the time of contract signing. ["Project Open Data Metadata Schema" was available at <https://project-open-data.cio.gov/v1.1/schema/> as of March 1, 2016.] In addition, the archive you choose must support the creation and maintenance of persistent identifiers and must provide for maintenance of those identifiers throughout the preservation lifecycle of the data. Your plan should address how your archiving and preservation choices meet these requirements.

*Possible questions to answer in this section of your DMP:*

- What archive will the data be stored in and why was it chosen?
- What is the persistent identifier type used by the archive?
- How will the data be archived for preservation and long-term access?
- How long should it be retained (e.g., 3-5 years, 10-20 years, permanently)?
- What file formats? Are they long-lived?
- Are there data archives that my data is appropriate for (subject-based? Or institutional)?
- Who will maintain my data for the long-term?





# Data Management Plan (DMP) for Iowa DOT Research Projects

## Research Data Management Documentation Table

Name of Principle Investigator(s) or Contractor(s) and ORCID number	REDACTED (SUPPLIED & REDACTED) REDACTED REDACTED (SUPPLIED & REDACTED)	
Current Project Title, and all previous project titles	SUPPLIED & REDACTED	
Iowa DOT Project Manager	REDACTED	
Iowa DOT Project number	SUPPLIED & REDACTED	
Other contract or grant numbers	N/A	
Iowa DOT Research-assigned project Digital Object Identifier (DOI), or researcher acquired DOI	N/A	
TRB Research in Progress (RiP) Title, Accession Number, and URL	N/A	
Project Duration (projected)	Start Date: SUPPLIED & REDACTED	End date: SUPPLIED & REDACTED
Do the data management requirements of the US DOT "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research" apply to this project	Yes or No; and if No, why not: Yes	
Name(s) of Federal funder(s), Funding Program Name(s), Agency Code(s) and/or Contract/Grant numbers	Federal Highway Administration	
DMP Version	1.0	
Date DMP amended, if any	N/A	
Name and ORCID number of each author	REDACTED (SUPPLIED & REDACTED) REDACTED REDACTED (SUPPLIED & REDACTED)	
Persistent links or identifiers assigned to this project, datasets, reports, or peer reviewed publications generated by this project	N/A	
Name and URL of all peer reviewed publications which have been generated from this project	N/A	

# Our First DMP: Section 1



## Missing:

- ORCID for author #2

This table is to be filled out as completely as possible before the beginning of the project, and updated as needed, including at the end of the of the project, and after, as derivative publications are created.



# Our First DMP:

## Section 2 Description of the Data

One of the project tasks is structural testing of concrete slabs overlaid by Ultra-High Performance Concrete (UHPC) material. From this testing, **two sets of data** will be generated that which include information such as load, displacement, and strain . This data will be used to **quantify the performance** of the UHPC overlay, in particular its bond strength at the interface with the concrete slabs.

### Missing:

- Data's "value"
- Audience
- Why two sets?
- Retention length



# Our First DMP: Section 3 Standards to be Used

The data will be stored in **ASCII delimited text format** . This is a platform-independent and non-proprietary format.

## **Missing:**

- File formats (not encoding)
- Standards info



# Our First DMP:

## Section 4 Data Organization and Description

Along with the datasets, a **metadata file** that contains the test setup, location of the sensors, naming scheme, and any relevant information will be generated. Each of the **data file** will consist of a brief explanation of what it contains.

### Missing:

- File names/organization
- Schema info



# Our First DMP: Section 5 Policies for Access

The nature of the data collected in this research is **not private nor confidential**. The obtained data will be made available for public use.

## Missing:

- When?
- Embargos?
- Any usage concerns.



# Our First DMP:

## Section 6 Policies for Re-Use, Re-Distribution, and Derivative Products

The intellectual property rights of the data will be held by Iowa State University. However, permission to distribute data will be given to appropriate personnel and data archiving centers.

### Missing:

- Reuse conditions
- Licenses
- Who's responsibility?
- Discoverability



# Our First DMP:

## Section 7 Plans for Archiving and Preservation

During the project period, the data collected will be stored in an online data archival system called CyBox, which is backed up daily. This is a secured data storage within the Iowa State University and the main server is located in the US. In the long term, the data will be curated and archived in the National Transportation Library (NTL) data archival system.



≠ preservation

### Missing:

- Which data?
- Repository
- Persistent identifiers
- Length of retention
- Responsibility



# Choosing a Repository

1. Harvard Dataverse
2. Zenodo (CERN! Will states allow storage in Europe?)
3. Odum Institute (Dataverse install. Could move up following summer 2016 upgrade.)
4. figshare (Not yet on NTL's list of conformant repositories.)
5. Dryad (Journal focused, not friendly to technical reports)
6. Inter-university Consortium for Political and Social Research (ICPSR) (Should be number 1. However, transportation research declared out-of-scope by ICPSR staff)





# Choosing a Repository: Toxic Terms of Service

From Harvard Dataverse “General Terms of Use”

<http://dataverse.org/best-practices/harvard-dataverse-general-terms-use>

## **Indemnification**

“You will indemnify and hold Harvard Dataverse harmless from and against any and all loss, cost, expense, liability, or damage, including, without limitation, all reasonable attorneys’ fees and court costs, arising from i) your use or misuse of the Service; (ii) your access to the Site...”

**Problem:** This clause violates Iowa state law and therefore we cannot use Harvard Dataverse as chosen repository!

(Other governmental bodies will have similar clauses...)

# Future Steps

- Outreach and Teach!
- Study Repository Terms of Service
- Select short-term repository
- Look for long-term solution