

Center for State, Tribal, Local, and Territorial Support (CSTLTS)

Data Management Plan (DMP)

Extramural Guidance



This Extramural Guidance is maintained by the CSTLTS Science Unit. If you have any questions or concerns regarding the information included in this document, please contact your CDC/CSTLTS technical monitor (TM) or project officer (PO).

Science Unit, Center for State, Tribal, Local, and Territorial Support
August 2019



Table of Contents

| | |
|--|-----------|
| Introduction..... | 3 |
| Purpose and Scope..... | 3 |
| Determining If a DMP is Necessary | 4 |
| Defining Public Health Data | 4 |
| Conducting Evaluations and Assessments | 5 |
| Expanding Existing Public Health Datasets..... | 6 |
| Secondary Data Analysis..... | 7 |
| Drafting a DMP | 7 |
| Submitting a DMP | 8 |
| Appendix A: DMP Examples | 9 |
| Appendix B: Frequently Asked Questions (FAQs) | 13 |

Introduction

A data management plan (DMP) is a written description of the processes involved for the collection, protection, sharing, and long-term preservation of public health data. It is a blueprint that will assist in planning for data management and sharing in advance of the actual data generation and collection. The DMP is a *living document*, meaning it must be updated and revised as the project evolves and throughout the lifecycle of the data collected.

What is public health data?

Public health data is defined as “digitally recorded factual material commonly accepted in the scientific community as a basis for public health findings, conclusions, and implementation.”

All CDC contracts, awards, and continuations from FY2017 forward should include a DMP or a statement that a DMP is not needed because the project will not collect or generate public health data. Recipients should follow the Additional Requirement 25 (AR-25) included in the Notice of Funding Opportunity (NOFO). While there are several publicly available online tools for generating a DMP that many may find useful, an example is provided and may be used by recipients and/or contractors in [Appendix A](#) (DMP Examples) of this document. Other links to examples of DMP tools are provided in [Appendix B](#) (Frequently Asked Questions).

Purpose and Scope

In general, all CDC-funded public health datasets are expected to be made freely available to the public in a timely manner. DMPs ensure that CDC (including all CDC-funded data collection projects) follows all applicable laws, regulations, directives, and guidelines while managing public health data and providing appropriate access to the data for public use. In a DMP, recipients are expected to describe how they intend to manage, preserve, and make accessible data generated or collected with CDC funds. The DMP should be developed during the project planning phase prior to initiating data generation or collection activities.

Determining if a DMP is Necessary

Data that do necessitate a DMP in accordance with [CDC's policy](#) and this guidance include those that **contain public health data and meet at least one of the following criteria:**

- Collected or generated by CDC;
- Collected or generated by other agencies or organizations funded or co-funded by CDC (e.g., through grants, cooperative agreements, contracts, similar mechanisms); and/or
- Reported to CDC by another entity (e.g., by state health departments) that become a part of a CDC data collection system (e.g., CDC surveillance systems such as the Cancer Registry).

Data that do not necessitate a DMP in accordance with CDC's policy or this guidance include those:

- Collected and generated by other organizations but that are shared for informational use with CDC (i.e., data not funded or co-funded by CDC);
- Provided to CDC through an agreement that contains restrictions on data usage and sharing (e.g., a data-transfer agreement, data-use agreement, memorandum of understanding); and/or
- Provided to CDC by another federal agency (e.g., the Centers for Medicare & Medicaid Services) under restricted terms of usage and sharing.

Defining Public Health Data

Under CDC's [Policy on Public Health Research and Nonresearch Data Management and Access \(https://www.cdc.gov/maso/policy/policy385.pdf\)](https://www.cdc.gov/maso/policy/policy385.pdf), all public health datasets collected or generated using federal funds must have a data management plan (DMP). *Public health data* is defined as—

Digitally recorded factual material commonly accepted in the scientific community as a basis for public health findings, conclusions, and implementation. Public health data includes those from research and nonresearch activities.

Public health data do not include preliminary analyses, drafts of scientific papers, plans for future research, reports, recipient progress reports, communications with colleagues, financial/ administrative data, laboratory quality assurance data, laboratory training, laboratory emergency response exercises, and physical objects (e.g., laboratory notebooks, laboratory specimens.)

Conducting Evaluations and Assessments

When determining whether a non-research project involves the collection or generation of public health data, the purpose(s) of the data collection activity and the intended use of the data collected must also be considered. This is especially important when determining whether the policy applies to assessments and evaluations.

If the purpose of an assessment or evaluation is to understand the public health impact of a program, policy, or intervention, and/or if the results from the data collection will be used to make policy-related decisions, then the data are considered to be public health data, and a DMP is required.

Example assessment activity that requires a DMP:

In a project funded through one of CSTLTS's cooperative agreements or grants, the following activity is described: "Department of Health (DOH) staff will update data infrastructure (to include at a minimum, measurement for foreign-born and LGBT populations) and data related to the HP2020 objectives. DOH will prepare, receive, and analyze data for inclusion in meetings for policy decision making and publications. DOH will interactively visualize data and other information included in [state] HP2020 and post online for public use."

In nearly all circumstances, data collection activities included in quality improvement projects; organizational performance measurement and management projects; formative or process evaluations; customer satisfaction surveys; and needs assessments will not require a DMP.

If the specific purpose of an assessment or evaluation is to improve the design or operations of a program, process, system, or service delivery mechanism, and the data are not used to determine public health impact, the data are not considered public health data and a DMP is not required.

Example evaluation activity that does not require a DMP:

In a project funded through one of CSTLTS's cooperative agreements or grants, the following activity is described: "The recipient [Partner A] is working with other agencies to implement the Massachusetts Approach to Partnership in Parenting (MAPP) training. The recipient plans to implement a pre- and post-assessment to all participants of the training to evaluate the training's effectiveness at reaching intended learning objectives, and plan program improvements for future training events."

Expanding Existing Public Health Datasets

In some circumstances, CDC funds are used by cooperative agreement and grant recipients to expand an existing data collection activity that was not previously subjected to this policy. An existing data collection is defined as an activity that has been conducted previously and is typically repeated on a regular basis (e.g., an annual survey). Expansion of an existing data collection activity may include a) adding new variables or data elements, and/or b) including new subjects, cases, populations, or sites to the existing dataset.

If CDC funds are used to add new variables or data elements to an existing data collection, and the expanded dataset is consistent with the definition of public health data, a DMP focused on the new or additional variables is required.

Example activity that requires a DMP for new data variables/elements:

In a project funded through one of CSTLTS's cooperative agreements or grants, the following data collection activity is described: "[State] will expand current Behavioral Risk Factor Surveillance System (BRFSS) data collection activities by constructing a state-specific questionnaire. These additional questions are designed to meet the state's specific data needs. The BRFSS coordinator will closely monitor the contractor to ensure adherence to all BRFSS surveillance methodologies and protocols."

In some cases, CDC funds are used to expand an existing data collection activity by adding new subjects, cases, populations, and/or sites.

If CDC funds are used to add new subjects, populations, or sites to an existing data collection and the expanded dataset is consistent with the definition of public health data, then a DMP covering the entire dataset is required.

Example activity that requires a DMP for the entire dataset:

In a project funded through one of CSTLTS's cooperative agreements or grants, the following data collection activity is described: "[State] will expand a current data collection project to include additional cases to a dataset that is currently collected and maintained through private funding. This data collection project aims to develop a typology of American health values across the state, which will inform specific marketing, policy, and program decisions in the future. The implementation of the data collection activity will be expanded to include a larger random sample of the population across the state, increasing the sample size from 250 participants to 300 participants."

Secondary Data Analysis

Some projects may require the acquisition and use of secondary data. In most cases, these data collection activities will not require a DMP.

If secondary data are coded and/or combined with another dataset (primary or secondary data collections) for the purposes of creating a new public health dataset, a DMP for the new dataset is required.

Example secondary data analysis activity that requires a DMP:

In a project funded through one of CSTLTS's cooperative agreements or grants, the following data collection activity is described: "The recipient has collected evaluation information which includes accreditation status. The original use of this data collection activity was for program and process improvements and not considered *public health data*. This dataset will be combined with publicly available community indicators to create a new dataset. Analysis on this new dataset aims to evaluate the impact of health department accreditation on select community indicators (e.g., employment rates, violent crime, homelessness)."

Drafting a DMP

Recipients of CDC funding are responsible for creating a DMP when a project is initiated and updating it as appropriate throughout the life cycle of the data. CDC currently does not have a standard form to use when creating a DMP. However, DMPs should include all of the following information:

- A. Description of the data to be collected or generated in the proposed project.
- B. Mechanisms for, or limitations to, providing access to the data, including a description of provisions for the protection of privacy, confidentiality, security, intellectual property, and other rights.
- C. Statement of the use of data standards that ensure all released data have appropriate documentation that describes the method of collection, what the data represent, and potential limitations for use.
- D. Plans for archiving and long-term preservation of the data, or an explanation why long-term preservation and access are not justified.

Determining Data Access Level

Unless there is a strong reason not to do so, datasets are expected to be made publicly available in a timely manner in a nonproprietary format. However, CDC and recipients have a responsibility to protect individuals' privacy and ensure confidentiality of the data collected. Therefore, some datasets cannot be made public but should be available

on a restricted basis, or if necessary, not available. Final, adjudicated data files with no personally identifiable information (PII) or information in identifiable form (IIF) are eligible for public data release.

There are 3 access options for data:

1. **Freely accessible to the public** (e.g., can be downloaded from an internet site)—
These datasets are freely available to the public and do not require a data-sharing agreement. Public-use data files can be categorized into the following types:
 - Full dataset—access to the full scope of the data collected (excluding PII or IIF) is provided
 - Aggregate dataset—summary data are publicly available without restriction
 - Ad-hoc requests – the full dataset or an aggregate dataset is publicly available upon request
2. **Restricted access** (e.g., can be accessed with a data-sharing or use agreement)—
These datasets are available for use under certain provisions, restrictions, and/or agreements.
3. **No access**—Datasets identified as *no access* are not available for public use.

If at least one of the considerations below do apply to the dataset, the data access level should be identified as either *restricted* or *nonpublic* (e.g., no access):

- Country/jurisdiction owns the data with protections under its laws and regulations
- Not shareable for protection of intellectual property or trade secrets (e.g., proprietary rights); precluded by licensing or other agreement
- Removal of identifiers renders the remaining data of no value; data cannot be shared without compromising subjects' privacy
- Cost of sharing the dataset outweighs the expected benefit
- Data quality is poor/inadequate

For additional information and guidance on determining appropriate public access level, please contact your TM or PO.

Submitting a DMP

A DMP is a *living document* that must be updated, submitted, and approved for accuracy as plans solidify or change during the project's period of performance. Submit the DMP to your CSTLTS TM/PO via email. Your TM/PO will review the submitted DMP and follow up with you as necessary.

Appendix A: DMP Examples

EXTRAMURAL RECIPIENT DATA MANAGEMENT PLAN: CSTLTS (With Example Data)

This example is designed to help extramural recipients develop a data management plan for any type of CDC-funded public health data collection activity, including nonresearch (public health practice) data.

Section 1: Project Information

Agency/Organization Name: [Organization A](#)

Funding Mechanism Type:

- Cooperative Agreement, please specify announcement name and #: [Click or tap here to enter text.](#)
- Grant, please specify announcement name and #: [Grant Name/Title and number 123456789](#)
- Contract, please specify name and #: [Click or tap here to enter text.](#)
- Other, please specify: [Click or tap here to enter text.](#)

Project Contact (POC)

POC Name: [Jane Doe](#)

POC Organization: [Organization A](#)

POC Email Address: jane_doe@organization_a

POC Phone Number: 555-555-5555

Section 2: Dataset Information

Date of DMP creation: [10/16/2019](#)

Dataset Title – *In plain English with sufficient detail to facilitate search and discovery of the dataset, name of project data/data collection activity. This should remain the same throughout the life of the project.* [Dataset Title A](#)

Dataset Description – *Briefly describe the data captured for this project and its purpose. This should have sufficient detail that enables readers to quickly understand whether the project or dataset is of interest to them. An interactive dataset that includes measures of the extent to which evidence-based policies and practices are in place for US states to address 34 public health concerns.*

Tags/Keywords – *Terms to help users discover the project and its datasets; please include terms that would be used by technical and nontechnical users. For tags, refer to Mesh Terms, which are loaded annually from the [US National Library of Medicine](#).*

[Public health, epidemiology, preventive medicine](#)

Data Publisher/Owner

Discuss which entities own the data, who the publishing entity is, and which entities have custodial responsibility and/or give permission to share the information collected or obtained from this project.

Owner of the Data: [Organization A will collect the raw data and provide the final, cleaned dataset excluding PII and IIF, or identifying information to CDC's TM assigned to this project.](#)

Publishing Entity: CDC; all publications resulting from this data collection will be in collaboration with and as directed by the CDC TM assigned to this project.

Custodial Responsibility: CDC; all copies of data collected and custodial responsibility of Organization A will be relinquished upon CDC's receipt of the final dataset.

Section 3: Description of the Data and Data Quality

Data Description

Data Collection Start Date – Date the staff will begin collection of data. [2/6/2020](#)

Data Collection End Date – Date the staff will complete the collection of data. [6/30/2020](#)

Methods – Briefly describe the study/project design and methods, including the approach and plan to meet the objectives, such as interventions, procedures, target population or respondent recruitment, screening, and enrollment. [Click or tap here to enter text.](#)

Data Collection Instrument(s) and Protocol – Provide a brief description with reference to a document or website that provides detailed information. Include information such as how often the data will be collected. [Click or tap here to enter text..](#)

Data Management and Quality

Data Management Protocol – Briefly describe, with reference to physical location(s) or system(s), where data will be housed (e.g., CDC shared network drive, data host system name, SQL database, etc.) [Click or tap here to enter text..](#)

Process for Omitting Identifying Information – Describe what identifiers are in the database, how they will be removed, and by whom. [Click or tap here to enter text..](#)

Data Quality Protocol – Describe methods for data validation and error resolution; removal or shielding of any proprietary information; removal or shielding of sensitive information; removal or shielding of any individually identifying information including indirect identification. [Click or tap here to enter text..](#)

Section 4: Data Access and Protection

Proposed Public Access Level – Briefly describe who will have access to stored information. [Entire dataset in original form will be released for general public use without restriction](#)

Select an access level below. Select Public if the data collected/generated in this project will be released to the public in either microdata or aggregated format; select Restricted if the project data will be shared with restrictions or via CDC Research Data Center (RDC); select Nonpublic if the project data will not be released to or shared with the public

PUBLIC Release

- Public release – Full dataset
(Dataset can be made available without restrictions; data steward no longer controls data. **This should be the default selection for all datasets unless justified otherwise.**)
- Public release – Aggregate data
(Underlying dataset cannot be released or shared, but aggregate/summary data can be made available to public access without restriction.)
- Public release – Release by ad-hoc request
(Metadata will be released and the dataset is available by ad-hoc request; data requests CANNOT be denied; no data use agreement or restrictions; data steward no longer controls data.)

RESTRICTED Release

- Restricted use data sharing
(Dataset is available to particular parties under certain use restrictions or use agreement; data not always under CDC custody.)
- Restricted access data sharing
(Dataset is available only in an RDC; data need to remain under CDC custody.)

NONPUBLIC

- No release or data sharing

Access Rights/Restrictions

Public Access Justification – For a Restricted Release or Nonpublic dataset, provide an appropriate justification for why the data collected/generated in this project cannot be released to/shared with the public.

- Country/Jurisdiction owns the data with protections under its laws and regulations
- Not sharable for protection of intellectual property or trade secrets
- Removal of identifiers renders the remaining data of no value
- Other, please specify: [Click or tap here to enter text.](#)

Data Use Type – For a Restricted Release dataset, select the type of data use agreement that must be in place in order to release this dataset.

- Data-Sharing Agreement
- Data Transfer Agreement
- Joint Statement of Understanding
- Memorandum of Understanding
- Other, please specify: [Click or tap here to enter text.](#)

Data Use Type URL – For a Restricted Release dataset, this is the website where the process for requesting access to the dataset can be found. [Click or tap here to enter text.](#)

Data Use Contact – For a Restricted Release dataset, this is a name of a data steward/contact from whom external investigators can request access to the dataset. [Click or tap here to enter text.](#)

Section 5: Data Release and Documentation

Estimated Date(s) of Data Release: 2021

Data Release Format – Recommend to use nonproprietary format when possible, such as CSV, XML, JSON, etc. Also specify data dictionary file format. CSV

External Access URL, If Known

Please provide the URL for external access to the documentation associated with this project. The information is to include protocol, data dictionary (e.g., variable names, definitions), data collection instrument, and other relevant information.

www.cdc.gov/organization_a

Download URL(s), If Known

Provide the URL to a downloadable file of the dataset(s). www.cdc.gov/organization_a

Type of Data Released – Will the released data be the data tables associated with the publication, or will it be the raw data in either microdata or aggregated format?

- Aggregated data (e.g., summary data or statistics)
- Microdata (e.g., information at the level of individual respondents)
- Tables

Data Release Documentation – Identify the documents that will be provided to users (e.g., variable definitions, codebook, metadata file, guidance on data use). Check all that apply.

What documents will be provided/available with the datasets?

- Variable definitions
- Codebook
- Data collection instrument
- Metadata file
- Guidance on data use
- Description of the population studied
- Methodology
- Description of dataset, such as response rates and limitations/caveats
- Other, describe here: [Click or tap here to enter text.](#)

Describe Long-Term Preservation Plan: Describe the plan for archiving and long-term preservation of the data, or explain why this is not justified. The final dataset (no PII or IIF) will be shared with the CDC TM for backup / preservation purposes and use. The raw and final dataset will be stored on Organization A's shared drive network while a backup copy will be stored on an onsite Network Attached Storage (NAS) supported by Organization A. The backup will be synced on routine schedule outlined in Organization A's SOP for backup procedures. A review of the dataset to ensure availability and usability will be conducted by Organization A on an annual basis. Both the network dataset and the backup dataset will be password protected and accessible only to the approved program staff listed in the workplan. The dataset and backup dataset will be maintained by Organization A for a minimum of 3 years following the completion of this project (estimated date is June 2027). Following this date, the dataset will be archived indefinitely.

Appendix B: Frequently Asked Questions (FAQs)

1. What does a DMP look like?

CDC does not currently have a standard form to use when creating a DMP. The DMP may be created in a checklist, paragraph, or other format. However, CSTLTS has developed a template that shows suggested fields and field descriptions as a resource for recipients. In addition, extramural recipients are referred to the following websites for examples of how to draft a DMP:

- [University of California](#)
- [United States Geological Survey \(USGS\)](#) (scroll to the “Templates and Examples” section)
- [Inter-university Consortium for Political and Social Research \(ICPSR\)](#)

2. What topics does the DMP need to cover?

The DMP should cover, at a minimum, the following topic areas:

- Description of data to be generated or collected
- Standards to be used in generating/collecting data
- Mechanisms and limitations for access to the data
- Standards for data release
- Plan for archiving and long-term preservation

3. What if project data cannot be made available due to security, confidentiality, or privacy concerns?

It is understood that not all data can be made publicly available due to security, confidentiality, or privacy concerns. Regardless of whether the data are made available as restricted or as a public-use dataset, recipients are responsible for planning and implementing timely access to the data as described in the DMP. Should the recipient determine that the data cannot be made available for public use, a written justification is required.

4. What is considered public health data?

Public health data means digitally recorded factual material commonly accepted in the scientific community as a basis for public health findings, conclusions, and implementation.

Public health data does not include grantee progress reports, process monitoring data, administrative data, preliminary analyses, drafts of scientific papers, plans for future research, reports, communications with colleagues, or physical objects, such as laboratory notebooks or laboratory specimens.

5. My project is new. What if all details of the project design are not yet certain?

All descriptions can be general if project design is not yet certain. DMPs should indicate that the extramural awardee understands the criteria and the DMP’s purpose. By the end of the project, the final DMP should contain all details and be precise.