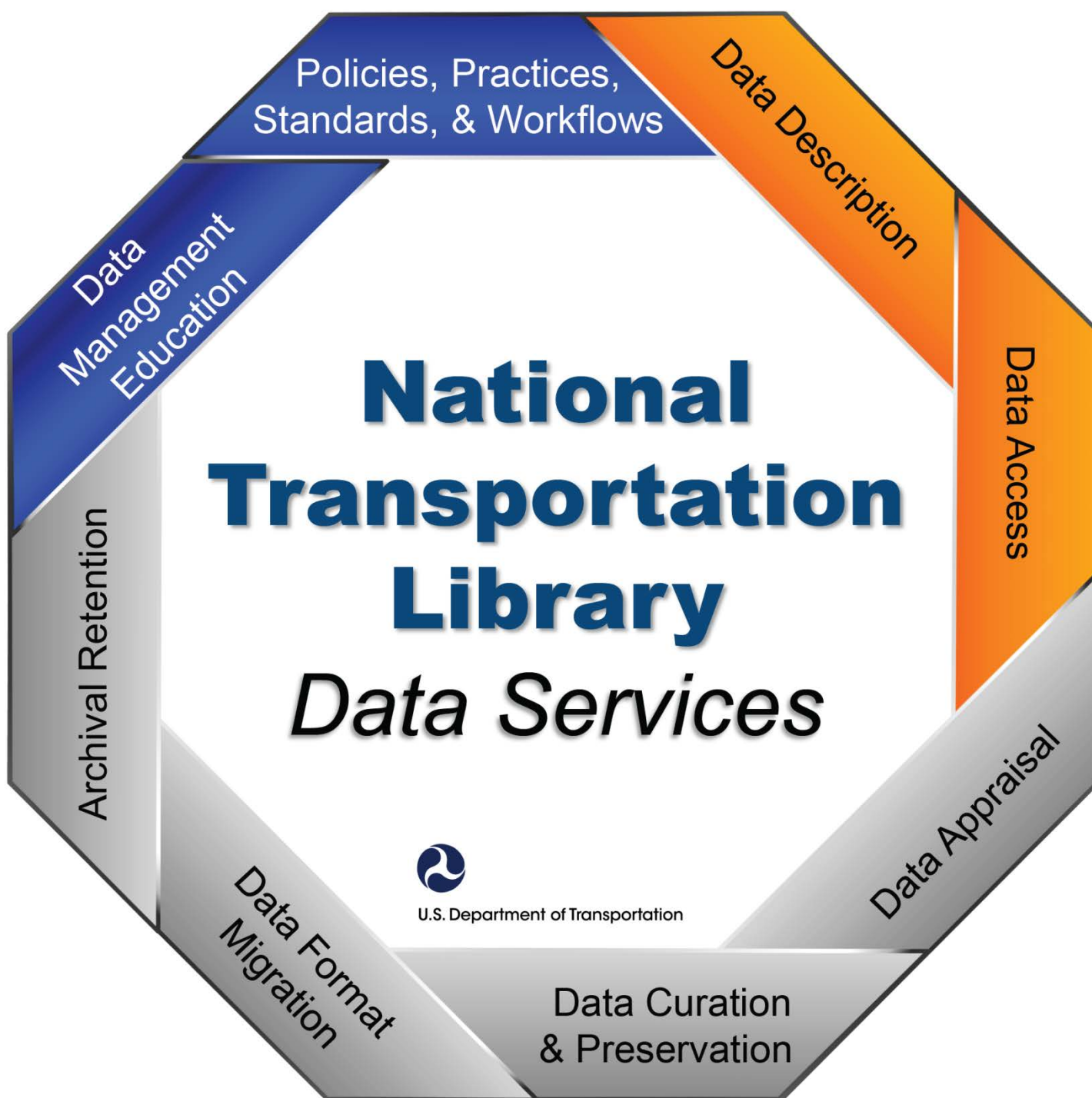




Bureau of Transportation of Statistics Data Services & Data Curation

Leighton L Christiansen [0000-0002-0534-4268](https://www.transportation.gov/0000-0002-0534-4268)
Data Curator, National Transportation Library
ntldatacurator@dot.gov





Some Key Terms

- **Data Management¹**
 - deliberate planning, creation, storage, access and preservation of data produced from a given investigation¹
- **Data Curation²**
 - enables data discovery and retrieval, maintains data quality, adds value, and provides for re-use over time²
- **Data Science³**
 - drawing useful conclusions from large and diverse data sets through exploration, prediction, and inference³



Linked Processes

Data Management \in Data Curation

Data Curation \Rightarrow Data Science



U.S. DOT Public Access Requirements

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

Researchers and authors are subject to U.S. DOT Public Access requirements if:

- U.S. DOT funded scientific research;
- Newly funded or extended on or after January 1, 2016.

Researchers and authors will need to submit:

- ORCiDs
- 2 – 3 page Data Management Plan (with proposal);
- Final peer-reviewed manuscript;
- Final Digital Datasets; and,
- Any other written outputs (final reports, technical reports, tech summaries, etc.)



Public Access Guidance

http://ntl.bts.gov/publicaccess/creatingaDMP_extramural.html

United States Department of Transportation

Office of the Assistant Secretary for Research and Technology | About OST-R | Press Room | Programs | OST-R Publications | Library | Contact Us

National Transportation Library

About NTL | Public Access | Networking | Research/Tools | FAQs | Ask-a-Librarian | Site Map

OST-R > BTS > NTL > Public Access

Printable Version

Creating Data Management Plans for Extramural Research

This page serves to provide guidance for how to prepare a Data Management Plan (DMP). The guidance outlined here does not constitute an approved government form or template. Those preparing DMPs for submission to the U.S. Department of Transportation (USDOT) should use their best judgment in determining what information to include. USDOT has identified five (5) broad areas that should be addressed in a DMP, but is not requiring any specific information to be included in any submitted DMP. USDOT may, at its discretion, establish an Office of Management and Budget-approved information collection. Once approved, the information collection will become a form with a control number, and certain DMP elements may become mandatory.

Data Management Plans (DMPs) Content Overview

A data management plan (DMP) describes how researchers will handle digital data both during and after a research project. DMPs will describe how the research process will generate, collect, store, manage, share, and preserve data. A DMP should also describe how the data will be used, distributed, and archived.

5 Sections in DMP

- Data Description
- Standards Used
- Access Policies
- Re-Use, Redistribution, and Derivative Products Policies
- Archiving and Preservation Plans

Standards Used: 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. Identify the metadata standards you will use to describe the data.

Data Services:

Education & Cheerleading

- **PA Compliance Guidance**
- **Data Management Planning training & consulting**
- **DMP Sufficiency Checklist and Training**
- **Data Citation Recommendations**
- **Best Practice sharing**

Future Actions

- **Data Curation & Preservation**
- **FAIR Data Principles**
- **Repository**





DATA MANAGEMENT PLANNING “GREENOUT” BINGO SELF-ASSESSMENT TOOL

This tool is designed to help assess your current data management practices. Pick a dataset or research project, sit down with your data collection team, and discuss each prompt below. This tool can guide DMP improvement by revealing best practices that you are not using or information that you may not have gathered.

Dataset Name: _____

Evaluators: _____

Date: _____



Red: Not current practice, Unknown, or No.

Yellow: We are planning on this or are working on it.

Green: Practice enacted, or Yes.

Blue: Not Applicable.

Data Description	Standards Employed	Access Policies	Re-use & Derivative Policies	Archiving & Preservation Plans
Dataset is named, well described, and linked to a research project or data program.	If data is created in proprietary formats, copies in open formats are also provided.	Data is publicly accessible.	The intellectual property rights to this dataset have been established.	A data repository dedicated to long-term preservation has been chosen for this data.
The types of data generated by the research or program are known.	Data formats are standard for our field.	What data will be shared, and how, is known and recorded.	Data is licensed under the most open license possible, such as Public Domain or Creative Commons 0.	A minimum data retention period has been agreed upon with chosen archive.
The size of data is known and organization has capacity for files.	Directory and file naming conventions are documented & used at time of data collection.	Researchers have received training to protect PII and other rights.	If there is a data embargo period, it is as short as possible.	Persistent identifiers (such as DOIs) are used to link to the data.
Long-term value of the data to organization and public has been discussed.	Metadata is used to contextualize the data, making it comprehensible to us and others.	Personally Identifiable Information (PII) is protected or anonymized.	Possible reuse audiences for this data, now and later, have been considered.	Back-up and disaster recovery plans are in place.
Parties responsible for managing data are known and listed.	Published metadata schemas are employed, & are standard to the field.	Embargo periods to protect PII and business sensitive information have been established.	Special tools needed to work with the data are documented in metadata.	Staff has been assigned to migrate data files as data formats change over time.

In-House Training

DMP Self-Assessment

- Gives visual feedback on current state of DMP best practices
- Tested with each BTS office
- Used to begin DMP process for 2 BTS datasets
 - Transborder Freight
 - State Transportation Stats



2. Data Description:

Data Description Narrative Evaluation Prompts		Explained Fully	Partially Explained	No information	Not applicable
2.01	The DMP names the data, data collection project, or data producing program.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.02	The DMP describes the purpose of the research or data collection.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.03	The DMP describes the data generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.04	The DMP describe methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; digital data such as images and video; etc.).	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.05	The DMP discusses the period of time data will be collected and frequency of update.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2.06	The DMP describes the relationship between the new data collected for this effort and any existing data also used..	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2.07	The DMP lists potential users of the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.08	The DMP discusses the potential value the data have over the long-term for not only U.S. DOT, but also for the public.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2.09	If the DMP contains a request permission to not make the data publicly accessible, it explains the rationale for lack of public access.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2.10	The DMP indicates the party responsible for managing the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.11	The DMP describes how project leads will check for adherence to this data management plan.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Total of checked boxes for each column, out of 11:		6	1	2	2

Evaluation questions:

- Did a majority of the prompts rate “Explained Fully”?
- Do you have a complete picture of: What the data will be gathered; How much data to expect; and, Who is responsible for managing data, and how the data will be managed?

If not, this section may not be sufficiently detailed.

In-House Training

DMP Sufficiency Evaluation

- Gives quantitative feedback to bolster subjective judgement
- Precise feedback to researcher on what is missing
- Better DMPs improve likelihood of long-term data preservation for public re-use



Data on the Web Best Practices⁴

Data Description & Standards

Metadata
Data Licenses
Data Provenance
Data Quality
Data Versioning
Data Identifiers

Access Policies

Data Formats
Data Vocabularies
Data Access
Data Access APIs

Data Reuse & Derivatives

Feedback
Data Enrichment
Republication

Archiving & Preservation

Data Preservation

Benefits



Reuse



Trust



Comprehension



Access



Linkability



Interoperability



Discoverability



Processability

NTL's Repository & Open Science Access Portal (*Rosa P*)

United States Department of Transportation

National Transportation Library

Home Collections Recent Additions Public Access Submit Content About ROSA P

rosap Repository & Open Science Access Portal

All Collections Enter keyword or phrase... Search Advanced Search

About ROSA P

ROSA P Collection Content Types	
	59.0%
	14.9%
	23.6%
	0.4%
	0.4%
Other	1.6%

Founded as an all-digital library program, NTL's collections in ROSA P are full-text electronic publications, datasets, and other resources. The repository is also designated as the full-text repository for USDOT-funded research under the USDOT Public Access Plan. NTL provides its collections freely to transportation researchers, statistical organizations, the media, and the general public.

Collection Overview

To meet the requirements outlined in its legislative mandate, NTL collects resources across the transportation domain, specifically focused on information from USDOT, state DOTs, and other transportation organizations. Content types found in ROSA P include text, links to websites, datasets, images, video, other multimedia, and maps. These resources have value to federal, state, and local transportation decision makers, transportation analysts, and researchers. As of December 2016, the repository has approximately 55,000 records and serves as the full-text archive for the transportation research ecosystem.

All resources in ROSA P are in the public domain and/or explicit permission has been provided by the rights holder to NTL to make their materials available for free over the web.

For more information, please see the [ROSA P Collection Development Policy](#)

Persistence Statement

All resources deposited into ROSA P will be assigned a [digital object identifier](#), or DOI, as a persistent identifier. This globally unique link will always lead users to a landing page containing the resource and its metadata or a description documenting the alteration and/or destruction of the information source, if applicable. The DOI will always resolve to a landing page with metadata describing the curation of the resource, even if the resource itself is removed. NTL's plan for organizational persistence and succession can be located in its [Digital Curation Policy](#)

Immediate Steps

- Launch in April 2017
- Based on the CDC Public Access platform, which hosts the *CDC Stacks* and *NOAA Institutional Repository* digital libraries
- AWS cloud storage
- Publications & Data
- DOIs

Future Steps

- Trusted repository status



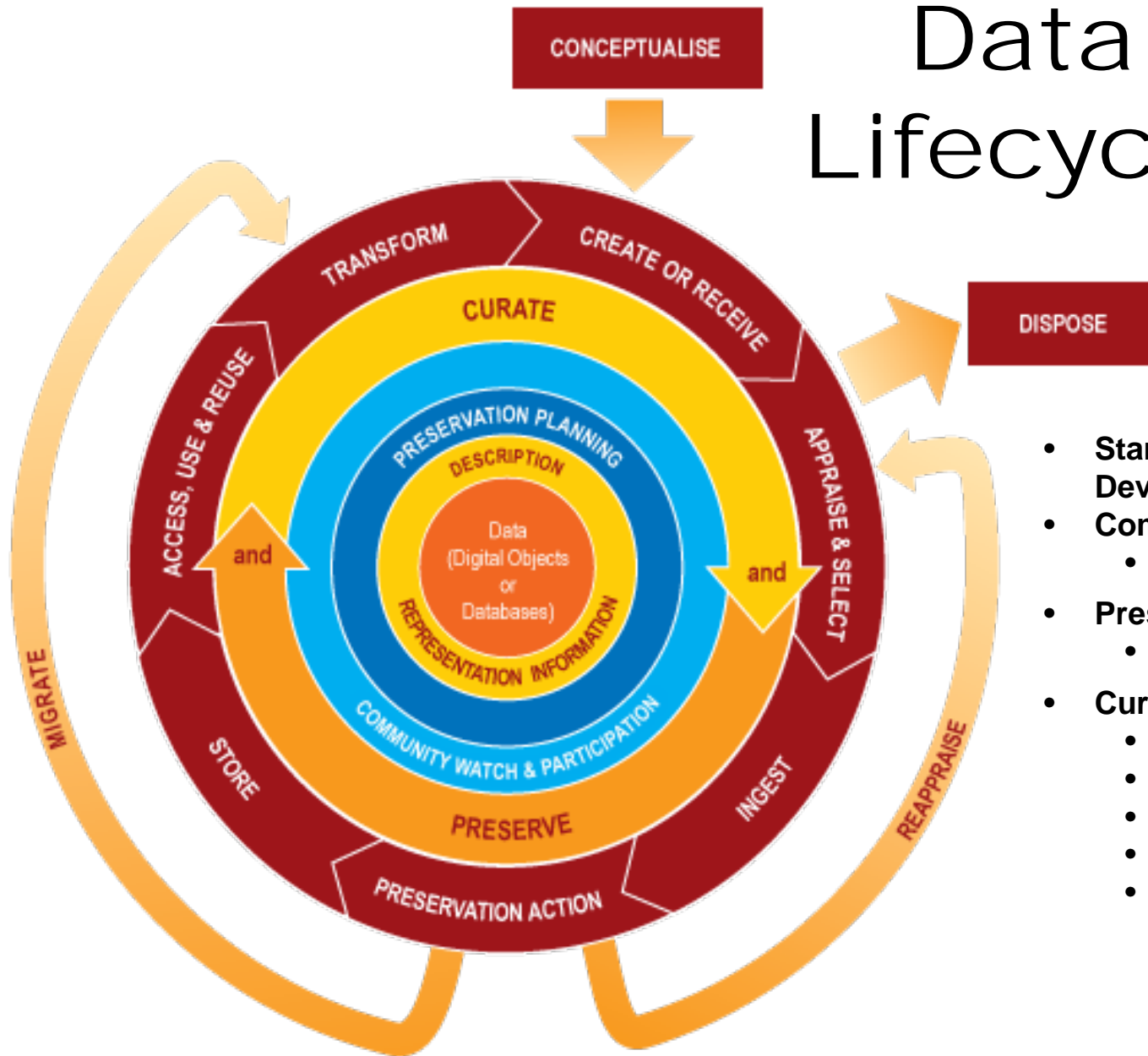
Trusted Repository Status & Data Curation⁵

Data Seal of Approval – World Data Systems Repository Audit and Certification

- 1. The data can be found on the Internet**
- 2. The data are accessible (clear rights and licences)**
- 3. The data are in a usable format**
- 4. The data are reliable**
- 5. The data are identified in a unique and persistent way
so that they can be referred to**



Data Curation Lifecycle Model⁶



- Standards & Workflow Development & Implementation
- Complete Description
 - Metadata Standards
- Preservation Planning
 - DM Training
- Curation & Preservation
 - Access & Reuse
 - Repository Ingest
 - Preservation
 - Migration
 - Disposition



Benefits of Data Curation

- Protects Unique Data from Loss
- Improves Data Search & Retrieval
- Enables Reuse
- Facilitates Longitudinal and/or Meta Analyses
- Avoids Duplication of Effort & Spending
- Increases Verifiability
- Opens New Lines of Scientific Discovery
- Satisfies Public Access & Open Government & Legal Requirements



Questions?

ntldatacurator@dot.gov



References

1. University Library, Texas A&M University. “Data Management Defined- Research Data Management - Guides at Texas A&M University.” Research Data Management, October 1, 2013. <http://guides.library.tamu.edu/DataManagement>
2. Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. “Specialization in Data Curation,” 2013. http://www.lis.illinois.edu/academics/programs/specializations/data_curation
3. Definition based on Ani Adhikari and John DeNero, “The Foundations of Data Science” <http://www.inferentialthinking.com/index.html> “What is Data Science” <http://www.inferentialthinking.com/chapter1/what-is-data-science.html>
4. Adapted from: W3C. 2016. “Data on the Web Best Practices.” <https://www.w3.org/TR/dwbp/>
5. Repository Audit and Certification DSA–WDS Partnership WG Recommendations <https://www.rd-alliance.org/group/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg/outcomes/dsa-wds-partnership>
6. Digital Curation Centre. Data Curation Lifecycle Model. <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

