A Guide to Ensure Access to the Results of Federally Funded Transportation Research

Carol A. Flannagan
Jared Lyle
Jake Carlson
University of Michigan
Ann Arbor, MI

Denise A. D. Bedford
Georgetown University
Washington D.C.

June 2018

DISCLAIMER
The opinions and conclusions expressed or implied in this document are those of the researchers who performed the research. They are not necessarily those of the program sponsors; the FHWA; the Transportation Research Board; or the National Academies of Sciences, Engineering, and Medicine. The information contained in this document was taken directly from the submission of the authors. This material has not been edited by the Transportation Research Board.

SPECIAL NOTE: This document IS NOT an official publication of the Transportation Research Board or the National Academies of Sciences, Engineering, and Medicine. A final, edited version of this document will be released at a later date.
NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed, and implementable research is the most effective way to solve many problems facing state departments of transportation (DOTs) administrators and engineers. Often, highway problems are of local or regional interest and can best be studied by state DOTs individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation results in increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

Recognizing this need, the leadership of the American Association of State Highway and Transportation Officials (AASHTO) in 1962 initiated an objective national highway research program using modern scientific techniques—the National Cooperative Highway Research Program (NCHRP). NCHRP is supported on a continuing basis by funds from participating member states of AASHTO and receives the full cooperation and support of the Federal Highway Administration (FHWA), United States Department of Transportation, under Agreement No. 693JJ31950003.

COPYRIGHT INFORMATION

Authors herein are responsible for the authenticity of their materials and for obtaining written permissions from publishers or persons who own the copyright to any previously published or copyrighted material used herein.

Cooperative Research Programs (CRP) grants permission to reproduce material in this publication for classroom and not-for-profit purposes. Permission is given with the understanding that none of the material will be used to imply endorsement by TRB and any of its program sponsors of a particular product, method, or practice. It is expected that those reproducing the material in this document for educational and not-for-profit uses will give appropriate acknowledgment of the source of any reprinted or reproduced material. For other uses of the material, request permission from CRP.

DISCLAIMER

To facilitate more timely dissemination of research findings, this pre-publication document is taken directly from the submission of the research agency. The material has not been edited by TRB. The opinions and conclusions expressed or implied in this document are those of the researchers who performed the research. They are not necessarily those of the Transportation Research Board; the National Academies of Sciences, Engineering, and Medicine; the FHWA; or the program sponsors.

The Transportation Research Board, the National Academies, and the sponsors of the National Cooperative Highway Research Program do not endorse products or manufacturers. Trade or manufacturers’ names appear herein solely because they are considered essential to the object of the report.

This pre-publication document IS NOT an official publication of the Cooperative Research Programs; the Transportation Research Board; or the National Academies of Sciences, Engineering, and Medicine.

# Contents

List of Figures ................................................................. iv

List of Tables ............................................................... iv

What are We Trying to Accomplish? ......................................... 1
  Policy Sources and Directives
  The Guide’s Structure and Conventions
  Essential Requirements — the Current State
  Evolving Landscape — Going Beyond
  Chapter Checklist

Understanding Essential Requirements ...................................... 7
  Introduction
  The National Transportation Library’s Role
  The Research Project Lifecycle and Compliance
  Essential Requirements
  Public Access Plan Effective Date
  Sources of Research Funding
  Finding Your Submitted Research Products and Data
  Chapter Checklist

Going Beyond the Minimum in an Evolving Landscape .................. 17
  Principles of Accessibility and Availability
  Curation: Definition, Landscape, and Solutions
  Going Beyond the Minimum — Creating an Open Science Culture
  Chapter Checklist

Developing Strategies and Policies .......................................... 25
  Getting Started
  Planning for Essential Requirements
  Implementing a Short-term Strategy
  Going Beyond: Create a Holistic Plan & Long-term Goals
  Implementing a Long-term Vision
  Develop Practical Policies and Strategies
  Measuring Progress
  Chapter Checklist

Roles & Responsibilities ..................................................... 35
  Key Roles for Essential Requirements
  Going Beyond: Expanding Stakeholders, Roles and Responsibilities
  Aligning Stakeholders with Roles and Responsibilities
  Building a Research Preservation-Conscious Culture
  Chapter Checklist
Managing Research Publications ...................................................... 43
- Determining Essential Requirements Eligibility
- Tracking and Identifying Eligible Research Products
- USDOT Registries and Repositories
- Going Beyond Essential Compliance: Including All Research Products
- Understanding Publication Preservation Basics
- Understanding Metadata Standards and Practices
- Understanding Legal and Copyright Strategies
- Chapter Checklist

Managing Research Data ................................................................. 53
- Definition of Research Data
- Explaining Essential Requirements for Research Data
- Going Beyond: Research Data Management and Access
- Understanding Data Preservation
- Deciding What to Preserve: Data Scope & Coverage
- Deciding Which Formats to Preserve
- Managing Quality of Research Data
- Understanding Metadata Standards and Metadata for Transportation Data
- Deciding Where to Preserve Data
- Understanding How Long to Preserve Data
- Chapter Checklist

Data Management Plans ................................................................. 67
- Developing and Maintaining Data Management Plans (DMPs)
- Essential Requirements for Developing Sections of Data Management Plans
- Going Beyond the Minimum for Data Management Plans
- Chapter Checklist

Building Blocks of the Solution ...................................................... 75
- The Big Picture — USDOT and State DOTs
- USDOT Systems and Services Today
- Data Repository and Management Solutions
- Non-USDOT Registries and Repositories for Written Research Products
- Chapter Checklist

Learning and Training ................................................................. 81
- Learning Strategies for the Research Organization
- Training Design and Delivery
- Baseline Awareness Training for Everyone
- Training for Researchers
- Training for Executive and Management Roles
- Training for Research Support Roles
- Chapter Checklist
Estimating and Managing Costs ................................................................. 91
  Balancing Costs and Benefits
  Cost Factors: Essential Requirements for the Research Organization
  Cost Factors: Essential Requirements for Individual Researchers
  Cost Factors: Going Beyond in Repository Management
  Chapter Checklist

Assessing and Managing Progress .......................................................... 103
  Assessment of Essential Requirements
  Assessment of Process and Support for Essential Requirements
  Assessment for Going Beyond
  Chapter Checklist

Putting it all Together .............................................................................. 111
  How to Use the Checklists
  Essential Requirements Checklist
  Using the Checklist to Go Beyond Essential Requirements
List of Figures and Tables

Figures

Figure 1. Overview of the Research Process and Public Access ................................. 8
Figure 2. Current National and International Landscape of Curation .......................... 19
Figure 3. What is Open Science Anyway? ................................................................. 22
Figure 4. US Geological Survey’s Research Model .................................................. 27
Figure 5. Communication Network ........................................................................... 45
Figure 6. Digital Preservation: A Real World Example ............................................. 56
Figure 7. Research Data Repository Options ............................................................. 63
Figure 8. The Big Picture ......................................................................................... 76
Figure 9. Tools to Help: Planning for Costs ............................................................... 94
Figure 10. Approximate Activity Data Costs ............................................................ 100
Figure 11. FAIR Guiding Principles ......................................................................... 108

Tables

Table 1. Whose Responsibility Is It? USDOT vs. Local Research Organizations (e.g., State DOT) ................................................................. 27
Table 2. Examples of Policy Short-term Objectives ................................................... 30
Table 3. Worksheet to Help Identify Stakeholders for the Research Project and/or Dataset ........................................................................... 39
Table 4. Project Open Data Metadata Schema ......................................................... 51
Table 5. Long-term Preservation of Data Types and File Formats .......................... 59
Table 6. Cost Factors for Meeting Essential Requirements .................................... 92
Table 7. Costs of Essential Requirements and Possible Activities with their Associated Costs ................................................................. 96
Table 8. Costs Associated with Creating a Repository ........................................... 101
Table 9. Tracking ..................................................................................................... 104
CHAPTER 1.

What are We Trying to Accomplish?

In This Section
» Policy Sources and Directives
» The Guide’s Structure and Conventions
» Essential Requirements — the Current State
» Evolving Landscape — Going Beyond
» Chapter Checklist

Policy Sources and Directives

In February 2013, the White House Office of Science and Technology Policy (OSTP) released a memorandum titled “Increasing Access to the Results of Federally Funded Research (Holdren Memo),” which directed federal agencies to develop plans to make the publications resulting from federally funded research freely available to the public within one year of publication. The memorandum also required researchers to better account for and manage the digital data resulting from federally funded scientific research, with the goal of making these data publicly accessible, too. The memorandum directed the heads of executive departments and agencies with over $100 million in annual extramural research and development (R&D) budgets to develop a public access plan for the results of research, both publications and digital scientific data, directly arising from their funds.

In response to the OSTP Memorandum, the US Department of Transportation (USDOT) developed their Public Access Plan, which was published on December 15, 2015, and became effective as of Jan 1, 2016. The USDOT Public Access Plan can be found on the USDOT’s website. In addition, “Executive Order 13642, Making Open and Machine Readable the New Default for Government Information” (Obama, 2013), and “Office of Management and Budget (OMB) Memorandum M-13-13, Open Data Policy — Managing Information as an Asset” (Burwell et al., 2013) were released in 2013. Individually and collectively, these directives established the mandates for the federal government to transform data and information into usable and accessible digital artifacts and promote and accelerate their release, subject to certain limitations imposed by privacy, confidentiality, and national security considerations.

USDOT Public Access Plan:
1. WHAT ARE WE TRYING TO ACCOMPLISH?

The purpose of this Guide to Ensure Access to the Results of Federally Funded Transportation Research (hereafter referred to as “the Guide”) is to help state DOTs, as well as other organizations that do transportation research, better understand and consider how they will comply with the USDOT policy. Compliance is discussed in the context of the essential requirements that are mandated for organizations to continue to receive federal research funding in the future. It is also discussed in the broader context of going beyond to achieve the wider goal of Open Science. Policy becomes process and practice through interpretation and guidance.

Historically, the USDOT has provided access to intramural and extramural research in progress and technical reports, as well as many final publications through partnerships with organizations such as the Transportation Research Board (TRB). USDOT’s intramural research programs also have a long history of making data available to the public (e.g., Fatality Analysis Reporting System, or FARS database). On an incremental basis, USDOT’s extramural research programs are taking steps to increase data sharing. The purpose of the Public Access Plan is to scale and institutionalize those research and development access practices across the department.

The USDOT’s public access process and practice have continued to evolve during the course of this project. Essential requirements for compliance are now well documented on the USDOT website. In addition to describing essential requirements and how to comply with them, this Guide provides background information and explanations pertinent to key questions and concerns of researchers, administrators, information service providers, and information technology managers. Understanding the context and challenges, as well as participating in the discourse surrounding public access and Open Science is important for everyone who generates, funds, manages, or uses transportation research.

While essential compliance to these new policies and requirements contributes to the goal of Open Science, the full vision will be realized when state DOTs and other organizations go beyond the minimum. The Guide explains both how to comply with the essential requirements and how to go beyond them to achieve the benefits of Open Science. Open Science is the movement to make scientific research, data, and dissemination accessible to all levels of an inquiring society — amateur or professional.

In general, Open Science strives to make it easier to publish and communicate scientific knowledge. It is a long-range vision which goes beyond the requirements of the USDOT’s Public Access Plan. However, Open Science is the practice of science in such a way that: (1) others can collaborate and contribute, and (2) research data, lab notes, and other research processes are freely available under terms that enable reuse, redistribution, and reproduction of the research and its underlying data and methods.

Essential requirements for compliance:
https://ntl.bts.gov/public-access/plan-executive-summary
initiatives such as the Public Access Plan make the Open Science vision achievable. For more information on Open Science, see “What is Open Science Anyway?” on page 22.

The primary audience for the Guide is staff in state Departments of Transportation (DOTs) who will be responsible for complying with the requirements for their own federally funded research and ensuring compliance when such funds are used internally or passed to contractors. Federally funded transportation-related research is often conducted by organizations beyond state DOTs, but state DOTs will be responsible for communicating requirements and ensuring compliance. In addition to state DOT staff, researchers and research staff at universities and other contract organizations should find this helpful in understanding their role in compliance with the new requirements.

The Guide’s Structure and Conventions

The Guide can be read in different ways. Administrators looking for task-oriented guidance may want to begin by reviewing the checklists

SURVEYING THE LANDSCAPE

Chapter 1: What are We Trying to Accomplish?  
Chapter 2: Understanding Essential Requirements  
Chapter 3: Going Beyond the Minimum

CREATING A ROADMAP

Chapter 4: Developing Strategies & Policies  
Chapter 5: Roles & Responsibilities

BEGINNING THE JOURNEY

Chapter 6: Managing Research Publications  
Chapter 7: Managing Research Data  
Chapter 8: Data Management Plans  
Chapter 9: Building Blocks of Your Plan

STAYING THE COURSE

Chapter 10: Learning & Training  
Chapter 11: Estimating & Managing Costs  
Chapter 12: Assessing & Managing Progress  
Chapter 13: Putting it All Together
1. WHAT ARE WE TRYING TO ACCOMPLISH?

in Chapter 13, and then refer back to earlier chapters for explanations. This is the first edition of the Guide. As practices at state DOTs and research institutions grow, so will the community’s understanding of how to adopt and adapt the practice and guidance to a particular context.

Each Chapter is structured to address:

- Main Ideas and Concepts
- Essential Requirements
- Going Beyond the Minimum
- Chapter Review and Checklists

Within each chapter the Guide uses design elements for emphasis and identification including:

- Key Websites
- Common Questions and Answers
- Definitions
- Tips, Tricks, and Good Ideas

Essential Requirements — the Current State

Compliance with essential requirements is mandated for organizations seeking federal funding for transportation-related research after December 16, 2015. The USDOT will include assessments of compliance in the review of funding proposals. Proposals from noncompliant organizations will not be eligible for consideration.

Since the issuance of the Public Access Plan, the National Transportation Library (NTL) and the TRB have defined the practices that satisfy essential compliance, described and explained in detail in
1. WHAT ARE WE TRYING TO ACCOMPLISH?

Chapter 2. While most of the essential requirements are straightforward and reflect current practice, the practices surrounding the requirement to preserve and provide access to research data supporting research results are evolving. The Guide helps navigate the requirements and explains good practices for developing an institutional strategy to satisfy the USDOT requirements.

**Evolving Landscape — Going Beyond**

Essential requirements, now established, enable state DOTs and other transportation research organizations to receive federal research funds. However, the practice of preservation, access, and sharing of research products and data is progressing over time. The Guide also describes and explains evolving practices, enabling organizations to go beyond the minimum and keep up with this evolution. In particular, going beyond allows organizations and researchers to take full advantage of the Open Science vision.

**Chapter Checklist**

From this chapter, you should be able to:

- Identify and track the national and state DOT level policies and authorities
- Understand the purpose and structure of the guide
- Understand concept of essential requirements
- Understand the consequences of noncompliance
- Understand the concept of “going beyond minimum”
1. WHAT ARE WE TRYING TO ACCOMPLISH?
CHAPTER 2. Understanding Essential Requirements

In This Section
» The National Transportation Library’s Role
» The Research Project Lifecycle and Compliance
» Essential Requirements
  » Before, During, and After a Research Project
  » Research Products
  » Open Researcher and Contributor ID (ORCID)
  » Data Management Plans (DMP)
  » Consequences of Noncompliance
» Public Access Plan Effective Date
» Sources of Research Funding
» Finding Your Submitted Research Products and Data
» Chapter Checklist

Introduction
This chapter covers the USDOT’s essential requirements for researchers and research institutions requesting and receiving transportation-related federal research funds. Public Access Plan compliance guidance and recommendations are available in the “DOT Public Access” webpages, hosted by the National Transportation Library, at https://ntl.bts.gov/public-access.

The National Transportation Library’s Role
The USDOT’s infrastructure builds upon existing research lifecycle tracking arrangements between the Transportation Research Board’s Research in Progress (RiP) database, the USDOT Research Hub, NTL Digital Library, and a new interface with the USDOT Public Data Listing.

The USDOT Research Hub (Research Hub) is a publicly available, web-based, searchable database of the department’s research portfolio, containing project records that describe the department’s current and past research activities at the project level. When projects are completed, project deliverables (typically technical reports) are deposited into the NTL Digital Library and added as NTL-based URLs to associated project records.
Figure 1. Overview of the Research Process and Public Access

Some terms and concepts may be unfamiliar but they will be covered in detail in the chapters of this Guide.

The Research Project Lifecycle and Compliance

Compliance is best understood in the context of a research project’s lifecycle. Figure 1 illustrates the USDOT’s conceptualization of the lifecycle and notes key points of compliance in the cycle. Some terms and concepts in the figure may be unfamiliar but they will be covered in detail both below and in the chapters of this Guide.

Essential Requirements

Requirements: Before, During, and After a Research Project

Before a Research Project

The “Submit Research Proposal” stage of the lifecycle embodies activities needed to develop and start a research project, and include:

- Develop a Data Management Plan (DMP) to define what steps will be taken over the course of the project to administer the data and prepare it for eventual public dissemination and long-term preservation. More information on DMPs and their use is presented later in this chapter and in Chapter 8.
- Obtain approval of the DMP with your organization and your funding agency as a component of the research project proposal. See Chapter 8.
- Register an Open Researcher and Contributor ID (ORCID) for all project investigators and contributors (https://orcid.org/trademark-and-id-display-guidelines). An ORCID is a unique identifier that is easily registered and supports researchers in receiving proper attribution and credit for their work. More details on ORCID identifiers and their use is presented later in this chapter and in Chapter 4.
Be sure to note the RH (Research Hub) ID in the Research Hub project record (https://researchhub.bts.gov/faq).

Obtain and note the funding agreement number (this number is the sponsoring agency’s contract or project number).

**During a Research Project**

Once funding for research has been received and the research is underway, researchers should ensure:

- Update the project record in RiP regularly and consistently for the duration of the project (e.g., if there are any changes to funding amount, project end dates, etc.). These records will be automatically updated in the Research Hub.
- Follow data management practices outlined in the DMP. Revise and update the DMP as needed. See Chapter 8.

**After a Research Project**

When their research project is complete, researchers should:

- Ensure that the project status is changed in RiP from “Active” to “Completed” within two months of completion.
- Make sure that someone retrieves the RH Display IDs for each project record.
- Package and archive data as outlined in the DMP.
- Include ORCID identifiers and funding agreement number(s) on the Techno Report documentation page and in any peer-reviewed publication submissions. Be sure to format ORCID identifiers correctly in your final reports (https://orci.org/tradmark-and-id-display-guidelines).
- Ensure that all rights under copyright are nonexclusively retained by the USDOT and that the terms and conditions of publication to peer-reviewed journals and other outlets do not impair the obligation of the authors to comply with the plan. See Chapter 6.
- Send one email to the USDOT Research Hub (Research.Hub@dot.gov, NTL (NTLDigitalSubmissions@dot.gov) and TRB (TRIS-TRB@nas.edu):
  - Final Report URL(s) or PDFs for any resulting publications
  - URL(s) to and associated descriptive metadata for any final datasets and the arising from the research project
  - The funding agreement number of the project
  - The RH Display ID for the project
Requirements: Research Products

The USDOT’s Public Access Plan is focused on two types of research products: (1) reports and publications, and (2) data.

Research Reports and Publications

The USDOT defines research publications to include:

- Any final peer-reviewed manuscript accepted for publication
- Any intramural technical or final reports
- Any scientific research project’s written deliverable (e.g., technical/final reports) that arises from extramural research funded, either fully or partially, by federal funds awarded through a USDOT-managed contract, grant, or other agreement
- Written deliverables published by the USDOT, with no embargo
- Written deliverables published externally by USDOT scientists or extramural USDOT-funded scientists, with embargo

Research Reports

In the field of transportation, final project and technical reports will represent a larger percentage of research products than formal peer reviewed research articles and publications. Final and technical reports may be more challenging to track because they may be less easily discovered than journal publications.

The researcher is the key actor in this process, since often only she/he knows when the final report or technical report has been submitted to the funding agency. It is important for researchers and their research organizations to establish internal processes for ensuring that research projects are known, that project status is tracked and that final and technical reports are identified and submitted to the TRB and NTL repositories. All reports must be submitted to these repositories within 12 months of issuance.

Emargo is a time-limited restriction on publication
2. ESSENTIAL REQUIREMENTS

Requirements for Submission of Final Research Publications

All final peer-reviewed manuscripts accepted for publication, intramural, technical, or final reports, and any scientific research project written deliverables (e.g., technical and final reports) produced under a USDOT contract or grant must be submitted to NTL under a nonexclusive license agreement. These publications will be made publicly available after an embargo period of 12 months following publication.

Research Data

Research data is, as defined in the USDOT’s Public Access Plan, all scientific data collected through research projects funded, either fully or partially, by federal funds awarded through a USDOT contract, grant, or other agreement or collected by USDOT employees. Not all data are eligible for public access, though. Research data that are covered under the Public Access Plan are those data that were used to support the research conclusions in a publication or a research report. The public access policy does NOT include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects such as laboratory specimens.

Requirements for Identifying and Preserving Final Research Data

Researchers are encouraged to use publicly accessible repositories for the deposit of their data, where appropriate and available. Currently, the list of recommended repositories on the NTL website includes:

- Dryad (http://www.datadryad.org)
- Harvard Dataverse (https://dataverse.harvard.edu)
- Inter-university Consortium for Political and Social Research (ICPSR) (http://www.icpsr.umich.edu/)
- Odum Institute Data Archive (http://www.odum.unc.edu/odum/)
- Zenodo (http://zenodo.org)

As this is an evolving field and practice, the list of conforming repositories will be updated periodically. Researchers may use the Registry of Research Data Repositories’ searchable listing of data repositories (https://www.re3data.org/) to locate potential archiving options for their data. If their organization includes a library, then a librarian may be available to assist researchers in identifying repositories that meet their specific data needs.

Repositories include:

datadryad.org
dataverse.harvard.edu
icpsr.umich.edu
odum.unc.edu
zenodo.org

Locate potential data archiving options

https://www.re3data.org

Researchers may use the Registry of Research Data Repositories’ searchable listing of data repositories (https://www.re3data.org/) to locate potential archiving options for their data.
If state DOTs choose to use another repository, that repository should be able to:

- Provide unique permanent identifiers for digital datasets and the standards that are defined for managing identifiers
- Support interaction among the awardee(s), the data repository, and the state DOT or USDOT grant manager to enable all parties to verify that:
  - Data meet minimum quality standards
  - Data are appropriately evaluated for and secured to prevent disclosure of personally identifiable information
  - Proprietary interests and confidentiality are protected
  - Intellectual property rights are respected
  - Data are licensed in a manner encouraging access, availability, and reuse
- Support long-term access and preservation of the research data

There may be situations in which multiple repositories would potentially be suitable for meeting USDOT requirements. In this case the researcher may need to do some additional investigation to assess the capabilities and services of the repositories and supporting organizations to determine which is a best fit for their needs.

Requirements for Describing Preserved Research Data

The USDOT requires Project Open Data metadata standards for scientific data. Per the National Information Standards Organization, metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Archives not operated by the USDOT are required to be able to generate the common core and extensible metadata so that the USDOT can harvest inventory records for inclusion in its own public data listing and enterprise data inventory. Metadata is discussed in detail in Chapter 6 and Chapter 7.

Requirements: Open Researcher and Contributor ID (ORCID)

ORCID stands for Open Researcher and Contributor ID. ORCID provides a persistent digital identifier for each researcher that supports linkages between researchers, their work, and their professional activities. The USDOT Public Access Plan requires that researchers should obtain and use their ORCID identifiers when submitting research funding proposals.

ORCID (Open Researcher and Contributor ID) is a nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors.
2. ESSENTIAL REQUIREMENTS

Requirements: Data Management Plans (DMP)

Researchers and research organizations (e.g., state DOTs) applying for research funds from the USDOT under contract, grant, or cooperative agreement are required to submit a data management plan (DMP) for approval. The DMP must identify a repository for preserving data that is accessible by the NTL and must include the following elements:

☑ Project description
☑ Explanation of the data to be collected, including collection methods, formats, metadata and other standards followed
☑ Short-term storage and access plan during the project
☑ Treatment of any legal issues, including intellectual property and personally identifiable information
☑ Preservation and archiving strategies after the project
☑ Unique permanent identifiers for all publications (DOI), authors (ORCID identifiers), and datasets for the linking and correlation of authors and relevant data.

Data Management Plans will be covered in detail in Chapter 8.

Consequences of Noncompliance

State DOTs and other organizations will be held accountable for the compliance of all of their researchers. If a researcher in the Planning Department is not in compliance with the minimum requirements, noncompliance will also impact the funding proposal of a researcher from the Bridge Administration in the same DOT. Organizations with researchers seeking funding from any of the sources listed on the next page should ensure that they have tracked and accounted for compliance across the board.

Public Access Plan Effective Date

The USDOT’s Public Access Plan does not apply to written deliverables or data submitted for publication prior to the plan’s effective date, or to digital data generated prior to the plan’s effective date. The effective dates can be no sooner than the publication date of the department’s final plan (December 16, 2015).
Sources of Research Funding

Primary Funding Sources
USDOT funding sources may include, but are not limited to the following:

- Federal Highway Administration (FHWA)
- Federal Transit Administration (FTA)
- Research and Innovative Technology Administration (RITA)
- National Highway Traffic Safety Administration (NHTSA)
- Federal Aviation Administration (FAA)
- Office of the Secretary of Transportation (OST)
- Federal Railroad Administration (FRA)
- Federal Motor Carrier Safety Administration (FMCSA)
- Federal Maritime Administration (MARAD)
- Pipeline and Hazardous Materials Safety Administration (PHMSA)

It is important to note that while these sources of research funding are included under the Public Access Plan, not all funding types from these sources are currently included in the essential requirements to participate. In particular, at this time the SP&R (State Project & Research) funds are not covered under the essential requirements unless there are matching UTC funds. SP&R funds may be eligible at a later time, so it is important to regularly check the Public Access Plan website for updates in what is covered and what constitutes essential compliance.

Funding Sources Beyond the USDOT
While transportation-related research is primarily funded by the USDOT, it is also funded by other federal departments and agencies. Each department or agency will have its own requirements and processes for making research products available which will need to be followed by the researchers and organization receiving the award. Transportation-related research is also supported by, but not limited to the following:

- Department of Energy (DOE)
- National Science Foundation (NSF)
- Environmental Protection Agency (EPA)
- Department of Homeland Security (DHS)
- Department of Agriculture (DOA)
- Department of Education (DOE)
- National Institutes of Health (NIH)
Finding Your Submitted Research Products and Data

The NTL currently serves as the permanent archive of technical reports. Links for technical reports listed in the Research Hub are provided for copies deposited in NTL's archive. The NTL will also provide a searchable database of DMPs submitted as a part of the USDOT's Public Access Plan requirements. Reports housed or accessible through the NTL are available without charge.

The USDOT Public Data Listing will be expanded to add descriptions of and links to datasets resulting from scientific research. This listing will serve as a single access point for the USDOT's intramural and extramural public datasets, which are also reported to data.gov. Once the Public Access Plan is implemented, the Research Hub will serve as the linking mechanism for scientific publications and their underlying research data by requiring that links for digital datasets (as referenced in the USDOT Public Data Listing) are submitted along with other project deliverables as part of the project close-out process. This will create one publicly-accessible record for each completed research project that contains the research project description and links to any associated publications and data.

Chapter Checklist

From this chapter you should be able to:

- Understand what research is eligible under the policy
- Know how to track and identify your funded research projects
- Understand what to do when in the research project lifecycle
- Provide guidance to researchers in registering ORCID identifiers
- Know when and how to submit a Data Management Plan
- Know where to submit your eligible research publications
- Know where to submit your eligible research reports
- Know where and how to preserve your eligible research data
- Know how to provide a unique permanent identifier

Endnotes

CHAPTER 3.
Going Beyond the Minimum in an Evolving Landscape

In This Section
» Principles of Accessibility and Availability
» Curation: Definition, Landscape, and Solutions
» Going Beyond the Minimum — Creating an Open Science Culture
» Chapter Checklist

Principles of Accessibility and Availability

“Availability” and “accessibility” refer to the ways in which the USDOT and its policies ensure or strive to ensure that citizens have the means to find, read, understand, trust, and otherwise use research results that have been partially or fully funded by the USDOT. To achieve the Public Access Plan’s goals, it is important to focus on both availability and accessibility.

Availability means the public is aware of the research products or can determine that research exists. In other words, available research products can be found and acquired by the public. Federally-funded research results are available if they meet the following criteria:

- There is traceability from the funding sources to the research produced from those funds
- Research results are registered in the NTL’s ROSA P (Repository and Open Science Access Portal) registry: https://rosap.ntl.bts.gov/
- The publications registry supports searching for research publications and their underlying research data
- Search and discovery are grounded on metadata that describes both the research publications and their underlying research data
- There is a persistent and reliable link to the:
  - publication from its metadata (e.g., DOI)
  - research data from its research publication (DOI)
  - researcher conducting the research (e.g., ORCID identifier)
- The research publication and data have been stored in facilities that guarantee their future archival preservation and access
Accessibility means the data resource is understandable to and usable by its intended audience. It also means that its research products are understandable to and usable by as many people as possible who wish to use it. Federally funded research results are considered accessible if the following criteria are met:

- Data creation and transformation are documented
- Data versions are documented
- Programs or algorithms used to manipulate data are documented
- Technical documentation explaining how data are designated and labeled are available, including codebooks, dictionary files, data collection instruments, data maps, errata files, frequency files, cross-tabulation files, user guides, manuals, appendices, reports, record layout files, or rectangular files

The steps to achieving accessibility and availability of transportation research products and data are discussed in more detail in the remaining chapters of this Guide. They include concrete actions that state DOTs and researchers can (and should) take to achieve these ends.

Curation: Definition, Landscape, and Solutions

Curation is the planning and actions taken to add value to the data through enabling its use beyond its immediate purpose, making it accessible to others and ensuring that the data retain their value over time. In other words, it can be broadly defined (ARL 2017) as the “active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities.” As noted by ICPSR, an organization dedicated to curating social science data, the work done to curate data has similarities to the work done by an art museum curator. Data are organized, described, cleaned, enhanced, and preserved for public use through the curation process, just as work is done on paintings or other art works to make them understandable accessible to the public now and in the future.

The curation landscape is constantly evolving and the demand for access to data will only increase as time goes on. Curation can be handled in two different ways: local curation or off-site curation. Local curation solutions, while a feasible option for organizations that produce research data on a large scale, may be challenging given the dynamic and evolving nature of the practice. Several repositories offer varying levels of off-site curation. For more information, see “Deciding Where to Preserve Data” on page 63.

List of off-site repositories:
Satisfying essential requirements is the immediate concern of researchers and their research organizations. As noted earlier, the National Transportation Library (NTL) and the Transportation Research Board (TRB) provide guidance and services that enable state DOTs to comply. However, rapidly changing technologies and increased capabilities to handle diverse types of data, along with the increased number of publishers and funders requiring data sharing have created new demands for access. In the long term, research organizations should expect the demand for access to their research results and data to increase. State DOTs that produce research on a large scale may find that local curation solutions facilitate tracking, managing, access, and preservation of their research assets. Local curation solutions, though, are challenging for most state DOTs today given the dynamic and evolving nature of the practice.

For a state DOT considering a local curation solution, it will be helpful to gather more information about the basic elements of research products and data curation. The box to the right provides a list of basic elements of research product and data curation.

In addition, some federal agencies have a historical foundation of managing research products and a historical mandate or practice that promotes sharing, dissemination, and reuse. These include the US Geological Service, Department of Agriculture, National Oceanic and Atmospheric Administration (NOAA), Department of Energy, National Science Foundation (NSF), and National Institutes of Health (NIH).

To stay abreast of progress, state DOTs may wish to visit the individual websites of the organizations listed above. In considering how to apply the OSTP memorandum, agencies are striving to reflect the needs of the research communities that they serve. Therefore, state DOTs may want to consider actively participating in the process of forming policies and practices by getting involved in discussions and partnering with agencies on these subjects when opportunities arise to do so.

Basic Elements of Research Products and Data Curation

- Data Deposit Practices and Agreements
- Metadata Creation and Management
- Data Access and Availability Services
- Research Registry Design and Standards
- Research Repository Design and Standards
- Persistent Identifiers for Research Products
- Curation of Sensitive Data
- Data Processing and Review Support
- Data Appraisal Services
- Preservation Services
Going Beyond the Minimum — Creating an Open Science Culture

Complying with the USDOT’s essential requirements is a critical first step towards achieving the Open Science vision. An Open Science culture is another critical success factor. Understanding the factors that contribute to such a culture will facilitate compliance and allow organizations and researchers to go beyond the minimum. Organizational policies and processes should be supported by strategies that suit the context, including consideration of researcher behaviors and practices; the state DOT or other research institution; the research community; research norms; and research data infrastructures. The purpose of this section of the Guide is to raise awareness of the factors that may influence compliance in addition to policies.

Factors Related to Researchers

Although it is relatively easy to describe how and why researchers should comply with the policy, other factors may influence compliance and should be considered:

- Consider sociodemographic factors such as age, seniority and career goals, and research practices. Older researchers may be more willing to share their knowledge and research with others because their reputation is established, and they are at a point in their careers where they can see how shared use will advance their own work. Younger researchers may be more hesitant to share if there is a chance that doing so will impact their advancement.

- Consider degree of control, including knowledge of the data requester, involvement in the decision to share, and priority rights for publication. Researchers want to have a say in or an understanding of how their research and data may be shared; they want to know who is using their research and how.

- Consider resources needed to make data available, including time and effort, skills and knowledge, and financial resources. Researchers may be concerned about the amount of time and effort required to make their research available. Some of this concern may be related to how the research data were originally curated, and the time and effort required to make them usable today.

- Consider non-monetary benefits to researchers such as formal recognition, professional exchange, and quality improvements. Academic researchers are concerned about their reputation and recognition for their work. Researchers are unlikely to willingly share where research and data sharing requires significant time, detracts from other work, and does not add in some way to a researcher’s performance metrics.
Factors Related to Research State DOTs and Other Transportation Research Organizations

Organizations should support their researchers by having and promoting data sharing policies, and encouraging and incentivizing a knowledge- and data-sharing culture. This means having organizational-level rewards and recognitions for researchers who actively share. Organizations should also ensure there is a deep understanding of the policies and requirements of all agencies from whom an organization requests research funds. Organizations whose researchers are not in compliance run the risk of having their research funding proposals turned down.

When funding agency policies and requirements are tracked and interpreted at the research institution level, cost efficiencies are achieved. When organizations do not provide this support, researchers incur comparative disadvantages — time taken away from doing research.

Factors Related to the Research Community

While the research community is beyond the control of individual organizations or researchers, the community’s culture and practices influence how researchers share knowledge and data.

• Different research communities have different knowledge and data sharing cultures. An organization may need to further incentivize researchers if the community does not openly share knowledge and data.

• Some research communities accustomed to sharing knowledge and data have established metadata standards, data formats, and interoperability standards. Where these are not available, organizations may need to provide additional guidance and support to researchers.

• The scientific value of the research, including rate of scientific progress, scientific exchange practices, and scientific review practices may influence how accustomed researchers are to sharing research and data.

• Peer review publishing in high-impact academic journals is a critical element of how academics are judged by their institutions. Where a research community’s journals have established data sharing policies, researchers will be more inclined to comply with organizational and funding agency policies.
The Open Science Vision

Open Science is a movement to make scientific research, data, and dissemination available and accessible to all levels of society — from citizen science to academically produced research to research conducted in the private sector. It involves publishing open research, advocating for open access, encouraging scientists to share their methodologies, and generally making it easier to leverage existing scientific knowledge and research. Open Science is not a new concept — its roots date back to the earliest academic journals of the 17th century. Further, there is no one definition of Open Science. However, the USDOT's Public Access Plan supports the Open Science vision by requiring that federally-funded transportation related research is accessible and available to the public by ensuring that:

- The public is aware of the research publications and datasets generated, fully or partially, through federally-funded scientific research (Principle of Availability).
- The public is able to download and analyze unclassified publications and/or digital datasets unless specifically precluded by privacy, confidentiality, or security concerns where access may be restricted to subsets of the public, or other means as necessary (Principle of Accessibility).

Value of Open Science to the Field of Transportation

By supporting the Open Science vision, the USDOT generates value for those working in the transportation field and the general public. Open Science in transportation:

- Increases the public's understanding of all modes of transportation
- Informs and supports appropriate decision making and planning for transportation at all levels of government and by the private sector
- Supports the reproducibility of science, and enables scientists and engineers to move beyond immediate projects to leverage knowledge and build upon the work of other researchers
- Ensures that information about advances in transportation safety, mobility, and economic development are available to all consumers
- Improves the quality and transparency of research by providing access to the underlying data used to derive conclusions
- Generally maximizes the impact of federal research funding in transportation
Factors Related to Research Norms

Research norms include ethical and legal considerations. These issues are addressed in and essential to the USDOT’s Public Access Plan.

Ethical Norms

Ethical norms focus on confidentiality, informed consent practices, and considerations of potential harm resulting from public access to research data. In the USDOT policy, organizations are responsible for ensuring that ethical norms are observed and respected. A critical internal partner for ensuring compliance in academic organizations is the Institutional Review Board (IRB). Ethical standards and concerns are part of the internal IRB review process, and can act as a compliance checkpoint.

Legal Norms

Legal norms include issues such as ownership and right of use, privacy, contractual consent, and copyright practices. Unresolved legal issues can adversely affect knowledge and data sharing. State DOTs and research institutions need clearly defined policies and processes to ensure that legal norms are observed. Private sector organizations may need to consider legal norms surrounding copyrights and patents for their research products. In such cases, research supported by funds from federal agencies should be managed according to federal policy.

Factors Pertaining to Data Users and Consumers

While data users and consumers are generally beyond the control of individual organizations or researchers, their use of data can influence researchers’ willingness to comply.

Adverse Use of Data

Consider the adverse use of research data and its effect on the researcher. A universal concern for researchers is the adverse use of their research knowledge or data, including falsification, commercial misuse, competitive misuse, flawed interpretation, and unclear intent of the use.

It is equally important for a research institution to incentivize appropriate use and citation of others research data, as it is for an organization to incentivize those same researchers to share.

Security and Protection of Data

The security and protection of research data are significant concerns for researchers. In addition to a general culture of respect for and acknowledgment of other researchers’ data, organizations whose researchers actively use others’ research data should ensure
the security conditions for those data, and commercial or public organizational requirements for public access.

Factors Pertaining to Data Infrastructures
State DOTs are required by the USDOT Public Access Plan to preserve and safeguard the research data upon which research conclusions depend. Data infrastructures are one way to fulfill this requirement. However, state DOTs are responsible for identifying a trustworthy data infrastructure. Data sharing within and across state DOTs will be affected by the following:

- Architectures that support access, performance, storage, data quality review, and data security
- Usability of tools and applications, as well as technical support available to support availability, access to, and use of research data
- Management software that supports data documentation and metadata standards and tells researchers about the quality, nature, and trustworthiness of the research data.

Chapter Checklist
From this chapter you should be able to:

- Explain the goal of Open Science and translate the goal to your organization
- Understand the difference between availability and accessibility
- Understand how your organization supports availability and accessibility
- Gain a familiarity of current landscape of other federal agency and departmental practices
- Gain a familiarity with national and international landscape of research preservation
- Understand organizational factors that may present challenge
CHAPTER 4.
Developing Strategies and Policies

In This Section
» Planning for Essential Requirements
» Implementing a Short-term Strategy
» Going Beyond: Create a Holistic Plan & Long-term Goals
» Implementing a Long-term Vision
» Develop Practical Policies and Strategies
» Measuring Progress
» Chapter Checklist

Getting Started

Starting a completely new program of sharing and preserving research products can seem daunting, but it can be done in small steps over time. In order to successfully navigate the small steps, it is essential to begin with a vision.

The starting point for a state DOT research preservation program must be the policies and essential requirements established by the USDOT Public Access Plan, and supplemented by the guidance web pages hosted by NTL. These essential requirements, described in more detail in Chapter 2, include:

- Create a Data Management Plan (DMP) for every project covered under the USDOT Public Access Plan and follow, revise, and update DMPs as needed throughout the projects
- Have an ORCID identifier for each researcher and research contributor on every research project covered under the Public Access Plan
- Report research projects to TRB’s Research in Progress (RiP) database and update the record over the duration of the projects, particularly ensuring that every project’s status is changed from “Active” to “Completed” within two months of completion
- Package and archive data as outlined in the DMP
- Submit research reports and peer-reviewed publications to the NTL’s ROSA P repository and Research Hub with all appropriate documentation

RiP database:
http://rip.trb.org
While the goals and purview of the USDOT Public Access Plan are specified, the way state DOT programs might get there is not. Moreover, by “Going Beyond the Minimum” on page 17, state DOTs can envision policies that not only are more comprehensive or inclusive than the minimum, but also help the state DOT and its researchers move towards the Open Science vision.

The state DOT’s vision for research preservation will need to include:

- Governance and oversight
- Reach of program — will it include only what is required by the USDOT or will it extend to state-funded programs?
- Policy and provision for training
- A process for re-evaluating goals and progress

**Planning for Essential Requirements**

While the steps that lead to compliance have been clearly laid out by the USDOT (Chapter 2), each state DOT must interpret and institutionalize a process to ensure compliance. The impacts of noncompliance are felt by the whole organization, not just individual researchers. For this reason, it is important to have a short-term plan that will ensure that the state DOT is in good standing on all of its research projects funded after January 2016.

**Short-term Objectives Focusing on Essential Requirements**

*Short-term goals should focus on the USDOT’s essential requirements.* As Chapter 2 indicates, achieving compliance means taking specific actions before, during, and after a research project is completed. While researchers are critical to achieving compliance, it is the organization’s (e.g., state DOT’s) responsibility to make sure all of their researchers are in compliance. This means that people other than researchers need to be involved in planning and carrying out the plan. Who these people are will vary by state DOT. The Guide describes functional roles that you can use to map to your organization in Chapter 5. The table entitled “Whose Responsibility Is It?” on page 27 identifies which plan aspects the USDOT is currently responsible for and which aspects the local organization is responsible for. As the table suggests, organizations must develop plans to translate and track eligible research, explain how the plan will be translated to a process,

**Staying abreast of changes to essential requirements**

Essential compliance requirements are consistently being updated and can be found at [http://ntl.bts.gov/publicaccess/](http://ntl.bts.gov/publicaccess/).
determine who will be responsible for implementing the plan, and determine where research data will be managed and stored. Programs that meet and support these goals will also facilitate transition to the longer-term goals of Open Science, as described in Going Beyond. Successful programs will communicate clearly with researchers, detailing what falls under the policy, why it should be followed, and how.

**Key Questions to Ask when Developing Strategies**

A set of key questions to consider when developing strategies and policies is provided below. By answering these questions, it will become clear what structures are missing and what are already in place.

**Data Management Plans**

- Is there a strategy for creating DMPs?
- Do researchers have experience developing DMPs?
- Is there a need for additional short-term training or is the USDOT guidance on DMPs sufficient for the immediate future?

**ORCID Identifiers**

- Is there an institutional policy for recording the ORCID identifiers of each researcher and contributor?
- How many researchers currently have ORCID identifiers?
- Is there a need to raise awareness about ORCID identifiers across the organization?

---

<table>
<thead>
<tr>
<th>Issue</th>
<th>Responsible for Essential Compliance</th>
<th>Responsible for Going Forward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why? Goals</td>
<td>USDOT</td>
<td>USDOT &amp; State DOT</td>
</tr>
<tr>
<td>What? Eligible Research &amp; Data</td>
<td>USDOT &amp; State DOT</td>
<td>USDOT &amp; State DOT</td>
</tr>
<tr>
<td>When? Timelines</td>
<td>USDOT</td>
<td>USDOT &amp; State DOT</td>
</tr>
<tr>
<td>How? Defining a Process</td>
<td>USDOT &amp; State DOT</td>
<td>USDOT &amp; State DOT</td>
</tr>
<tr>
<td>Who? Functional Roles</td>
<td>State DOT</td>
<td>State DOT</td>
</tr>
<tr>
<td>Where? Registries &amp; Repositories</td>
<td>USDOT &amp; State DOT</td>
<td>USDOT &amp; State DOT</td>
</tr>
</tbody>
</table>

---

Table 1. Whose Responsibility Is It? USDOT vs. Local Research Organizations (e.g., State DOT)
4. DEVELOPING STRATEGIES & POLICIES

Registering & Depositing ...

Questions to Ask
Registering and depositing research publications and reports

Eligible Research Publications & Reports

- Research funding proposals:
  - Who creates research funding proposals at the state DOT?
  - Are those who manage research funds and proposals familiar with TRB's RiP database and with Form 1700.17?
  - If individual researchers are responsible, are they aware of the new requirements?
  - Is there a template for funding proposals? If yes, do researchers know how to add new information to the template?

- Project updates: Are research fund managers routinely informed of project updates during the project lifecycle? Are all updates registered in the TRB RiP database?

- Project closeout: Who is responsible for closing out the research project?

- Final reports and publications:
  - Are final project reports, final technical reports, or any other research publications generated by research projects routinely submitted to the NTL, the TRB RiP, and the USDOT's Research Hub? If so, how can this role be leveraged in the future?
  - How often does the state DOT submit research products to these sources? If the current practice is maintained, will the state DOT be in compliance with the new requirement one year after project closeout? If not, what needs to be changed?

Questions to Ask
Registering and depositing research data

Eligible Research Data

- Is research data currently submitted along with research publications and reports to any sources? Is this a process that can be leveraged?

- Do researchers have a clear understanding of what constitutes eligible research data? If not, who is the best person to provide this clarification?

- Data packages (data and metadata packaged for deposit):
  - Are researchers familiar with data packages?
  - Is there someone at your state DOT who understands how to create a data package?

- Repositories:
  - Are the repository choices recommended by the USDOT acceptable and affordable for the state DOT?
  - Is there a preferred repository for the state DOT to manage the preservation of research data?
Will the repository of choice be approved by the USDOT if it is included in the DMP?

Will the repository choices be approved by the USDOT if it is included in the DMP?

If the repository choices are not affordable, what other options are available? Who is the best person at your state DOT to assess and select a repository solution if one has not yet been identified?

What is the strategy for adding data management cost factors into the DMPs?

Who is the best person in the state DOT to advise on cost estimates?

Questions to Ask
Registering and depositing research data

Implementing a Short-term Strategy

Communicating the short-term strategy to the research institution is important for ensuring compliance. Research institutions might consider memos to staff, extended memos to research funds and proposal managers, and targeted communications to researchers. Research institutions that also have Institutional Review Boards (IRBs) or Sponsored Program/Research Offices might want to consider including explanations and process changes in their procedures documentation and websites. Implementation is most effective, though, when the requirements have been translated directly to functional roles and responsibilities. This means explaining requirements to:

• **Researchers** about ORCID identifiers, research tracking, DMPs, research products registration and deposit

• **Research fund and IRB managers** about new DMP requirements and costs for research proposals

• **Research program managers** for proposal tracking in TRB RiP and noncompliance

• **Technologists** about data repository requirements and assessments

• **Contracts managers** about data management support agreements where new arrangements need to be made

• **Librarians and internal publishing offices** for publication tracking and the provisioning of advice on institutional repositories

Implementing a short-term policy gives research institutions time to gain a better understanding of the issues, to make and learn from incremental progress, and to engage in conversation with these emerging and evolving fields. It is highly recommended that research institutions focusing only on essential requirements identify individuals who can serve as liaisons to the communities of practice engaged in research product and data management.
### 4. Developing Strategies & Policies

#### Table 2. Examples of Policy Short-term Objectives

<table>
<thead>
<tr>
<th>Policy</th>
<th>Short-term Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservation Requirements</td>
<td>Projects: Start with the USDOT essential requirements for which funding types require compliance with the requirement. They are specified on the Public Access Plan website: <a href="https://ntl.bts.gov/public-access/plan-executive-summary">https://ntl.bts.gov/public-access/plan-executive-summary</a>.</td>
</tr>
<tr>
<td>Tracking</td>
<td>Start with internal systems that track submission of funding proposals and see if they can be adapted to track elements of compliance. Focus on tracking of DMPs, ORCID identifiers, and submission of reports and data to the NTL at project completion. Also leverage information in the TRB RIF record.</td>
</tr>
<tr>
<td>Storage</td>
<td>Start with one or more of the suggested repositories listed by the NTL at <a href="https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan">https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan</a>.</td>
</tr>
<tr>
<td>Confidentiality</td>
<td>Data that could endanger research subjects’ confidentiality, including those containing direct identifiers and/or indirect identifiers, will be considered confidential and restricted-use. These should be shared only through the following methods (requiring approval from the data archive): secure online analysis, restricted-use data agreement, physical data enclave, anonymization, and/or virtual data enclave.</td>
</tr>
<tr>
<td>Oversight</td>
<td>Identify an oversight committee that includes at least one high-level manager and stakeholders from among researchers, research administrators, and data managers as participants. At least one person should be assigned to be up-to-date on essential requirements and any changes to the USDOT requirements and/or available support for compliance.</td>
</tr>
<tr>
<td>Training</td>
<td>Start by requiring training for regular research and anyone responsible for compliance in the state DOT or other transportation research institution. These are people who need to implement in practice early on. For state DOTs, any other Principal Investigators (PIs) outside the state DOT who have projects with relevant funding (covered by the Public Access Plan) should be required to complete training before funds can be allocated. Going Beyond, all staff involved in research should have exposure to preservation process and policy for longer-term culture change.</td>
</tr>
<tr>
<td>Costs</td>
<td>The USDOT has explicitly stated that costs of preservation can be included in research budgets. Thus, a policy that will enable essential compliance is to require researchers to incorporate some costs related to preservation in their budgets. The cost of general infrastructure such as training, tracking, and oversight will likely need to come from non-project budgets, so early planning for these costs (especially initial setup costs) will be important.</td>
</tr>
</tbody>
</table>
4. DEVELOPING STRATEGIES & POLICIES

Going Beyond: Create a Holistic Plan & Long-term Goals

This section is for organizations that already have an institutional policy in place or are interested in creating an institutional policy that goes beyond the minimum. In Going Beyond essential requirements, state DOTs and research institutions will determine their goals (why), the scope and coverage of the plans (what), internal timelines (when), organization-wide processes (how), organizational roles (who), and organizational registries, repositories, and tracking systems.

Establishing Long-term Goals for Preservation

The long-term vision is grounded on the research institution’s own goals for Open Science, which would then be translated into a set of concrete strategies, which should include the following:

- Assess the current state of goals, strategies, roles, and resources
- Define specific goals for a three- to five-year plan, including strategies, roles, and resources specific to accomplishing each goal
- Plan how to fill gaps or define new action items for each goal
- Assess methods and metrics to measure progress against goals

Implementing a Long-term Vision

Assess the Current Situation

Once the long-term vision is established, the next step involves assessing where the research institution is with respect to research preservation. Many institutions have no program in place, so it’s fine to start at the beginning, with asking questions such as:

- **Policy**: What is the policy on scope and coverage of research projects and data to be preserved?
- **People**: Are there defined roles and responsibilities that align with the full research lifecycle?
- **Process**: What is the process for tracking and managing research projects? Will it scale to the long-term goals of scope and coverage?
- **Systems**: What are the institutional registries and repositories? If not institutionally supported, are there arrangements in place with external registries and repositories?

Early planning to cover non-project infrastructure costs is important.
4. DEVELOPING STRATEGIES & POLICIES

Develop Practical Policies and Strategies

A research preservation system cannot be built overnight. Practical policies must be understandable to the people implementing them and manageable with the available resources (staff, time, budget, etc). Concrete policies and strategies must address the items listed below. These topics, as they relate to Going Beyond, are covered in detail in later chapters of the Guide, which are indicated.

- **Preservation Requirements**
  - **Projects:** What research projects are included in the USDOT’s preservation requirement? (Chapter 6)
  - **Data:** What data must be preserved? (Chapter 7)

- **Storage:** Where may/must data be stored and for how long? (Chapter 7)

- **Confidentiality:** Which data are considered confidential and how will confidential data be preserved and shared? (Chapter 7)

- **Oversight:** Who will provide oversight to ensure that the requirements are being met? (Chapter 5)

- **Tracking:** How will applicable research project and compliance be tracked? (Chapter 6)

- **Training:** What training will be required and for whom? (Chapter 10)

- **Data Management Plan:** Does the DMP contain the elements needed to successfully enable researchers to plan for the eventual sharing and preservation of the data? (Chapter 8)

- **Costs:** How will preservation costs be handled? (Chapter 11)

Measuring Progress

Measuring progress is an essential component of being able to move from a short-term focus of compliance to a long-term focus of developing policies that are fully aligned and embedded within the state DOT’s culture and work practices. Although progress may be slow in the beginning, developing and putting these policies into place will pick up over time as researchers and staff become more acclimated to what they are being asked to do. As a part of facilitating this transition from the old way of doing things to the new, careful monitoring of the state DOT’s progress will be needed to determine what elements of the policies might need more attention or adjustment. Building in measurements of the policies is also important to determine cost-effectiveness and the rate and extent of culture change that is (or is not) taking place. If measurements reveal that the policies and supports are not achieving the desired results or compliance is leveling off it may be that adjustments in time, resources, and attention need to be made.
Chapter 12 contains a more detailed set of suggested ways to measure progress, including periodic surveys, tracking compliance, and so on. However, below is a set of basic measures that can be tied to policies:

- Count projects that fall under the new public access policy and where actions have been taken to address this policy
- Count projects with active DMPs
- Count projects that have preserved reports
- Count projects that have preserved data
- Count people involved in research and what percent have ORCID identifiers
- Count the number of people who have gone through training
- Measure and account for the costs of complying with the Public Access Plan including DMP oversight, training, preserving research products, etc.
- Survey stakeholders on their knowledge, actions, expectations, and needs over time

**Chapter Checklist**

From this chapter, you should be able to:

- Understand the USDOT’s and the research organization’s roles and responsibilities
- Develop a strategy to comply with essential requirements
- Understand the differences between short- and long-term strategies
- Understand how to grow short-term into long-term strategies
- Consider a research organization’s long-term interest in Open Science
- Importance of making incremental progress
- Importance of monitoring the work of communities of practice
4. DEVELOPING STRATEGIES & POLICIES
CHAPTER 5.
Roles & Responsibilities

In This Section
» Key Roles for Essential Requirements
» Going Beyond: Expanding Stakeholders, Roles and Responsibilities
» Aligning Stakeholders with Roles and Responsibilities
» Building a Research Preservation-Conscious Culture
» Chapter Checklist

This chapter provides guidance on identifying and operationalizing roles, including those necessary to meeting essential requirements and those needed to go beyond the minimum. State DOTs and other organizations conducting research should take the opportunity to consider whether they want to play a direct role in achieving the Open Science vision, and, if so, what the extent and nature of that role might be. An expanded vision will mean expanded roles across research institutions.

Key Roles for Essential Requirements

As discussed in Chapter 4, the USDOT has established and taken responsibility for several key roles in achieving essential compliance. Specifically, the USDOT has defined the vision and the goal (why), specified what should be submitted, when it should be submitted and where. Research institutions are responsible for determining who should be involved in compliance at their organizations, and how compliance will be achieved. The table below identifies a set of organizational roles that may already exist within the organization and that are relevant for each essential-compliance task.

<table>
<thead>
<tr>
<th>Essential Compliance Tasks</th>
<th>Possible Existing Organizational Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track Research Funding Proposals</td>
<td>Institutional Research Board, Research and Sponsored Programs Officers, Research Managers, Researchers</td>
</tr>
<tr>
<td>Develop Data Management Plans</td>
<td>Data Librarians, Institutional Repository Managers, Data Services Support Teams, Researchers, Research Teams</td>
</tr>
<tr>
<td>Estimate Costs/Budgets for Research Data Management</td>
<td>Institutional Repository Managers</td>
</tr>
<tr>
<td>Register Researcher ORCID identifiers</td>
<td>Researchers</td>
</tr>
<tr>
<td>Identify and Approve Repositories for Research Data Management</td>
<td>Institutional Repository Managers, Information Technology Managers, Research Managers, Librarians</td>
</tr>
<tr>
<td>Track and Submit Research Publications</td>
<td>Researchers, Reference and Research Librarians</td>
</tr>
<tr>
<td>Track and Submit Research Final and Technical Reports</td>
<td>Research and Sponsored Program Officers, Research Managers, Researchers</td>
</tr>
<tr>
<td>Close-out and Audit Research Project Compliance</td>
<td>Researchers, Research and Sponsored Program Officers</td>
</tr>
</tbody>
</table>
Research organizations and state DOTs vary in their structures, roles, responsibilities, and in the nature and extent of their research. Each research institution will need to interpret and adapt the roles and responsibilities to suit their situation. Where research is an internal service, a research institution may need to work closely with management and researchers to develop a plan for essential compliance.

Going Beyond: Expanding Stakeholders, Roles and Responsibilities

Going Beyond means expanding an organization’s understanding of stakeholders to include both producers and consumers of their research assets. Open Science adds value to all aspects of an organization’s research. The challenge is to determine the level of investment in managing and making those research assets available and accessible.

Going Beyond: Stakeholders

There are several types of stakeholders for publications and data, and they can assume a variety of roles within and outside of a research institution or state DOT. Some will provide support or resources that are needed to develop, manage, share, curate, and preserve the data. Others will require information and actions from people in the research institution or state DOT to ensure that requirements and expectations for the data are being met.

Specific stakeholders for data and publications will vary, depending upon factors that include the following:

- Nature of the data being developed or worked with
- Intended purpose of the data
- Intended audience for the data
- The institution and their requirements or needs for the data
- Requirements made by funding agencies or other external bodies for the data
- Access to resources (staff, money, time, etc.)

Identifying stakeholders early on in the research, ideally prior to generating any data, and mapping when and how interactions will occur will help take full advantage of the support they can provide and keep the work moving forward without delays.
Going Beyond: Expanding Roles and Responsibilities

The research lifecycle model is an important tool for state DOTs or research institutions that aim for the long term and expand their goals for Open Science. This lifecycle model is crucial because it helps visualize the full range of research products an organization is generating. It also provides a framework for identifying internal and external stakeholders, and for determining whether all of the support roles and resources are in place to support the vision. Lifecycles are specific to a research project, as the steps taken and the data generated are unique to the nature and intent of the research. However, some common categories can be identified, which may be useful as a starting point. Lifecycle models, such as the US Geological Survey’s model, shown in Figure 4, depict the stages of development that a research project, particularly the dataset, will pass through in its development. In other words, they define what happens when.

Defining the stages of the research lifecycle and what activities will take place in each stage will help organizations identify who will be providing support for the work, and who may require information and action from the organization. In other words who will do what, when, or who will need what, when. This provides a springboard for defining the roles and responsibilities for achieving an expanded vision.

The USGS lifecycle model describes six stages, each of which includes activities related to preservation:

- **Plan** refers to the early work of designing the project, getting it funded, and once funded, getting it started. This is the stage at which a data management plan would be drafted and part of the process could include establishing who is responsible for contributing to, drafting, approving and communicating the DMP. In addition, high-level information about the types of data that will be generated, what support or infrastructure is in place or will be needed to enact the DMP (such as storage, security or other equipment), and who will be responsible for ensuring this support or infrastructure is in place.

**Figure 4. US Geological Survey’s Research Model**
5. ROLES & RESPONSIBILITIES

• **Acquire** covers the actions taken by staff to support generating or collecting the data for the project. How is each type of data collected in the project, by whom and through what method? In addition, what documentation of the data collection process will be needed for others to understand and use the data, and who is responsible for ensuring proper documentation?

• **Process** refers to the step between data collection and data analysis, in which data are typically transformed into an analyzable (and eventually shareable) form. Here, documentation of the resulting dataset, including developing a codebook, will be critical.

• **Analyze** refers to the actions taken by staff to analyze the data to answer the research questions of the project. This stage could include information on the analysis procedures, how they affect or impact the data and inform the findings derived from the data, and who is responsible for performing and documenting the analysis.

• **Preserve** covers the outputs of the research (papers and data), how/where they will be shared, what actions will be needed for preparing these outputs for sharing and who will be responsible for the process.

### Table 3. Worksheet to Help Identify Stakeholders for the Research Project and/or Dataset

<table>
<thead>
<tr>
<th>Role</th>
<th>Name (or Title if name is not known)</th>
<th>Support Provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal Investigator</td>
<td>Martha Johnson</td>
<td>Budgets and ensures time for preservation within the project; authors DMP; submits all eligible publications and data to NTL</td>
</tr>
<tr>
<td>Researcher</td>
<td>Joseph Garcia</td>
<td>Overseeing gathering of survey data</td>
</tr>
<tr>
<td>Researcher</td>
<td>Rebecca Aoki</td>
<td>Overseeing data collection on noise levels at intersections</td>
</tr>
<tr>
<td>Lab Manager</td>
<td>Jeff King</td>
<td>Stores and processes the noise level data</td>
</tr>
<tr>
<td>Compliance Personnel</td>
<td>Debra Strauss</td>
<td>Reporting to USDOT; informing personnel of USDOT requirements</td>
</tr>
<tr>
<td>Data Repository</td>
<td>Support staff at ICPSR to deposit the data; support staff at ROSA P to deposit articles and register the data.</td>
<td>Guidelines on preparing data for deposit, sharing and archiving</td>
</tr>
</tbody>
</table>
5. ROLES & RESPONSIBILITIES

- **Publish/Share** refers to the actions and infrastructure needed to ensure long-term access to articles, data and other products of the research project as determined by the needs of the organization conducting the research and the requirements of the funding agency.

An important goal of creating a research lifecycle model is to identify the considerations and actions needed in order to be able to share and archive the data generated from a research project. Remember to complete the information in the research lifecycle table with an eye towards the perspective of the data, giving due consideration to its eventual public release and preservation.

In addition to those roles and responsibilities required to support compliance with essential requirements, Going Beyond may also include new roles and expanded responsibilities for existing roles. For example, stakeholders interested in consuming an organization’s research products and data may include funding agencies, other research institutions, research communities, external researchers or researchers in other domains, departments, and schools, scholarly societies, publishers, entrepreneurs and innovators within and beyond the transportation industry, and the general public. Internal

<table>
<thead>
<tr>
<th>Information Required</th>
<th>Time Frame/Stage they Need to be Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the requirements for compliance; what makes a good DMP</td>
<td>All</td>
</tr>
<tr>
<td>What documentation is needed to develop a codebook; IRB guidelines and how they impact sharing and archiving data</td>
<td>All</td>
</tr>
<tr>
<td>What documentation is needed and how it should be structured</td>
<td>All</td>
</tr>
<tr>
<td>Who should have access to the data and when; what elements of the data should be archived</td>
<td>Data collection; data analysis; data archiving</td>
</tr>
<tr>
<td>Documentation that the USDOT requirements are being met</td>
<td>Proposal planning; sharing results; project wrap up</td>
</tr>
<tr>
<td>Deposit of the data</td>
<td>Sharing results; data archiving</td>
</tr>
</tbody>
</table>

---
stakeholders might now expand to include researchers conducting research without funding, funding from internal sources, or funding from private foundations. Depending on what stakeholders are seeking, Going Beyond may also include preservation of research products produced in earlier stages of the lifecycle.

Wherever the scope and coverage of research expands, responsibilities will also expand. Research funds and proposal managers may find that they now have responsibilities for tracking all research projects and all of the products they generate. Compliance managers may have expanded responsibilities for tracking research across the full organization. Information technology managers may now have expanded responsibilities for managing registries and repositories for what was previously stored on departmental drives or personal workstations. Laboratory managers may have new responsibilities for ensuring that laboratory data quality is assured or that laboratory products are supported by full documentation. Librarians’ responsibilities may expand to include management of gray literature (i.e., reports not published in academic journals) research products, and the establishment of institutional information management policies and practices. Going Beyond the essential requirements will have a significant impact on roles and responsibilities and should be well anchored in the plan discussed in Chapter 4.

Aligning Stakeholders with Roles and Responsibilities

The next step is to identify who the stakeholders for a specific research project are or will be in the future. This is done by identifying who in the state DOT or research institution performs the roles that were described in the previous step. The “Worksheet to Help Identify Stakeholders for the Research Project and/or Dataset” on page 39 provides a template for assigning specific people to the roles and responsibilities related to preserving research products. In attaching people to stages in the research lifecycle, be sure to identify any gaps in the support available at your organization. This will help in determining whether additional staff or additional training for existing staff are needed.

Building a Research Preservation-Conscious Culture

Expanding stakeholders, roles and responsibilities will have an impact on your organization’s culture. It is important to consider how the team will handle managing, sharing, curating, and preserving the data, and how all of these roles and responsibilities will align and work together. The new culture will be influenced by internal and
external requirements — each expansion of responsibilities may involve bringing in new perspectives and principles. In addition, communication among and across all of these roles is critical to achieving the vision. Any plan for Going Beyond should be sure to address how stakeholders will communicate with each other and how often. Write out plans and share them with stakeholders as appropriate. Communication of expectations of roles and responsibilities (and how they connect across the team) is critical. All of these are important pieces of the long-term strategic plan.

Chapter Checklist

This chapter will help you to:

☑ Identify key roles for compliance with essential requirements
☑ Map roles to job titles
☑ Review current responsibilities
☑ Define expanded stakeholder model for Going Beyond
☑ Define expanded roles and responsibilities for Going Beyond
☑ Anticipate impacts to organizational culture
☑ Anticipate new communication patterns
☑ Design and implement new governance models and methods
CHAPTER 6.
Managing Research Publications

In This Section
» Determining Essential Requirements Eligibility
  » Research Publications and Reports Eligible under Essential Requirements
  » Publications and Reports Not Eligible under Essential Requirements
» Tracking and Identifying Eligible Research Products
  » Technical and Final Research Project Reports
  » Peer-reviewed Scholarly Publications
» USDOT Registries and Repositories
» Going Beyond Essential Compliance: Including All Research Products
» Understanding Publication Preservation Basics
» Understanding Metadata Standards and Practices
» Understanding Legal and Copyright Strategies
» Chapter Checklist

Determining Essential Requirements Eligibility

Research Publications and Reports Eligible under Essential Requirements
As described in Chapter 2, research publications are defined by the USDOT Public Access Plan as follows:

• Any intramural technical or final reports
• Any federally funded scientific extramural research project’s written deliverables (e.g., technical/final reports)
• Written deliverables that are published by the USDOT, with no embargo period
• Written deliverables published externally by USDOT scientists or extramural USDOT-funded scientists, with an embargo period.
• Any final peer-reviewed manuscript accepted for publication

Embargo refers to a period of time, usually not more than a year, where access to the article or other published material is restricted by the publisher to subscribers of the journal or other publication. After the embargo period expires, the manuscript for the article or other material is made publicly available by the publisher or through an open access repository such as the one provided by the National Transportation Library (NTL).
In a nutshell, this means any final written records of the research results are eligible for essential compliance. The Guide breaks the eligible products into two basic categories: (1) formally published and peer-reviewed research publications, and (2) final project and technical reports. The distinction between the two is important for understanding how to track and where to submit to ensure essential compliance.

**Publications and Reports Not Eligible under Essential Requirements**

The Public Access Plan does not apply to the following:

- Written deliverables or data submitted for publication prior to the plan’s effective date or to digital data generated prior to December 16, 2015 (the effective date of the USDOT’s final plan)
- Preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects (e.g., laboratory samples) associated with conducting research
- Classified research
- Sensitive or personally identifiable information

**Tracking and Identifying Eligible Research Products**

**Technical and Final Research Project Reports**

The majority of research products resulting from federally funded transportation related research will be final project reports and technical reports. These are typically referred to as gray literature as they are made available outside of traditional commercial or academic publishers. Academic libraries have established policies and practices for managing peer-reviewed, published research results such as journal articles, books, chapters, or series. In contrast, gray literature is largely managed by the researchers who produce them or technical libraries charged to collect them. Practices for managing gray literature vary across organizations and within the information management domain.

The USDOT and the TRB have provided easy-to-follow methods for submitting these reports. In collaboration, they also apply consistent and effective information management practices to ensure discovery, availability, and access — including the creation of metadata (discussed in detail later in this chapter).

Research final reports and technical reports are more likely than peer-reviewed publications to contain sensitive or personally-identifiable information, since these issues are likely to be caught in the peer-review process. Where there is no peer-review process, research organizations must take responsibility for ensuring that classified,
sensitive, or personal identification information is not made publicly available. Where final project and technical reports are classified, established practices may allow the sharing of metadata publicly, but only allow authorized or permissioned access to the publication and data. This also involves the securing of any explanatory materials such as a data dictionary or code book, using encryption methods to deliver or transfer content to authorized users, and ensuring that the research publication and sensitive content is not stored in a publicly accessible repository. Generally, these materials should not be submitted for public access.

**Peer-reviewed Scholarly Publications**

State DOTs and research institutions are responsible for tracking and submitting those scholarly peer-reviewed publications that result from research projects that are eligible under the USDOT Public Access Plan (see Chapter 2). Research publications that are deliverables of the project must be submitted within one year of the project closeout. Ideally, research institutions will establish proactive methods for tracking and submitting publications as they are produced. The best strategy involves establishing good communication pipelines between researchers and research fund managers and information professionals. Librarians and information professionals have professional relationships with the National Transportation Library and can help to track publications. When an article, chapter, or book is accepted for publication, the researcher should either submit the publication to NTL themselves or work with a librarian to ensure it is submitted according to guidelines. Depending on the research organization’s internal roles, responsibilities, and processes, the research funds manager or institutional review board liaison may also be included in the communications. A supported communication model is illustrated below.

**Figure 5. Communication Network**

![Communication Network Diagram]

The best strategy involves establishing good communication pipelines between researchers and research fund managers and information professionals.
Tracking Eligible Publications Retroactively

Research funding organizations face significant challenges tracking eligible publications when they are not submitted in a timely manner by researchers or other designated individuals. Retroactive discovery of eligible publications is particularly challenging because scholarly publications frequently do not include funding source information. Retroactive tracking of eligible publications is a complex and labor-intensive process when performed in the open web or against commercial databases, and many research funding organizations do not have the resources to support such an effort. Funding organizations may consider two important sources that provide quality-controlled access to funding information: CrossRef (http://search.crossref.org/) and its Funders Registry (https://www.crossref.org/services/funder-registry/). CrossRef and its Funders Registry, which work together as a collaborative project of scholarly publishers and funding agencies, supports a standard way of reporting funding sources for published scholarly research, and thus tracking that research back to its source (and associated requirements for public access).

USDOT Registries and Repositories

The USDOT essential requirements for submitting written research products are simple and easy to follow:

- Send a single email addressed to three recipients, including the USDOT Research Hub (Research.Hub@dot.gov, NTL (NTLDigitalSubmissions@dot.gov) and TRB (TRIS-TRB@nas.edu).
- The email should include the following information:
  - Final Report URL(s) or PDFs for any resulting publications
  - URL(s) to and associated descriptive metadata for any final datasets and the arising from the research project
  - The funding agreement number of the project
  - The Research Hub (RH) Display ID for the project (from the TRB RiP database entry created at the start of the project)
  - ORCID identifiers (unique researcher IDs) for all publication author(s) and contributor(s)
  - Any documented project outputs or outcomes resulting from the research project
- Append to the email any final peer-reviewed manuscripts, final project or final technical reports. Remember, ROSA-P serves as a persistent archive for publications, so it is important to provide a copy of the report or manuscript to support this goal.
- Ensure that the email includes the URLs for the eligible final research data.
Sending a single message to the three recipients listed above will allow USDOT and TRB to coordinate the tracking and management of the research products. USDOT supports both Research Hub (RH) and NTL's Repository & Open Science Access Portal (ROSA P). RH is a searchable database of past and current USDOT-sponsored research, development, and technology project records. ROSA P is a final report repository and data registry for USDOT-funded and -created research. The TRB manages the Transportation Research Information Database (TRID) which serves as a registry for all final project and technical reports. The TRB also maintains a mirror version of the Research Hub which stores an archival copy of the final project and technical reports. Links for technical reports listed in the USDOT Research Hub are also visible for copies deposited in NTL's archive.

The NTL's ROSA P and TRB's TRID both provide searchable access for all written research projects. Going forward, the Public Access Plan calls on the National Transportation Library to host a searchable database of DMPs. This will take the form of a collection within ROSA P, called USDOT Public Access Data Management Plans. Reports housed or accessible through the NTL are available without charge. For public access, NTL's functionality will be expanded to include the capability for managing the embargo and compliance processes required by this Public Access Plan. ROSA P does the following:

- Supports nonproprietary preservation standards and archival formats for publications and their associated content to help make research products more accessible to future generations
- Provides practical backup, migration, and technology-refreshing strategies to ensure sustainable access to scientific research
- Enables sharing of publication archives across the federal, academic, and business communities

**Going Beyond Essential Compliance: Including All Research Products**

Going beyond essential compliance provides state DOTs and other research institutions with an opportunity to expand availability and access to a broader range of research products. Today, essential compliance pertains only to final research publications, technical reports, and final project reports that result from federally funded research projects. Organizations conducting the research may generate research products that extend beyond those funded by the USDOT. As noted earlier, they include unfunded research, research funded by internal sources, and research that is funded by private sources or foundations. A logical next step — and one that supports
the Open Science vision — would be for research institutions to consider managing all of their research products in a way that aligns with the good practices and vision set forth by the OSTP memo. Some benefits of this to state DOTs include:

- Improved access to and tracking of all of the research publications of the state DOT for state DOT use
- Less redundancy of research projects funded and more efficient use of results already supported
- Greater impact of state DOT research on the broader transportation community
- Increasing expectations of funding organizations for more rigorous management and access to research results
- Increasing expectations for access and reuse from researchers, research communities, private sector businesses and industries, and the general public
- Increasing expectations from publishers to submit supporting research data as part of the peer-review process

Going beyond to increase the scope and coverage of research means that state DOTs and research institutions must assume some of the responsibility for the services that are provided by NTL and TRB. This means identifying or establishing a local registry to track publications, final project, and technical reports, and identifying or building a repository solution to store written research products. Research institutions that are affiliated with a university will have access to research library services and sources. For these institutions, Going Beyond means understanding:

- Preservation basics for written research products, and engaging archivists and preservation experts to ensure long-term preservation options are supported and are appropriate for gray literature products
- Metadata standards and practices, and ensuring that cataloging and metadata librarians are familiar with metadata schemes and practices that support subject domains or full project lifecycles (e.g., going beyond metadata standards for publications)
- Collection development policies to reflect new scope and coverage of gray literature
- Information management lifecycle and providing lifecycle information management training for metadata and cataloging librarians accustomed to working with formal publications
Research institutions in state DOTs, though, may not have access to either library services or sources. For a state DOT, Going Beyond means:

- Understanding metadata schema
- Understanding how to create metadata
- Having a registry or library catalog to maintain metadata
- Having a policy in place for choice of persistent identifiers
- Understanding how to generate a persistent identifier for written research products

Pursuing this knowledge requires a deeper dive than this Guide can provide, but the next section gives an introduction to publication preservation to help with further study.

Understanding Publication Preservation Basics

Digital preservation takes active management to insure that content is usable and understandable over time. Successfully preserving digital content is more than just finding a hard drive to store it. This means that for textual publications, open formats (e.g., Open Office) and PDF (preferably PDF/A) are highly preferred over non-proprietary binary formats that may be difficult to render in the future. The Library of Congress maintains a Recommended Formats list (https://www.loc.gov/preservation/resources/rfs/) that can help in selecting the optimal long-term preservation format for publications. Another component of preservation involves the accounting for and tracking of any changes made to the digital content over time. Researchers need to be able to trust that the content was not altered deliberately or accidentally since the content was published, released or otherwise deemed to be complete.

Understanding Metadata Standards and Practices

Metadata plays a number of important roles in sharing and preserving information including:

- **Identifying the object and distinguishing it from others** by describing who created it, what it is called, when it was created, who published it, the format in which it is offered, its version or edition, report number, contract number, size or length, the language in which it is written, etc.

- **Describing the nature of the content**, including its subject or topic, its geographical relevance, and the type of content (report, chapter, article, etc.).
6. MANAGING RESEARCH PUBLICATIONS

- **Supporting discovery of the object**, including searching or browsing for it in a repository or other digital library
- **Providing information to support or constrain access**, including where the publication can be found, its copyright status, its security classification, its disclosure policy and status, and detailed information on who can access it
- **Chronicking its preservation and management history**, including the records retention and disposition schedule that pertains, any transformations or format conversions that have been completed, and any redactions that have been made to the content, etc.

Metadata creation is a professional task. A recommended best practice is for research organizations to assign the role of metadata creator and publication submitter to an information professional who is familiar with metadata standards, the authoritative reference sources that support them, and the submission process. This individual will create metadata, and/or review metadata created by researchers. In addition to ensuring that the research organization is in compliance with the USDOT’s Project Open Data requirements, information professionals can ensure that publications and their associated datasets are registered in organizational or domain-specific sources. Information professionals should also be kept informed of changes to publications to ensure that the metadata descriptions are accurate and relevant.

**There are many different metadata standards and schema. For essential requirements, researchers should include** a .json metadata file with their final reports, following the Project Open Data metadata schema, with their submission to NTL, TRID, and RH. This is a good standard to consider for Going Beyond. However, research organizations should consult with their researchers on the metadata standards that support their research communities.

Research organizations should strive to provide good quality metadata to manage their written research products. A good quality metadata record does the following:

- Has accurate, consistent, and complete values for all required metadata fields
- Has complete and sufficient values to support discovery by other researchers and any member of the public who may wish to find or access it
- Complies with specifications for metadata values where they are available, i.e., distinct subfields for repeatable fields (multiple authors), full and standard date format for publication date
- Conforms with data entry instructions (last name, first name, middle name versus single string of first name middle name and last name
• Leverages metadata values from authoritative reference sources and standard vocabularies wherever they exist — to support quality and to represent the language of the potential audience

• Understand the metadata which must be submitted to NTL, TRID, and RH

• Contains actionable authorization values (i.e., limiting access to authorized individuals or groups)

### Table 4. Project Open Data Metadata Schema

<table>
<thead>
<tr>
<th>Field</th>
<th>Label</th>
<th>Definition</th>
<th>Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>Title</td>
<td>Human-readable name of the asset.</td>
<td>Always</td>
</tr>
<tr>
<td>description</td>
<td>Description</td>
<td>Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest</td>
<td>Always</td>
</tr>
<tr>
<td>keyword</td>
<td>Tags</td>
<td>Tags (or keywords) help users discover the dataset</td>
<td>Always</td>
</tr>
<tr>
<td>modified</td>
<td>Last Update</td>
<td>Most recent date on which the dataset was changed, updated, or modified</td>
<td>Always</td>
</tr>
<tr>
<td>publisher</td>
<td>Publisher</td>
<td>The publishing entity and optionally their parent organization(s).</td>
<td>Always</td>
</tr>
<tr>
<td>contactPoint</td>
<td>Contact Name and Email</td>
<td>Contact person's name and email for the asset.</td>
<td>Always</td>
</tr>
<tr>
<td>identifier</td>
<td>Unique Identifier</td>
<td>A unique identifier for the dataset or API as maintained within an Agency catalog or database.</td>
<td>Always</td>
</tr>
<tr>
<td>accessLevel</td>
<td>Public Access Level</td>
<td>The degree to which this dataset could be made publicly-available, regardless of whether it has been made available. Choices: public (Data asset is or could be made publicly available to all without restrictions), restricted public (Data asset is available under certain use restrictions), or nonpublic (Data asset is not available to members of the public).</td>
<td>Always</td>
</tr>
</tbody>
</table>
6. MANAGING RESEARCH PUBLICATIONS

- Contains actionable values that define the legal ownership and liability constraints associated with using the content
- Contains actionable values pertaining to any distribution constraints and liabilities

Understanding Legal and Copyright Strategies

To Go Beyond essential requirements, research organizations should have defined and clearly communicated policies and practices for legal and copyright strategies. Research institutions affiliated with academic institutions will have guidance and practices to draw from. An expansion of scope and coverage, though, may introduce new challenges for existing policies and should be reviewed by legal experts. Additionally, librarians and information professionals will be good sources of guidance on copyright strategies. State DOTs and research institutions not affiliated with a library or copyright office should consult with their legal teams.

Chapter Checklist

From this chapter you should be able to:

☑ Identify eligible written research products
☑ Distinguish formal publications and manuscripts and final project and technical reports
☑ Understand how to submit written research products to NTL and TRB
☑ Understand what information to provide to NTL and TRB
☑ Understand when to submit to NTL and TRB
☑ Understand the metadata services provided by NTL and TRB
☑ Understand what Going Beyond means for state DOTs and research institutions
CHAPTER 7.
Managing Research Data

In This Section
» Definition of Research Data
» Explaining Essential Requirements for Research Data
» Going Beyond: Research Data Management and Access
» Understanding Data Preservation
» Deciding What to Preserve: Data Scope & Coverage
» Deciding Which Formats to Preserve
» Managing Quality of Research Data
» Understanding Metadata Standards and Metadata for Transportation Data
» Deciding Where to Preserve Data
» Understanding How Long to Preserve Data
» Chapter Checklist

Definition of Research Data

The USDOT is a complex research environment, containing a range of data. A literature review of the Transport Research International Documentation (TRID) database produces a rich set of references to operational research data management practices in transportation. The results illustrate the many facets of transportation data and identify some natural research communities within the larger domain. These include:

- General Data Management Issues
- Air Transport Data
- Bridge Asset Data
- Construction Data
- Crash and Safety Data
- Driver Data
- Engineering Data
- Freight and Cargo Data
- Environmental and Land Use Data
- GIS Data
- Intelligent Transport System Data
- Marine Transport Data
- Materials Science Data
- Planning and Design Data
- Railway Data
- Road Data
- General Statistical Data
- Traffic Data
- Transit Data
- Transportation Information Systems and Signage Data
- Vehicle and Asset Data
- Video and Photogrammetric Data
- Weather and Climate Data
Explaining Essential Requirements for Research Data

This aspect of essential compliance focuses on managing, accessing, and using the data that supports federally funded transportation research. The USDOT has provided guidance, but left some important decisions and actions to research organizations. For written research products, the USDOT and TRB have provided the registry for discovery and the repository for storage. However, it is the responsibility of the organization to select a registry and repository for the supporting research data. This means that to be in compliance, state DOTs and other research organizations must make the key decision of which registry and repository to use.

The USDOT has provided guidelines for selecting a data registry and repository for preserving and providing access to data. Additionally, the USDOT has provided a list of repositories that meet essential requirements. This means that organizations must do the following:

- Ensure the data preserved are those used to draw research conclusions in the written research product
- Guarantee that data is stored in an open format, or describe in DMP which proprietary formats are used, and why
- Certify the quality of the data and that data is interpretable, understandable, and usable by providing explanatory materials within the data package
- Choose a compliant data registry and repository to store data and make sure that those chosen:
  - will meet the guideline of conformance to the USDOT Public Access Plan at https://ntl.bts.gov/public-access/guidelines-evaluating-repositories
  - are able to generate and maintain a persistent identifier for the data
  - has a long-term preservation strategy
- Create metadata for the research data to support discovery, availability, and access

Evaluating Repositories for Conformance with the Public Access Plan

The USDOT has provided a list of repositories that currently meet their essential requirements (see page 11 for a listing). As this is an evolving field, research organizations are also advised to use the Registry of Research Data Repositories’ searchable listing of data repositories as a starting point to locate potential archiving options for their data. Researchers evaluating local or other data repositories as the option for storing and preserving their data should ensure the repository will:

- Promote an explicit mission of digital data archiving
- Ensure compliance with legal regulations and maintain all applicable licenses covering data access and use, including,
if applicable, mechanisms to protect privacy rights and the confidentiality of respondents

• Have a documented plan for long-term preservation of its holdings
• Apply documented processes in managing data storage
• Perform archiving according to explicit workflows across the data lifecycle
• Enable users to discover and use the data, and refer to data in a persistent way through proper citation
• Enable reuse of data, ensuring appropriate formats and application of metadata
• Ensure data integrity and authenticity
• Be adequately funded and staffed and have a system of governance in place
• Possess a technical infrastructure that explicitly supports the tasks and functions described in internationally accepted archival standards, like Open Archival Information System (OAIS)

Complying with essential requirements means that state DOTs and other research institutions have to make choices with both short- and long-term implications. A short-term strategy to meet essential requirements means selecting an option that will be conformant with the USDOT Plan. An organization’s choice of repository is an essential part of the DMP, which is now part of every research funding proposal. Given the impacts of this choice, it is important for state DOTs and other research institutions to make an “institutional” rather than individual decision. Allowing researchers to make individual decisions may result in: (1) increased and redundant institutional costs if multiple repositories are chosen, (2) a scattering of the institutions’ data across different repositories, and (3) a challenging foundation from which to define a long-term strategy. Institutions should make short-term strategic choices that will also support their long-term goals.

**Key Websites**

https://www.re3data.org

**Going Beyond: Research Data Management and Access**

There are several ways an institution can move beyond essential compliance in the area of data management and access. Some of these include the following:

• Expand the scope and coverage of research data by including data from fields other than transportation, including data from
sources other than USDOT, and including more of the data than just what was used to produce analyses for deliverables
• Broaden the scope and coverage of formats supported by the solution
• Include preservation services in the solution
• Expand data quality management and assurance
• Choose metadata standards and provide metadata services
• Build an institutional repository solution

Understanding Data Preservation

Digital preservation is not equal to basic data storage, but is separate and unique. While basic data storage addresses short-term access and business continuity needs, the data are at high risk of loss over the long term without the active intervention provided by data preservation.

Unfortunately, few data appear to be deposited into a long-lived repository with dedicated digital preservation systems and staff — including hardware, metadata, and organizational — ensuring the data are available and understandable long-term. Devices fail, formats become obsolete, and accidents or disasters can destroy data.

A key aspect of digital preservation is guaranteeing the integrity of content over time, by ensuring that a file's essential elements are preserved, context is documented, and content is traceable to its

Figure 6. Digital Preservation: A Real World Example

Digital preservation is “the active management of digital content over time to ensure ongoing access.” Whereas hard copy materials require minimum maintenance to last decades or centuries, digital content requires active management to make sure they can be accessed over the long-term. Consider, for example, books created centuries ago. Most can still be viewed and understood by modern eyes. The paper may yellow or become brittle, but the contents would still be readable with the naked eye. In contrast, consider a 9-track tape just a few decades old. To read, you need specialized equipment and trained staff. The media itself is often much more fragile and can become corrupted or destroyed more readily than simple paper.
point of creation. While the digital preservation process can differ considerably depending upon the type of data being preserved, it is essential that the integrity of the information be the foremost goal.

Deciding What to Preserve: Data Scope & Coverage

Researchers generate many types and versions of data during the course of a research project. Going Beyond, state DOTs and research institutions should have a set of guidelines that can help determine which data from the research lifecycle to preserve. Not all data can or should be preserved. Storage can be expensive and resource-intensive. At a minimum, researchers should preserve all data necessary to replicate the findings of a published or significant study. Researchers may go beyond the minimum by identifying and organizing the data prior to depositing with a repository, rather than relying on others to determine the important and useful data. Key questions to consider when appraising data include the following:

How significant are the data for research?
Factors to consider include:
- Substantive value of the collected information
- Time frame of the information
- Uniqueness of the collected information
- Relationship to previous studies
- Scope of the data
- Influence of these data in the transportation fields
- Data collection methodology
- Ability to use the collected information for secondary studies

How significant is the source and context of the data, particularly in regard to scientific progress and society?
Data must have demonstrated importance to the community as determined by the following:
1. Substantive value and its influence on scientific knowledge
2. Likely value to science and/or society over time
3. Uniqueness

It is also important to place a high value on data that permit policy analysis and research addressing broad public policy issues or transportation policy more specifically, safety countermeasure evaluation and recommended practice, relationships between transportation policy and other impacts such as environment, mobility, and equity, and transportation economics.
Is the information unique?
Determine whether the data are the only source or are the most complete source for significant information. Data that contain information not available in other sources are more likely to warrant permanent retention than records containing data duplicated in other sources. Even if data are unique, however, they may not warrant continued preservation depending on the other appraisal criteria.

How usable are the data?
Consider how the usability of the data is affected by the way they were gathered, organized, presented, or analyzed. For example, does the scope of the data cover a national population sample or a representative subsample of the population? Do the data offer enough depth and breadth of information to support a wide range of research methodologies?

Consider how the technical considerations affect the usability of the data. For example, some electronic records may pose such technological challenges that extraordinary measures may be required to recover the information, while other records containing similar documentation (either electronic records or records in another format) may be usable with much less effort.

Consider how the physical condition of the preservation media affect the usability of the data. For example, some media may have deteriorated such that the data contained is unreadable.

Are the data related to data in other repositories?
Data that add significantly to the meaning or value of other data already archived are more likely to warrant retention than data lacking such a relationship. Examples would be data that fill substantive gaps, that round out existing subject area concentrations, or that are new version of or additions to data collections in the holdings. Data that are chronological continuations of data already held by the archives are likely to warrant permanent retention, particularly if the older segments of the data are used often.

What are the cost considerations for long-term maintenance?
This consideration should play a significant role only in marginal cases. In such cases, an appraisal should balance the anticipated research potential of the data with the resource implications of retaining them permanently. If data carry significant costs for acquisition, processing, archiving and distribution, the value of the data must clearly outweigh the costs. Other things being equal, data with low long-term cost implications are more likely to warrant permanent retention than those data with high long-term costs.
What is the volume of data?
Data that are clearly of value based on the guidelines listed above should be designated for permanent retention regardless of the size/volume of the data. The size/volume of a collection should be a factor in the decision making only when the permanent value is marginal.

Deciding Which Formats to Preserve

While transportation data come in many formats, interviews with individual researchers within the field found quantitative tabular data the most frequently used. The table “Long-term Preservation of Data Types and File Formats” lists commonly used data formats and provides a high-level reference to data types and formats recommended for long-term preservation purposes. The format used to store data can impact a repository’s ability to preserve the content for long-term access and as such should be carefully considered. Generally, formats that are open source rather than proprietary in nature are better suited for preservation. Consult the repository you intend to submit your data to for guidance on selecting formats.

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Acceptable formats for sharing, reuse, and preservation</th>
<th>Other acceptable formats for data preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative tabular data with extensive metadata (e.g., Dataset with variable/code labels and defined missing values)</td>
<td>SPSS portable format (.por) Delimited text and command (‘setup’) file (SPSS, Stata, SAS, etc.) with metadata information Structured text or markup file of metadata information, (e.g., DDI XML file)</td>
<td>Proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)</td>
</tr>
<tr>
<td>Quantitative tabular data with minimum metadata (Ex. Data with or without column headings or variable names, and no other metadata or labeling)</td>
<td>Comma-separated values (.csv) Tab-delimited file (.tab) Delimited text with SQL data definition statements where appropriate</td>
<td>Delimited text (.txt) of given character set (only characters not present in the data should be used as delimiters) Widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods)</td>
</tr>
<tr>
<td>Type of data</td>
<td>Acceptable formats for sharing, reuse, and preservation</td>
<td>Other acceptable formats for data preservation</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Geospatial data</strong></td>
<td>ESRI Shapefile • Essential: .shp, .shx, .dbf • Optional: .prj, .sbx, .sbn Geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) Tabular GIS attribute data</td>
<td>ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Markup Language (KML) (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) Binary formats of GIS and CAD packages</td>
</tr>
<tr>
<td><strong>Qualitative textual data</strong></td>
<td>eXtensible Markup Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml) Rich Text Format (.rtf) Plain text data, ASCII (.txt)</td>
<td>Hypertext Markup Language (HTML) (.html) Widely-used proprietary formats: MS Word (.doc/.docx) some Proprietary/software-specific formats: NUD*IST, NVivo and ATLAS.ti</td>
</tr>
<tr>
<td><strong>Digital audio data</strong></td>
<td>Free Lossless Audio Codec (FLAC) (.flac)</td>
<td>MPEG-1 Audio Layer 3 (.mp3) but only if originally created in this format Audio Interchange File Format (AIFF) (.aif) Waveform Audio Format (WAV) (.wav)</td>
</tr>
<tr>
<td><strong>Digital image data</strong></td>
<td>TIFF version 6 uncompressed (.tif)</td>
<td>JPEG (.jpeg, .jpg) but only if originally created in this format TIFF (other versions) (.tif, .tiff) Adobe Portable Document Format (PDF/A, PDF) (.pdf) Standard applicable RAW image format (.raw) Photoshop files (.psd)</td>
</tr>
<tr>
<td><strong>Digital video data</strong></td>
<td>MPEG-4 (.mp4) motion JPEG 2000 (.mj2)</td>
<td>Plain text (.txt) Some widely-used proprietary formats: MS Word (.doc/.docx) or MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g., XHMTL 1.0</td>
</tr>
<tr>
<td><strong>Documentation and scripts</strong></td>
<td>Rich Text Format (.rtf) PDF/A or PDF (.pdf) HTML (.htm) OpenDocument Text (.odt)</td>
<td></td>
</tr>
</tbody>
</table>

This table should primarily be used to understand the variety of data types and possible formats. All of these types and formats are pertinent to transportation research.
For more information on data types and file formats, go to: http://www.data-archive.ac.uk/create-manage/format/formats-table

At a minimum, data should be stored in preferred formats to ensure long-term access and reuse. If preferred preservation formats are not available, acceptable formats may be used. Researchers may go beyond the minimum by creating the preferred formats from the beginning of the project rather than waiting until the end to transfer file formats.

Managing Quality of Research Data

Generally accepted digital preservation standards are not the only features to consider when assessing where to deposit data. Data curation is another important aspect. Curation enhances collections so they are complete and self-explanatory for future users. That is, through curation, those responsible for preserving content “ensure that the preserved information is independently understandable to the user community, in the sense that the information can be understood by users without the assistance of the information producer.”

During curation, data are reviewed and cleaned for accuracy and completeness. This may include recoding missing or out-of-range codes, as well as enhancing and adding labels, metadata, and other documentation. Repositories may offer a broad range of curatorial options.

Understanding Metadata Standards and Metadata for Transportation Data

Two important metadata standards often cited in data preservation literature include Dublin Core and the Project Open Data Metadata Scheme. The former was developed with publications in mind and provides a few generic attributes. These attributes may not be sufficient to enable researchers to learn enough about data. In addition, this scheme was designed to support formal and final publications, rather than project documentation or project reports. For research publications generated from contract research or research projects, the Project Open Data Metadata Scheme may be preferred. In addition, there are domain-specific metadata schemas that are relevant to types of research, including geospatial data, environmental science data, biological data, and so on.
Metadata is the area that is most often supported by intermediary services. A common option is to provide a form to enable researchers to generate metadata. These tools support metadata capture, and can also provide easy access to master data vocabularies (e.g., geographical locations, International Organization for Standardization (ISO) country names, Multipurpose Internet Mail Extensions (MIME type), etc.). They can also provide easy access to authoritative vocabularies such as the Transportation Research Thesaurus. Good metadata are critical for basic access and discovery, so it may prove useful to have an intermediary review metadata once the form is complete.

Metadata guidance and services is a current gap in the transportation research management lifecycle. In data collection, transportation researchers overwhelmingly indicated that they used no metadata standards to describe their research products and data. Researchers also noted that metadata creation services were often not available, and even when they were available, they were not sufficient to meet their research needs. The discrepancies between the literature devoted to metadata standards and services and their use by transportation researchers is noteworthy. This is a critical success factor in the operationalization of the USDOT’s Public Access Plan, and it appears to pose a significant gap.

At a minimum, researchers should provide basic description of their data, including using simple Project Open Data Metadata Standard attributes such as the following:

- Title
- Creator
- Description
- Subject terms
- Geographic coverage

Researchers may go beyond the minimum by using established metadata standards used by their disciplines. Several online lists provide guidance on selecting a disciplinary metadata format. Disciplinary repositories can also help generate the preferred metadata.

Need help choosing a disciplinary metadata format? Go to one of the following:

- http://www.dcc.ac.uk/resources/metadata-standards
Deciding Where to Preserve Data

Several options exist for preserving research data (see “Research Data Repository Options”), ranging from metadata registries to focused disciplinary repositories. While it is wonderful to have so many options about where to deposit data, at the least, researchers should select long-lived repositories that commit to core digital preservation standards and provide essential curation services. “Chapter 9” details issues to consider when selecting a repository. Researchers may go beyond the minimum by selecting repositories that offer additional features to enhance the immediate and long-term reuse of the data. The National Transportation Library maintains a regularly updated list of data repositories conformant with the DOT Public Access Plan (https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan). The re3data Registry of Research Data Repository (https://www.re3data.org/) is an online tool to help identify existing international repositories for research data.5

Simple Data Registry of Metadata

A simple data registry of metadata is a catalog of information that links to the actual data collections typically stored elsewhere. Such a registry provides a central inventory of collections relevant to or sponsored by an organization. The registry can be maintained through collecting and refreshing metadata, but without the need to host the actual data. In essence, registries can outsource the digital preservation heavy lifting to long-lived repositories while simply maintaining a catalogue pointing to the data sources. The challenge of such a registry is that it hosts no data. If metadata aren’t automatically harvested, it can be difficult to maintain (add, update) the records for external content.

The ROSA P registry at the NTL hosts documents but does not host datasets, and is thus serving the important purpose of providing a centralized simple data registry for transportation data under the USDOT Public Access Plan. This means that research organizations must find a repository in which to deposit their data so that it will be preserved for the long term and so that metadata can be updated in the ROSA P registry as needed. What follows are descriptions of available repository options, including core strengths and weaknesses.

Figure 7. Research Data Repository Options

<table>
<thead>
<tr>
<th>General Research Data Repository (Breadth)</th>
<th>Institutional Data Repository (Breadth and Local)</th>
<th>Domain-Specific Data Repository (Depth and Curation)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Data Repository Icon" /></td>
<td><img src="image2" alt="Data Repository Icon" /></td>
<td><img src="image3" alt="Data Repository Icon" /></td>
</tr>
</tbody>
</table>
OPTION 1: General Research Data Repository

Beyond a simple registry is the general research data repository. This is a repository that accepts a wide variety of data formats from a wide variety of disciplines. Because of the heterogeneity of data accepted, general research data repositories typically provide basic metadata and take minimum curation or enhancement actions. Two examples of this type of registry are Zenodo and figshare.

General repositories make it very easy to deposit content, although the heterogeneity of collections can make browsing and discovery challenging. Likewise, the broad base of data covered may mean that tools and services, such as online analysis packages, may be lacking for the type of data to be archived.

OPTION 2: Institutional Data Repository

Like general research repositories, institutional data repositories, such as the University of Michigan Deep Blue Data repository, typically cover a wide range of data formats and subject areas, with minimum and broad metadata coverage. Again, this makes data deposit easy, but removes some descriptive power that aids search and discovery.

That said, institutional data repositories have several strengths. They are often supported by the institution’s library, which means they are durable and persist for centuries. They are neutral and interdisciplinary, making them especially attuned to diverse data types data without a home in a more specialized repository. Institutional data repositories are local, and can provide in-person services and resources that more distant repositories cannot.

OPTION 3: Domain-Specific Data Repository

Yet another repository option is a domain-specific repository, such as ICPSR or the Protein Data Bank. Domain repositories “serve a scientific community, which may be a traditional academic discipline, a sub-discipline, or an interdisciplinary network of scientists, united by a common focus.” It offers specialized metadata and curatorial enhancements specific to the domain or specialty. ICPSR, for instance, uses the Data Documentation Initiative (DDI) metadata standard, which records details specific to surveys and other social science data collections typically archived at ICPSR, as well as encodes values at the granular variable level. Disciplinary repositories can also provide other specialized features, including disclosure expertise for human subject research, customized preservation services such as format validations and migrations, and specialized tools to enhance the user experience. Additionally, “they seek to know what [their] community wants and expects in terms of content, format, delivery options, support, and training”? Because disciplinary repositories offer focused, deep collections, users may have an easier time searching and browsing for relevant content.
Before entering into third party agreements with any repository providers, be sure to consider the legal terms carefully. If there are any concerns, be sure to consult with your counsel.

Understanding How Long to Preserve Data

A question that is often asked is how long should the data be preserved? The short answer is generally for as long as the data has value for its community of users. This can be a challenging question to answer as it is difficult to predict how long a data set will have value into the future.

Instead of focusing on the maximum amount of time for preservation, it might be more effective to consider what the minimum amount of time for preserving the data should be and developing a checklist for determining the value of the dataset after the minimum time has elapsed.

Funding and publishing agencies will sometimes define their expectations for the retention or availability of a dataset. Currently the NSF Engineering Directorate states that the “Minimum data retention of research data is three years after conclusion of the award or three years after public release, whichever is later.” If such an expectation is stated, it should be used to inform the minimum preservation period.

A checklist to determine if preservation should continue will naturally vary depending upon the type of data, the needs of the researchers who make use of the data, and the specific value the data has for the community. The questions listed in the “Deciding What to Preserve” on page 57 are also relevant for deciding if preservation should continue. Other key questions might include:

- How much and what kind of use has the data received over time?
- Who depends upon having access to the data? How would they be impacted by not having access to the data?
- Is the data connected to publications or other research outputs? Would not having access to the data harm someone’s ability to understand or trust these publications?

If data preservation will take place through a third party repository (see “Deciding Where to Preserve Data” on page 63), be sure that the preservation services meet the needs of the research organization by carefully reviewing the terms of service, including how long they will commit to preserving the data.
Chapter Checklist

From this chapter, you should be able to:

- Define essential requirements for research data as described by the USDOT
- Interpret essential requirements for your research organization’s transportation research
- Understand the research organization’s roles and responsibilities for data management and preservation
- Understand the difference between a short-term compliant strategy and Going Beyond
- Understand the issues involved in building or supporting a local solution for research data preservation

Endnotes

1 http://www.ihsn.org/sites/default/files/resources/IHSN-WP003.pdf
3 http://www.dpconline.org/docs/lavoie_OAIS.pdf
4 See: http://rd-alliance.github.io/metadata-directory/standards/ or http://www.dcc.ac.uk/resources/metadata-standards
5 https://www.re3data.org/
8 https://nsf.gov/eng/general/ENG_DMP_Policy.pdf
9 http://www.digitalpreservation.gov/about/
CHAPTER 8.
Data Management Plans

In This Section
» Developing and Maintaining Data Management Plans (DMPs)
» Essential Requirements for Developing Sections of Data Management Plans
  » The Data Section of the DMP
  » The Standards Section of the DMP
  » The Access Section of the DMP
  » The Use Section of the DMP
  » The Archiving/Preserving Section of the DMP
» Going Beyond the Minimum for Data Management Plans
» Chapter Checklist

Developing and Maintaining Data Management Plans (DMPs)

Over the past 20 years or so, there has been an increasing interest in making research data more accessible outside of the research project in which they were generated. Recognizing this interest, the National Science Foundation (NSF) in 2010 announced that beginning in January 2011, all grant applications would have to include a two-page Data Management Plan (DMP). The USDOT followed suit and described their expectations for researchers to create and submit a DMP for any funded research in the USDOT Public Access Plan.

As described by funding agencies, DMPs are the means through which researchers: (1) describe the data they will generate as a part of their research process and (2) detail how they will make their data accessible to those outside the project team. However, DMPs can be more than just this. They are a useful exercise that can help guarantee that all stakeholders have a shared understanding of how data are to be developed and handled throughout its lifecycle. Having a DMP facilitates connections and communication between project stakeholders. It can also make sharing data easier not only with peers and potential collaborators, but with research partners and current stakeholders as well. And finally, a DMP can help ensure that research data retains its value to the organization over time.

A Significant Shift

While DMPs are only one piece of the effort to enable research data to be shared in ways that retain or enhance value, they are a critical component. Reviewing DMPs can reveal how well researchers understand what they are being asked to do and how prepared they are to do it. Recent studies indicate that there is work to be done on both fronts. An early study of researchers’ responses to the DMP requirements done in 2012 surveyed principle investigators
(PIs) of NSF grant submissions. These PIs reported widespread uncertainty over the new DMP requirements and actions needed to adequately respond. Later studies have provided some additional details. The results show:

- There is a lack of understanding of what constitutes public access to data
- Researchers only have a moderate level of awareness of the regulatory environment around issues of data management and sharing
- Data type and format generated varies widely by discipline
- Researchers desire support for their data management needs
- Strategies for sharing data varies dramatically among researchers
- Research assistants handle the majority of data management related tasks (with the exception of data sharing, which is addressed by the PI)
- When planning for training and outreach, consider the needs of all roles and responsibilities

Changing the culture of how research is practiced in any field will take time as well as dedicated resources, support, and education. Efforts to promote culture change will need to consider how research is currently practiced, articulate measurable benefits for individual researchers and the field (while recognizing potential near- and long-term costs or drawbacks), and identify the gaps between current practices and desired outcomes with actionable plans to close these gaps.

**Tools that May Help**

Various tools have been developed in an effort to educate and assist researchers as they navigate the regulatory environment. For instance, the DMPTool walks researchers through the process of developing a DMP for their research. The Massachusetts Institute of Technology (MIT) Libraries, along with other academic libraries, have produced extensive guides on managing, sharing, and preserving data. Data repositories, such as the Inter-university Consortium for Political and Social Research (ICPSR), the UK Data Service, and the Australian National Data Service, have developed guides on not only their data services, but on funding agency requirements and expectations. Finally, the United States Geological Survey (USGS) has created a useful webpage that includes information on and examples of DMPs.

Although the effect of these tools on researcher compliance has not been studied, state DOTs and other research organizations should emphasize training and support for researchers to meet the DMP requirement. Recommending useful tools, providing DMP support,
and emphasizing the role of DMPs in training are all good ways to help researchers incorporate DMPs in their standard research practices and initiate culture change around data sharing. In addition, researchers who have access to an academic library with data services should contact them for help.

**Essential Requirements for Developing Sections of Data Management Plans**

At a minimum, researchers need to generate a DMP for their projects that will meet the expectations of the funding agency to which they are applying. Funding agencies have different expectations of what they want to see in a DMP and so it is important for researchers to carefully review the call for proposals to be sure they understand what information needs to be included.

However, many DMP requirements, including the USDOT’s requirement, are based on the criteria developed by the National Science Foundation and are often comprised of five sections:

1. **Data**: A description of the data to be generated in the project
2. **Standards**: A description of the standards that would be used in developing and structuring the data
3. **Access**: A policy statement about how others would be able to gain access to the data
4. **Use**: A policy statement about what others would be permitted to do with the data
5. **Archiving/Preserving**: A statement on how the data would be archived

Some considerations to keep in mind while addressing these five basic sections of a DMP are listed below.

**The Data Section of the DMP**

In describing the data that will be generated over the course of the project, the primary focus should be on identifying the parts or portions of the data that will be shared with others outside of the project or organization that collected (and owns) the data. Not all data has to be shared necessarily, but the DMP should include a statement as to why not (including any ethical or privacy issues). Include information about how the data will be stored as it is being developed. It is also advisable to include as much information about the characteristics of the data as you can, including formats, volume, size, number of files, etc. Rough estimations are sufficient.
The Standards Section of the DMP

This is where information about how the data will be prepared according to community expectations. This section primarily focuses on data formats and the metadata that will accompany the data.

Formats

Consider what formats would make the data most useful or more accessible for others in your field or the general public. Note that open formats are usually better options than proprietary formats.

Metadata

Consider what information others in your field or the public would need in order to find, understand, and make use of the dataset. If there is a metadata standard used in your discipline or for the kind of data you are generating, consider using it for your data. If there is not, consider adapting a general standard such as Project Open Metadata or Dublin Core and applying it to the data.

See Chapter 7 for a more detailed discussion on formats and metadata.

Although it’s not mentioned by the NSF or USDOT directly, including information as to how the data will be documented in your DMP is a good idea.

The Access Section of the DMP

This describes how the data will be made accessible to people outside of the organization that collected the data. If a repository for the data has been identified, it should be described in this section. If the organization intends to host the data and make it accessible themselves, a description of how this will be accomplished and for how long will be needed. More information about data repositories can be found in Chapter 7.

The DMP should also clearly state the intended audience for the data. What research communities are expected to be interested in the data? Is there an expectation that the general public will be making use of the data?

Data citation and attribution is not only a best practice, but also a good rule of thumb.
The Use Section of the DMP

Once someone is able to access and acquire the data, what will they allowed to do with it? This is often a difficult section to write as many researchers and research organizations have not thought critically about how they would like their data to be used once it has been made available. Some elements to consider in writing this section are:

- **Reuse:** Can people reuse the data for research, commercial, or other purposes? Can people create derivatives of your dataset or use your dataset to develop new datasets?
- **Attribution:** Is citation or another form of attribution required if the data are used by others?
- **Licensing:** What license will be applied to the data to govern how it can be used? Creative Commons licenses are often applied to content in repositories, but they may be somewhat problematic to apply to data as CC licenses depend on copyright law and datasets may not be covered. Other licenses for data to consider are those provided by the Open Data Commons.

The Archiving/Preserving Section of the DMP

Agencies want to be sure that your data are not just available to others for a limited amount of time, but that steps are taken to ensure that the data are available for the long term. Some questions to address in the DMP are:

- **What is the long-term value of the data and how long should it be made available?** Some agencies have expectations that the data will be available for a certain amount of time, so be sure to read the requirements carefully.
- **What related information (documentation, software, reports, etc.) needs to be preserved along side the data?**
- **What steps will be taken to preserve the data by the repository selected to host and make the data accessible?** If a repository is not being used to host the data, what steps will the organization take to preserve it?

Chapter 7 provides additional information about preserving research data.
Going Beyond the Minimum for Data Management Plans

DMPs submitted to funding agencies are often limited in how long they should be. The NSF, for example, limits the DMP to two pages. Therefore, it is important to be succinct and direct in drafting the DMP that will be submitted as a part of the grant application. However, as DMPs are meant to be actionable documents, the version of the DMP used by project personnel may need to expand beyond the initial two page limitation to be useful.

More importantly, the DMPs being created should go beyond the considerations of federal and private funding agencies to consider and address local needs. Ideally, DMPs ought to connect with the policies of the state DOT or research institution as they describe the organization as a whole. In contrast to the high-level focus of policies, DMPs are useful in describing and informing the actual and specific data management, sharing, curation, and preservation practices taking place for individual datasets being developed within the organization.

A fully developed DMP can be an extremely useful tool in helping research teams come to a consensus on the work to be done in developing rich datasets, defining expectations for how the work will be done and by whom, and in holding researchers and their research organizations accountable for providing needed resources and completing the work. Moreover, well crafted DMPs can help change cultures of practice in organizations to help them accomplish more and build their reputation. As more and more DMPs are developed, patterns will begin to emerge that can be used to help state DOTs and other research institutions better understand current practices with data and to inform where additional support and resources may be needed to ensure that organizational goals are being met.

Some additional questions to consider including while crafting locally focused DMPs:

- What additional details are needed beyond funding agency requirements to make this DMP fully actionable and useful to the organization?
- Who are the stakeholders for the research project and/or the dataset(s) that are covered by this DMP? What roles and responsibilities will these stakeholders have in managing, sharing, curating, and/or preserving the data? (See Chapter 5 for more discussion on Roles and Responsibilities)
- What quality control or quality assurance measures will be taken to ensure the data are reliable and trusted?
- What resources within the organization will be needed for the dataset(s) being developed? Are there any special needs for the
dataset(s) requiring resources beyond what the organization currently provides? If needed resources are not available currently, how will they be acquired or what alternatives can be put into place?

- What costs might be incurred in managing, sharing, curating, or preserving these specific dataset(s)? If these costs are outside of the norm, how will they be covered?
- How do the DMPs link together or relate to each other? Are there elements of one research project that can be used to inform or improve upon another?

Creating DMPs will become easier over time as they become more commonplace. It is expected that some elements of the DMP will be similar or the same for every DMP and, as such, templates can be created with defined options to choose from as a starting point.

**Chapter Checklist**

From this chapter, you should be able to:

- Know the USDOT requirements for DMPs
- Understand that other agencies may have their own specific requirements
- Know the five common elements of DMPs and what goes into them: 1) a description of the data, 2) standards to be used, 3) making data accessible, 4) provisioning use of the data and 5) preservation of the data.
- Identify tools and templates that can help with writing a DMP
- Know how to Go Beyond with DMPs by adding information related to local research communities and interests
8. DATA MANAGEMENT PLANS
CHAPTER 9.

Building Blocks of the Solution

In This Section

» The Big Picture — USDOT and State DOTs
» USDOT Systems and Services Today
» Data Repository and Management Solutions
  » Selecting the Best Solution
» Non-USDOT Registries and Repositories for Written Research Products
» Chapter Checklist

The Big Picture — USDOT and State DOTs

Previous chapters of the Guide have identified and explained essential requirements for state DOTs and other research institutions, and options that research institutions have for going beyond the minimum in the future. This chapter pulls together those requirements and choices from a “big picture” systems perspective. The goal of this chapter is to provide a simple visual representation of the three essential systems and services components (see the “The Big Picture” on page 76).

“The Big Picture” identifies three sets of systems and services, including those:

- Provided by the USDOT to state DOTs and other research institutions to satisfy essential requirements today
- Related to data repository and management solutions to satisfy both essential requirements and long-term goals
- Related to local institutional registries and repositories for written research products to satisfy long-term goals

USDOT Systems and Services Today

The USDOT provides a registry to support discovery, availability, and access to support long-term storage and preservation of written research publications (ROSA P). The TRB provides a registry and repository for discovery, availability, and access to support long-term storage and preservation of written final project and technical reports.
At the time the Guide is written, the USDOT does not demand a cookie-cutter, one-size-fits-all solution for discovery, availability, access, storage, or preservation of research data. Instead, the Public Access Plan provides flexibility to meet local needs. The research organization must make its choice for both short and long term.
Data Repository and Management Solutions

Selecting a place to store and manage data for public access is the most complex aspect of meeting essential requirements. There are four essential choices:

- Researcher stores on local drive or on one-off storage devices
- Institutional solution — local repositories
- Open solutions — including the list provided by the USDOT
- Domain specific solutions — based on research community

Selecting the Best Solution

Most often, in today’s research climate, it is the researcher who decides where and how to store data. Academic organizations tend to have some storage and preservation guidelines, but unfortunately, these have requirements of limited access. This means there needs to be a shift from researcher-based to organization-based decisions on where and how to store federally funded research. The question to ask moving forward is this: Who makes the choice on where and how data is stored?

Perhaps the most significant question is this: should the state DOT (or other transportation research institution) use an existing repository or work with a repository to create their own collection? If the state DOT is selecting among existing repositories, a good starting point for evaluation of the choices is to ensure that the preservation process meets the FAIR guiding principles: Findable, Accessible, Interoperable, and Reusable. Repository assessments should reflect these principles and can be used by a state DOT or research institution to select one or more repositories for use. For more information on assessing repositories, see Chapter 12, Assessing and Managing Progress. If an state DOT elects to create their own collection, it is more involved at the beginning but could be more efficient in the long term. One way to do this is to contract with an experienced repository to use existing tools and expertise while providing a dedicated site for their data and other research products. If researchers or the state DOT choose to use a repository that is run outside of their organization, it will be very important to closely review the repository’s terms of service and ascertain if the repository is in compliance with the USDOT Public Access Plan. Any questions that arise from this review should be discussed with repository personnel.
Non-USDOT Registries and Repositories for Written Research Products

State DOTs and other research institutions are not required to have a local solution for written research products management to meet essential requirements. However, proceeding without a local or USDOT solution in place means relegating the organization doing the research to an indirect rather than direct role in achieving the Open Science vision. When an organization chooses to establish or use a local solution, they then have the ability to internally track and manage their research assets, and, as such, have a direct role in accomplishing the Open Science vision. Regardless, use of a registry or repository is essential to achieving discovery, use, and sharing within and without the research organization. Such systems enable interdisciplinary collaboration — across research communities and subject areas.

How should a state DOT approach the decision to support local management of research products? How do these systems and services translate to the local environment? The first step is to determine what you have to work with. Below are some key questions that might help determine what your organization has in place now.

- Is there a content or document management system?
- Is there a records management or archiving system or function?
- Is there an institutional or are there local search systems?
- Is there a library (contract or in-house) designated to support your research organization (e.g., state transportation library)?
- Is there a research services and support function?

Content, document, records management systems, as well as archiving systems can serve as a starting point for capturing, organizing, searching, using, and preserving written research products. Each of these systems serves a different function, but those functions can be adapted to support a general strategy across the organization. Having a large-scale data management system (e.g., business enterprise system) might also be a starting point, but it must be supported by a policy and process to store and preserve content. Similarly, library catalogs and bibliographic systems may support discovery but unless there is an underlying repository or storage solution, they will not solve the access and use challenge.
Chapter Checklist

From this chapter, you should be able to:

- ✔ Describe the USDOT and TRB systems and services that support essential requirements
- ✔ Understand local practices for managing written research products and research data assets
- ✔ Describe research institutions’ choices for research data repositories and registries to meet essential requirements
- ✔ Describe research institutions’ choices for long-term research data discovery, availability, access, and preservation
- ✔ Identify and describe existing local systems and services for managing written research products
- ✔ Describe long-term local options for management of written research products
- ✔ Use the Guide to find more detailed discussions of these issues

Endnotes

9. BUILDING BLOCKS OF THE SOLUTION
CHAPTER 10.
Learning and Training

In This Section
» Learning Strategies for the Research Organization
» Training Design and Delivery
   » Essential Requirements
   » Going Beyond
» Baseline Awareness Training for Everyone
» Training for Researchers
» Training for Executive and Management Roles
» Training for Research Support Roles
   » Lab Managers/Data Curators
   » Compliance Personnel
   » IT Personnel
» Chapter Checklist

Learning Strategies for the Research Organization

Merely developing a strategy to address the requirements of making publications and data accessible and archived is not enough. Instead, all stakeholders must be equipped to understand and be prepared to carry out their responsibilities. To that end, all stakeholders need training appropriate to his/her organizational role, which in turn, requires that the state DOT or other transportation research institution has well-defined stakeholder roles and clear identification of the expertise, knowledge, and skills each role needs for success.

Following good practice in data management involves more than the introduction of new tools; it also requires a cultural shift in how research is practiced. Training should go beyond introducing the products and tools used to support data management, sharing, and curation; it should also explain how they fit within the larger context. Prioritize learning within the research organization by doing the following:

- Spell out goals and objectives for the research organization as a whole forming a culture of practice around learning in the organization
- Provide resources and support for staff (both trainers and trainees)
- For trainers: Ensure adequate time is provided to teach, and teaching is recognized as a valuable contribution to the organization.
- For trainees: Ensure time is provided to learn, and learning is recognized as a requirement of the position and an expectation of the organization.

Different people learn in different ways. Multiple learning options should be provided.
• Offer training materials to anyone in the research organization. Connections are made between stakeholders in training programs so that they understand how their roles are connected to other stakeholders and the organization as a whole.

• Recognize that different people learn in different ways, and provide multiple learning options and opportunities, not just face-to-face, in-person instruction. Educational offerings and delivery should also consider available time and capacity of the target audiences.

Educational programs need to include best practices and guidelines of relevant research disciplines, as well as the open access and data curation community. They also should include expectations and requirements of external actors such as publishers, funding agencies, and off-site repositories. Programs also need to consider the current culture of practice around research within the organization and its goals and objectives, as well as the organization’s needs surrounding the management, sharing, and preservation of research materials and data. These external and internal elements should be in alignment for training to be successful.

Training Design and Delivery

Although multiple training programs and resources exist, there are gaps in what is available and what is needed for researchers at research institutions. Training programs in these areas are often developed by agencies that support the research activities of a particular discipline. The Inter-university Consortium for Political and Social Research (ICPSR), for example, has educational programming for social scientists, DataONE has programs for environmental scientists, and the United States Geological Survey (USGS) supports geologists and others doing earth science research. At a high level, the content provided by these sources addresses similar topics and provides general guidance that would be applicable for most types of research being conducted. However, there are disciplinary and other differences in how research is practiced. Any effective training program must: (1) account for the tools, software, and equipment used to generate and analyze data, (2) address how findings are published and shared, (3) speak to the cultures and practices of researchers at research institutions, and (4) specifically convey the requirements of the USDOT policy.

That said, existing programs can still serve as potential models for the structure and content for developing training programs for research institutions. The USGS in particular is an example in providing useful resources and guidance, particularly in their use of a defined research lifecycle as a means to ground their educational programs, their ability to break down concepts to make them easier to understand, and their assistance helping researchers make connections between these concepts. Whether from the USGS or other organizations, lifecycle models and other concepts can be appropriated and modified to serve the needs of researchers in research institutions.
Developing Training for Essential Requirements

Training programs should be based upon the USDOT Public Access Plan with a focus on making the essential requirements understandable and actionable by researchers and their support staff. Among other things, the Public Access Plan requires researchers to do the following:

☑ Make publications available after 12 months (unless an embargo is enacted)
☑ Submit publications to the USDOT NTL digital repository
☑ Ensure public access to final research data, subject to necessary restrictions such as security, individual privacy, or confidentiality
☑ Make their data accessible for search, retrieval, and analysis
☑ Develop Data Management Plans (DMPs) that describe their strategies for making their data publicly available
☑ Ensure that research project descriptions are submitted to Transportation Research Board's Research-in-Progress (RiP) database and are updated throughout the project

A possible curriculum for training programs that are focused on meeting essential requirements is described in the “Baseline Training for Everyone” and “Training for Researchers” sections below.

Developing Training for Going Beyond

Even if researchers are familiar with these and other requirements, it is unlikely they will possess an innate understanding of them to identify the necessary considerations or steps to take to satisfy these requirements. Looking beyond the researchers themselves, additional training programs will be needed to help research managers, compliance officers, and others who provide support for research understand and respond to these requirements. As with researchers, a more targeted training program informed by the culture and practice of USDOT research will be needed for all stakeholders within state DOTs and other transportation research institutions.

Over time the foundations of the training programs created or adopted by state DOTs and other transportation research institutions should extend beyond satisfying the USDOT to addressing the needs and goals of the research organization itself. It may take some time for a research organization to consider and articulate fully what its goals and needs are in managing, sharing, curating, and preserving data as these areas may not have been directly explored prior to the requirements of federal agencies on managing and sharing articles and data. Developing training programs in concurrence with developing the goals of the organization can provide opportunities to discuss stakeholder assumptions, confirm understandings of goals, and to explore what is needed to ensure that the goals of the organization will be achieved.
The Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) tool can help research organizations identify specific training needs of researchers and staff based on an assessment of current knowledge and practice (see the call out box on page 87 for more information).

A possible curriculum for training programs that are focused on going beyond essential requirements is described in the “Training for Researchers,” “Training for Executives and Management Roles,” and “Training for Research Support Roles” sections below.

Baseline Awareness Training for Everyone

Everyone in the state DOT or other research organization needs to have a baseline understanding of their organization's expectations for the management, sharing, and preservation of research materials, data, and publications. Instruction in what research and data management are and why they are important is also needed for all stakeholders to see the big picture of the research lifecycle. This general training helps stakeholders to develop a common vocabulary, a shared understanding of stakeholder roles, and an understanding of how they are all connected.

Baseline curricula could include the following:

☑ Introduction to the importance of public access
☑ State of research/data in the 21st century: general goals and benefits to managing, sharing, curation, and preservation
☑ Introduction to external requirements and expectations

Resources to Inform Training Programs

**USGS Data Lifecycle Overview:** The website consolidates data management resources into a single point of access and helps users understand best practices for various aspects of data management. [https://www2.usgs.gov/datamanagement/](https://www2.usgs.gov/datamanagement/)

**DataONE Education Modules:** A series of slides on data management topics that can be enhanced and reused for training purposes (based in earth and environmental sciences, but adaptable/reusable by other disciplines). [https://www.dataone.org/education-modules](https://www.dataone.org/education-modules)

**ICPSR — Guide to Social Science Data Preservation and Archiving:** A detailed, step-by-step manual for Social Science researchers interested in developing a dataset that can be deposited into the ICPSR repository, or simply understood and used by others. [https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/](https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/)
Introduction to article and data management/sharing plan requirements for:
- Funding agencies: what is being asked/required?
- Important journals in the field: what is being asked/required?
- Introduction to local needs, expectations, and practices
- Goals/benefits for the research organization
- Local culture(s) of practice in working with data (current and/or evolving)
- Defining stakeholders in managing, sharing, curating, and preserving data: who has what role in the organization, who can support you, and who will you need to support?
- Tools/resources — communication

Training for Researchers

As the producers of the data, researchers will require training in how to develop datasets that can be discovered, understood, and used by others outside of their immediate research team. This includes developing a DMP when applying for a grant — both to satisfy the requirements of the USDOT or other funding agency — and developing longer plans that articulate how the data will be managed throughout its lifecycle.

Researchers will require more detail than the baseline courses will cover. At minimum they will need the following:

- In-depth training on publication and DMP requirements for funding agencies — how to address in application, during project, reporting at end?
  - USDOT essential requirements
  - Tools/resources
- In-depth training on publication and DMP requirements for important journals in the field
  - How to identify the journals
  - What are the journals’ open-access policies and how do they conform with funding agency requirements?
  - What are the journals’ data-sharing policies?
  - Common elements of data-sharing policies
  - Guidance on complying
  - Tools/resources

Checklist: Baseline Curricula

Checklist: Minimum Researcher Curricula
More advanced information that builds off of the DMP requirement should be added into training programs as quickly as possible. This includes:

- Introduction to a data lifecycle model that best suits the data and organizational culture of the state DOT (or other research institution)
  - Use of a lifecycle to plan for managing, sharing, and preserving
- Overview on managing and organizing data
- How to educate others on how the different elements of the data relate to one another
- How to document and describe data through metadata
- How to share data
  - Selecting a data repository
  - Connecting data to publications
  - Handling presentations and other outputs
  - Giving data rights to those using their data
- How to support curation and preservation of data beyond the life of the project
- How to use data generated by others effectively and ethically
- How to cite data

This training will also help researchers and other stakeholders further connect with each other as areas where support is provided or needed will likely be made more clearly visible.

Training for Executive and Management Roles

In addition to the basic training for everyone, research administrators will need to devote time and energy into learning:

- A broad overview of the data lifecycle models used by researchers in the research organization and what should happen at each stage
  - What support is needed?
  - Who provides the support?
  - When is the support provided?
- In-depth training on compliance with:
  - publisher requirements
  - funding agency requirements
  - other external actor requirements
- In-depth training on how compliance needs to be demonstrated
- How to set requirements and expectations internally
- How to create a culture of data management
Within a research organization, there are multiple support roles that provide support to researchers and the research process. Each type of role will require specialized training directly relevant to their areas of responsibility, particularly as it pertains to making data and materials accessible. Ideally, researchers and those who support research should engage in shared training. The training required will naturally vary by role, but some specialized training areas could include those below.

**Lab Managers/Data Curators**

These individuals will need to work closely with researchers to help them develop and manage materials and data in ways that enable the resulting publications and datasets to be easily discovered, understood, trusted, and used by others in their field. This work is carried out in conjunction with researchers, however managers’/curators’ training may need to go beyond the broad training researchers receive and should involve additional training on best practices.

---

**Tools and Resources: CARDIO**

The Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) tool is designed to help departments, research groups or organizations within higher education institutions assess their infrastructure, staff skill levels, support of management, and other resources in assuring that data are adequately managed. The tool is administered by gathering information to determine a maturity rating in 30 relevant areas covering organization, technology and resources aspects of managing and curating data.

[http://cardio.dcc.ac.uk/](http://cardio.dcc.ac.uk/)

---

**Training for Research Support Roles**

Ideally, researchers and those who support research should engage in shared training.
Training topics beyond basic courses would depend upon the nature of the work being performed but could include the following:

- Documentation/metadata for sharing and preservation
- Techniques for organizing/managing data
- Quality control practices for sharing and preservation
- Policies of funders, publishers, and/or repositories and how to prepare publications/data for submission to repositories.

**Compliance Personnel**

Depending on the nature and commercial value of the research being done, intellectual property issues may arise. Legal staff may require training on navigating federal and other requirements in sharing and preserving data, in light of commercialization and other intellectual property issues.

Legal staff may also require training on licensing shared materials, data in particular. The state DOT will need to consider who should have access to the materials and data once they are released, and what people will be allowed to do with them. For example, good questions to ask may include:

- Must people provide attribution and, if so, how?
- Will people be allowed to redistribute the data?
- To what extent are people allowed to reuse the data?

Once these questions have been answered, the legal staff will need to select a license that best represents their decisions.

Compliance personnel will need training in how to demonstrate compliance to funding agencies and other bodies as needed.

Personnel who oversee ethical issues in research (such as working with human subjects) will require training on how to reconcile ethical requirements of conducting research with requirements on how to share and preserve the resulting data.

**IT Personnel**

IT personnel are often responsible for providing the technical infrastructures used in conducting research and managing data. This usually includes providing storage for data and other materials that allows access to authorized personnel while keeping unauthorized personnel out. Depending on the data or the specific requirements specialized training may be needed.

This could also include assistance in transferring the data from an environment where it is actively being developed to a repository.
where it will be shared, curated and preserved. Depending upon the organization’s strategy for sharing and preserving data, this may require training on the following:

- **Internal Repositories:** Building an in-house publications or data repository. What existing options are there? What are existing resources’ strengths and weaknesses in meeting federal and other requirements? What investment and support is needed to sustain a repository.

- **External Repositories:** What is the best way to select and work with an external repository? This could include assessing a repository's policies and services, potential issues in submitting and transferring data to an external repository.

**Chapter Checklist**

From this chapter, you should be able to:

- Identify the contents of training needed for USDOT essential requirements
- Understand how to extend training to go beyond by addressing local organizational needs and goals
- Identify key contents of baseline awareness training for everyone
- Identify key contents of basic researcher training
- Identify key contents of advanced researcher training
- Identify key contents of executive and management training
- Identify key contents of training for research support roles
10. LEARNING & TRAINING
CHAPTER 11.
Estimating and Managing Costs

In This Section
» Balancing Costs and Benefits
» Cost Factors:
  » Essential Requirements for the Research Organization
  » Essential Requirements for Individual Researchers
  » Going Beyond in Repository Management
» Chapter Checklist

Balancing Costs and Benefits

Meeting essential requirements introduces both costs and benefits. State DOTs and other transportation research institutions should aim to balance the costs with the benefits they expect to derive from achieving both essential compliance with essential requirements and going beyond minimum. The balance point is the scale and scope of an organization’s research effort. If an organization has only one or two researchers doing federally funded research, the focus should be on meeting essential requirements in the most cost effective way. In this case, the costing strategy might focus solely on those individual researchers. Organizations that have larger research teams operating on funding from federal and other sources should have an organization-level strategy for calculating costs and benefits. This chapter will help organizations understand which cost factors apply and how to use such factors to create a cost model. This chapter presents cost models in a way that allows an organization to tailor a solution to their specific situation.

Benefits of Managing Research Products and Data

The benefits of achieving essential compliance with federally funded research requirements is straightforward — it can be simply calculated as the dollar value of any future research proposal submitted. Where an organization is not compliant, future proposals may not be funded. The long-term benefits of applied and theoretical research are difficult to quantify, particularly at project inception or the proposal stage. Research organizations should begin their cost-benefit analysis by making their best estimate of the benefits they expect to accrue based on level of research production and engagement in the Open Science model.

Estimating the Cost of Meeting Essential Requirements

There are two levels at which an organization must think about and manage costs: (1) the individual researcher level, which must include cost estimates for managing research products in their DMPs; and (2) the organization level, which must include cost estimates across all research projects.
11. ESTIMATING & MANAGING COSTS

Cost Factors: Essential Requirements for the Research Organization

The most effective strategy starts at the organization level and focuses on meeting essential requirements. State DOTs and research institutions should begin by estimating how much it will cost to meet essential requirements over a period of three to five years. Organization level cost estimates will depend on the organization’s choices. As such, cost estimates can be derived by aligning choices with cost factors and building those cost factors into a formal cost model. Once the organization level cost models are established, the project-level cost estimates can be modeled upon them. Having a stable formula for researchers to use will ensure consistency across projects and facilitate acceptance by funding agencies, whether public or private. Furthermore, once the cost models are established, the organization will have a well-defined cost-recovery strategy for supporting research products management.

Cost Factors
The table below provides a high level description of the types of costs a research organization should expect to incur while complying with USDOT essential requirements. Organizations should consider their

<table>
<thead>
<tr>
<th>Step</th>
<th>Focus</th>
<th>Cost Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Raise Awareness of Federal Policies</td>
<td>• Communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Content Development Costs</td>
</tr>
<tr>
<td>2</td>
<td>Raise Awareness of Essential Requirements</td>
<td>• Training Development &amp; Delivery Costs</td>
</tr>
<tr>
<td>3</td>
<td>Raise Awareness of Open Science Vision &amp; Going</td>
<td>• Communication Costs</td>
</tr>
<tr>
<td></td>
<td>Beyond the Minimum</td>
<td>• Content Development Costs</td>
</tr>
<tr>
<td>4</td>
<td>Develop Long- and Short-Term Strategies</td>
<td>• Administrative Overhead Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Meeting and Consultation Costs</td>
</tr>
<tr>
<td>5</td>
<td>Identify Stakeholders and Assign Roles</td>
<td>• Human Resources Costs</td>
</tr>
<tr>
<td></td>
<td>and Responsibilities</td>
<td>• New Personnel Costs</td>
</tr>
<tr>
<td>6</td>
<td>Manage Research Text Products</td>
<td>• Training Development &amp; Delivery Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Content Data Management Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• New Personnel Costs</td>
</tr>
<tr>
<td>7</td>
<td>Manage Research Data</td>
<td>• Training Development &amp; Delivery Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• New Personnel Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Repository Subscribe, Buy or Make Costs</td>
</tr>
<tr>
<td>8</td>
<td>Develop Data Management Plans</td>
<td>• Training Development &amp; Delivery Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Content Data Management Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• New Personnel Costs</td>
</tr>
<tr>
<td>9</td>
<td>Develop Costing Strategies</td>
<td>• Administrative Overhead Costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Meetings and Consultation Costs</td>
</tr>
<tr>
<td>10</td>
<td>Assess Progress Against Requirements</td>
<td>• New Personnel Costs</td>
</tr>
</tbody>
</table>
essential compliance strategy that either focuses on the individual researcher or the research organization to identify the costs incurred. The costs will be specific to the organization as a whole and the short- and long-term goals.

Organizational Cost Models
Organizational cost models should be designed around the cost factors identified in the strategy and account for the scale of the investment needed to achieve its goals. While not all state DOTs and transportation research institutions will have sufficient research efforts to justify a cost model and budget for a full-service solution, every state DOT and research institution should have a good understanding of the cost factors and plan for managing costs and funding at all levels. There are eight cost factors, for which consistent costing strategies should be developed. Each factor is described below.

Administrative Overhead Costs are one-time sunk costs. Cost estimates for this factor should be available from human resources, calculated using administrative salaries and project time estimates.

Communication Costs are one-time sunk costs. These costs will include communication methods and systems, as well as salaries/time of communication specialists.

Policy Content Development Costs are one-time sunk costs. These costs would be derived from human resources, calculated using administrative and analysts salaries.

Human Resources Costs are one-time sunk costs. Human resources personnel will be engaged in revising job descriptions to include new responsibilities associated with essential requirements.

Meeting and Consultation Costs are one-time sunk costs. In some organizations, developing the local policy and the local practice will mean consulting with stakeholders within and across the organization. The time and costs to conduct these meetings should be built into the cost model.

New Personnel Costs are recurring operational costs. Because new roles or expanded responsibilities have been defined, the research organization should expect to incur additional ongoing operating costs.

Data Repository Subscribe, Buy, or Make Costs are recurring operational costs. Most state DOTs will chose to subscribe or buy a membership in an existing repository rather than to create or buy a repository solution. At a minimum, the state DOT should expect to incur annual subscription fees, and per project or per publication...
charges for data storage. Repositories may offer discounts for per project or per dataset storage to organizations paying an annual subscription or membership fee. In some cases, fees for individual dataset storage may be waived. As these pricing models may vary over time, the Guide suggests that state DOTs consult directly with the repositories. The Dryad repository, referenced earlier, provides a clear description of their costing and pricing models. This is a good place to start to understand the options.

### Figure 9. Tools to Help: Planning for Costs

Several generic cost spreadsheets are available online, and can be examined to understand principles for calculating repository costs. These include:

**NASA Cost Estimation Tool**


The Cost Estimation Toolkit (CET) is designed to provide NASA budget estimators, PIs, project managers, and resource planners with the capability to generate lifecycle cost estimates for implementing, operating and maintaining a science data system.

**UC3 Total Cost of Preservation (TCP): Cost Modeling for Sustainable Services**

https://confluence.ucop.edu/display/Curation/Cost+Modeling

Only the costs pertaining to preservation service providers are considered.... The TCP analysis encompasses the full economic costs associated with the long-term preservation of digital assets, although the resulting price models can be easily customized to deal only with various subsets of those costs as a matter of local policy.

**Curation Costs Exchange**

http://www.curationexchange.org/

The Curation Costs Exchange (CCEx) is a community owned platform that helps organizations of any kind assess the costs of curation practices through comparison and analysis. The CCEx aims to provide real information about costs to help make more informed investments in digital curation.

See also the **OPF Digital Preservation and Data Curation Costing and Cost Modeling** website: http://wiki.opf-labs.org/display/CDP/Home
Training Development and Delivery Costs are both one-time sunk and recurring operational costs. The development of training materials is primarily a one-time upfront cost. However, as policies and practices expand there will be a need to update training. Training delivery is an ongoing cost — as new researchers join the organization they will require training. All support roles will also require training. Tracking training compliance is also an ongoing cost the organization should expect to incur.

Project and Proposal Cost Estimates

There are two basic types of costs associated with meeting essential requirements: one-time sunk costs and recurring project costs. Depending on the scale and scope of your organization’s research effort, the one-time sunk costs may be minimal. State DOTs and research institutions may be able to absorb these costs without building them into the cost recovery proposals for research projects. If the scale of research is greater, though, these organizations will need to build these upfront and ongoing costs into a cost recovery strategy for individual research projects. From the general cost model, a cost calculator or estimation tool might be provided for researchers to determine costs for data management plans. In developing the cost calculator, organizations should strive for cost estimates that will cover the organization’s costs and will be considered reasonable by the funding agencies. Research institutions should consult with personnel from research and sponsored project management units for guidance on calculating costs for individual research projects. Existing models may be in place and available for adaptation.

Cost Factors: Essential Requirements for Individual Researchers

Cost Factors for Report Preservation

Since the NTL provides a registry and repository for all reports that fall under the USDOT Public Access Plan purview, the cost of compliance most likely will be low. In addition, planning ahead significantly reduces costs, while registering the project and researcher ORCID identifiers with the RiP database at the beginning of the project enables tracking to start early and project identifiers to be created and linked. All project deliverables are subject to the USDOT requirement, and these can be identified at the project start. Additionally, a DMP identifies the items to be preserved and details how they should be handled throughout the project. Ultimately, the time spent writing and following the DMP will save time later.

For more detail on cost factors, visit:
https://www.ukdataservice.ac.uk/manage-data/plan/costing
### Table 7. Costs of Essential Requirements and Possible Activities with their Associated Costs

<table>
<thead>
<tr>
<th>Activity</th>
<th>Questions to Consider</th>
</tr>
</thead>
</table>
| **DATA DESCRIPTION**   | • Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions, etc.?  
                          • Do textual data like interview transcripts need description of context, (e.g., included as a heading page)?                                                                                       |
| **DATA CLEANING**      | • Do quantitative data need to be cleaned, checked, or verified before sharing, (e.g., check code validity, anomalous values)?  
                          • Will data match documentation (e.g., same number of variables, cases, records, files)?                                                                                                                  |
| **DOCUMENTATION**      | • Do you have documentation for the data that describes the context and methodology of how data were gathered, created, processed and quality controlled?                                                                 |
| **METADATA**           | • What metadata need to be created when data are shared via a selected data center or archive (e.g., completing a deposit form for the data repository)?                                                                 |
| **FORMATTING & ORGANIZING** | • Are your data files, spreadsheets, interview transcripts, records, etc. all in a uniform format or style?  
                            • Are files, records, and items in the collection clearly named with unique file names and well organized?                                                                                     |
| **TRANSCRIPTION**      | • Will you transcribe qualitative data (e.g., recorded interviews or focus group sessions) as part of your research; or will you need to do this specifically so data can be more easily shared and reused?  
                            • Is full or partial transcription needed?  
                            • Is translation needed?  
                            • Will you need to develop a standard transcription template or transcription guidelines to ensure consistent formatting?                                                              |
<p>| <strong>DIGITIZATION</strong>       | • Do analogue or paper-based research data (maps, newspaper clippings, photographs, images, text) need to be digitized to increase their potential for sharing?                                                                 |
| <strong>FILE FORMAT</strong>        | • Do data need to be converted to a standard or open format with long-term validity for long-term preservation?                                                                                                     |</p>
<table>
<thead>
<tr>
<th>Associated Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low to no additional cost</strong> incurred if data description is carried out as part of data creation, input, or transcription.</td>
</tr>
<tr>
<td><strong>Higher cost</strong> incurred if added afterwards.</td>
</tr>
<tr>
<td>✅ Tip: codebooks for datasets can often be easily exported from software packages.</td>
</tr>
<tr>
<td><strong>Low to no additional cost</strong> incurred if carried out as part of data entry and prep before analysis.</td>
</tr>
<tr>
<td><strong>Higher cost</strong> incurred if needed afterwards.</td>
</tr>
<tr>
<td><strong>Low to no additional cost</strong> incurred if all data creation steps are well documented and documentation is kept well organized during research.</td>
</tr>
<tr>
<td><strong>Higher cost</strong> incurred if documentation is written or compiled specifically afterwards.</td>
</tr>
<tr>
<td>✅ Tip: often essential contextual and methods documentation will be written up in publications and reports.</td>
</tr>
<tr>
<td><strong>Low to no additional cost</strong> incurred. Completing a repository deposit form may take 1-2 hours.</td>
</tr>
<tr>
<td>✅ Tip: data repositories will generally have their own metadata forms, so it is important to select a repository and know the requirements early in the process.</td>
</tr>
<tr>
<td><strong>Low to no additional cost</strong> incurred if planned beforehand by developing templates and forms for individual data files (transcripts, spreadsheets, databases) and constructing clear file structures.</td>
</tr>
<tr>
<td><strong>Higher cost</strong> incurred if added afterwards.</td>
</tr>
<tr>
<td>✅ Tip: free software exists for batch file renaming to harmonize file names.</td>
</tr>
<tr>
<td><strong>Higher cost</strong> incurred if not planned as part of research practice.</td>
</tr>
<tr>
<td>✅ Tip: (1) consider cost of (time needed for) developing procedures, templates, and guidance for transcribers; and (2) calculate time needed for transcription — four to eight hours per hour recording (3) use transcribing calculator: <a href="http://www.socialsciences.manchester.ac.uk/morgancentre/methods-and-resources/toolkits/toolkit-8">http://www.socialsciences.manchester.ac.uk/morgancentre/methods-and-resources/toolkits/toolkit-8</a></td>
</tr>
<tr>
<td><strong>Low to no additional cost</strong> incurred for simple image scanning of text. Remember to include cost of time to scan.</td>
</tr>
<tr>
<td><strong>Higher cost</strong> incurred If (1) additional equipment or software is needed for scanning or conversion, (2) Optical Character Recognition is required, with manual checking for accuracy (revising entire scanned text), or (3) manual data entry or typing is needed (e.g., to digitize tabular data)</td>
</tr>
<tr>
<td><strong>Higher cost</strong> incurred for audio-visual data (converting to open digital formats can be time-consuming or require special equipment and/or software) and for databases, conversions may require checking for truncation, loss of metadata or annotation, loss of relationships, etc.</td>
</tr>
</tbody>
</table>
### Activity | Questions to Consider
--- | ---
**DATA STORAGE** | • How much data storage space is needed for the duration of the project?  
• For long-term storage, decide which data will be kept long term, which storage volume this represents and how long data will be stored and preserved.

**DATA BACKUP** | • Does the institution provide regular backup or not?  
• Consider how frequently backups should be done, how many backups should be stored.

**DATA SECURITY** | • Protect data from unauthorized access or use or from disclosure

**CONSENT FOR DATA SHARING** | • Do you need to ask participants for their consent for data to be shared?  
• Consent is essential for qualitative interviews; less so in quantitative surveys where data can be more easily anonymized.

**ANONYMIZATION** | • Do you need to remove identifying information or conceal the identity of participants (e.g. using pseudonyms) before data can be shared?  
**Tip:** Anonymization needs to be consistent throughout a data collection.

**COPYRIGHT** | • Do other parties hold copyright in the data?  
• Do you need to seek copyright clearance before sharing data?

**OPERATIONALIZING DATA MANAGEMENT** | • What measures are needed to implement and operationalize data management throughout the research lifecycle?

**DATA SHARING** | • Will data be deposited with a data center or institutional repository?  
• Which requirements exist to prepare data to particular standards (e.g., regarding documentation or format)?  
• Will journal publishers require deposit of data supporting article findings?
### Associated Costs

**Low to no additional cost** incurred if storage provided by the institution — cost is included in standard indirect costs or overheads.

**Higher cost** incurred if additional storage is needed.

*Tip: consider cost of server/disk space, as well as the cost of setup and maintenance.*

**Low to no additional cost** incurred if institutional backup (included in standard indirect cost).

**Higher cost** incurred if additional backup needed — cost according to number of copies to be kept, frequency of backup and storage media needed.

**Higher cost** incurred if for confidential or sensitive data. Ask if security can be arranged by institutional IT services or is extra software/hardware needed?

*Tip: (1) determining conditions for controlling access to shared data may require extra time and discussion, (2) data files may need encrypting before storage or transfers.*

**Low to no additional cost** when consent for data sharing is considered as part of standard consent procedures early in research.

**Higher cost** when participants need to be re-contacted after research has finished to obtain retroactive consent for data sharing. Questions to ask: does this require extra preparation of information sheets and consent forms; extra time for consent discussions; or training of interviewers?

**Low to no additional cost** incurred if (1) only removal of names is required, (2) planned before data collection or transcription/digitization, (3) for quantitative data (e.g., survey data) if identifiers are a priori excluded from data files, easy to remove, or identifiable variables are coded to avoid disclosure; cost may be higher if variables need recoding afterwards, or (4) for qualitative textual data (e.g., interview transcripts) if carried out or at least highlighted during transcription.

**Higher cost** incurred if (1) variables need recoding afterwards to avoid disclosure, (2) for qualitative textual data (e.g., interview transcripts) if entire texts need to be read and checked for identifying information, (3) for audio-visual data — anonymizing/editing voices or faces can be very costly and reduces the usefulness of data, and (4) pseudonymization requires more time.

**Cost factor questions:** Is time required to seek copyright clearance? Is legal advice required?

**Cost factor questions:** Do you need extra time/resources to implement data management throughout your research, (e.g., regular team meetings, setting up a collaborative research environment?)? Do you need staff training (if required) or a dedicated data manager (if needed)?

**Cost factor questions:** What is the cost of data deposit and/or longer-term storage? Find out from data center/repository/journal whether charges apply — costs can be included in project budgets if planned for ahead of time. What is the cost of data curation by a repository? Find out from data center/repository whether data enhancement or data cleaning charges apply. What is the cost in time and effort needed to prepare data for sharing and preservation? This can be included in project budgets.
Cost Factors for Data Management

Active data management is required for essential compliance. Related costs accrue starting with data creation and continue through describing, cleaning, organizing, storing, and preparing the data for long-term preservation. The UK Data Service’s data management checklist is an excellent resource for estimating costs of good data management. The table on the next page is adapted from that checklist and includes a description of possible activities and associated costs. In particular, the table highlights the fact that planning for public access ahead of time by preparing and documenting data as the research is conducted is more cost-effective than going back afterwards and documenting to comply with essential requirements. For more detail on cost factors, the UK Data Service website is an excellent resource. https://www.ukdataservice.ac.uk/manage-data/plan/costing

Cost Factors: Going Beyond in Repository Management

Going Beyond generally involves an expanded scope of eligible research and, for a limited few organizations, establishing their own repository for preservation. State DOTs and other research institutions that elect to establish their own repository may choose to either (1) create their own repository or registry from scratch, or (2) work with an existing repository to establish a collection dedicated to their own organization.

For those choosing to establish their own repository, costs may accrue under the activities and sub-activities listed in the table “Costs Associated with Creating a Repository” on page 101.

Studies have shown that the majority of costs (about half of lifetime costs) accrue at the time of acquisition and ingest. Preservation makes up about a sixth of the remaining costs, although they decline over time. Access rounds out the other third.

Personnel are the highest costs encountered when maintaining a data repository. One study found 70 percent or more of the “costs of preservation services in the case studies relate to staff costs and historically these have always been seen as the major component of preservation costs.”

Repositories need skilled personnel to manage infrastructure and preservation.
## Table 8. Costs Associated with Creating a Repository

<table>
<thead>
<tr>
<th>ACTIVITY</th>
<th>SUB-ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition</td>
<td>• Acquisition, evaluation and selection relating to preservation</td>
</tr>
<tr>
<td></td>
<td>• Submission agreement</td>
</tr>
<tr>
<td></td>
<td>• Ordering and invoicing</td>
</tr>
<tr>
<td></td>
<td>• Depositor support</td>
</tr>
<tr>
<td>Ingest</td>
<td>• Receiving submission</td>
</tr>
<tr>
<td></td>
<td>• Performing quality assurance</td>
</tr>
<tr>
<td></td>
<td>• Generation of metadata and documentation</td>
</tr>
<tr>
<td></td>
<td>• Reference linking</td>
</tr>
<tr>
<td></td>
<td>• Validation</td>
</tr>
<tr>
<td></td>
<td>• Deposit</td>
</tr>
<tr>
<td></td>
<td>• Holdings update</td>
</tr>
<tr>
<td>Archive Storage</td>
<td>• Receiving data from ingest</td>
</tr>
<tr>
<td></td>
<td>• Managing storage hierarchy</td>
</tr>
<tr>
<td></td>
<td>• Replacing media</td>
</tr>
<tr>
<td></td>
<td>• Disaster recovery</td>
</tr>
<tr>
<td></td>
<td>• Error checking</td>
</tr>
<tr>
<td></td>
<td>• Providing copies to access</td>
</tr>
<tr>
<td>Preservation Planning</td>
<td>• Monitoring of designated user community</td>
</tr>
<tr>
<td></td>
<td>• Monitoring technology</td>
</tr>
<tr>
<td></td>
<td>• Developing and implementing preservation strategies and standards</td>
</tr>
<tr>
<td></td>
<td>• Developing packaging designs and migration plans</td>
</tr>
<tr>
<td>Preservation Action</td>
<td>• Generation of preservation metadata</td>
</tr>
<tr>
<td></td>
<td>• Refreshment</td>
</tr>
<tr>
<td></td>
<td>• Backup</td>
</tr>
<tr>
<td></td>
<td>• Re-ingest</td>
</tr>
<tr>
<td></td>
<td>• Inspection</td>
</tr>
<tr>
<td></td>
<td>• Disposal</td>
</tr>
<tr>
<td>Data Management</td>
<td>• Administering database</td>
</tr>
<tr>
<td></td>
<td>• Performing queries and generating reports</td>
</tr>
<tr>
<td></td>
<td>• Receiving database updates</td>
</tr>
<tr>
<td></td>
<td>• Loading new descriptive information</td>
</tr>
<tr>
<td></td>
<td>• Loading new archive administrative data</td>
</tr>
<tr>
<td>Access</td>
<td>• Access provision</td>
</tr>
<tr>
<td></td>
<td>• Access control</td>
</tr>
<tr>
<td></td>
<td>• User support</td>
</tr>
<tr>
<td></td>
<td>• Rights management</td>
</tr>
<tr>
<td>Administration &amp; Services</td>
<td>• General management</td>
</tr>
<tr>
<td></td>
<td>• Administrative support</td>
</tr>
<tr>
<td></td>
<td>• IT support</td>
</tr>
<tr>
<td></td>
<td>• Physical security</td>
</tr>
<tr>
<td></td>
<td>• Utilities</td>
</tr>
<tr>
<td></td>
<td>• Supplies inventory and logistics</td>
</tr>
<tr>
<td></td>
<td>• Staff training and development</td>
</tr>
</tbody>
</table>
tasks over time. Staffing includes all levels: management, technical support, domain expertise, and administrative support. Personnel need adequate training to perform all preservation tasks and maintain changing technologies. In-house training should be developed for staff to learn the organization’s process and procedures for managing, sharing, curating, and preserving data and how to carry out their responsibilities. Training begins upon hire and should continue throughout a staff member’s career as requirements will evolve and personnel must be prepared to face new challenges. Funding should also be budgeted for staff to participate in conferences and other events to continue building upon their knowledge and skills.

Chapter Checklist

From this chapter, you should be able to:

- Understand the kinds of activities that may incur costs associated with meeting essential requirements for reports and data preservation
- Know how to reduce costs by planning ahead and making preservation part of the research project activities (rather than doing it afterwards)
- Know the basic factors that influence the cost of setting up your own repository
- Know where to find tools to help estimate costs of preservation

Endnotes

1 UK Data Service (2013). Data management costing tool. UK Data Archive, University of Essex. https://www.ukdataservice.ac.uk/manage-data/plan/costing


3 https://www.cessda.eu/eng/content/download/1430/20402/file/03_Costs_Factsheet_final.pdf

4 For a summary, see: https://confluence.ucop.edu/download/attachments/163610649/TCP-total-cost-of-preservation.pdf
CHAPTER 12.
Assessing and Managing Progress

In This Section
» Assessment of Essential Requirements
» Assessment of Process and Support for Essential Requirements
  » Assessing Quality of Interaction
  » Assessing Policy and Process
» Assessment for Going Beyond
  » Assessing Impact
  » Developing the Assessment Process
» Chapter Checklist

Once a system has been implemented for preserving the products of research, it will be critical to incorporate metrics that identify progress in implementation and success in meeting essential requirements and achieving reuse. These metrics can be used to provide feedback for changes to improve how the system functions. Initially, assessment will help the state DOT ensure that its researchers are complying with the USDOT essential requirements. However, over time, regular assessment can help states in Going Beyond as they move towards the broader Open Science vision.

Assessment of Essential Requirements

Chapter 2 describes essential requirements in detail, but for assessment purposes, they break down as follows:

1. Research projects should be reported to TRB’s RiP database shortly after being awarded, and should be updated during the project.
2. Every project should have an approved Data Management Plan (DMP).
3. Every researcher involved in eligible research should have an ORCID identifier.
4. RiP project status should be changed from “Active” to “Completed” within two months of completion.
5. ORCID identifiers and funding agreement number(s) should be included on the Technical Report documentation page and in any peer-reviewed publication submissions.
6. Research data should be packaged with metadata and deposited in an approved repository.

7. Within one year of the end of the project, an email should be sent to the USDOT Research Hub, NTL, and TRB including the following information:
   - Final Report URL(s) or PDFs for any resulting publications
   - URL(s) to and associated descriptive metadata for any final datasets and the arising from the research project
   - The funding agreement number of the project
   - The RH Display ID for the project

### Table 9. Tracking

<table>
<thead>
<tr>
<th>What is Tracked?</th>
<th>Assessment Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number and status of eligible funded projects (tracked at time of funding)</td>
<td>Total count, total in progress, total completed</td>
</tr>
<tr>
<td>Number of projects entered in RiP database</td>
<td>Number entered/Total projects</td>
</tr>
<tr>
<td>Number of projects with approved DMP</td>
<td>Number DMP/Total projects</td>
</tr>
<tr>
<td>Number of projects labeled “Completed” in RiP</td>
<td>Number marked completed/Total completed projects</td>
</tr>
<tr>
<td>Number of data deposits</td>
<td>Number deposited/Total completed projects</td>
</tr>
<tr>
<td>Number of correct technical documentation pages</td>
<td>Number correct/Number of eligible reports, Number correct/Total completed projects</td>
</tr>
<tr>
<td>Number of successfully deposited reports in NTL registry</td>
<td>Number deposited/Total reports eligible, Number deposited/Total completed projects</td>
</tr>
</tbody>
</table>
• ORCID identifiers (unique researcher IDs) for all publication author(s)
• Any documented project outputs or outcomes resulting from the research project

Assessing compliance with essential requirements for a state DOT or other research institution requires tracking from the beginning of projects. In doing so, all eligible projects (eligibility criteria are discussed in Chapter 2) should be identified and assessed on the seven items above throughout the course of the project. The table below shows each item to be tracked, one or more relevant metrics for assessment, and target levels.

<table>
<thead>
<tr>
<th>Target Level</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>Shows how the Public Access Plan affects research and how the research organization contributes to public access</td>
</tr>
<tr>
<td>Always near 100% because projects must be entered when funded</td>
<td>If low, researcher training or communication with researchers may be needed</td>
</tr>
<tr>
<td>Always near 100% because DMPs are required near the start of each project</td>
<td>If low, researcher training or communication with researchers may be needed</td>
</tr>
<tr>
<td>Near 100% with 2-month window</td>
<td>If low, compliance personnel may not be advising researchers of the need to mark completion</td>
</tr>
<tr>
<td>Should increase over time</td>
<td>Data packing and deposit is one of the most complex parts of the process. This metric will help the research organization track how well its support and training for data documentation and deposit are working.</td>
</tr>
<tr>
<td>Should increase over time, but counting technical documentation can be more difficult than tracking reports or deposits. This metric requires either submission of documentation to the research's organization (e.g., state DOT) in addition to the sponsor or reporting by the researcher that the task was completed.</td>
<td>Support for documentation and communication with researchers about the documentation requirement will be critical if this number is low.</td>
</tr>
<tr>
<td>Should increase over time. Near 100% for percent of completed projects should be possible, since it should be feasible to deposit one report per project. However, percent of eligible reports may be more challenging since some eligible reports may be missed early.</td>
<td>The ideal metric is the percent of eligible reports that are deposited, but tracking eligible reports is more difficult than tracking completed projects. Low deposit rates indicates a need for more training and communication with researchers.</td>
</tr>
</tbody>
</table>
Assessment of Process and Support for Essential Requirements

The metrics listed above directly address the essential requirements for compliance with the USDOT Public Access Plan. However, the research organization’s structures including tracking, training, and oversight can have a big influence on the levels of compliance seen. Thus, it is also worth having regular assessments of the process and support structures themselves to identify areas for improvement.

Assessing Quality of Interaction

To assess quality of interaction (of stakeholders in the data preservation process), metrics will rely primarily on user surveys. An ideal preservation process should minimize the burden on the researcher to make it more likely that they will comply. Regular surveys may uncover specific areas for improvement and additional support to researchers or personnel ensuring compliance.

Specific topics for survey questions should include:

- Quality and accessibility of training
- Availability of and usefulness of DMP templates
- Time spent by researchers to complete steps of the process (e.g., DMP completion, data upload, reporting compliance, etc)

In addition to surveys, some specific measurements will also be useful. For example, these may include:

- Time spent by compliance personnel interacting with each researcher
- Time spent in training, attendance at training sessions, number of training sessions

Assessing Policy and Process

Like quality of interaction, assessment of preservation policy and processes will rely primarily on surveys. Here, instead of questions about how smoothly the system works in practice, the questions are concerned with whether the high-level policies are appropriate and consistently applied. Specific topics for survey questions should include:

- Clarity of stated policies and communication of those policies
- Clarity and reasonableness of consequences of noncompliance
- Relationship between policy and practice (i.e., Do they match?)
Assessment for Going Beyond

Once a structure is in place to track compliance with essential requirements, additional measures can be considered for Going Beyond. Going Beyond the minimum includes expanding the scope of publications that the state DOT considers eligible. The state DOT may also go beyond by establishing its own data and report repository internally or contracting with an outside repository. The assessments described before apply to Going Beyond as well. However, the scope of tracking will increase.

Once a state DOT decides to Go Beyond, however, assessment of the impact of that choice will become particularly important. The motivation for essential compliance may be to avoid consequences of noncompliance, but the motivation for doing more will be to support the goals of Open Science and for the state DOT’s work to have a greater impact on transportation science and practice. Thus, assessment for Going Beyond will focus on assessment of impact.

Assessing Impact

A good starting point for assessment of impact is to ensure that the preservation process meets the FAIR guiding principles (see “FAIR Guiding Principles” on page 108): Findable, Accessible, Interoperable, and Reusable.

The one FAIR guiding principle implemented primarily by the researcher is describing the data with rich metadata. The rest of the time, most FAIR guiding principles are directly implemented by a trusted, long-lived repository. Researchers do not need to procure, for instance, persistent identifiers themselves, although they should still confirm that a repository conforms to the FAIR guiding principles.

Assessment of these guiding principles (whether implemented by an external repository or the state DOT’s own internal repository) will rely on user survey. The survey should ask about the user’s experience with reuse of other datasets, when problems with metadata will become more obvious.

A repository that conforms to the FAIR guiding principles well will tend to maximize impact. However, it is also important to directly measure impact, through tracking of dataset downloads and citations. The number of datasets reused and the number of papers or datasets cited is the gold standard metric for how much impact a data sharing system has on the field. As researchers in all organizations doing transportation research become aware of the availability of data and reports, the rate of reuse should increase and impact should accelerate.
### Figure 11. FAIR Guiding Principles

<table>
<thead>
<tr>
<th>Findable</th>
<th>Accessible</th>
<th>Interoperable</th>
<th>Reusable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data and metadata are assigned a globally unique, eternally persistent identifier.</td>
<td>1. (Meta)data are retrievable by their identifier using a standardized communications protocol.</td>
<td>1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.</td>
<td>1. (Meta)data have a plurality of accurate and relevant attributes.</td>
</tr>
<tr>
<td>2. Data are described with rich metadata.</td>
<td>• The protocol is open, free, and universally implementable.</td>
<td>2. (Meta)data use vocabularies that follow FAIR principles.</td>
<td>• (Meta)data are released with a clear and accessible data usage license.</td>
</tr>
<tr>
<td>3. (Meta)data are associated with their provenance.</td>
<td>• The protocol allows for an authentication and authorization procedure, where necessary.</td>
<td>3. (Meta)data include qualified references to other (meta)data.</td>
<td>• (Meta)data are associated with their provenance.</td>
</tr>
<tr>
<td>4. Metadata specify the data identifier.</td>
<td>2. Metadata are accessible, even when data are no longer available.</td>
<td></td>
<td>• (Meta)data meet domain-relevant community standards.</td>
</tr>
</tbody>
</table>

### Developing the Assessment Process

Good metrics require early establishment of good tracking system with the metrics in mind. These metrics particularly call for attention at three stages of the system: 1) selection criteria for recommended or approved repositories, 2) tracking system established when researcher is first awarded a project using federal funds, and 3) embedding the survey into standard operating procedures at all levels.

Survey questions should address the suggested topics above, but should also focus on elements of the process that the organization is focused on at each stage. We have suggested taking small steps to start and then building from there. The survey, in particular, should mirror these steps so that early on, questions focus on the elements that have been implemented most recently and might be most in need of adjustment.
Chapter Checklist

From this chapter you should be able to:

☑ Identify metrics to track each element of the essential requirements
☑ Set up a system that tracks eligible projects from the start to support assessment
☑ Develop questions for a user survey that will address policy and process from the user point of view
☑ Identify metrics to assess impact
12. ASSESSING & MANAGING PROGRESS
CHAPTER 13.
Putting it all Together

In This Section
» How to Use the Checklists
» Essential Requirements Checklist
» Using the Checklist to Go Beyond Essential Requirements

The Guide presents two checklists: one for helping state DOTs and other transportation research institutions to track their progress against essential requirements, and one to plan for going beyond the minimum. Checklists can serve as a foundation for more advanced management of people, processes, and resources. The checklists presented in this chapter are a synopsis of the tasks and issues that have been discussed throughout the Guide. Transportation research organizations can begin their journey by using the checklists as a roadmap, and learn more about each task by referring back to the individual chapter. The checklists can be used over time — as needed — to ensure your organization and researchers are in compliance with USDOT policy and processes.

How to Use the Checklists
Chapter by chapter, the Guide has walked you through the key issues and tasks. Some of these were tied to essential requirements and some were tied to going beyond the essential requirements in the future. Because the Guide will be used by organizations that are at different stages of compliance, you may want to work with the whole checklist or focus only on the essential requirements checklist. The checklist can provide a guide to assessing where your organization is now in complying with the essential requirements. Organizations can use the checklist as a working tool to do the following:

- Assess and assign responsibility for determining the current state of the organization in complying with essential requirements
- Formulate a strategy for current future implementation of compliance
- Involve all stakeholders, both current and future, in the practicalities of planning and accomplishing compliance

The Guide suggests that organizations use the checklist to assess the status of each requirement by noting whether it is (1) Complete; (2) In Progress; or (3) Not Yet Started. The checklist can then be used to assign and track responsibility for each requirement. To assign responsibilities, organizations might find it helpful to use the RASCI approach — assigning a...
task to those who are likely responsible or accountable (R, A) will prove the most effective as a starting point. Organizations might also use the checklist to identify stakeholders who might be expected to support (S) the requirement, or who should be consulted or informed (C, I).

Essential Requirements Checklist

The essential requirements checklist includes 12 steps. The first three steps are one-time steps which establish a foundation for both short- and long-term strategies. The remainder of the steps pertain to meeting essential requirements.

STEP 1. FOUNDATIONAL — Raise Awareness of Federal Policies

Identify and Track the National and USDOT Level Policies and Authorities

<table>
<thead>
<tr>
<th>Status of Activity:</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action Required (if any):</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individual Responsible:

Individual(s) to be Consulted:

Understand the Purpose and Structure of the Guide

<table>
<thead>
<tr>
<th>Status of Activity:</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action Required (if any):</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individual Responsible:

Individual(s) to be Consulted:

Understand Essential Requirements

<table>
<thead>
<tr>
<th>Status of Activity:</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action Required (if any):</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individual Responsible:

Individual(s) to be Consulted:
**Step 1 (cont’d)**

<table>
<thead>
<tr>
<th>Understand the Consequences of Noncompliance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Understand the Concept of “Going Beyond” Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

**STEP 2. FOUNDATIONAL — Raise Awareness of Essential Requirements**

<table>
<thead>
<tr>
<th>Understand What Research is Eligible Under the Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Know How to Track and Identify Your Funded Research Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>
### Understand What to Do When in the Research Project Lifecycle

**Status of Activity:**  
- Complete  
- In Progress  
- Not Yet Started

**Action Required (if any):**

**Individual Responsible:**  

**Individual(s) to be Consulted:**  

### Provide Guidance to Researchers in Registering ORCID identifiers

**Status of Activity:**  
- Complete  
- In Progress  
- Not Yet Started

**Action Required (if any):**

**Individual Responsible:**  

**Individual(s) to be Consulted:**  

### Know When and How to Submit a Data Management Plan (DMP)

**Status of Activity:**  
- Complete  
- In Progress  
- Not Yet Started

**Action Required (if any):**

**Individual Responsible:**  

**Individual(s) to be Consulted:**  

### Know Where to Submit Your Eligible Research Reports & Research Publications

**Status of Activity:**  
- Complete  
- In Progress  
- Not Yet Started

**Action Required (if any):**

**Individual Responsible:**  

**Individual(s) to be Consulted:**
### Step 2 (cont’d)

#### Know Where and How to Preserve Your Eligible Research Data

**Status of Activity:**  
- □ Complete  
- □ In Progress  
- □ Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

#### Know How to Provide a Persistent Identifier to Preserved Research Data

**Status of Activity:**  
- □ Complete  
- □ In Progress  
- □ Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

#### STEP 3. FOUNDATIONAL — Raise Awareness of Open Science Vision & Going Beyond the Minimum

**Understand the USDOT’s and the Local Research Organization’s Roles and Responsibilities**

**Status of Activity:**  
- □ Complete  
- □ In Progress  
- □ Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

**Develop a Strategy to Achieve Essential Requirements**

**Status of Activity:**  
- □ Complete  
- □ In Progress  
- □ Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**
Using the Checklist to Go Beyond Essential Requirements

Building on the Existing Foundation
Assuming the state DOT has developed a strategy for meeting essential requirements, there should be a good understanding of the USDOT policy and requirements, a deep understanding across the state DOT of essential requirements, and a sense of the state DOT’s expectations for engaging in Open Science. For those state DOTs that want to go beyond the essential requirements, the checklist begins with Step 4 — Developing the Long-Term Strategy. The long-term strategy should be a continuation of, rather than a radical departure from, the short-term strategy. However, going beyond the minimum is a significant step for a state DOT. A formal project plan would be a better tool to use to realize that long-term vision.

The remaining checklists cover specific steps in the overall research preservation process and correspond to the chapters in the Guide. Each step includes some activities that address details related to essential requirements and some that relate to going beyond the minimum.

Step 4. Develop Short- and Long-Term Strategies
The essential requirements checklist is supplemented to include the following tasks for Step 4:

- Consider a research institution’s long-term interest in Open Science
- Recognize the importance of making incremental progress
- Monitor the work of communities of practice

<table>
<thead>
<tr>
<th>Status of Activity: □ Complete □ In Progress □ Not Yet Started</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Action Required (if any):</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Individual Responsible:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Individual(s) to be Consulted:</th>
</tr>
</thead>
</table>
### Design a Short-Term Strategy that Can Grow into Long-Term Strategy

**Status of Activity:**
- 🔴 Complete
- 🔴 In Progress
- 🔴 Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

### Define the Organization’s Long-Term interest in Open Science

**Status of Activity:**
- 🔴 Complete
- 🔴 In Progress
- 🔴 Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

### Develop a Plan For Incremental Progress

**Status of Activity:**
- 🔴 Complete
- 🔴 In Progress
- 🔴 Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

### Define Method for Monitoring Progress of External Communities

**Status of Activity:**
- 🔴 Complete
- 🔴 In Progress
- 🔴 Not Yet Started

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**
Step 5. Identifying and Supporting Stakeholders, Roles and Responsibilities

For Step 5, the essential requirements checklist is augmented to include:

- Define expanded stakeholder model for Going Forward
- Define expanded roles and responsibilities for Going Forward
- Anticipate impacts to organizational culture
- Anticipate new communication patterns
- Design and implement new governance models and methods

### STEP 5. Identify Stakeholders and Assign Roles and Responsibilities

#### Identify Roles and Responsibilities for Meeting Essential Requirements

<table>
<thead>
<tr>
<th>Status of Activity:</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
</table>

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

#### Map Roles to Organization’s Job Titles

<table>
<thead>
<tr>
<th>Status of Activity:</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
</table>

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**

#### Review Current Responsibilities

<table>
<thead>
<tr>
<th>Status of Activity:</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
</table>

**Action Required (if any):**

**Individual Responsible:**

**Individual(s) to be Consulted:**
**Step 6. Managing Eligible Research Publications and Reports**

For Step 6, the essential requirements checklist is augmented to include tasks that pertain to a broader scope and coverage of research publications and reports at the local organization level. This may lead to a new project description pertaining to information and archives management across the organization. In this case the checklist may expand to include development of local content registries and repositories.

**STEP 6. Manage Research Text Products**

<table>
<thead>
<tr>
<th>Identify Eligible Written Research Products</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distinguish Formal Publications and Manuscripts and Final Project and Technical Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Understand How to Submit Written Research Products to NTL, TRB, and RH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>
13. PUTTING IT ALL TOGETHER

**Step 6 (cont’d)**

<table>
<thead>
<tr>
<th>Understand What information to Provide to NTL, TRB, AND RH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong>  ❑ Complete  ❑ In Progress  ❑ Not Yet Started</td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Understand When to Submit to NTL, TRB, AND RH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong>  ❑ Complete  ❑ In Progress  ❑ Not Yet Started</td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Understand the Metadata Services Provided by NTL, TRB, and RH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong>  ❑ Complete  ❑ In Progress  ❑ Not Yet Started</td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>
Step 7. Managing Eligible Research Data
The Guide assumes that most research institutions will not have a local solution for managing research data in place. The Guide also acknowledges that the design and implementation of a local solution is a significant undertaking. For those organizations that do opt for a local solution over the longer term, a full project plan with full cost models should be considered.

### STEP 7. Manage Research Data

**Define Essential Requirements for Research Data as Described by the USDOT**

<table>
<thead>
<tr>
<th>Status of Activity</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Action Required (if any):</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Interpret Essential Requirements for Your Organization’s Transportation Research**

<table>
<thead>
<tr>
<th>Status of Activity</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Action Required (if any):</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Understand the Organization’s Roles and Responsibilities for Data Management and Preservation**

<table>
<thead>
<tr>
<th>Status of Activity</th>
<th>Complete</th>
<th>In Progress</th>
<th>Not Yet Started</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Action Required (if any):</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Step 8. Developing and Using Data Management Plans

Going Beyond in this step means expanding the role of DMPs for planning across the organization, building upon their short-term strategy. Extended use for planning, though, should be designed to suit each organization's unique environment.

<table>
<thead>
<tr>
<th>STEP 8. Develop DMPs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Understand the Purpose of Data Management Plans</strong></td>
</tr>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

| **Understand the Basic Components of a DMP** |
| **Status of Activity:** | ☐ Complete | ☐ In Progress | ☐ Not Yet Started |
| **Action Required (if any):** |
| **Individual Responsible:** |
| **Individual(s) to be Consulted:** |

| **Comply with Essential Requirements for Intramural Research DMPs** |
| **Status of Activity:** | ☐ Complete | ☐ In Progress | ☐ Not Yet Started |
| **Action Required (if any):** |
| **Individual Responsible:** |
| **Individual(s) to be Consulted:** |

| **Comply with Essential Requirements for Extramural Research DMPs** |
| **Status of Activity:** | ☐ Complete | ☐ In Progress | ☐ Not Yet Started |
| **Action Required (if any):** |
| **Individual Responsible:** |
| **Individual(s) to be Consulted:** |
Step 9. Big Picture Choices and Decisions
This step does not change when moving beyond essential requirements. The bigger picture remains the same, though the organization's choices may change. Having established essential compliance, an organization now has a framework for assessing the state of progress in an evolving field.

<table>
<thead>
<tr>
<th>STEP 9. Understand the Big Picture for Transportation Research Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe the USDOT and TRB Systems and Services that Support Essential Requirements</td>
</tr>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

Understand Local Practices for Managing Written Research Products and Research Data Assets

| **Status of Activity:** | ☐ Complete | ☐ In Progress | ☐ Not Yet Started |
| **Action Required (if any):** | | | |
| **Individual Responsible:** | | | |
| **Individual(s) to be Consulted:** | | | |

Describe Organizations’ Choices for Research Data Repositories and Registries to Meet Essential Requirements

| **Status of Activity:** | ☐ Complete | ☐ In Progress | ☐ Not Yet Started |
| **Action Required (if any):** | | | |
| **Individual Responsible:** | | | |
| **Individual(s) to be Consulted:** | | | |
Step 10. Learning and Training
This step does not change very much when moving beyond the essential requirements. The bigger picture remains the same, but the details of training materials may change somewhat. Stakeholders need to understand the broadened scope of the effort, and a possibly narrowed set of choices for data deposit (e.g., eliminating those that provide less effective long-term preservation). In addition, training for Going Beyond needs to focus especially on culture change and why stakeholders should embrace the Open Science vision.

<table>
<thead>
<tr>
<th><strong>STEP 10. Learning and Training</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify Stakeholder Groups and Training Needs</strong></td>
</tr>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
<tr>
<td><strong>Identify Resources Available from which to Develop Targeted Training Materials</strong></td>
</tr>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
<tr>
<td><strong>Developing a Training Program</strong></td>
</tr>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>
Step 11. Costs

Many costs come from compliance with essential requirements, and in particular, the cost to an individual project is unlikely to change substantially. However, the cost to the organization will change in the following ways:

- Increase in total scope of preservation will increase the total cost
- Increasing the level of services provided by a repository or internal support staff will increase cost (e.g., bibliographic search, long-term data preservation and hosting)

STEP 11. Determine Costs, Budgets, and Economic Value

Identify Cost Factors

Status of Activity:  □ Complete  □ In Progress  □ Not Yet Started

Action Required (if any):

Individual Responsible:

Individual(s) to be Consulted:

Calculate Essential Compliance Costs at the Organization Level

Status of Activity:  □ Complete  □ In Progress  □ Not Yet Started

Action Required (if any):

Individual Responsible:

Individual(s) to be Consulted:

Calculate Essential Compliance Costs at the Project/Proposal Level

Status of Activity:  □ Complete  □ In Progress  □ Not Yet Started

Action Required (if any):

Individual Responsible:

Individual(s) to be Consulted:
**Step 12. Assessment**

For Going Beyond, the most important change in assessment is to add a focus on metrics to assess impact.

All other metrics and assessment structures should remain in place and be consulted regularly to check how well the system is operating.

### STEP 12. Assess Progress Against Essential Compliance

<table>
<thead>
<tr>
<th>Identify Metrics to Track Each Element of the Essential Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set Up a System that Tracks Eligible Projects from the Start to Support Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Develop Questions for a User Survey that will Address Policy and Process from the User Point of View</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Status of Activity:</strong></td>
</tr>
<tr>
<td><strong>Action Required (if any):</strong></td>
</tr>
<tr>
<td><strong>Individual Responsible:</strong></td>
</tr>
<tr>
<td><strong>Individual(s) to be Consulted:</strong></td>
</tr>
</tbody>
</table>