

SECTION 4: DATA MANAGEMENT PLAN AND METADATA SCHEMA

Chapter 11 Federal and State Requirements

Chapter 12 Data Management Policies for the WYDOT Research Center

CHAPTER 11: FEDERAL AND STATE REQUIREMENTS

Since the turn of the century, the Federal government and the State of Wyoming began implementing rules and regulations regarding digital data, data storage, data archiving, and open data sharing plans. This Chapter outlines the Federal and state requirements for data management and digital data.

11.1 American COMPETES Reauthorization Act of 2010

The *American COMPETES Reauthorization Act of 2010* (ACRA; Public Law 111-358)¹, was signed into law in 2011, and required the director of Office of Science and Technology Policy (OSTP) to establish a working group under the National Science and Technology Council (NSTC) direction to coordinate Federal science agency research and policies related to the dissemination of long term stewardship of the results of unclassified research, including digital data.

11.2 Office of Science and Technology Policy Memorandum, dated February 22, 2013

In February of 2013, OSTP released a memorandum entitled *Increasing Access to the Results of Federally Funded Scientific Research*² (OSTP 2013 Memorandum), which directed the heads of executive departments to develop a plan to support increased public access to the published results of research funded by Federal funding. The OSTP 2013 Memorandum provided guidance on federally funded scientific research, including peer-reviewed publications and digital data. The OSTP 2013 Memorandum further sets out that “Federal agencies investing in research and development must have clear and coordinated policies for increasing such access” and a plan should be developed to “support increased public access to the results of research funded by the Federal Government.”

This plan directs all agencies to implement:

- A strategy for leveraging existing archives, where appropriate.
- A strategy for improving the public’s ability to locate and access digital data.
- An approach for optimizing search, archival, and dissemination features, while ensuring long-term stewardship of the results.
- A plan for notifying awardees and other federally funded scientific researchers of their obligations.
- An agency strategy for measuring and enforcing compliance with its plan.

¹ <http://www.nsf.gov/statistics/about/BILLS-111hr5116enr.pdf>

² http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

11.3 Executive Order 13642 dated May 9, 2013

The May, 2013, Executive Order 13642, entitled, *Making Open and Machine Readable the New Default for Government Information*³ (EO 13642 Memorandum), states that government information shall be “managed as an **asset** throughout its life cycle to promote interoperability and openness... and to ensure that data are released to the public in ways that make the data easy to find, accessible and usable.” [emphasis added] In the EO 13642 Memorandum, the Office of Management and Budget (OMB) was directed to issue an Open Data Policy to advance the management of Government information. Agencies, including the USDOT, were directed to implement the requirements of the OMB Open Data Policy.

11.4 Office of Management and Budget Memorandum dated, May 9, 2013

In May of 2013, the OMB issued a memorandum entitled *Open Data Policy for Managing Information as an Asset* (M-13-13).⁴ This memorandum required executive departments and agencies (Agencies) to “collect and create information in a way that supports downstream information processing and dissemination activities”, which included using “machine-readable and open formats, data standards, and common core and extensible metadata for all new information creation and collection efforts.” Agencies, including the USDOT, were required to ensure the use of “open licenses and review of information for privacy, confidentiality, security, or other restrictions to release.” M-13-13 is in line with the *Paperwork Reduction Act*,⁵ the *Privacy Act of 1974*,⁶ the *Federal Information Security Management Act of 2002 (FISMA)*,⁷ the *Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)*,⁸ the *Freedom of Information Act*,⁹ and the existing OMB and the OSTP guidance policies. M-13-13 applies to all new information collection, creation, and system development efforts as well as major modernization projects that update or re-design existing information systems beginning in the early stages of the planning process, and management should continue throughout the projects life cycle.

11.5 USDOT Plan to Increase Public Access to the Results of Federally-Funded Scientific Research Results Version 1.1, dated December 16, 2015

The USDOT *Plan to Increase Public Access to the Results of Federally-Funded Scientific Research Results Version 1.1, dated December 16, 2015* (USDOT Public Access Plan)¹⁰, was issued in response to the OSTP 2013 Memorandum. The USDOT Public Access Plan requires:

³ <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>

⁴ <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

⁵ Pub. L. No. 107-347, 116 Stat. 2899 (2002) (codified as 44 U.S.C. § 3501 note).

⁶ 5 U.S.C. §552a

⁷ 44 U.S.C. §3541, *et seq.*

⁸ Section 503(a), Pub. L. No. 107-347, 116 Stat. 2899 (2002) (codified as 44 U.S.C. § 3501 note); see also Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), available at http://www.whitehouse.gov/sites/default/files/omb/assets/omb/fedreg/2007/061507_cipsea_guidance.pdf

⁹ 5 U.S.C. 552(a)(2)

¹⁰ <https://www.transportation.gov/sites/dot.gov/files/docs/Official%20DOT%20Public%20Access%20Plan%20ver%201.1.pdf>

- Authors to submit to the NTL digital repository all publications that meet the scope criteria as set out on page 2 of the USDOT Public Access Plan.
- Publications to be freely available to the public no later than 12 months following publication.
- Public access to final research data, subject to restrictions that pertain to privacy, confidentiality, and homeland security.
- All DOT funded research proposals to include a supplementary document labeled “Data Management Plan” (DMP), and include long-term digital data set preservations and storage location information.
- Repositories to enable and allow public access and sharing.
- Full access to the DOT’s scientific publications and digital data requirements for tracking of the research and data.
- Data to be entered into the NTL repository that serves as the permanent archive of technical reports. The NTL will provide a searchable database of data management plans.
- The USDOT Research Hub to serve as the linking mechanism for scientific publications and their underlying research data.
- All research project descriptions to be submitted to the Transportation Research Board Research-in-Progress database and ensure that the projects are updated.

The plan applies to all DOT employees and awardees from non-DOT organizations that publish scientific material or compile digital data sets resulting from research and development programs conducted under a DOT grant, contract, or other agreement.

11.6 FHWA - Data Capture and Management: Needs and Gaps in the Operation and Coordination of USDOT Data Capture and Management Programs

In 2010, the FHWA Office of Operations, Office of Transportation Management (HOTM) commissioned a white paper entitled, *Data Capture and Management: Needs and Gaps in the Operation and Coordination of U.S. DOT Data Capture and Management Programs* (HOTM White Paper). The HOTM White Paper focused on infrastructure, travel, and climate, and looked at data capture and management activities across various U.S. DOT program areas, and identified gaps and potential opportunities for filling the gaps and managing program activities. The HOTM White Paper recommended that HOTM should develop a data business plan (DBP) to address the gaps and to improve coordination among real time data capture programs within U.S. DOTs.

In 2011, HOTM commissioned a DBP study to address gaps; to serve as a prototype for other U.S. DOT offices; and to provide leadership by offering or suggesting best data collection practices. In 2012, the FHWA Data Governance Council signed a charter that provided strategic review and oversight of all FHWA data collections efforts. The council signaled the need for improved data management within FHWA and its programs and allowed stakeholders to benefit from data programs.

11.7 USDOT - U.S. DOT Roadway Transportation Data Business Plan (Phase 1)

In 2013, *U.S. DOT Roadway Transportation Data Business Plan (Phase 1)* (U.S. DOT DBP), was written and focused on roadway travel mobility data, but also reviewed the above-mentioned papers and their recommendations and conclusions. The U.S. DOT DBP sets out practices and recommendations for the U.S. DOT based on experiences of other national and state agencies data management plans. The recommendations set out in U.S. DOT DBP followed best practices outlined in *NCHRP 666, Volume II: Guide for Target-Setting and Data Management*, Chapter 2: Guide for Data Management.

11.8 National Transportation Library DOT Public Access Guidelines

The NTL is the repository for USDOT information, serves as the portal for transportation data, and provides library services to USDOT employees. The NTL sets out that DMPs should conform to the USDOT Public Access Plan and that each plan shall include a narrative describing:

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

11.9 AASHTO

American Association of State Highway and Transportation Officials (AASHTO) is working on data management plans for research data, and the subcommittee of data is working on the core data principles proposed for use by state DOTs¹¹ regarding data management. The core data principles as set out by AASHTO are:

- Valuable: Data is an asset.
- Available: Data is open, accessible, transparent, and shared.
- Reliable: Data quality and extent is fit for a variety of applications.
- Authorized: Data is secure and compliant with regulations.
- Clear: There is a common vocabulary and data definition.
- Efficient: Data is not duplicated.
- Accountable: Decisions maximize the benefit of data.

¹¹ Stickel, Jack; Anita Vandervalk, (2015) Data Business Plans and Governance Programs, TRR 2460. <http://trrjournalonline.trb.org/doi/abs/10.3141/2460-17>; AASHTO Core Data Principles, <http://planning.transportation.org/Pages/Data.aspx>

CHAPTER 12: DATA MANAGEMENT PLAN POLICIES FOR THE WYDOT RESEARCH CENTER

Data is considered an asset and is valuable to Principle Investigators, WYDOT, stakeholders and the public. As such, Principle Investigators, Project Champions, and WYDOT become accountable for the quality of the data generated and they must ensure the data is accompanied with enough metadata to assist the public in accessing the data. Managing the data in a formal manner ensures that duplication of the data does not occur and that the research dollars are spent wisely.

This Chapter outlines the policies and procedures the Principle Investigator and Project Champion will follow when establishing their Data Management Plan (DMP). In developing the DMP and the policies, the Research Center balanced insuring the data was open to the public, the needs of the Principle Investigators and WYDOT, and insuring confidentiality and privacy. The Research Center is implementing the following policies and guidelines to be used for all research projects funded by the Research Center.

12.1 Steps for data management

All research projects shall follow the below timeline.

12.1.1 Prior to beginning a research project

A kickoff meeting shall be set up prior to or just as the contract is initiated. The kickoff meeting shall include reviewing the proposed DMP and necessary metadata to ensure all required data is archived and all parties understand how the data will be managed. Any changes to the DMP must be approved by the Research Center. The kickoff meeting will also ensure that the Principle Investigator(s) understands the deliverables set out in the contract. At this time, the parties will determine whether there needs to be disclosures or legal protection for the report and the data (PHI, PII, and/or sensitive data), and whether data needs to be scrubbed and/or deidentified. At this meeting, all rights to the data and/or datasets shall be determined. See Appendix 2, Template for Data Management Plan.

12.1.2 During the research project

During the research project, the Principle Investigator shall ensure that the data is maintained and backed up in a secure manner, and it shall be the Principle Investigator's responsibility to determine storage requirements. The Principle Investigator shall update the Metadata Schema (Appendix 3) during the life of the research project, as needed. The Principle Investigator shall also apply for an Open Research and Contributor ID (ORCID) number. Prior to providing the final report or publishing any publication based on the contracted research, the Principle Investigator shall discuss with the Research Center the need of a digital object identifier (DOI) for the report and/or data, and what DOIs have already been attached to derivative reports.

12.1.3 After the final report is written

The Principle Investigator, Project Champion, and WYDOT shall set the embargo period for the data; appraise the data inventory; organize the data and content files; process the data; ensure the Metadata Schema is complete; determine who will manage the data once the project is complete; determine what the retention period will be; determine who will budget for the retention of data; set out how to name the files and datasets; determine whether there are any outliers; determine how to cite the data and/or datasets; and determine what the data parameters and/or units of measure will be. If the data is sensitive, privileged, or confidential, the Principle Investigator and the Research Center shall determine who can access the data, who can reuse the data, and how to ensure confidentiality. It should be kept in mind that data formats may be different during the submission, distribution, and preservation phases.

12.2 What Data Are Archived and How is it Determined

For purposes of these Guidelines, the Research Center will use the following primary definition for data:

“the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.”¹²

When determining what data should be archived, and what data should be linked to the final report, the Research Center, Project Champion and the Principle Investigator will review the data available, and determine what data and datasets to archive and link. Determining factors of what data to archive include:

- Will there be a demand for the data.
- How difficult will the data be to replicate.
- Are there barriers to future use.
- Is the data stored or archived somewhere that cannot guarantee long-term storage.
- Is the data the only copy.

The following forms of data/datasets should be reviewed when determining what data should be archived:

- Primary data used in the production of the report: Raw, verified data that has been obtained directly from a source. It can be captured through experimental, surveys, interviews, focus groups, or other direct interactions with individuals in the field. Does not include analysis data.
- Unpublished datasets: Materials and methods; clear description of the variables presented; supported by unpublished reports; and any other relevant material.
- Secondary Data: Pre-existing data not gathered or collected by the authors. Usually collected by another organizations or source.

¹² https://www.whitehouse.gov/omb/fedreg_a110-finalnotice

- Metadata: Set of data that describes and gives information about the dataset – cataloging information.
- Dataset description document: Describes all variables in the dataset and the measurement units used.
- Codebook: A list of variable names, variable labels, and label values. Should specify the data position of each variable, describe the contents of each variable, and identify the range of possible codes and the meanings of the codes.
- Questionnaires: An unused copy of the questionnaire.
- Handbooks, guides, and manuals derived from research.

When determining what counts as data and what should be archived, it will depend on the Principle Investigator's knowledge of the data, and what he/she believes is valuable.

NOTE: It should be noted that data gathered and/or archived, will be determined on a case-by-case basis for each research project.

12.3 What Data Are Not Included

Data that does not need to be archived or saved includes preliminary analyses of a project, drafts, plans for future research, peer reviews, interoffice communications, e-mails, letter, or other forms of correspondence when looking at what data to archive. The Principle Investigator and the Project Champion will have the opportunity to discuss what data and/or other digital material should be excluded prior to finalizing the project.

12.4 Data Management Plan Template for Research Projects

Principle Investigators are required to complete a DMP prior to beginning a research project, and update the DMP throughout the life of the project. See Appendix 2.

12.5 Metadata Format and Content Template

All datasets, data and reports may need a Metadata Schema. Determination of what data will need a complete Metadata Schema will be determined on a case-by-case basis. Full details on the Metadata Schema can be found in the MARC21 Concise at <http://www.loc.gov/marc/marcdocz.html>, and the requirement from the National Transportation Library (NTL) can be found at <http://ntl.bts.gov/publicaccess/index.html>. The Research Center Metadata Schema can be found in Appendix 3.

12.6 Data Sharing Agreements

The Research Center and the Principle Investigator will execute a data sharing agreement, if one is required, or for intellectual property reasons. The data sharing agreement may include clauses on cleaning of the data; data sharing requirements; forms for sharing; formats of data; software

requirement; and metadata requirements. Templates for data sharing during the time of the research project, and after completion of the project are available at the Research Center.

12.7 Locating and Accessing Digital Data Materials, Data, and Databases

For research projects that were completed prior to October 1, 2014, the Research Center shall link digital materials, data, and/or databases, if possible. If the data was not been turned over to the Research Center, and the Principle Investigator and/or Project Champion still has access to the data/datasets, the Research Center may ask for a copy for its files.

For research projects completed after October 1, 2014, the Research Manager shall work with the Principle Investigator and the Project Champion to ensure the Research Center receives an inventory of digital material, data and/or datasets that are relevant to the research project. All digital material, data and datasets will be copyrighted, if necessary. From the inventory, the Project Champion, Principle Investigator, and the Research Center will determine what data will be linked to the report.

12.8 Ensure the Public Can Read, Access, Download and Analyze Digital Data

Pursuant to the *Freedom of Information Act*¹³, the *Wyoming Public Records Act*¹⁴, and the numerous State of Wyoming digital data management rules and regulations, the Research Center is required to ensure that the public can read, access, download, and analyze digital data. The goal of the Research Center is thus to work with the Principle Investigator, stakeholders, the Project Champion and the RAC to determine the best ways to accomplish this goal.

The Research Center shall archive data in a repository chosen by the Research Center, and link the data to the report that is stored in electronic form in the Wyoming State Library or though the NTL site. This will ensure the report and data are available to the public immediately. The Research Center does not wish to duplication storage efforts, and if data will be stored by the Principle Investigator or the Project Champion for the duration of its life cycle, the Research Center proposes to leave the data in those databases until such time as it is determine that the data should be either deleted or archived in a predetermined repository. The storage facility will be publicly accessible and will allow for linkage between the report and the data/dataset.

12.9 Data Formats and Dissemination

Before data is archived, the Principle Investigator, Project Champion and Research Center shall determine what the best format for archiving the data is. It must be kept in mind that formatting may change from collection to archiving and as such, the Principle Investigator should be prepared to adjust the formatting where necessary. A plan for formatting and archiving should be set up either prior to or during the research process.

¹³ 5 U.S.C. §552

¹⁴ *Wyo. Stat. §16-4-201 et. seq.*

12.10 Storage, Archiving and Backup

All data gathered from research projects should be considered “open”. This means that the data and/or datasets should be available to the public, unless there is a compelling reason not to release the information. Archiving will begin in the early stages of the research process and shall continue during the course of the project. When storing data, a backup must always be kept; security for the data must be incorporated during the project phase; and data must be organized in such a manner that it is easily accessible.

Archiving of the data could be a time specific endeavor or could be long term. Data must remain available for a minimum of three years after a project is complete, and it may be necessary to retain the data indefinitely. To ensure the availability of data, each set of data and report must have persistent identifiers or digital object identifiers (DOI) as set out by CrossRef.org. Data must also be formatted in such a manner that it can be stored and retrieved easily.

If the Principle Investigator and/or the Project Champion will be archiving and/or storing data, the DMP shall describe physical and cyber resources, and facilities that will be used for the effective preservation and storage of research data. In collaborative proposals or proposals involving sub-awards, the lead Principle Investigator is responsible for assuring data storage and access for data gathered by the subrecipient.

If there is sensitive and/or confidential data gathered, the Principle Investigator is tasked with ensure that the storage, backup and security systems meet all Health Insurance Portable and Accountability Act (HIPAA)¹⁵ and Health Information Technology for Economic and Clinical Health Act (HITECH)¹⁶ requirements. When archiving sensitive and/or confidential data, please provide the following:

- A brief description of the data.
- Whether the data was original or was gathered from an existing database.
- Security and privacy information.
- Intellectual property rights for any data that comes from an existing database.
- How the data were managed during the project, with information about versions.
- What outliers are present in the data.

12.11 Intellectual Property Rights

The Research Center wants to ensure that we are able to collect data while at the same time protecting intellectual property rights of agencies and the private sector. The Research Manager, in conjunction with the Attorney General’s office, will set out guidelines on who owns the intellectual property rights and proprietary interests for data, datasets, and other documents. The major questions to be kept in mind are:

¹⁵ <http://www.hhs.gov/hipaa/for-professionals/privacy/>

¹⁶ <http://www.hhs.gov/hipaa/for-professionals/special-topics/HITECH-act-enforcement-interim-final-rule/index.html>

- Who owns or has a valid license to the intellectual property.
- What steps should be taken to protect the intellectual property.
- When sharing data, how will the intellectual property rights be protected.

Issues such as disclaimers, open licenses and copyrights will be reviewed to ensure all data, datasets, etc. contain the proper coverage. All contracts for research projects will have specific language on respective rights and the obligations of all parties regarding the use, reproduction, and disclosure of the data, and will set out what data will be delivered to the Research Center. The type, quantity or quality of the data to be delivered will be set out in the DMP.

Data, which is first produced in the performance of the contract, must be copyrighted at the first instance. The copyright will list the contractor, the State of Wyoming and WYDOT as right owners, with all rights reserved. WYDOT retains an “unlimited, and irrevocable license to reproduce, publish, or otherwise use and to authorize others to use, for federal and state purposes: a) the copyright in any work developed during the contract period; and b) any rights of copyright to which the Principle Investigator and/or contractor purchases ownership using funds awarded”. (Wyoming Contract Template)

The Programming Engineer and the Research Manger, or his/her designee, shall be responsible for handling intellectual property issues within the Research Center.

The Research Center, and the Principle Investigator will determine what intellectual property will be released to the public, if there is an issue with confidentiality, trade secrets, etc. The Research Center and Principle Investigator shall monitor and audit any agreements (e.g. Cooperative Research and Development Agreement (CRADA)) as well as any licensing compliance issues to identify intellectual property assets that are produced, used, or purchased as a result of the project. This shall include a review of any licenses and agreements with third parties to ensure that sufficient agreements are in place.

The Research Center will set up an intellectual property catalog for all research reports. This will provide a comprehensive assessment of all forms of intellectual property owned by WYDOT. This should include expiration dates, maintenance fee due dates, license agreements with licensees, conception data, and potential intellectual property.

The Research Center will work with the Attorney General’s office to determine if the intellectual property needs to be registered. The Attorney General’s office will have the final say on registration of any intellectual property.

The Principle Investigator and the Research Center will make all determinations on whether the intellectual property should be excluded from use by others, if there is a need to control future use, licensing agreements, and how to keep the intellectual property in the public domain. As a note, if the report or data is release to the public domain without claiming intellectual property rights, the Principle Investigator and WYDOT could forfeit their rights.

Management decisions on any source codes will be made by the Wyoming ETS Department, the Principle Investigator, and the Research Center.

12.12 Disclaimer

When necessary, a disclaimer form will be provided to the RAC with the proposal and updated forms will be attached to the progress reports. Disclaimer forms will provide an overview and description of the invention, concept, work or idea proposed in the research project or set out in the research report with sufficient information to decide appropriately whether to pursue intellectual property protection. The disclaimer form shall include:

- A description of the technology, invention or creative work.
- What does it do. Key features and benefits.
- Are there any national security concerns.
- What is the funding source.
- The technology, invention or creative work to be protected by Intellectual Property.

Principle Investigator, Project Champion, the Research Manager will determine which data and/or datasets will require disclaimers. Disclaimers may be required for confidentiality issues, intellectual property rights, or other issues. Examples of Disclaimers can be obtained from the Research Center.

12.13 Selection and Retention Periods

Data from research projects can be large and cumbersome. The Research Center and Principle Investigator shall collaborate to determine what data should be retained and what data should be linked to the research report. All data collected during the research process shall be retained until the Principle Investigator, Project Champion and the Research Manager have an opportunity to discuss the data available and what will be needed for the final report. See Sections 12.2 and 12.3 for guidelines on what data should and should not be archived. Any data which is archived, should be machine-readable and in an open format. The Principle Investigator shall provide the Research Center with a data inventory, and if the data is not housed by the Research Center, the location and the custodian of the data should be provided to the Research Center. Further, the formatting of the data will be determined by the Principle Investigator, WYDOT, and/or the chosen data repository.

Retention periods for data shall be on a case-by-case basis. The minimal retention period is three years after conclusion of the award or three years after public release, whichever is later. Exceptions requiring longer retention periods may happen when data supports patents, when questions arise from inquiries or investigations with respect to research, or when a post-graduate student is involved. Research data that supports patents shall be retained for the entire term of the patent. The Federal Government, State of Wyoming, and ETS all require that data, records and other items be made available to the general public pursuant to *Section 508 of the Rehabilitation Act of 1973*.¹⁷

¹⁷ <https://www.fcc.gov/general/section-508-rehabilitation-act>

When determining data retention, data quality should be a major factor. The data should accurately represent the research project; have consistently defined data elements; be within acceptable ranges defined by the project; be provided to WYDOT in a timely manner; and be easily accessible, understandable and usable. The Principle Investigator will be required to provide clean data and data logs for any outliers.

Retention policies for all data, datasets, and other documents are approved by the WYDOT Records Department and the Attorney General's office, with recommendations from the Principle Investigator and the Research Center. At a minimum, all data, datasets, and documents generated from the project must be retained for three years. All other time limits shall be jointly determined by the Research Center and Principle Investigator.

If any data is deemed confidential or sensitive, it shall be retained in a secure storage center which adheres to the guidelines set out in HITECH and/or HIPAA, and shall only be retained as long as necessary. All sensitive and confidential data shall be disposed of in a manner that complies with either HIPAA or HITECH.

12.14 Embargo Periods

The Principle Investigator shall be allowed to maintain the data without releasing it to the public for a stated period not to exceed 1 year. This is known as an embargo period. On a case-by-case basis, the Research Center and Principle Investigator will entertain requests from outside sources for the release of data within the embargo period. Clear and convincing evidence of the need for the data must be presented to the Research Center in order to request copies of the data during the embargo period.

12.15 Ethics and Privacy

Principle Investigator, Project Champion and the Research Center must ensure ethics requirements, and data privacy rules and regulations are followed when working on research projects. All information gathering rules regarding data, datasets, software, etc must be followed. Though data gathered for these research projects should be transparent, this does not mean the rules for confidentiality and privacy should be ignored. If the project gathers data that is considered personal health information (PHI), personally identifiable information (PII), or sensitive material, precautions must be in place prior to beginning the project, and HIPAA and HITECH guidelines must be followed. Further, all legal permissions must be obtained and property storage must be acquired. Informed consent forms must be retained for the life of the data. Internal Review Boards may be necessary for any project where protected health information (PHI), personally identifiable information (PII) or sensitive materials are involved.

All individuals who will have access to any confidential and/or sensitive information must receive training on privacy, security, safeguards, encryption, breach notifications and mitigation procedures. Safeguards must be in place to ensure no unauthorized user has access to the data, and authorized users must use secure passwords to access confidential and sensitive data. The Principle Investigator must set out policies on how to identify the identity of all users.

12.16 Budget

Data storage shall be determined on a case-by-case basis. Data may be stored by the Principle Investigator, Project Champion, the Research Center, or an approved outside repository. The length of time for retention, the need for the data on a daily basis, and the type of data will dictate where the data is retained, and as such, the cost of storing the data. The entity which stores the data shall be responsible to keep the data operational and available, and for the cost of storage.

12.17 Digital Object Identifier (DOI), Open Research, and Contributor ID (ORCID)

Prior to publishing any paper or report using data or conclusions drawn from a research project, the Principle Investigator(s) must work with the Research Center to ensure that there is no duplication or conflicts with digital object identifiers. Derivative copies of reports from funded research projects must be provided to the Research Center with the reports DOI.

Principle Investigator and authors named on any report or paper must register for an ORCID number at <http://orcid.org/> prior to publishing any work that is derived from a research project. This number helps distinguish researchers who have similar names and allows Principle Investigator to attach their work to other works. The ORCID registry is available free.

12.18 Data Citation Index

Principle Investigators are required to provide the Research Center with a data citation index (Index). The Index should include all data gathered and datasets built during the life of the project. This will enable the Principle Investigator, Program Champion(s) and the Research Manager to easily access the data and/or datasets. It will also enable the Research Center to better archive the data. The Index should include an attribution for the data object to the persons and/or institutes creating the data, and a standard form of citation for each piece of data.

12.19 Annual review of data management plan template

During the October RAC meeting, the Research Center will provide the RAC with an updated data management plan to be reviewed. Changes and comments to the plan will be discussed during the January RAC meeting each year.

12.20 Annual Review of Project Specific Data Management Plans, Data Clean Up, Scrubbing

The Principle Investigator and Project Champion shall meet yearly to discuss changes in the original and any revised DMPs. If there are any changes to the DMP, the Principle Investigator

shall provide the Research Center with a copy of the revisions. The Principle Investigator shall be responsible for detecting, removing, and/or correcting dirty data (i.e. data that is incorrect, out of data, redundant, incomplete, or formatted incorrectly). Data cleansing allows for consistency in the data and can include data auditing, workflow specification, workflow execution, and post-processing and control.

12.21 Licensing

Terms and conditions for software, codes, and data are essential to ensure ownership and rights. All research projects, which are funded specifically for building or the maintenance of software, codes and some other types of database, must have a license attached to the report before the software is made publically available. All licenses must:

- Have full attribution for the Principle Investigator, WYDOT and FHWA.
- Must be open source.
- Must be considered copyleft.
- Must be for non-commercial purposes.
- Must be approved by WYDOT prior to finalizing.

The best form of license to use is a General Public License (GNU).¹⁸ When attaching the license, a copy must be attached to the research report as an appendix. For all electronic copies of the data and/or software, there must be a rights statement that appears prior to the stakeholder obtaining the data and/or software. Rights statements are very short and usually state that the database and/or software is made available under the license number “X”, which text can be found at (specific location).

¹⁸ <http://www.gnu.org/licenses/gpl-3.0.en.html>