Guide to Ensuring Access to the Publications and Data of Federally Funded Transportation Research (2020)

DETAILS

138 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-67266-5 | DOI 10.17226/25704

CONTRIBUTORS

Carol A. Flannagan, Jared Lyle, Jacob Carlson, University of Michigan and Georgetown University Denise Bedford; National Cooperative Highway Research Program; Transportation Research Board; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2020. *Guide to Ensuring Access to the Publications and Data of Federally Funded Transportation Research*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25704.

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NCHRP RESEARCH REPORT 936

Guide to Ensuring Access to the Publications and Data of Federally Funded Transportation Research

Carol A. Flannagan
Jared Lyle
Jacob Carlson
UNIVERSITY OF MICHIGAN
Ann Arbor, MI

Denise Bedford
GEORGETOWN UNIVERSITY
Washington, DC

Subscriber Categories
Administration and Management • Data and Information Technology • Research

Research sponsored by the American Association of State Highway and Transportation Officials
in cooperation with the Federal Highway Administration

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

TRB
TRANSPORTATION RESEARCH BOARD

2020

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.

The Transportation Research Board (TRB) of the National Academies of Sciences, Engineering, and Medicine was requested by AASHTO to administer the research program because of TRB's recognized objectivity and understanding of modern research practices. TRB is uniquely suited for this purpose for many reasons: TRB maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; TRB possesses avenues of communications and cooperation with federal, state, and local governmental agencies, universities, and industry; TRB's relationship to the National Academies is an insurance of objectivity; and TRB maintains a full-time staff of specialists in highway transportation matters to bring the findings of research directly to those in a position to use them.

The program is developed on the basis of research needs identified by chief administrators and other staff of the highway and transportation departments, by committees of AASHTO, and by the FHWA. Topics of the highest merit are selected by the AASHTO Special Committee on Research and Innovation (R&I), and each year R&I's recommendations are proposed to the AASHTO Board of Directors and the National Academies. Research projects to address these topics are defined by NCHRP, and qualified research agencies are selected from submitted proposals. Administration and surveillance of research contracts are the responsibilities of the National Academies and TRB.

The needs for highway research are many, and NCHRP can make significant contributions to solving highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement, rather than to substitute for or duplicate, other highway research programs.

NCHRP RESEARCH REPORT 936

Project 20-110
ISSN 2572-3766 (Print)
ISSN 2572-3774 (Online)
ISBN 978-0-309-48158-8
Library of Congress Control Number 2020938326

COPYRIGHT INFORMATION

NOTICE

The research report was reviewed by the technical panel and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the National Academies of Sciences, Engineering, and Medicine.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and 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 of Sciences, Engineering, and Medicine; 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.

Published research reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from

Transportation Research Board
Business Office
500 Fifth Street, NW
Washington, DC 20001

and can be ordered through the Internet by going to

https://www.nationalacademies.org

and then searching for TRB

Printed in the United States of America

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, non-governmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

The **Transportation Research Board** is one of seven major programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation improvements and innovation through trusted, timely, impartial, and evidence-based information exchange, research, and advice regarding all modes of transportation. The Board's varied activities annually engage about 8,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

Learn more about the Transportation Research Board at **www.TRB.org**.

COOPERATIVE RESEARCH PROGRAMS

CRP STAFF FOR NCHRP RESEARCH REPORT 936

Christopher J. Hedges, *Director, Cooperative Research Programs*
Lori L. Sundstrom, *Deputy Director, Cooperative Research Programs*
Stephan A. Parker, *Senior Program Officer*
Stephanie Campbell, *Senior Program Assistant*
Eileen P. Delaney, *Director of Publications*
Natalie Barnes, *Associate Director of Publications*
Janet M. McNaughton, *Senior Editor*

NCHRP PROJECT 20-110 PANEL
Area Twenty: Special Projects

John M. Mason, *Pennsylvania State University–Harrisburg, Middletown, PA* (Chair)
Cynthia L. Jones, *Ohio DOT, Columbus, OH*
Majed N. Al-Ghandour, *North Carolina DOT, Raleigh, NC*
Leighton L. Christiansen, *U.S. DOT, Washington, DC*
Karen G. Kasuba, *California DOT, Sacramento, CA*
Robert "Bob" Meade, *University of California, Richmond, CA*
Ivan Mutis, *Illinois Institute of Technology, Chicago, IL*
Ellen M. "Leni" Oman, *Washington State DOT, Olympia, WA*
Richard Y. Woo, *Maryland State Highway Administration, Baltimore, MD*
Tim Schmidt, *FHWA Liaison*
Johanna Levene, *National Renewable Energy Laboratory Liaison*
James W. Bryant, Jr., *TRB Liaison*
Lisa Loyo, *TRB Liaison*

FOREWORD

By Stephan A. Parker
Staff Officer
Transportation Research Board

NCHRP Research Report 936: Guide to Ensuring Access to the Publications and Data of Federally Funded Transportation Research is a guide to help state departments of transportation (DOTs) and other organizations that receive transportation-related federal funds to comply with recent U.S. DOT requirements for the publication and communication of scientific knowledge. Centered on issues of data management and open data access, the report will be of immediate use to research program managers.

The transportation sector has only recently begun to look at issues of long-term preservation of research data and open access to that data, although there are a number of transportation agencies and organizations that have experience with quite large research data sets. Some federal agencies and organizations that receive federal funding have had data management requirements in place for several years and have created a pool of knowledge and experience from which the transportation research community can draw.

Under NCHRP Project 20-110, the University of Michigan was asked to develop guidance and activities to assist state DOTs, other public agencies, and transportation research organizations to efficiently and effectively ensure access to the results of federally funded transportation-related research.

Led by principal investigator Carol A. Flannagan and co-principal investigators Denise Bedford, Jared Lyle, and Jacob Carlson, the research team assembled a thorough literature review; identified stakeholder information requirements and needs; reviewed data management plans; and reviewed available training materials. While the research was in progress, the National Transportation Library published a set of essential requirements related to data management and data access with which the recipients of federal research funding from the U.S. DOT must comply. These requirements and the associated infrastructure substantially influenced the framework for the guidance that resulted from this project. The research team framed the guide around meeting essential requirements and provided a framework to keep state DOTs or other research organizations in compliance with the U.S. DOT policy. In addition, the team provided optional guidance for "going beyond the minimum" for organizations that desire to implement policies and processes that support the larger, more long-term vision of Open Science.

The report has 13 chapters that cover four strategic areas: surveying the landscape, creating a roadmap, beginning the journey, and staying the course. Chapter 13 combines the elements from all of the chapters and provides a series of practical steps and activities to create an implementable system of research product preservation. Each of the steps corresponds to chapters in the Guide, and the activities are provided in a checklist format.

NCHRP Research Report 936 is available on the TRB website at www.trb.org, along with a project report, *NCHRP Web-Only Document 270: Developing a Guide to Ensuring Access to the Publications and Data of Federally Funded Transportation Research*, which describes the information-gathering process of the project as a whole as well as how that information informed the guidance for state DOTs.

Contents

Acronyms

CARDIO	Collaborative Assessment of Research Data Infrastructure and Objectives
DMP	data management plan
DOT	department of transportation
FAIR	findable, accessible, interoperable, and reusable
ICPSR	Inter-university Consortium for Political and Social Research
IRB	institutional review board
ISO	International Organization for Standardization
NTL	National Transportation Library
OAIS	Open Archival Information System
ORCID	Open Researcher and Contributor ID
OSTP	Office of Science and Technology Policy
R&D	research and development
RASCI	responsible, accountable, supported, consulted, informed
RH	Research Hub
RiP	Research in Progress
ROSA P	Repository and Open Science Access Portal
SP&R	State Planning and Research
TRID	Transport Research International Documentation
U.S. DOT	U.S. Department of Transportation
USGS	U.S. Geological Survey

CHAPTER 1.

What are We Trying to Accomplish?

In This Section

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 1 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 more than $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 U.S. Department of Transportation (U.S. DOT) developed its Public Access Plan (http://ntl.bts.gov/public-access), which was published on December 15, 2015, and became effective as of January 1, 2016. In addition, "Executive Order 13642, Making Open and Machine Readable the New Default for Government Information"[1] and "Office of Management and Budget Memorandum M-13-13, Open Data Policy — Managing Information as an Asset"[2] 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.

U.S. DOT Public Access Plan:

https://www.transportation.gov/mission/open/official-dot-public-access-plan-v11

The purpose of this *Guide to Ensuring 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 U.S. DOT 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 U.S. DOT 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). The U.S. DOT's intramural research programs also have a long history of making data available to the public (e.g., the Fatality Analysis Reporting System, or FARS database). On an incremental basis, the U.S. DOT'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.

Essential requirements for compliance:

https://ntl.bts.gov/public-access/plan-executive-summary

The U.S. DOT'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 U.S. DOT 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 with 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 U.S. DOT'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

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.

Open Science strives to make it easier to publish and communicate scientific knowledge.

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

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 and Policies
- Chapter 5: Roles and Responsibilities

BEGINNING THE JOURNEY

- Chapter 6: Managing Research Publications
- Chapter 7: Managing Research Data
- Chapter 8: Data Management Plans
- Chapter 9: Building Blocks of the Solution

STAYING THE COURSE

- Chapter 10: Learning and Training
- Chapter 11: Estimating and Managing Costs
- Chapter 12: Assessing and Managing Progress
- Chapter 13: Putting It All Together

checklists in Chapter 13, and then referring 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 U.S. DOT 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 TRB have defined the practices that satisfy essential compliance, which are described and explained

in detail in 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 U.S. DOT 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 to enable 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 the concept of essential requirements.
- ☑ Understand the consequences of noncompliance.
- ☑ Understand the concept of "going beyond minimum."

Endnotes

1 Exec. Order No. 13642 of May 9, 2013. 78 FR 28111. https://www.govinfo.gov/content/pkg/CFR-2014-title3-vol1/pdf/CFR-2014-title3-vol1-eo13642.pdf.

2 Sylvia M. Burwell, Steven VanRoekel, Todd Park, and Dominic J. Mancini. Open Data Policy—Managing Information as an Asset. M-13-13. Executive Office of the President, Office of Management and Budget, Washington, DC, May 9, 2013. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf.

CHAPTER 2.

Understanding Essential Requirements

In This Section

Introduction

This chapter covers the U.S. DOT'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" web pages hosted by the National Transportation Library at https://ntl.bts.gov/public-access.

How to Comply with Public Access:

https://ntl.bts.gov/public-access

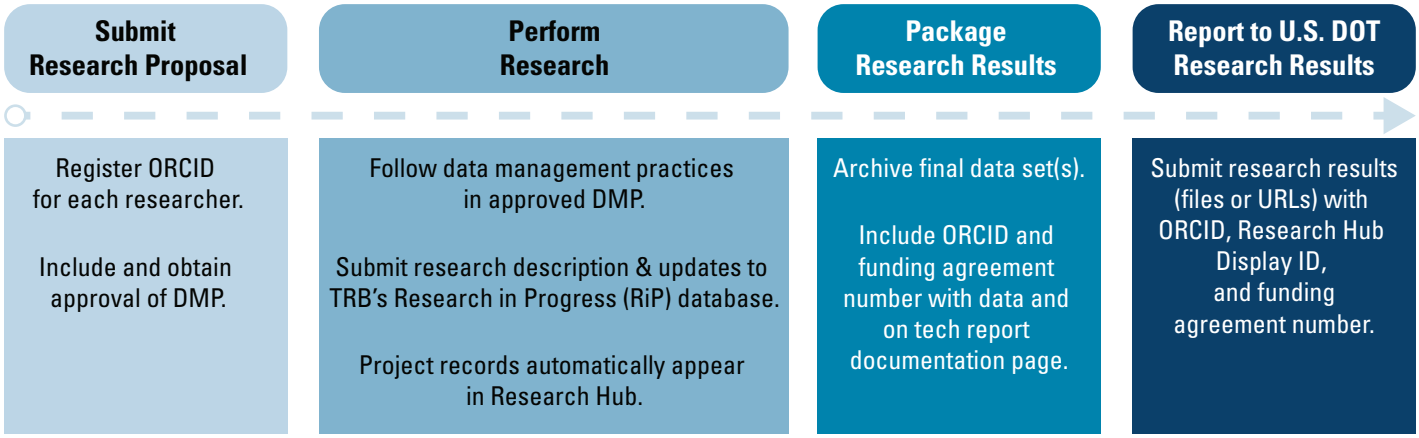
The National Transportation Library's Role

The U.S. DOT's infrastructure builds upon existing research life-cycle tracking arrangements between the TRB Research in Progress (RiP) database, the U.S. DOT Research Hub, the National Transportation Library (NTL) Digital Library, and a new interface with the U.S. DOT Public Data Listing.

The U.S. DOT Research Hub (RH) 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 are covered in detail in the chapters of this Guide.



The Research Project Life Cycle and Compliance

Compliance is best understood in the context of a research project's life cycle. Figure 1 illustrates the U.S. DOT's conceptualization of the life cycle and notes key points of compliance in the cycle. Some terms and concepts in the figure may be unfamiliar, but they are 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 life cycle embodies activities needed to develop and start a research project, and includes the following:

Checklist:

BEFORE the Research Project

- ☑ 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 from 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 are presented later in this chapter and in Chapter 4.

- ☑ Be sure to note the RH ID in the RH 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).

(cont'd)
BEFORE

During a Research Project

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

- ☑ Report your research project to TRB's RiP database at: http://rip.trb.org. See Chapter 6.
- ☑ 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). 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.

Checklist:

DURING the Research Project

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 2 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 Technical Report documentation page and in any peer-reviewed publication submissions. Be sure to format ORCID identifiers correctly in your final reports (https://orcid.org/trademark-and-id-display-guidelines).
- ☑ Ensure that all rights under copyright are nonexclusively retained by the U.S. DOT 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 U.S. DOT 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) and associated descriptive metadata for any final data sets arising from the research project.
  - ☑ The funding agreement number of the project.
  - ☑ The RH Display ID for the project.

Checklist:

AFTER the Research Project

(cont'd)
AFTER

- ☑ ORCID identifiers (unique researcher IDs) for all publication author(s).
- ☑ Any documented project outputs or outcomes resulting from the research project.
- ☑ Final reports and data set URLs will be appended to your project records in the U.S. DOT Research Hub.

Requirements: Research Products

The U.S. DOT's Public Access Plan is focused on two types of research products: (1) reports and publications and (2) data.

Research Reports and Publications

The U.S. DOT 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 U.S. DOT–managed contract, grant, or other agreement;
- Written deliverables published by the U.S. DOT, with no **embargo**; and
- Written deliverables published externally by U.S. DOT scientists or extramural U.S. DOT–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.

Embargo is a time-limited restriction on publication.

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 U.S. DOT 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

As defined in the U.S. DOT's Public Access Plan, research data are all scientific data collected through research projects funded, either fully or partially, by federal funds awarded through a U.S. DOT contract, grant, or other agreement or collected by U.S. DOT 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/), and
- 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 do the following:

- Provide unique permanent identifiers for digital data sets and the standards that are defined for managing identifiers.
- Support interaction between the awardee(s), the data repository, and the state DOT or U.S. DOT grant manager to enable all parties to verify that
    - Data meet minimum quality standards,
    - Data are appropriately evaluated and secured to prevent disclosure of personally identifiable information,
    - Proprietary interests and confidentiality are protected,
    - Intellectual property rights are respected, and
    - 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 U.S. DOT 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 her/his needs.

Project Open Data Metadata Schema:

https://project-open-data.cio.gov/

Requirements for Describing Preserved Research Data

The U.S. DOT requires Project Open Data metadata standards for scientific data. According to 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.[1] Archives not operated by the U.S. DOT are required to be able to generate common core and extensible metadata so that the U.S. DOT 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

The Open Researcher and Contributor ID (ORCID) provides each researcher a persistent digital identifier that supports linkages between researchers, their work, and their professional activities. The U.S. DOT Public Access Plan requires that researchers 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.

Requirements: Data Management Plan

Researchers and research organizations (e.g., state DOTs) applying for research funds from the U.S. DOT 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 (Digital Object Identifier, or DOI), authors (ORCID identifiers), and data sets for the linking and correlation of authors and relevant data.

Data management plans are covered in detail in Chapter 8.

Checklist:

Data management plan

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 affect 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 U.S. DOT'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 date can be no sooner than the publication date of the department's final plan (i.e., December 16, 2015).

It is important to regularly check the DOT Public Access guidance web pages for updates on what research is covered and what constitutes compliance.

https://ntl.bts.gov/public-access

Sources of Research Funding

Primary Funding Sources

U.S. DOT funding sources may include, but are not limited to, the following agencies:

- 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), and
- 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, State Planning and Research (SP&R) funds are not covered under the essential requirements unless there are matching University Transportation Center 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 U.S. DOT

While transportation-related research is primarily funded by the U.S. DOT, 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 that 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 agencies:

- 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), and
- National Institutes of Health (NIH).

Finding Your Submitted Research Products and Data

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. NTL will also provide a searchable database of DMPs submitted as a part of the U.S. DOT's Public Access Plan requirements. Reports housed or accessible through NTL are available without charge.

The U.S. DOT Public Data Listing will be expanded to add descriptions of and links to data sets resulting from scientific research. This listing will serve as a single access point for the U.S. DOT's intramural and extramural public data sets, 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 data sets (as referenced in the U.S. DOT 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 life cycle.
- ☑ 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

1 National Information Standards Organization (NISO). *Understanding Metadata*. Bethesda, MD: NISO Press, 2004, p. 1. http://www.niso.org/publications/press/UnderstandingMetadata.pdf.

CHAPTER 3.

Going Beyond the Minimum in an Evolving Landscape

In This Section

Principles of Accessibility and Availability

"Availability" and "accessibility" refer to the ways in which the U.S. DOT 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 U.S. DOT. 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:

Availability means the public is aware of the research data or can determine that research exists.

- There is traceability from the funding sources to the research produced from those funds.
- Research results are registered in the NTL Repository and Open Science Access Portal (ROSA P) 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), and
    - 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 and also as many people as possible who wish to use it.

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 as the "active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities."[1] As noted by the Inter-university Consortium for Political and Social Research (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 understandably accessible to the public now and in the future.

List of off-site repositories:

https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan

The curation landscape is constantly evolving, and the demand for access to data will only increase as time goes on (Figure 2). 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.

Figure 2. Current National and International Landscape of Curation

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 (see the box to the right).

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 U.S. Geological Survey (USGS), the U.S. Department of Agriculture (USDA), the National Oceanic and Atmospheric Administration (NOAA), the U.S. Department of Energy (DOE), the National Science Foundation (NSF), and the 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 Office of Science and Technology Policy 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

Understanding the factors that contribute to an Open Science culture allows organizations and researchers to go beyond the minimum.

Going Beyond the Minimum — Creating an Open Science Culture

Complying with the U.S. DOT's essential requirements is a critical first step toward achieving the Open Science vision (see Figure 3). 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:

- **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 affect their advancement.
- **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.
- **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.
- **Nonmonetary 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 when 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 which 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-reviewed 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.

Figure 3. What is Open Science Anyway?

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 U.S. DOT'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 data sets generated, fully or partially, through federally funded scientific research (Principle of Availability) and
- The public is **able to download and analyze** unclassified publications and/or digital data sets 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 U.S. DOT 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; and
- 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 U.S. DOT'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 U.S. DOT 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.

Procedures should be established to ensure legal norms are observed.

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 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 the commercial or public organizational requirements for public access.

Factors Pertaining to Data Infrastructures

State DOTs are required by the U.S. DOT 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; and
- 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 familiarity with the current landscape of other federal agency and departmental practices.
- ☑ Gain familiarity with the national and international landscape of research preservation.
- ☑ Understand organizational factors that may present challenges.

Endnotes

1 Cynthia Hudson-Vitale, Heidi Imker, Lisa R. Johnston, Jake Carlson, Wendy Kozlowski, Robert Olendorf, and Claire Stewart. *SPEC Kit# 354: Data Curation*. Association of Research Libraries, Washington, DC, 2017. https://publications.arl.org/Data-Curation-SPEC-Kit-354/.

CHAPTER 4.

Developing Strategies and Policies

In This Section

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 U.S. DOT Public Access Plan and supplemented by the guidance web pages hosted by the National Transportation Library (NTL). These essential requirements, described in more detail in Chapter 2, include the following:

- Create a data management plan (DMP) for every project covered under the U.S. DOT Public Access Plan and follow, revise, and update DMPs as needed throughout the projects.
- Have an Open Researcher and Contributor ID (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 project; in particular, ensure that every project's status is changed from "active" to "completed" within 2 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 U.S. DOT Public Access Plan are specified, the way state DOT programs might get there is not. Moreover, by "Going Beyond the Minimum" on page 17(Chapter 3), state DOTs can envision policies that not only are more comprehensive or inclusive than the minimum, but that also help the state DOT and its researchers move toward the Open Science vision.

The state DOT's vision for research preservation will need to include the following elements:

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

Planning for Essential Requirements

While the steps that lead to compliance have been clearly laid out by the U.S. DOT (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 U.S. DOT'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., the state DOT's) responsibility to make sure all of its 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. Table 1, "Whose Responsibility Is It?" on page 27), identifies which aspects of the plan the U.S. DOT 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

Staying abreast of changes to essential requirements

Essential compliance requirements are consistently being updated and can be found at **http://ntl.bts.gov/publicaccess/.**

Table 1. Whose Responsibility Is It? U.S. DOT vs. Local Research Organizations (e.g., State DOT)

Issue	Responsible for Essential Compliance	Responsible for Going Forward
Why? Goals	U.S. DOT	U.S. DOT & State DOT
What? Eligible Research & Data	U.S. DOT & State DOT	U.S. DOT & State DOT
When? Timelines	U.S. DOT	U.S. DOT & State DOT
How? Defining a Process	U.S. DOT & State DOT	U.S. DOT & State DOT
Who? Functional Roles	State DOT	State DOT
Where? Registries & Repositories	U.S. DOT & State DOT	U.S. DOT & State DOT

eligible research, explain how the plan will be translated to a process, 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 the section on going beyond on page 31. Successful programs will communicate clearly with researchers and detail 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 strategies and policies are being developed is provided below. In answering these questions, it will become clear what structures are missing and what structures are already in place.

Questions to Ask DMPs

Data Management Plans

- ☑ Is there a strategy for creating a DMP?
- ☑ Do researchers have experience developing DMPs?
- ☑ Is there a need for additional short-term training, or is the U.S. DOT guidance on DMPs sufficient for the immediate future?

Questions to Ask ORCIDs

ORCID Identifiers

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

Registering & Depositing

Questions to Ask
Registering and depositing research publications and reports

Eligible Research Publications and 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 life cycle? 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 U.S. DOT'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 1 year after project closeout? If not, what needs to be changed?

Questions to Ask
Registering and depositing research data

Eligible Research Data

- ☑ Are 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 U.S. DOT 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 U.S. DOT 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 DMP?
- ☑ Who is the best person in the state DOT to advise on cost estimates?

Questions to Ask *(continued)*
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, and 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; and
- **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. See Table 2 for examples of short-term policy objectives. 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.

Table 2. Examples of Short-Term Policy Objectives

Policy	Short-Term Objective
Preservation Requirement	Projects: Start with the U.S. DOT essential requirements for which funding types require compliance with the requirement. They are specified on the Public Access Plan website: https://ntl.bts.gov/public-access/plan-executive-summary.
Tracking	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 NTL at project completion. Also leverage information in the TRB RIP record.
Storage	Start with one or more of the suggested repositories listed by NTL at https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan.
Confidentiality	Data that could endanger research subjects' confidentiality, including those containing direct identifiers and/or indirect identifiers, will be considered confidential and for 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.
Oversight	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 on any changes to the U.S. DOT requirements and/or available support for compliance.
Training	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 the preservation process and policy for longer-term culture change.
DMP	Start with developing templates for creating DMPs in your research organization. See the U.S. DOT Public Access Plan guidance page "Creating Data Management Plans" (https://ntl.bts.gov/public-access/creating-data-management-plans) for the expected DMP sections and suggested points to cover. Additionally, a DOT DMP template is available at: http://guides.lib.umich.edu/c.php?g=283277&p=2138498.
Costs	The U.S. DOT 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 nonproject budgets, so early planning for these costs (especially initial setup costs) will be important.

Going Beyond: Creating a Holistic Plan and Long-Term Goals

Beyond minimum
Holistic plans & 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 that should include the following:

- Assess the current state of goals, strategies, roles, and resources.
- Define specific goals for a 3- to 5-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 is fine to start at the beginning, asking questions such as these:

- ☑ **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 life cycle?
- ☑ **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 there are none, are there arrangements in place with external registries and repositories?

Early planning to cover nonproject infrastructure costs is important.

Developing 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, as indicated.

- **Preservation Requirements**
  - **Projects:** What research projects are included in the U.S. DOT'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 projects 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 that 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 forth. However, the following set of basic measures 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 percentage 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, and preserving research products.
- Survey stakeholders on their knowledge, actions, expectations, and needs over time.

Measure and account for costs of compliance with the Public Access Plan.

Chapter Checklist

From this chapter, you should be able to

- ☑ Understand the U.S. DOT'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.
- ☑ Understand the importance of making incremental progress.
- ☑ Understand the importance of monitoring the work of communities of practice.

CHAPTER 5.

Roles and Responsibilities

In This Section

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 U.S. DOT has established and taken responsibility for several key roles in achieving essential compliance. Specifically, the U.S. DOT 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.

Essential Compliance Tasks	Possible Existing Organizational Roles
Track research funding proposals	Institutional research board, research and sponsored programs officers, research managers, researchers
Develop data management plans	Data librarians, institutional repository managers, data services support teams, researchers, research teams
Estimate costs/budgets for research data management	Institutional repository managers
Register researcher Open Researcher and Contributor ID (ORCID) identifiers	Researchers
Identify and approve repositories for research data management	Institutional repository managers, information technology managers, research managers, librarians
Track and submit research publications	Researchers, reference and research librarians
Track and submit research final and technical reports	Research and sponsored program officers, research managers, researchers
Close out and audit research project compliance	Researchers, research and sponsored program officers

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 its 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.

Beyond minimum
Expanding an organization's understanding of stakeholders

Going Beyond: Expanding Stakeholders and Roles and Responsibilities

Going Beyond means expanding an organization's understanding of stakeholders to include both producers and consumers of its 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 on 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 its requirements or needs for the data,
- Requirements made by funding agencies or other external bodies for the data, and
- Access to resources (e.g., staff, money, time).

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.

Identifying stakeholders early will help organizations plan well and avoid delays.

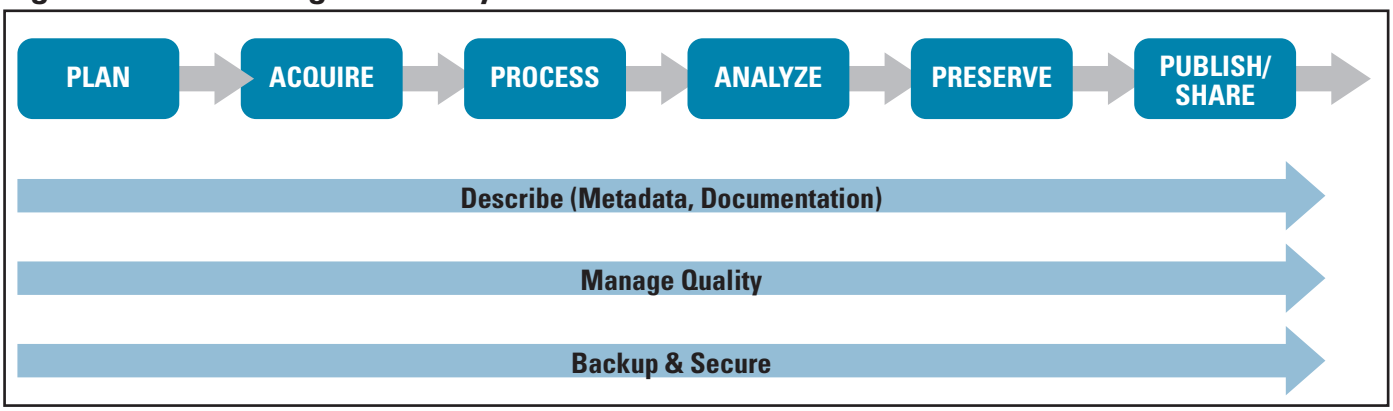
Going Beyond: Expanding Roles and Responsibilities

The research life-cycle 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 life-cycle 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. Life cycles 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. Life-cycle models, such as that of the U.S. Geological Survey (USGS), shown in Figure 4, depict the stages of development that a research project, particularly the data set, will pass through in its development. In other words, these models define *what* happens *when*. Defining the stages of the research life cycle 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 life-cycle model describes six stages, each of which includes activities related to preservation:

1. **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 data management plan (DMP). This stage also includes 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. U.S. Geological Survey's Research Model



Documentation is key to effective future use by other researchers.

2. **Acquire** covers the actions taken by staff to support generating or collecting the data for the project. How is each type of data in the project collected, 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?
3. **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 data set, including development of a codebook, will be critical.
4. **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 the data and inform the findings derived from the data, and who is responsible for performing and documenting the analysis.
5. **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 Data Set

Role	Name (or Title if name is not known)	Support Provided
Principal Investigator	Martha Johnson	Budgets and ensures time for preservation within the project; authors DMP; submits all eligible publications and data to the National Transportation Library (NTL)
Researcher	Joseph Garcia	Oversees gathering of survey data
Researcher	Rebecca Aoki	Oversees data collection on noise levels at intersections
Lab Manager	Jeff King	Stores and processes the noise level data
Compliance Personnel	Debra Strauss	Reports to U.S. DOT; informs personnel of U.S. DOT requirements
Data Repository	Support staff at the Inter-university Consortium for Political and Social Research (ICPSR) to deposit the data; support staff at ROSA P to deposit articles and register the data.	Provide guidelines on preparing data for deposit, sharing, and archiving

6. **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.

Beyond minimum
Going beyond may also include new roles and expanded responsibilities for existing roles.

An important goal of creating a research life-cycle model is to identify the considerations and actions needed to be able to share and archive the data generated from a research project. Remember to complete the information in the research life-cycle table with an eye toward 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

Information Required	Time Frame/Stage Person Needs to Be Involved
What are the requirements for compliance; what makes a good DMP	All
What documentation is needed to develop a codebook; institutional review board (IRB) guidelines and how they impact sharing and archiving data	All
What documentation is needed and how it should be structured	All
Who should have access to the data and when; what elements of the data should be archived	Data collection; data analysis; data archiving
Documentation that the U.S. DOT requirements are being met	Proposal planning; sharing results; project wrap up
Deposit of the data	Sharing results; data archiving

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 life cycle.

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

Beyond minimum
Any plan for going beyond should be sure to address how stakeholders communicate with each other and how often.

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. Table 3, "Worksheet to Help Identify Stakeholders for the Research Project and/or Data Set" (p. 38), 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 life cycle, 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 and 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 between 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 an 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

As described in Chapter 2, research publications are defined by the U.S. DOT 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 U.S. DOT, with no embargo period;
- Written deliverables published externally by U.S. DOT scientists or extramural U.S. DOT–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, during which 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.

Research Publications and Reports Not Eligible Under Essential Requirements

The Public Access Plan does not apply to the following items:

- Written deliverables or data submitted for publication prior to the plan's effective date or digital data generated prior to December 16, 2015 (the effective date of the U.S. DOT'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; and
- 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 it or technical libraries charged to collect it. Practices for managing gray literature vary across organizations and within the information management domain.

Research final reports and technical reports are more likely than publications to contain sensitive or personally identifiable information.

The U.S. DOT and 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 may be 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,

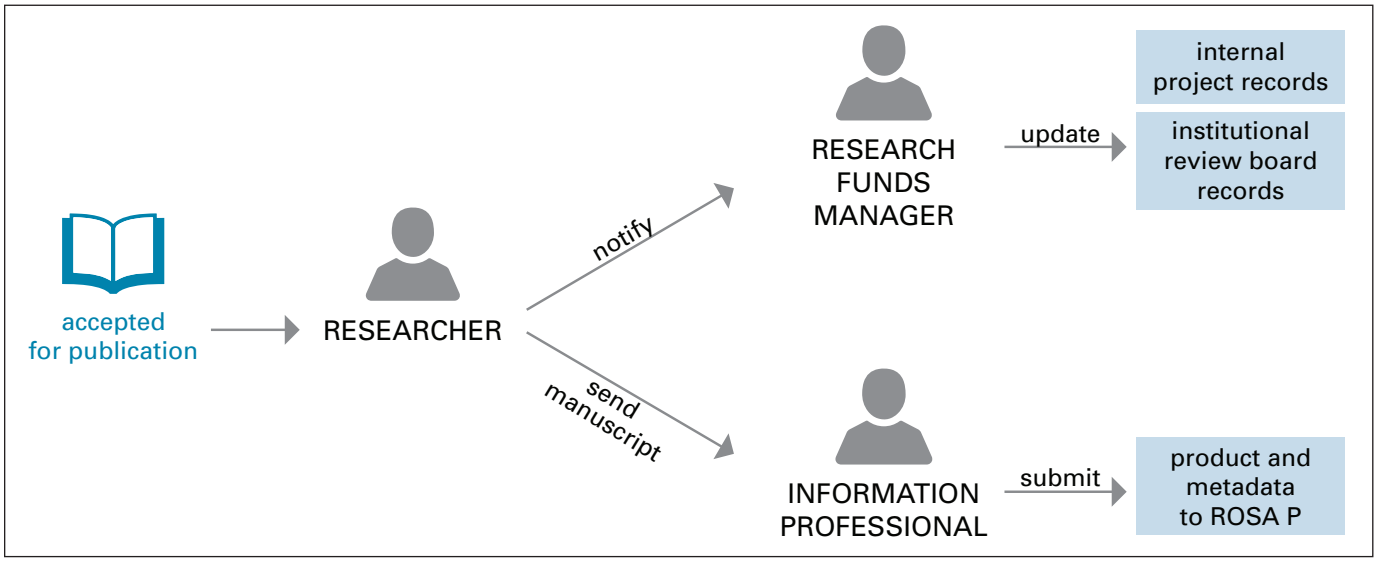
sensitive, or personal identification information is not made publicly available. Where final project and technical reports are classified, established practices may allow public sharing of metadata but only allow authorized or permissioned access to the publication and data. This also involves the securing 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 are 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 U.S. DOT Public Access Plan (see Chapter 2). Research publications that are deliverables of the project must be submitted within 1 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 NTL 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.

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

Figure 5. Communication Network



Funding sources for published research:

http://search.crossref.org

https://www.crossref.org/services/funder-registry

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, support 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).

U.S. DOT Registries and Repositories

Essential
Submitting written research products

The U.S. DOT essential requirements for submitting written research products are simple and easy to follow:

- ☑ Send a single email addressed to three recipients, including the U.S. DOT 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) and associated descriptive metadata for any final data sets arising from the research project,
  - ☑ The funding agreement number of the project,
  - ☑ The Research Hub (RH) Display ID for the project (from the TRB Research in Progress (RiP) database entry created at the start of the project),
  - ☑ Open Researcher and Contributor ID (ORCID) identifiers (unique researcher IDs) for all publication author(s) and contributor(s), and
  - ☑ Any documented project outputs or outcomes resulting from the research project.
- ☑ Append to the email any final peer-reviewed manuscripts and final project or final technical reports. Remember, the Repository and Open Science Access Portal (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 the U.S. DOT and TRB to coordinate the tracking and management of the research products. The U.S. DOT supports both the Research Hub (RH) and NTL's ROSA P. The RH is a searchable database of past and current U.S. DOT–sponsored research, development, and technology project records. ROSA P is a final report repository and data registry for U.S. DOT–funded and –created research. TRB manages the Transport Research International Documentation (TRID) database, which serves as a registry for all final project and technical reports. TRB also maintains a mirror version of the RH, which stores an archival copy of the final project and technical reports. Links for technical reports listed in the RH are also visible for copies deposited in NTL's archive.

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 NTL to host a searchable database of data management plans (DMPs). This will take the form of a collection within ROSA P called U.S. DOT Public Access Data Management Plans. Reports housed or accessible through 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

Beyond minimum
Managing all research products, including gray literature

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 U.S. DOT. As noted earlier, these products include those from 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 Office of Science and Technology Policy (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 by researchers, research communities, private-sector businesses and industries, and the general public; and
- Increasing publishers' expectations that supporting research data will be submitted as part of the peer-review process.

Beyond minimum for research organizations

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 and 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 life cycles (e.g., going beyond metadata standards for publications).
- Collection development policies to reflect new scope and coverage of gray literature.
- The information management life cycle and providing life-cycle 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 schemas,
- Understanding how to create metadata,
- Having a registry or library catalog to maintain metadata,
- Having a policy in place for choice of persistent identifiers, and
- Understanding how to generate a persistent identifier for written research products.

Beyond minimum for state DOTs

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 ensure 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 nonproprietary 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.

Selecting optimal long-term preservation formats:

https://www.loc.gov/preservation/resources/rfs/

Understanding Metadata Standards and Practices

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

- **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, and so forth.
- **Describing the nature of the content,** including its subject or topic, its geographical relevance, and the type of content (e.g., report, chapter, article).

A recommended best practice is to assign the role of metadata creator and publication submitter to an information professional.

- **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.
- **Chronicling 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.

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 U.S. DOT's Project Open Data requirements, information professionals can ensure that publications and their associated data sets 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.

Project Open Data Metadata Schema:

https://project-open-data.cio.gov/

There are many different metadata standards and schemas. For essential requirements, researchers should include a JSON metadata file with their final reports submitted to NTL, TRID, and RH. The JSON file should follow the Project Open Data Metadata Schema (Table 4). 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

- 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 (e.g., 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;
- Includes the metadata that 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

Field	Label	Definition	Required
title	Title	Human-readable name of the asset.	Always
description	Description	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	Always
keyword	Tags	Tags (or keywords) help users discover the data set.	Always
modified	Last Update	Most recent date on which the data set was changed, updated, or modified.	Always
publisher	Publisher	The publishing entity and, optionally, its parent organization(s).	Always
contactPoint	Contact Name and Email	Contact person's name and email for the asset.	Always
identifier	Unique Identifier	A unique identifier for the data set or application programming interface (API) as maintained within an agency catalog or database.	Always
accessLevel	Public Access Level	The degree to which this data set 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).	Always

- Contains actionable values that define the legal ownership and liability constraints associated with using the content; and
- 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, and
- ☑ Understand what going beyond means for state DOTs and research institutions.

CHAPTER 7.

Managing Research Data

In This Section

Definition of Research Data

The U.S. DOT 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

Essential
Research Data

Explaining Essential Requirements for Research Data

This aspect of essential compliance focuses on managing, accessing, and using the data that support federally funded transportation research. The U.S. DOT has provided guidance but left some important decisions and actions to research organizations. For written research products, the U.S. DOT 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 U.S. DOT has provided guidelines for selecting a data registry and repository for preserving and providing access to data. Additionally, the U.S. DOT 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 are stored in an open format or describe in the data management plan (DMP) which proprietary formats are used and why.
- Certify the quality of the data and that data are 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 U.S. DOT 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, and
    - Have 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 U.S. DOT has provided a list of repositories that currently meet its 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 (re3data.org) searchable listing of data repositories as a starting point for locating 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,

Key Websites

https://re3data.org
https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan

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 life cycle;
- 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; and
- Possess a technical infrastructure that explicitly supports the tasks and functions described in internationally accepted archival standards, such as the 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 for meeting essential requirements means selecting an option that will be conformant with the U.S. DOT 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 institution's 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.

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

Beyond minimum
Data management and access

Plan to use archival and open formats from the beginning of the research plan, if possible.

sources other than the U.S. DOT, 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.

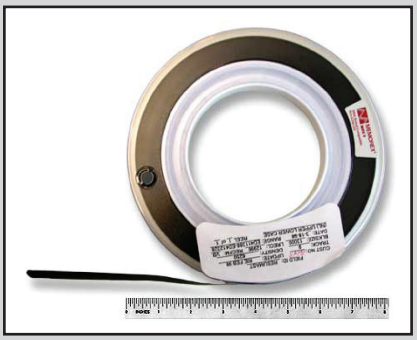
Understanding Data Preservation

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 (Figure 6).

Unfortunately, few data appear to be deposited into a long-lived repository with dedicated digital preservation systems and staff — including hardware, metadata, and organization — that ensure the data are available and understandable in the 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



An old 9-track tape

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 it 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 it, 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.[1] 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 and Coverage

Beyond minimum
What to preserve

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 life cycle 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:[2]

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, and
- Ability to use the collected information for secondary studies.

How significant are 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, and
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 affects the usability of the data. For example, some media may have deteriorated such that the data contained are 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 constitute a new version of or addition to data collections in the holdings. Data that are a chronological continuation 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 on the basis of 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. Table 5, "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 affect a repository's ability to preserve the content for long-term access and therefore 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.

Note: Table 5 should primarily be used to understand the variety of data types and possible formats. All of the data types and formats in this table are pertinent to transportation research.

Table 5. Long-Term Preservation of Data: Types and File Formats

Type of Data	Acceptable Formats for Sharing, Reuse, and Preservation	Other Acceptable Formats for Data Preservation
Quantitative tabular data *with extensive metadata* (e.g., data set with variable/code labels and defined missing values)	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)	Proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Quantitative tabular data *with minimum metadata* (e.g., data with or without column headings or variable names and no other metadata or labeling)	Comma-separated values (.csv) Tab-delimited file (.tab) Delimited text with SQL data definition statements where appropriate	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)

(continued on next page)

Table 5. ***(continued)***

Type of Data	Acceptable Formats for Sharing, Reuse, and Preservation	Other Acceptable Formats for Data Preservation
Geospatial data vector and raster data	ESRI Shapefile • Essential: .shp, .shx, .dbf • Optional: .prj, .sbx, .sbn Geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) Tabular GIS attribute data	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
Qualitative textual data	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)	Hypertext Markup Language (HTML) (.html) Widely used proprietary formats: MS Word (.doc/.docx) Proprietary/software-specific formats: NUD*IST, NVivo, and ATLAS.ti
Digital audio data	Free Lossless Audio Codec (FLAC) (.flac)	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)
Digital image data	TIFF version 6 uncompressed (.tif)	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)
Digital video data	MPEG-4 (.mp4) motion JPEG 2000 (.mj2)	
Documentation and scripts	Rich Text Format (.rtf) PDF/A or PDF (.pdf) HTML (.htm) OpenDocument Text (.odt)	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

For more information on data types and file formats, go to http://www.data-archive.ac.uk/

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 in the assessment of 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."[3]

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.

Data curation enhances collections so they are "complete and self-explanatory" for future users.

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 Schema. 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 Schema 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 forth.[4]

Project Open Data Metadata Schema:

https://project-open-data.cio.gov/v1.1/schema

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—for example, geographical locations, International Organization for Standardization (ISO) country names, and Multipurpose Internet Mail Extensions (MIME type). 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 life cycle. 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 often were not available and that even when they were available, they were not sufficient to meet the 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 U.S. DOT's Public Access Plan, and it appears to pose a significant gap.

Essential
Metadata

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

- Title,
- Creator,
- Description,
- Subject terms, and
- Geographic coverage.

Beyond minimum
Metadata

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://rd-alliance.github.io/metadata-directory/standards/
- http://www.dcc.ac.uk/resources/metadata-standards

Deciding Where to Preserve Data

Several options exist for preserving research data (see Figure 7, "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 a repository is being selected. 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 (NTL) maintains a regularly updated list of data repositories conformant with the U.S. DOT Public Access Plan (https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan). The Registry of Research Data Repositories (https://www.re3data.org/) is an online tool to help identify existing international repositories for research data.

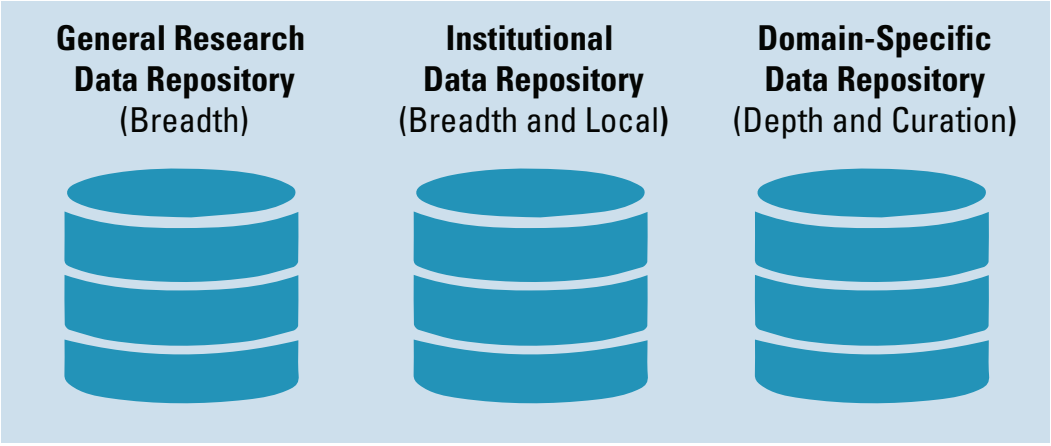
Beyond minimum
Enhancing data reuse

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 catalog pointing to the data sources. The challenge of such a registry is that it hosts no data. If metadata are not automatically harvested, it can be difficult to maintain (add, update) the records for external content.

The ROSA P registry at NTL hosts documents but does not host data sets and is thus serving the important purpose of providing a centralized simple data registry for transportation data under the U.S. DOT 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



Zenodo and figshare are examples of general research repositories.

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.

University of Michigan Deep Blue Data is an example of an institutional repository.

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, which makes them especially attuned to diverse 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.

ICPSR and Protein Data Bank are examples of domain repositories.

OPTION 3: Domain-Specific Data Repository

Yet another repository option is a domain-specific repository, such as the Inter-university Consortium for Political and Social Research (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."[5] 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."[6] 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 your counsel.

Understanding How Long to Preserve Data

A question that is often asked is how long data should 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 into the future a data set will have value.

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 to develop a checklist for determining the value of the data set after the minimum time has elapsed.

Funding and publishing agencies will sometimes define their expectations for the retention or availability of a data set. 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."[7] If such an expectation is stated, it should be used to inform the minimum preservation period.

A checklist for determining whether preservation should continue will naturally vary according to the type of data, the needs of the researchers who make use of the data, and the specific value the data have for the community. The questions listed in "Deciding What to Preserve" on page 57 are also relevant for deciding whether preservation should continue. Other key questions might include the following:

- How much and what kind of use have the data received over time?
- Who depends upon having access to the data? How would they be affected by not having access to the data?
- Are 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 the organization 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 U.S. DOT.
- ☑ 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 Inter-university Consortium for Political and Social Research (ICPSR). 2009. *Principles and Good Practice for Preserving Data*. International Household Survey Network, IHSN Working Paper No. 003. http://www.ihsn.org/sites/default/files/resources/IHSN-WP003.pdf.

2 Source: http://data-pass.org/sites/default/files/appraisal.pdf.

3 https://www.dpconline.org/handbook/institutional-strategies/standards-and-best-practice.

4 See the RDA Metadata Directory, http://rd-alliance.github.io/metadata-directory/standards/, or the Digital Curation Center, http://www.dcc.ac.uk/resources/metadata-standards.

5 Carol Ember and Robert Hanisch, "Sustaining Domain Repositories for Digital Data: A White Paper." Output of the workshop "Sustaining Domain Repositories for Digital Data," Ann Arbor, MI, June 24–25, 2013, doi:10.3886/SustainingDomainRepositoriesDigitalData.

6 Ann G. Green and Myron P. Gutmann, "Building Partnerships among Social Science Researchers, Institution-Based Repositories, and Domain Specific Data Archives." *OCLC Systems and Services: International Digital Library Perspectives* Vol. 23 No. 1, 2007, pp. 35–53. https://doi.org/10.1108/10650750710720757.

7 Directorate for Engineering Data Management Plans Guidance for Principal Investigators. Updated November 2018. https://nsf.gov/eng/general/ENG_DMP_Policy.pdf.

CHAPTER 8.

Data Management Plans

In This Section

Developing and Maintaining Data Management Plans

Over the past 20 years or so, there has been 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 U.S. DOT followed suit and described its expectations for researchers to create and submit a DMP for any funded research in the U.S. DOT 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 their life cycle. 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. Finally, a DMP can help ensure that research data retain their 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 principal investigators

(PIs) of NSF grant submissions. These PIs reported widespread uncertainty about the new DMP requirements and actions needed to respond adequately. Later studies have provided some additional details. The results show that

- 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 vary widely by discipline;
- Researchers desire support for their data management needs;
- Strategies for sharing data vary dramatically among researchers;
- Research assistants handle the majority of tasks related to data management (with the exception of data sharing, which is addressed by the PI); and
- 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.

DMPTool: https://dmptool.org

ICPSR: www.icpsr.umich.edu

UK Data Service: www.ukdataservice.ac.uk

Australian National Data Service: www.ands.org.au

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 not only on their data services, but on funding agency requirements and expectations. Finally, the U.S. 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

Essential DMPs

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 U.S. DOT's requirement, are based on the criteria developed by the National Science Foundation and often include 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 discussed below.

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 necessarily have to be shared, 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 they are 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, and so forth. Rough estimations are sufficient.

Including information as to how the data will be documented in your DMP is a good idea.

Standards Section of the DMP

This section contains information about how the data will be prepared according to community expectations. The standards 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 data set. 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 is not mentioned by NSF or the U.S. DOT directly, including information as to how the data will be documented in your DMP is a good idea.

Access Section of the DMP

This section 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.

Use Section of the DMP

Once someone is able to access and acquire the data, what will they be allowed to do with them? 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 the data have been made available. Following are some elements to consider in writing this section:

Checklist:

Elements to consider in the DMP Use Section

- ☑ **Reuse:** Can people reuse the data for research, commercial, or other purposes? Can people create derivatives of your data set or use it to develop new data sets?
- ☑ **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 they can be used? Creative Commons licenses are often applied to content in repositories, but they may be somewhat problematic to apply to data, as Creative Commons licenses depend on copyright law, and data sets may not be covered. Other licenses to consider for data are those provided by the Open Data Commons.

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. Following are some questions to address in the DMP:

Checklist:

Questions to address in the DMP Archive/Preserve Section

- ☑ What is the long-term value of the data and how long should they 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 (e.g., documentation, software, reports) needs to be preserved alongside 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.

Beyond minimum
DMPs

Going Beyond the Minimum for Data Management Plans

DMPs submitted to funding agencies are often limited in how long they should be. 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 data sets 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 data sets, 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 that can inform where additional support and resources may be needed to ensure that organizational goals are being met.

Following are some additional questions to consider including in DMPs that are locally focused:

- Beyond funding agency requirements, what additional details are needed to make this DMP fully actionable and useful to the organization?
- Who are the stakeholders for the research project and/or the data set(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 data set(s) being developed? Are there any special needs for the

data set(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 data set(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.

Creating DMPs will become easier over time as they become more commonplace.

Chapter Checklist

From this chapter, you should be able to

- ☑ Know the U.S. DOT 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 the data accessible, (4) provisions for 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.

CHAPTER 9.

Building Blocks of the Solution

In This Section

The Big Picture — U.S. DOT 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 three essential systems and services components (see Figure 8, "The Big Picture." on page 76).

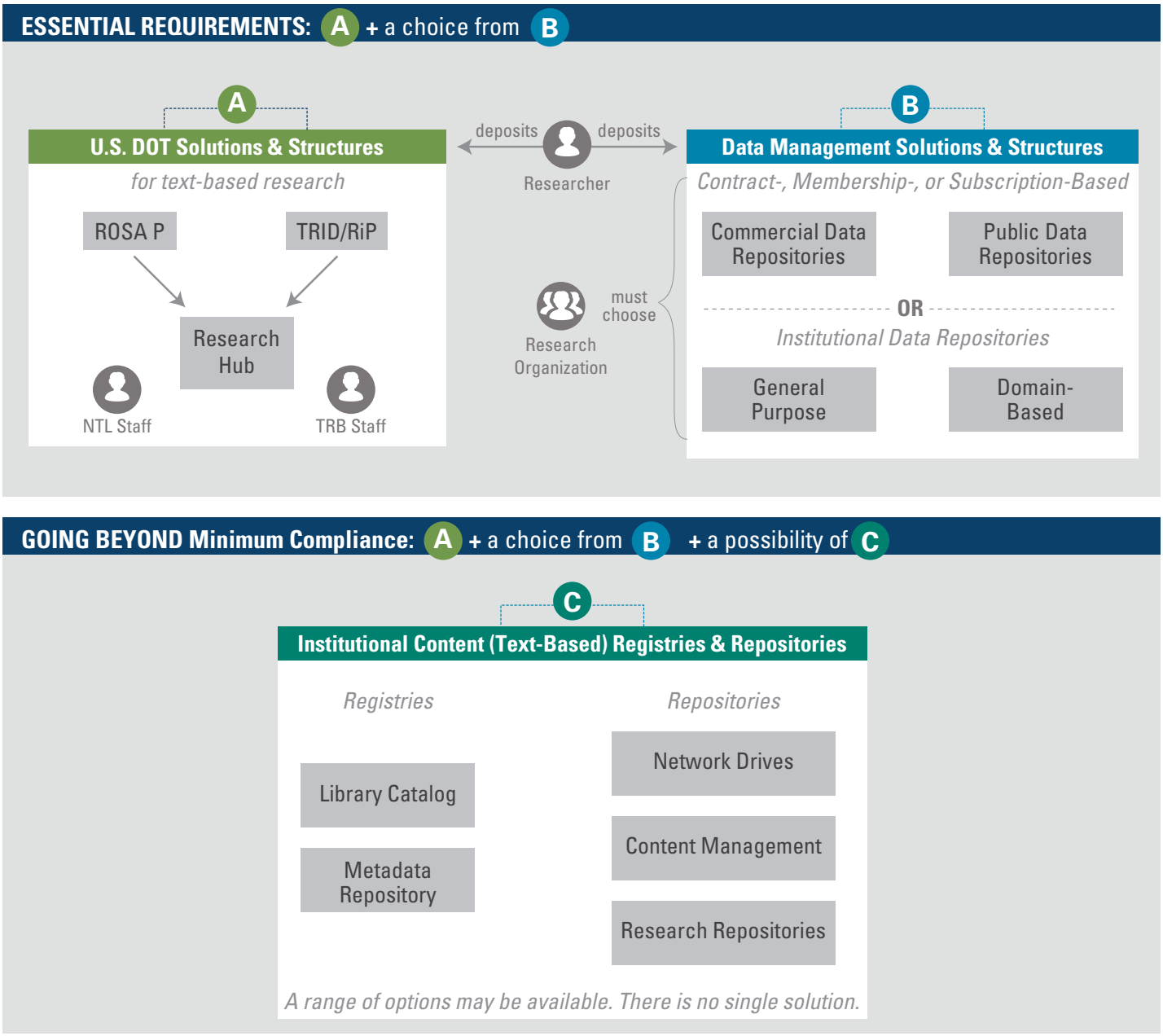
Figure 8 identifies three sets of systems and services, including

- Those provided by the U.S. DOT to state DOTs and other research institutions to satisfy essential requirements today,
- Those related to data repository and management solutions to satisfy both essential requirements and long-term goals, and
- Those related to local institutional registries and repositories for written research products to satisfy long-term goals.

U.S. DOT Systems and Services Today

The U.S. DOT 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.

Figure 8. The Big Picture



At the time this Guide was written, the U.S. DOT did 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 the short and the 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 U.S. DOT, and
- Domain-specific solutions based on the research community.

Needs to be a shift from research-based to institutional-based decisions.

Selecting the Best Solution

In today's research climate, it is most often 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, who makes the choice on where and how data are 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 its 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.[1] 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 a state DOT elects to create its own collection, the process 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 the 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 whether the repository is in compliance with the U.S. DOT Public Access Plan. Any questions that arise from this review should be discussed with repository personnel.

Non–U.S. DOT 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 U.S. DOT solution in place means that the organization doing the research is relegated to an indirect rather than direct role in achieving the Open Science vision. When an organization chooses to establish or use a local solution, it then has the ability to internally track and manage its research assets and, thus, to have a direct role in accomplishing the Open Science vision. Regardless, use of a registry or repository is essential to achieving discovery and use and to sharing within and without the research organization. Such systems enable interdisciplinary collaboration across research communities and subject areas.

Checklist:

Determining what an organization has in place

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 a local search system?
- ☑ 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, and 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 for storing and preserving 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 U.S. DOT 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

1 M. D., Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3(1):160018. http://dx.doi.org/10.1038/sdata.2016.18.

CHAPTER 10.

Learning and Training

In This Section

Learning Strategies for the Research Organization

Merely developing a strategy for addressing 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 their organizational roles, 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 to form a culture of practice around learning in the organization.
- Provide resources and support for staff (both trainers and trainees).
- For trainers: Ensure that adequate time is provided to teach and that teaching is recognized as a valuable contribution to the organization.
- For trainees: Ensure that time is provided to learn and that 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 to 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 the available time and capacity of the target audiences.

Educational programs need to include best practices and guidelines of relevant research disciplines as well as of the open access and data curation community. They also should include the 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 U.S. 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 U.S. DOT policy.

USGS life cycle:

https://www2.usgs.gov/datamanagement/

That said, existing programs can still serve as potential models for the structure and content in developing training programs for research institutions. The USGS, in particular, is an example in providing useful resources and guidance, particularly in its use of a defined research life cycle as a means to ground its educational programs, its ability to break down concepts to make them easier to understand, and its assistance to researchers in making connections between these concepts. Whether from the USGS or other organizations, life-cycle 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 on the U.S. DOT 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 U.S. DOT National Transportation Library's 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; and
- ☑ Ensure that research project descriptions are submitted to TRB'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 Awareness Training for Everyone" and "Training for Researchers" sections below.

Essential
Training Programs

Checklist of training program contents required by the Public Access Plan

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. Beyond training programs for the researchers themselves, additional 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 U.S. DOT 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 U.S. DOT 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 for managing and sharing articles and data. Developing training programs concurrently with developing the goals of the organization can provide opportunities to discuss stakeholder assumptions, confirm understandings of goals, and explore what is needed to ensure that the goals of the organization will be achieved.

Beyond minimum
Addressing the needs and goals of the organization

The Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) tool can help research organizations identify specific training needs of researchers and staff on the basis of an assessment of current knowledge and practice (see 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 Executive 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 the organization's expectations for the management, sharing, and preservation of research materials, data, and publications. To see the big picture of the research life cycle, all stakeholders need instruction in what research and data management are and why they are important. 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.

Checklist:

Baseline Curricula

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 Life-Cycle 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/**

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**

ICPSR — Guide to Social Science Data Preparation and Archiving: A detailed, step-by-step manual for social science researchers interested in developing a data set that can be deposited into the ICPSR repository or simply understood and used by others. **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.

Checklist:

Baseline Curricula
(continued)

Training for Researchers

As the producers of the data, researchers will require training in how to develop data sets 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, to satisfy the requirements of the U.S. DOT or other funding agency, and developing longer plans that articulate how the data will be managed throughout their life cycle.

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 the application, during the project, and reporting at end?
  - ☑ U.S. DOT essential requirements.
  - ☑ Tools/resources.
- ☑ In-depth training on publication and DMP requirements for important journals in the field —
  - ☑ How to identify the journals.
  - ☑ What the journals' open-access policies are and how they comply with funding agency requirements.
  - ☑ What the journals' data-sharing policies are.
  - ☑ Common elements of data-sharing policies.
  - ☑ Guidance on complying.
  - ☑ Tools/resources.

Checklist:

Minimum Researcher Curricula

Checklist:

Advanced Researcher Curricula

More advanced information that builds on the DMP requirement should be added to training programs as quickly as possible. This includes

- ☑ Introduction to a data life-cycle model that best suits the data and organizational culture of the state DOT (or other research institution) —
  - ☑ Use of a life cycle 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, and
  - ☑ 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

Checklist:

Executive and Management Curricula

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 life-cycle 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, and
  - ☑ 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.

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 institutions of higher education 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 the organization, technology, and resources aspects of managing and curating data.

http://dcc.ac.uk/resources/tools/cardio

Training for Research Support Roles

Within a research organization, there are multiple roles that provide support to researchers and the research process. Each type of role will require specialized training directly relevant to its 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.

Ideally, researchers and those who support research should engage in shared training.

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 data sets 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.

Training topics beyond basic courses would depend on 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; and
- 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. Good questions to ask may include the following:

Licensing: Good questions to ask

- 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.

Information Technology Personnel

Information technology (IT) personnel are often responsible for providing the technical infrastructure 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 on the organization's strategy for sharing and preserving data, this may require training on the following topics:

- **Internal Repositories:** Some questions to consider in building an in-house publications or data repository: What are the existing options? What are the existing resources' strengths and weaknesses in meeting federal and other requirements? What investment and support are 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 and 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 U.S. DOT 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.

CHAPTER 11.

Estimating and Managing Costs

In This Section

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 the 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 organizational 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 its 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. When 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 on the basis of the 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 data management plans (DMPs), and (2) the organizational level, which must include cost estimates across all research projects.

Essential

Cost Factors: Essential Requirements for the Research Organization

The most effective strategy starts at the organizational 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 3 to 5 years. Organizational 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 organizational level cost models are established, the project-level cost estimates can be modeled on 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

Table 6 provides a high-level description of the types of costs a research organization should expect to incur when complying with U.S. DOT essential requirements. Organizations should consider their essential

Table 6. Cost Factors for Meeting Essential Requirements

Step	Focus	Cost Factors
1	Raise awareness of federal policies	• Communication • Content development costs
2	Raise awareness of essential requirements	• Training development & delivery costs
3	Raise awareness of Open Science vision & going beyond the minimum	• Communication costs • Content development costs
4	Develop long- and short-term strategies	• Administrative overhead costs • Meeting and consultation costs
5	Identify stakeholders and assign roles and responsibilities	• Human resources costs • New personnel costs
6	Manage research text products	• Training development & delivery costs • Content data management costs • New personnel costs
7	Manage research data	• Training development & delivery costs • New personnel costs • Repository subscribe, buy, or make costs
8	Develop data management plans	• Training development & delivery costs • Content data management costs • New personnel costs
9	Develop costing strategies	• Administrative overhead costs • Meetings and consultation costs
10	Assess progress against requirements	• New personnel costs

compliance strategy, which focuses either 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.

There are eight cost factors for which costing strategies should be developed.

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 are calculated on the basis of administrative salaries and project time estimates and should be available from human resources.

Communications Costs are one-time sunk costs. These costs will include communication methods and systems as well as the salaries and time of communications specialists.

Policy Content Development Costs are one-time sunk costs. These costs are calculated on the basis of administrative and analyst salaries and would be available from human resources.

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 operational costs.

Data Repository Subscribe, Buy, or Make Costs are recurring operational costs. Most state DOTs will choose to subscribe to or buy a membership in an existing repository rather than 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 data set storage to organizations paying an annual subscription or membership fee. In some cases, fees for individual data set 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 (see Chapter 2) provides a clear description of its 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.

NASA Cost Estimation Tool

https://opensource.gsfc.nasa.gov/projects/CET/index.php

The Cost Estimation Toolkit (CET) is designed to provide NASA budget estimators, principal investigators, project managers, and resource planners with the capability to generate life-cycle 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 users 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 use in determining costs for DMPs. In developing the cost calculator, organizations should strive for cost estimates that will cover the organization's costs and will be considered reasonable by 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. (See Figure 9.)

Cost Factors: Essential Requirements for Individual Researchers

Cost Factors for Report Preservation

Since the National Transportation Library provides a registry and repository for all reports that fall under the purview of the U.S. DOT Public Access Plan, the cost of compliance most likely will be low. In addition, planning ahead significantly reduces costs, while registering the project and researcher Open Researcher and Contributor ID (ORCID) identifiers with the TRB Research in Progress (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 U.S. DOT 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 summarizes essential requirements and possible activities and their associated costs.

Table 7. Essential Requirements and Possible Activities with their Associated Costs

Activity	Questions to Consider
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 such as 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 the 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?
DIGITIZATION	• Do analog or paper-based research data (maps, newspaper clippings, photographs, images, text) need to be digitized to increase their potential for sharing?
FILE FORMAT	• Do data need to be converted to a standard or open format with long-term validity for long-term preservation?

Associated Costs

Low to no additional cost incurred if data description is carried out as part of data creation, input, or transcription.

Higher cost incurred if added afterward.

Tip: Codebooks for data sets can often be easily exported from software packages.

Low to no additional cost incurred if carried out as part of data entry and prep before analysis.

Higher cost incurred if needed afterward.

Low to no additional cost incurred if all data creation steps are well documented and documentation is kept well organized during research.

Higher cost incurred if documentation is written or compiled specifically afterward.

Tip: Often essential contextual and methods documentation will be written up in publications and reports.

Low to no additional cost incurred. Completing a repository deposit form may take 1-2 hours.

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.

Low to no additional cost incurred if planned beforehand by developing templates and forms for individual data files (transcripts, spreadsheets, databases) and constructing clear file structures.

Higher cost incurred if added afterward.

Tip: Free software exists for batch file renaming to harmonize file names.

Higher cost incurred if not planned as part of research practice.

Tip: (1) Consider cost of (time needed for) developing procedures, templates, and guidance for transcribers. (2) Calculate time needed for transcription — 4 to 8 hours per hour recording. (3) Use transcribing calculator: https://www.socialsciences.manchester.ac.uk/morgan-centre/research/resources/toolkits/toolkit-15/.

Low to no additional cost incurred for simple image scanning of text. Remember to include cost of time to scan.

Higher cost 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).

Higher cost incurred for audiovisual 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.).

Table 7. ***(continued)***

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 and how many backups should be stored.
DATA SECURITY	• Protect data from unauthorized access or use and 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, but less so in quantitative surveys where data can more easily be anonymized.
ANONYMIZATION	• Do you need to remove identifying information or conceal the identity of participants (e.g., by 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 life cycle?
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 and frequency of backup and storage media needed.

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

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 recontacted after research has finished to obtain retro-active 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 afterward); or (4) for qualitative textual data (e.g., interview transcripts) if carried out or at least highlighted during transcription.

Higher cost incurred (1) if variables need recoding afterward 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 audiovisual data (anonymizing/editing voices or faces can be very costly and reduces the usefulness of data). Also, 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?

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.

Beyond minimum

Cost Factors for Data Management

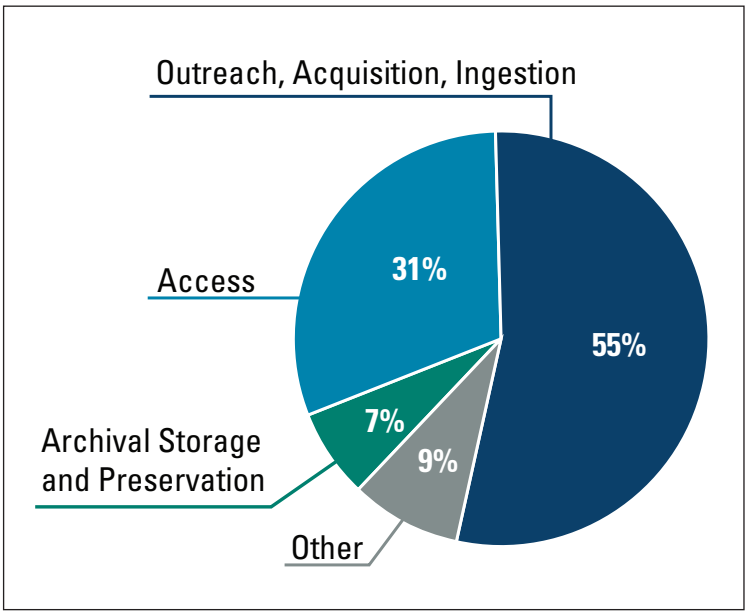
Active data management is required for essential compliance. Related costs accrue starting with data creation and continuing through describing, cleaning, organizing, storing, and preparing the data for long-term preservation. The UK Data Service website is an excellent resource, in particular its data management checklist for estimating costs of good data management (https://www.ukdataservice.ac.uk/manage-data/plan/costing). Table 8 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 afterward and documenting to comply with essential requirements.

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 subactivities listed in Table 8.[1]

Figure 10. Approximate Activity Data Costs



Studies have shown that the majority of costs (about half of lifetime costs) accrue at the time of acquisition and ingestion (Figure 10). Preservation makes up about a sixth of the remaining costs, although they decline over time. Access rounds out the other third.[2]

Personnel are the highest costs encountered in the maintenance of 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."[3]

Repositories need skilled personnel to manage infrastructure and preservation

Table 8. Costs Associated with Creating a Repository

ACTIVITY	SUBACTIVITY
Acquisition	• Acquisition, evaluation and selection relating to preservation • Submission agreement • Ordering and invoicing • Depositor support
Ingestion	• Receiving submission • Performing quality assurance • Generating metadata and documentation • Reference linking • Validation • Deposit • Holdings update
Archive storage	• Receiving data from ingestion • Managing storage hierarchy • Replacing media • Disaster recovery • Error checking • Providing copies to access
Preservation planning	• Monitoring of designated user community • Monitoring technology • Developing and implementing preservation strategies and standards • Developing packaging designs and migration plans
Preservation action	• Generation of preservation metadata • Refreshment • Backup • Re-ingestion • Inspection • Disposal
Data management	• Administering database • Performing queries and generating reports • Receiving database updates • Loading new descriptive information • Loading new archive administrative data
Access	• Access provision • Access control • User support • Rights management
Administration & services	• General management • Administrative support • IT support • Physical security • Utilities • Supplies inventory and logistics • Staff training and development

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 afterward).
- ☑ 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 See Alexandra Allen, *General Study 16 – Cost Benefit Models: Final Report*. InterPARES 3 Project, TEAM Canada, 2013, http://www.interpares.org/ip3/display_file.cfm?doc=ip3_canada_gs16_final_report.pdf. For another examination of costs, although by functional areas, see Neil Beagrie, Julia Chruszcz, and Brian Lavoie, *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities*, JISC Final Report, 2008, https://www.webarchive.org.uk/wayback/archive/20140613220103/http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf. See also *User Guide for Keeping Research Data Safe: Assessing Costs/Benefits of Research Data Management, Preservation and Re-Use*, v. 2.0, https://beagrie.com/static/resource/KeepingResearchDataSafe_UserGuide_v2.pdf.

2 *Costs Factsheet*, Cost-Benefit Advocacy Toolkit, CESSDA SaW, Charles Beagrie and CESSDA 2017, https://www.cessda.eu/eng/content/download/1430/20402/file/03_Costs_Factsheet_final.pdf.

3 Stephen Abrams, Patricia Cruse, John Kunze, and Michael Mundrane, *Total Cost of Preservation (TCP): Cost Modeling for Sustainable Services*, Rev. 0.16 – 2012-04-09, UC Curation Center, California Digital Library, University of California, Office of the President, https://confluence.ucop.edu/download/attachments/163610649/TCP-total-cost-of-preservation.pdf.

CHAPTER 12.

Assessing and Managing Progress

In This Section

Once a system for preserving the products of research has been implemented, 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 U.S. DOT's essential requirements. However, over time, regular assessment can help states in going beyond as they move toward 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 Research in Progress (RiP) database shortly after the award is received 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 Open Researcher and Contributor ID (ORCID) identifier.
4. RiP project status should be changed from "active" to "completed" within 2 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 1 year of the end of the project, an email including the following information should be sent to the U.S. DOT Research Hub (RH), the National Transportation Library (NTL), and TRB:
    - Final report URL(s) or PDFs for any resulting publications,
    - URL(s) and associated descriptive metadata for any final data sets arising from the research project,
    - The funding agreement number of the project,
    - The RH Display ID for the project,

Table 9. Tracking

What is Tracked?	Assessment Metric
Number and status of eligible funded projects (tracked at time of funding)	Total count, total in progress, total completed
Number of projects entered in RiP database	Number of entered/total projects
Number of projects with approved DMP	Number of DMPs/total projects
Number of projects labeled "completed" in RiP	Number of marked completed/total completed projects
Number of data deposits	Number of deposited/total completed projects
Number of correct technical documentation pages	Number of correct/eligible reports Number of correct/total completed projects
Number of successfully deposited reports in NTL registry	Number of deposited/total reports eligible Number of deposited/total completed projects

- ORCID identifiers (unique researcher IDs) for all publication author(s), and
- 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 listed above throughout the course of the project. Table 9 shows each item to be tracked, one or more relevant metrics for assessment, and target levels.

Target Level	Action
Not applicable	Shows how the Public Access Plan affects research and how the research organization contributes to public access.
Always near 100% because projects must be entered when funded	If low, researcher training or communication with researchers may be needed.
Always near 100% because DMPs are required near the start of each project	If low, researcher training or communication with researchers may be needed.
Near 100% with 2-month window	If low, compliance personnel may not be advising researchers of the need to mark completion.
Should increase over time	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.
Should increase over time, but counting technical documentation can be more difficult than tracking reports or deposits. This metric requires the researcher either to submit documentation to his or her organization (e.g., the state DOT), in addition to submitting it to the sponsor, or to report that the task was completed.	Support for documentation and communication with researchers about the documentation requirement will be critical if this number is low.
Should increase over time. For percentage of completed projects, near 100% should be possible, since it should be feasible to deposit one report per project. However, percentage of eligible reports may be more challenging, since some eligible reports may be missed early.	The ideal metric is the percentage of eligible reports that are deposited, but tracking eligible reports is more difficult than tracking completed projects. Low deposit rates indicate a need for more training and communication with researchers.

Assessment of Process and Support for Essential Requirements

The metrics listed above directly address the essential requirements for compliance with the U.S. DOT 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 the 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 researchers 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.

Checklist:

Interaction Survey: Question Topics

Specific topics for survey questions should include

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

In addition to surveys, some specific measurements will also be useful. These may include

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

Assessing Policy and Process

Like the 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

Checklist:

Policy & Process Survey Questions

- ☑ 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

Beyond minimum
Expanding the eligible publications

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 by contracting with an outside repository. The assessments previously described 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

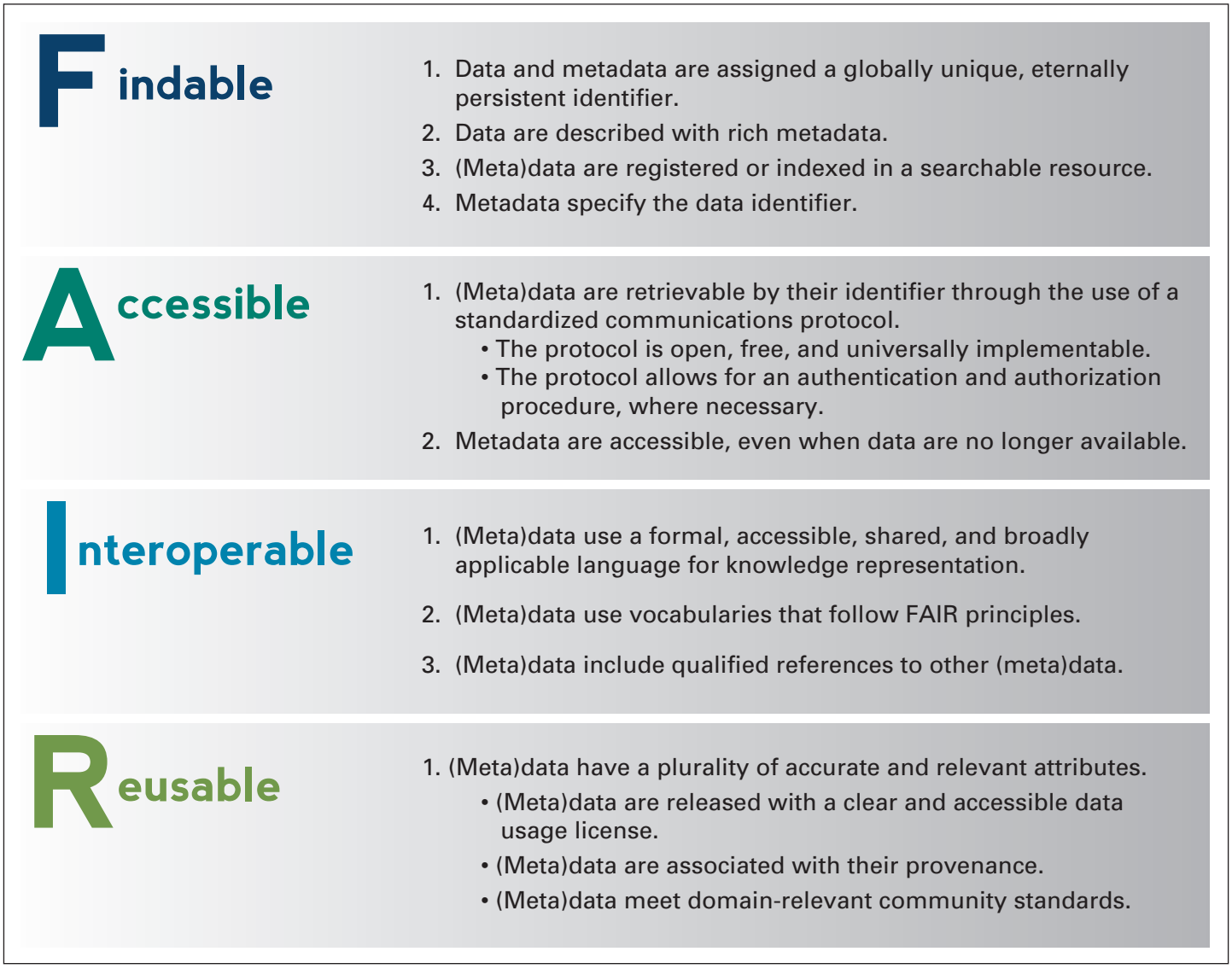
A good starting point for assessment of impact is to ensure that the preservation process meets the FAIR guiding principles: Findable, Accessible, Interoperable, and Reusable (see Figure 11, p. 108).

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. For instance, researchers do not need to procure 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 a user survey. The survey should ask about the user's experience with reuse of other data sets, 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 measure impact directly, through tracking of data set downloads and citations. The number of data sets reused and the number of papers or data sets 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



Developing the Assessment Process

Good metrics require early establishment of a 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) a tracking system established when the researcher is first awarded a project that uses federal funds, and (3) embedding the survey into standard operating procedures at all levels.

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

CHAPTER 13.

Putting It All Together

In This Section

Checklists can serve as a foundation for more advanced management of people, processes, and resources. This chapter presents 12 checklists to help state DOTs and other transportation research institutions track their progress against essential requirements. The first three checklists cover foundational steps, and the remaining nine checklists track the steps necessary to meet essential requirements. Each of these nine checklists is accompanied by additional tasks to undertake to go beyond the minimum. The 12 checklists together constitute 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 U.S. DOT 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 all 12 checklists or focus only on specific checklists. The checklists can provide a guide to assessing where your organization is now in complying with essential requirements. Organizations can use the checklists as a working tool to do the following:

A checklist is a simple aid for managing a complex set of tasks.

- 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; and
- Involve all stakeholders, both current and future, in the practicalities of planning and accomplishing compliance.

The Guide suggests that organizations use the checklists to assess the status of each requirement by noting whether it is (1) complete, (2) in progress, or (3) not yet started. The checklists 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 **r**esponsible or **a**ccountable (R, A) will prove the most effective as a starting point. Organizations might also use the checklists to identify stakeholders who might be expected to **s**upport (S) the requirement or who should be **c**onsulted or **i**nformed (C, I).

Foundational Checklists

The first three checklists cover one-time steps that establish a foundation for both short- and long-term strategies. The remaining checklists pertain to meeting essential requirements. The checklists for Steps 4 through 12 cover specific steps in the overall research preservation process and correspond to the chapters in the Guide.

Using the Checklists to Go Beyond Essential Requirements

Assuming the state DOT has developed a strategy for meeting essential requirements, there should be a good understanding of the U.S. DOT 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. Those state DOTs that want to go beyond the essential requirements should focus on the checklists for Steps 4 through 12. For each of these steps, activities that address going beyond the minimum are identified.

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.

Foundational Steps

Step 1. Raise Awareness of Federal Policies

Step 2. Raise Awareness of Essential Requirements

Step 3. Raise Awareness of Open Science Vision & Going Beyond the Minimum

STEP 1. FOUNDATIONAL — Raise Awareness of Federal Policies
Identify and Track the National- and U.S. DOT–Level Policies and Authorities
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Understand the Purpose and Structure of the Guide
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Understand Essential Requirements
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Understand the Consequences of Noncompliance
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 1 (continued)

Understand the Concept of Going Beyond the Minimum
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

STEP 2. FOUNDATIONAL — Raise Awareness of Essential Requirements
Understand What Research Is Eligible Under the Policy
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Know How to Track and Identify Your Funded Research Projects
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Understand *What* to Do *When* in the Research Project Life Cycle
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 2 (continued)

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 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:
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:

Step 2 (continued)

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 U.S. DOT'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:

Steps for Meeting Essential Requirements

Step 4. Develop Short- and Long-Term Strategies

To go beyond the minimum, include the following tasks:

- 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.

STEP 4. Develop Short- and Long-Term Strategies
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:
Design a Short-Term Strategy That Can Grow into a 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:

Step 4 (continued)

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 a Method for Monitoring the 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. Identify and Support Stakeholders and Assign Roles and Responsibilities

To go beyond the minimum, include the following tasks:

- Define an expanded stakeholder model for going forward.
- Define expanded roles and responsibilities for going forward.
- Anticipate impacts to the organizational culture.
- Anticipate new communication patterns.
- Design and implement new governance models and methods.

STEP 5. Identify and Support Stakeholders and Assign Roles and Responsibilities
Identify Roles and Responsibilities for Meeting Essential Requirements
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 5 (continued)

Map Roles to Organization's Job Titles
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Review Current Responsibilities
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 6. Manage Research Text Products

To go beyond, augment the Step 6 checklist by including tasks that pertain to a broader scope and coverage of research publications and reports at the local organizational 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 the development of local content registries and repositories.

STEP 6. Manage Research Text Products
Identify Eligible Written Research Products
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 6 (continued)

Distinguish Formal Publications and Manuscripts and Final Project and Technical Reports
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Understand How to Submit Written Research Products to NTL, TRB, and RH
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Understand What Information to Provide to NTL, TRB, and RH
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Understand When to Submit to NTL, TRB, and RH
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 6 (continued)

Understand the Metadata Services Provided by NTL, TRB, and RH
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 7. Manage 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. Those organizations that do opt for a local solution over the longer term should consider a full project plan with full cost models.

STEP 7. Manage Research Data
Define Essential Requirements for Research Data as Described by the U.S. DOT
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Interpret Essential Requirements for Your Organization's Transportation Research
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 7 (continued)

Understand the Organization's Roles and Responsibilities for Data Management and Preservation
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 8. Develop and Use Data Management Plans

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

STEP 8. Develop and Use Data Management Plans
Understand the Purpose of a DMP
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
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:

Step 8 (continued)

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. Understand the Big Picture for Transportation Research Management

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.

STEP 9. Understand the Big Picture for Transportation Research Management
Describe the U.S. DOT and TRB Systems and Services that Support Essential Requirements
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 9 (continued)

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 the Organization's 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 for organizations that want to go 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.

STEP 10. Learning and Training
Identify Stakeholder Groups and Training Needs
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 10 (continued)

Identify Resources Available from Which to Develop Targeted Training Materials
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Developing a Training Program
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 11. Determine Costs, Budgets, and Economic Value

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

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

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:

Step 11 (continued)

Calculate Essential Compliance Costs at the Organizational 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. Assess Progress Against Essential Compliance

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
Identify Metrics to Track Each Element of the Essential Requirements
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Step 12 (continued)

Set Up a System That Tracks Eligible Projects from the Start to Support Assessment
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:
Develop Questions for a User Survey That Will Address Policy and Process from the User Point of View
Status of Activity: ❑ Complete ❑ In Progress ❑ Not Yet Started
Action Required (if any):
Individual Responsible:
Individual(s) to be Consulted:

Abbreviations and acronyms used without definitions in TRB publications:

A4A	Airlines for America
AAAE	American Association of Airport Executives
AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
ACI–NA	Airports Council International–North America
ACRP	Airport Cooperative Research Program
ADA	Americans with Disabilities Act
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
ATA	American Trucking Associations
CTAA	Community Transportation Association of America
CTBSSP	Commercial Truck and Bus Safety Synthesis Program
DHS	Department of Homeland Security
DOE	Department of Energy
EPA	Environmental Protection Agency
FAA	Federal Aviation Administration
FAST	Fixing America's Surface Transportation Act (2015)
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
HMCRP	Hazardous Materials Cooperative Research Program
IEEE	Institute of Electrical and Electronics Engineers
ISTEA	Intermodal Surface Transportation Efficiency Act of 1991
ITE	Institute of Transportation Engineers
MAP-21	Moving Ahead for Progress in the 21st Century Act (2012)
NASA	National Aeronautics and Space Administration
NASAO	National Association of State Aviation Officials
NCFRP	National Cooperative Freight Research Program
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
PHMSA	Pipeline and Hazardous Materials Safety Administration
RITA	Research and Innovative Technology Administration
SAE	Society of Automotive Engineers
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (2005)
TCRP	Transit Cooperative Research Program
TDC	Transit Development Corporation
TEA-21	Transportation Equity Act for the 21st Century (1998)
TRB	Transportation Research Board
TSA	Transportation Security Administration
U.S. DOT	United States Department of Transportation

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001

ADDRESS SERVICE REQUESTED

NON-PROFIT ORG.
U.S. POSTAGE
PAID
COLUMBIA, MD
PERMIT NO. 88

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.
www.nationalacademies.org

ISBN 978-0-309-48158-8
90000
9 780309 481588