Centers for Disease Control and Prevention (CDC)

Notice of Funding Opportunity (NOFO) Data Management Plan (DMP) Template for Extramural Research

[Version: 2026-01-13]

Template Use and Policy Alignment Statement: This DMP template was created to capture information consistent with CDC Grants Notice of Funding Opportunity (NOFO) **Additional Requirement 25: Data Management and Access** [updated October 1, 2024 at https://www.cdc.gov/grants/additional-requirements/ar-25.html]. The completed DMP will be submitted as a part of the complete NOFO submission package.

Researchers responding to a NOFO should fill out this DMP as completely as possible, including using phrases such as "To be Determined", etc. as needed, to inform CDC DMP evaluators of intentions. Leave as few prompts empty as possible.

Frequent or Important Acronyms used in this document:

CDC: Centers for Disease Control and Prevention
CIO: A CDC Center, Institute, or Office
DMP: Data Management Plan
NOFO: Notice of Funding Opportunity
ORCID: Open Research and Contributor Identifier
RDC: CDC Research Data Center
ROR: Research Organization Registry
SAM: System Award Management

Introduction and Requirements Review

Adapted from CDC Grants **Additional Requirement 25: Data Management and Access:** https://www.cdc.gov/grants/additional-requirements/ar-25.html [as viewed on 2025-06-10] [with clarifications added as need 2026-01-12].

Overview

CDC requires recipients for projects that involve the collection or generation of data with federal funds to develop, submit and comply with a Data Management Plan (DMP) for each collection or generation of public health data undertaken as part of the award and, to the extent appropriate, provide access to, and archiving/long-term preservation of, collected or generated 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 includes those from research and non-research activities. Public health data may include the study of human or non-human subjects. Public health data may include research on organisms, or organizations and entities. Public health data could be quantitative, qualitative, imaging, or genomics output (for example, genome sequencing, arrays, gene expression, etc.). This description cannot list all possible types of Public health data. Instead, researchers receiving research funding from CDC should assume that their data are public health data, unless specifically directed otherwise.

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

Data Management Plan

Consistent with the terms of and activities expected under the notice of funding opportunity (NOFO), recipients must develop and submit a DMP generally during the **project planning phase**, but in any event, prior to the initiation of generation or collection of public health data. Accordingly, the DMP may be evaluated during the application, study proposal, or project review process or during other times in the period of performance (typically by the funding office). For NOFOs that involve already defined projects, which include data collection or generation at the time of application, applications submitted without the required DMP may be deemed non-responsive for award. For NOFOs where CDC specifies that submission of the DMP is deferred to a later period, funding restrictions may be imposed pending submission and evaluation of the DMP. For awards where data collection or generation activities may become necessary during the period of performance, DMPs will be required to be submitted and evaluated (typically by the funding office) during the period of performance of the award. These DMPs also will be required to comply with this Additional Requirement. In all instances described above, the reviewing officials must approve an acceptable DMP.

Costs associated with developing and implementing a DMP, including costs of sharing, archiving and long-term preservation, may be included in the budget submissions for grants and cooperative agreements.

A DMP for each collection and/or generation of public health data funded by this award should include **all** the following information:

- **A Description of the Data to be collected or generated in the proposed project. [Section 1 below]**
- **Standards** to be used for the collected or generated data. **[Section 2 below]**
- **Mechanisms for or Limitations to Providing Access to and Sharing of the Data** (include a description of provisions for the protection of privacy, confidentiality, security, intellectual property, or other rights). This section should address access to identifiable and de-identified data or justification for not making the data accessible (see below for additional information about access). **[Section 3 below]**
- 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. **[Section 4 below]**
- **Plans for Archiving and Long-term Preservation** of the data or explaining why long-term preservation and access are not justified. This section should address archiving and preservation of identifiable and de-identified data. [Long-term preservation, in this instance, means maintaining the dataset for the period that it remains relevant to, and of interest to, the scientific community. As a rule of thumb, at minimum 10 years or longer, possibly in perpetuity.] **[Section 5 below]**

Access to and archiving of the data

Recipients whose terms of award do not include submitting data to CDC are expected to plan and prepare for access to, and archiving/long-term preservation of, collected and/or generated data within the funding period, as set forth below. The final version of a collected and/or generated dataset intended for release or sharing should be made available **within thirty (30) months after the end of the data collection** or generation, **except surveillance data that should be made accessible within a year of the end of a collection cycle**. In addition, recipients should ensure the quality of data they make accessible and seek to provide the data in a **nonproprietary format**. If data cannot be made accessible, a justification for not doing so should be provided in the final DMP. Recipients who fail to release public health data in a timely fashion may be subject to procedures normally used to address lack of compliance consistent with applicable authorities, regulations, policies or terms of their award.

For public use de-identified (removal of sensitive identifiable or potentially identifiable information) datasets, an accompanying data dictionary, codes, and other documentation relevant to use of the dataset should be deposited in a sustainable repository to provide access to the data. Data that cannot be de-identified can be provided on request under a data use agreement.

Recipients will be required to inform the appropriate CDC point-of-contact identified in the award via an update to their DMP of the location of the deposited data. The DMP is a living document that should be updated throughout the lifecycle of data [for example: at each substantive change in or completion of project phases; changes in personnel; changes in scope or population; etc.].

For data underlying scientific publication [any type, any venue], recipients should make the data available [to the CDC and/or the public, as appropriate] coincident with publication of the paper, unless the dataset is already available via a release or sharing mechanism. At a minimum, release of the dataset should consist of a machine-readable version of the data tables shown in the paper.

Note: CDC Centers, Institutes, or Offices (CIOs) may have expectations for the prompts below, or may include DMP sections beyond those required in the text above. Please check the NOFO listing or with the posting CDC CIO for DMP creation guidance specific to their CIO.

Answering DMP Prompt Guidance

Answer each prompt below as completely as possible at this stage in the proposal:

- Not all answers may be known. Please indicate with "To be determined" if needed.
- Empty or unanswered prompts may make it more difficult to evaluate your DMP. Again, answer "To be determined" if needed.
- Answer to the best of your ability.

If a prompt offers multiple possible responses, but none fit your situation, choose "Other" and provide details to help evaluators understand your choice.

Required DMP Sections

1. **Description of the Data**
2. **Standards**
3. **Mechanisms for or Limitations to Providing Access to and Sharing of the Data**
4. **Appropriate Documentation**
5. **Plans for Data Archiving and Long-Term Preservation**

Additional Sections

Appendix A: Additional Section 1 Dataset Description for Projects with more than one Dataset

Appendix B: Links to Standards and Metadata Schema listed in Section 2 dropdown menus

NOFO Submission Information

S.01 NOFO Funding Opportunity Number from grants.gov:

S.02 NOFO URL from grants.gov:

S.03 NOFO Name:

S.04 CDC Center, Institute, or Office sponsoring NOFO:

S.05 Applicant Name [name of person completing this form]:

S.06 Applicant ORCID [persistent identifier for the person completing this form]:

S.07 Applicant Organization [name of organization applying for funding or where the Applicant is based]:

S.08 Applicant Organization Institutional Organization ID (System Award Management (SAM) or Research Organization Registry (ROR)):

S.09 Applicant Email:

S.10 Applicant Phone:

S.11 Applicant Role:

S.12 Will the organization provide a copy of the data to the CDC:

◯ Yes ◯ No ◯ Other

S.12.a If you chose the “No” or "Other" response above, please select the reason you will not be sharing data with the CDC:

S.12.b If you selected "Other", please specify:

S.13 Proposed timing of data availability to CDC, if known:

S.13.a If you selected "Other", please specify:

S.14 Will a Data Use Agreement (DUA) be required to share this data with the CDC?:

◯ Yes ◯ No

S.14.a DUA Identifier (if a DUA is already in place. If not, go to the next question):

S.15 Is a Data Management Plan (DMP) required [see "Overview" section on page 2]:

◯ Yes ◯ No, there will be no DMP for this project ◯ Other

S.15.a Justification for lack of DMP: If you chose the “No” or "Other" response please choose the reason you believe that a data management plan is not required for this project. There will be no DMP for this project because:

S.15.b If you selected "Other", please specify:

Project Information

P.01 Project Title:

P.02 Project Purpose or Objective:

Proceed to next page.

P.03 Contributing Team members, including PI, other primary researchers, and any Team contributing to data collection and preservation. Use as many or as few entries as necessary. Name, ORCID, Email, Organization, and Role are all required for each team member.

Team Member 1 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**
Team Member 2 Full Name [Given or First Name Middle Name(s) Family or Last Name]		**Team Member ORCID**
Email	**Organization**	**Primary Project Role**
Team Member 3 Full Name [Given or First Name Middle Name(s) Family or Last Name]		**Team Member ORCID**
Email	**Organization**	**Primary Project Role**
Team Member 4 Full Name [Given or First Name Middle Name(s) Family or Last Name]		**Team Member ORCID**
Email	**Organization**	**Primary Project Role**
Team Member 5 Full Name [Given or First Name Middle Name(s) Family or Last Name]		**Team Member ORCID**
Email	**Organization**	**Primary Project Role**
Team Member 6 Full Name [Given or First Name Middle Name(s) Family or Last Name]		**Team Member ORCID**
Email	**Organization**	**Primary Project Role**
Team Member 7 Full Name [Given or First Name Middle Name(s) Family or Last Name]		**Team Member ORCID**
Email	**Organization**	**Primary Project Role**
Team Member 8 Full Name [Given or First Name Middle Name(s) Family or Last Name]		**Team Member ORCID**
Email	**Organization**	**Primary Project Role**

Team Member 9 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

Team Member 10 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

Team Member 11 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

Team Member 12 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

Team Member 13 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

Team Member 14 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

Team Member 15 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

Team Member 16 Full Name [Given or First Name Middle Name(s) Family or Last Name]		Team Member ORCID
Email	**Organization**	**Primary Project Role**

1. Description of the Data

In this section, provide a description of the data to be collected, generated, obtained, or transfered in the proposed project. You may not have answers to some or many of the prompts in this section at proposal time. Use the "To be determined" response as needed in this section. Some projects will collect, generate, and/or reuse more than one (1) dataset. Use the spaces on this page to describe the primary or most important dataset. Then you should describe each additional dataset by filling out another version of Section 1. Description of Data in the space provided in Appendix A.

Dataset

1.01 Working Dataset Title, if known:

1.02 Please provide an initial Dataset Description:

1.03 Proposed Collection Start Date (YYYY-MM-DD):

◯ ◯ To be determined

1.04 Proposed Collection End Date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned / Continual without end

1.05 List expected Language(s) of the data (English, Spanish, Others):

1.06 List expected or proposed Tags or Keywords. Use controlled terms such as MeSH or your own:

Release

1.07 Expected initial public data release date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned

1.08 Expected type(s) of data released (ctrl+click for multiple selections):

Complete Raw Dataset
Research Replication Dataset
Aggregated Data
Microdata
Public-use Dataset
To be determined
Other

1.08.a If you selected "Other", please specify:

Frequency

1.09 Expected data update frequency, if known (ctrl+click for multiple selections):

One-time collection, no updated expected
Baseline and follow-up(s)
Ongoing Surveillance
Continually Updated
Updated Monthly
Updated Annually
Updated Irregularly
To be determined
Other

1.09.a If you selected "Other", please specify:

1.10 Proposed timing of data availability to the Public, if known:

1.10.a If you selected "Other", please specify:

Related Software, and Code

1.11 Expected software(s) to be used in the creation of the data (ctrl+click for multiple selections):

Electronic Health Records (EHR) interface
Fyre
Microsoft Excel
Microsoft Word
PDF
Plain Text Editor (.txt)
Qualtrics
REDCap (Research Electronic Data Capture)
RStudio
SAS
SPSS Statistics
To be determined
Other

1.11.a If you selected "Other", please specify:

1.12 Site name or URL to the external site where the script or code will be located or shared (if known):

1.13 Expected scripting language(s) to be used (ctrl+click for multiple selections):

Python
R
SAS
SPSS
To be determined
Other

1.13.a If you selected "Other", please specify:

Proceed to next page.

Spatiality

Include the geographic area that is the focus of the research. Be as specific as possible, using as few fields as needed.

1.14 Spatiality Details:

Country/Tribal Nation	State/Province	County/Region	City/Town	Specific Place Name / Other	Geographic Coordinates

2. Standards

In this section, describe the data and data documentation file format(s) and standard(s) to be used for the collected or generated data. You may not have answers to some or many of the prompts in this section at proposal time, and the standards used may become clear as the project moves forward. Use the "To be determined" response as needed in this section.

2.01 Please list the file format(s) and standard(s) expected to be used or generated in data collection (ctrl+click for multiple selections):

.CSV
.TSV
.TXT
Excel (.xlsx) (data tables)
.SAS (statistics)
.SPSS (statistics)
Analyze (medical imaging format)
Dicom (medical imaging format)
Minc (medical imaging format)
Nifti (medical imaging format)
.MP4 (moving images)
.JPG (lossy still images)
.PNG (lossless still images)
.TIFF (lossless still images)
HRA (high resolution audio)
WAVE (broadcast audio)
To be determined
Other

2.01.a If you selected "Other", please specify:

2.02 Will any of the data format(s) and standard(s) be proprietary (as opposed to open format such as .csv, etc.):

◯ Yes ◯ No ◯ To be determined ◯ Other

2.02.a If you selected the "Yes" response above, please state your rationale for using proprietary format(s) and standard(s) (as opposed to open format(s) and standard(s)). Otherwise, if you answered "To Be Determined" or "Other", please state the reason for your answer below:

2.03 Please name any expected metadata schema(s) to be used to describe the data,
if known (examples: NFDI4Health; Health-RI; HealthDCAT-AP; FGDC for geospatial metadata; others)
[Links to the Standards listed in the drop-down menus of 2.04 through 2.08 can be found in Appendix B.]
(ctrl+click for multiple selections):

Biotechnology - Requirements for data formatting and description in the life sciences (ISO 20691:20
Brain Imaging Data Structure (BIDS)
Core Information for metabolomics Reporting (CIMR)
Core Scientific Metadata Model (CSMD)
DCAT-AP-NL
Fast Healthcare Interoperability Resources (FHIR)
Genome Metadata
Genomics Informatics — Data elements and their metadata for describing structured clinical genom
Health-RI
Health Data CATalogue – Application Profile (HealthDCAT-AP)
Health informatics – Clinical knowledge resources (ISO 13119:2022)
Minimum Information about Any Sequence (MIxS)rces (ISO 13119:2022)
Minimum Information for Biological and Biomedical Investigations (MIBBI)
National Research Data Infrastructure for Personal Health Data (NFDI4Health)
NIH Common Metadata Elements for Cataloging Biomedical Datasets
NIH Protocol Data Element Definition for ClinicalTrials.gov
Open Microscopy Environment TIFF (OME-TIFF)
Open Microscopy Environment XML (OME-XML)
Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Fram
Recommended Metadata for Biological Images (REMBI)
Geographic information — Metadata (ISO 19115)
Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata
To be determined
Other

2.03.a If you selected "Other", please specify:

2.04 If the metadata schema is not standard for your field, discuss your rationale for using that schema:

2.05 Please name any standards used for persistent identifiers (PIDs) (examples: digital object identifiers (DOIs; handles, ARKs, others) (ctrl+click for multiple selections):

Digital Object Identifiers (DOI)
Archival Resource Key (ARK)
Handle System (handles)
ORCID (Open Researcher and Contributor Identifier)
Research Organization Registry (ROR
System Award Management (SAM) identifier
To be determined
Other

2.05.a If you selected "Other", please specify:

2.06 Select any standards used for creating associated data documentation such as: Data Dictionaries; README files; Codebooks. If you select "**Other**" as a response in **2.06.a - 2.06.c,** please specify other standards used for creating associated data documentation in **2.06.d**
2.06.a Data Dictionaries (ctrl+click for multiple selections):

FAIRCodeBook-Elixir
Open Science Framework
Smithsonian Data Management Best Practices: Describing You Data: Data Dictionaries
United States Geological Survey
University of Pennsylvania
To be determined
Other

2.06.b README files (ctrl+click for multiple selections):

ETS Bibliotech
National Transportation Library
To be determined
Other

2.06.c Codebooks (ctrl+click for multiple selections):

ICPSR
To be determined
Other

2.06.d Any other documentation not listed above (such as methodology, data use guidance, survey instruments, population descriptions, etc.):

2.07 Please list any standards used to ensure usability and interoperability of the data (including standardized variable names and controlled vocabularies such as: FIPS codes; US Census GEOIDs; ISO 8601 Date and time format; National Library of Medicine Medical Subject Headings (MeSH); Data Documentation Initiative (DDI); others) (ctrl+click for multiple selections):

Data Documentation Initiative (DDI) [survey data]
Date and time format ISO 8601
Geographic information — Metadata (ISO 19115)
Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (FGDC/CSI
Information Technology - Codes for the Identification of the States and Equivalent Areas within the
NLM Medical Subject Headings (MeSH) controlled vocabulary
U.S. Board on Geographic Names – Domestic Names Geographic Names Information System (GN
US Census American Indian, Alaska Native, and Native Hawaiian Areas (AIANNH) 2020 Census C
US Census Congressional Districts codes
US Census County and County Equivalent Entities
US Census GEOIDs
US Census North American Industry Classification System (NAICS)
US Census School Districts
US Census Voting Districts
To be determined
Other

2.07.a If you selected "Other", please specify:

3. Mechanisms for or Limitations to Providing Access to and Sharing of the Data

In this section, please respond to prompts about mechanisms for or limitations to providing access to and sharing of the data. Include a description of provisions for the protection of privacy, confidentiality, security, intellectual property, or other rights. This section should address access to identifiable and de-identified data or justification for not making the data accessible. You may not have answers to some or many of the prompts in this section at proposal time. Use the "To be determined" response as needed in this section.

3.01 Select a Public Data Access Level (ctrl+click for multiple selections):
Select "**Public**" if the data collected/generated in this project will be released to the public in either full, microdata, or aggregated form.
Select "**Restricted**" if the project data will be shared with restrictions or via CDC Research Data Center (RDC).
Select "**Non-Public**" if the project data will not be released to or shared with the public.

Public
Restricted
Non-Public

3.01.a If more than one Access Level is chosen, please provide your justification:

3.01.b If the response above is "Restricted" or "Non-Public" please provide a reason for not making data "Public". Select the single best response:

3.01.c If you selected "Other", please specify:

3.02 Select or describe how will you protect participant and or participating organization privacy and confidentiality, and de-identify your data before sharing. Choose a single primary response:

3.02.a If you selected "Other", please specify:

3.03 Select or describe how you will protect intellectual property rights before sharing data:

3.03.a If you selected "Other", please specify:

3.04 Will CDC have access to the PII: ◯ Yes ◯ No ◯ To be determined ◯ Not Applicable ◯ Other

3.04.a If you selected the "Yes" response above, enter the justification for CDC having access to the PII:

4. Appropriate Documentation

In this section, include statements to ensure all released data have appropriate documentation that describes the method of collection, what the data represent, and potential limitations for use. You may not have answers to some or many of the prompts in this section at proposal time. Use the "To be determined" response as needed in this section.

4.01 Select the data collection methods (ctrl+click for multiple selections):

Analytic Services (can be data/specimen TA for non-research, research, investigations)
Animal Model Research
Biomonitoring
Case-Control
Convenience Sample
Culture
Exposure Investigation
Focus Group / Discussion Group
Genetic Sequencing
Health Consultation
Health Education
Hybrid Study Design
Individual Interview (Quantitative)
Individual Interviews (Qualitative)
Method Development Research
Method/Device Evaluation
Modeling Study
Needs Assessment
Observation
Outbreak Investigation
Pathogen-specific Research
Prevalence (Cross-sectional) Surveys
Proficiency Testing
Prospective Cohort Study
Public Health Assessment
Quality Assurance/Quality Improvement
Randomized Clinical Trial
Randomized Controlled Trial
Record Review
Reference Diagnostics
Retrospective Cohort Study
Secondary Data Analysis
Secondary Specimen Analysis
Surveillance Support
Survey
Technical Assistance
To be determined
Toxicological Profiles
Other

4.01.a If you selected "Other", please specify:

4.02 Select the target population(s) included/represented (ctrl+click for multiple selections):

No Human Population
General US Population
Children
Fetuses
Neonates
Pregnant Women
Prisoners
American Indian or Alaska Native
Asian
Black or African American
Hispanic or Latino
Native Hawaiian or Other Pacific Islander
White
Female
Male
Adult 18-24 years
Older adults > 64 years
Businesses
Emancipated Minors
Farmers
Healthcare Provider
Impaired hearing or deaf
Impaired mental
Impaired physical
International
Immigrants or Refugees
Patient
Other

4.02.a If you selected "Other", please specify:

4.03 Describe the processes you will follow to provide informed consent to participants. Please provide Informed Consent waivers, and other protocol documents when sending this document if appropriate:

4.04 Has the Project been reviewed and approved by the appropriate IRB:

4.04.a If the IRB Review is complete, please provide the IRB #:

Note: The IRB process will not begin unless this proposal is accepted for funding.

4.05 If appropriate, describe how your research protocols align with other Ethical Protocols, such as the CARE Principles for Indigenous Data Governance; and/or local, regional, or national data collection ethical guidelines, especially in countries outside the United States:

4.06 Describe potential limitations of data use:

5. Plans for Data Archiving and Long-Term Preservation

In this section, you will indicate that you are planning for the archiving and long-term preservation of the data; or, you will indicate why long-term preservation and access are not justified. **You will check either response 5.01 or 5.02, but not both.**

○ **5.01** Data from this project will be archived and preserved for the long term. We are planning to use at least one of the following repositories for long-term preservation, as appropriate to the dataset content or file type (ctrl+click for multiple selections):

data.cdc.gov
onehealthportal.cdc.gov
icpsr.umich.edu
one or more National Institute of Health-supported repositories
Zenodo
Dryad
figshare
GitHub.com/CDCgov
Journal publisher repository
To be determined
Other

5.01.a If you selected "Other", please specify:

○ **5.02** Data from this project will NOT be archived and preserved for the long term. We do not feel long-term preservation of this data is justified for the following reason(s):

5.02.a If you selected "Other", please specify:

Notice of Funding Opportunity Data Management Plan (DMP)

This form documents the required electronic signatures for the Data Management Plan (DMP). The Authorized Organization Representative (AOR) must electronically sign the DMP at application and upon revisions, as required by the NOFO. The Principal Investigator/Project Director (PI/PD) should also acknowledge review of the plan.

Authorized Organization Representative (AOR) – Required

Name (typed):	
Title:	
Organization (Legal Name):	
Email:	
Phone:	
Electronic Signature (per ESIGN/GPEA):	
Signed by:	
Date/Time (UTC):	

Principal Investigator/Project Director (PI/PD) Review Complete – Recommended

Name (typed):	
Electronic Acknowledgment: I have reviewed and will follow the approved DMP	
Signed by:	
Date/Time (UTC):	

Certification: The undersigned AOR affirms that the DMP accurately describes how project data and metadata will be managed, protected, shared, and preserved in accordance with the Notice of Funding Opportunity and applicable federal policy. Material deviations from this plan require prior written approval by the Federal awarding agency.

Appendix A: Additional Section 1 Dataset Description for Projects with more than one Dataset

Some projects will collect, generate, and/or reuse more than one (1) dataset. If this is the case, you should describe each additional dataset by filling out another version of Section 1. Description of Data in the space provided in Appendix A. If your project has more than five (5) datasets, copy the pages of a blank version of the Section 1. Description of Data and insert them in this PDF.

AD1. Description of Additional Dataset 1

In this section, provide a description of the data to be collected, generated, obtained, or transfered in the proposed project. You may not have answers to some or many of the prompts in this section at proposal time. Use the "To be determined" response as needed in this section.

Additional Dataset 1

AD1.1.01 Working Dataset Title, if known:

AD1.1.02 Please provide an initial Dataset Description:

AD1.1.03 Proposed Collection Start Date (YYYY-MM-DD):

◯ ◯ To be determined

AD1.1.04 Proposed Collection End Date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned / Continual without end

AD1.1.05 List expected Language(s) of the data (English, Spanish, Others):

AD1.1.06 List expected or proposed Tags or Keywords:

Release

AD1.1.07 Expected initial public data release date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned

AD1.1.08 Expected type(s) of data released (ctrl+click for multiple selections):

Complete Raw Dataset
Research Replication Dataset
Aggregated Data
Microdata
Public-use Dataset
To be determined
Other

AD1.1.08.a If you selected "Other", please specify:

Frequency

AD1.1.09 Expected data update frequency, if known (ctrl+click for multiple selections):

One-time collection, no updated expected
Baseline and follow-up(s)
Ongoing Surveillance
Continually Updated
Updated Monthly
Updated Annually
Updated Irregularly
To be determined
Other

AD1.1.09.a If you selected "Other", please specify:

AD1.1.10 Proposed timing of data availability to the Public, if known:

AD1.1.10.a If you selected "Other", please specify:

Related Software, and Code

AD1.1.11 Expected software(s) to be used in the creation of the data (ctrl+click for multiple selections):

Electronic Health Records (EHR) interface
Fyre
Microsoft Excel
Microsoft Word
PDF
Plain Text Editor (.txt)
Qualtrics
REDCap (Research Electronic Data Capture)
RStudio
SAS
SPSS Statistics
To be determined
Other

AD1.1.11.a If you selected "Other", please specify:

AD1.1.12 Site name or URL to the external site where the script or code will be located or shared (if known):

AD1.1.13 Expected scripting language(s) to be used (ctrl+click for multiple selections):

Python
R
SAS
SPSS
To be determined
Other

AD1.1.13.a If you selected "Other", please specify:

Proceed to next page.

Spatiality

Include the geographic area that is the focus of the research. Be as specific as possible.

AD1.1.14 Spatiality Details:

Country/ Tribal Nation	State/ Province	County/ Region	City/ Town	Specific Place Name/ Other	Geographic Coordinates

Appendix A: Additional Section 1 Dataset Description for Projects with more than one Dataset

AD2. Description of Additional Dataset 2

In this section, provide a description of the data to be collected, generated, obtained, or transfered in the proposed project. You may not have answers to some or many of the prompts in this section at proposal time. Use the "To be determined" response as needed in this section.

Additional Dataset 2

AD2.1.01 Working Dataset Title, if known:

AD2.1.02 Please provide an initial Dataset Description:

AD2.1.03 Proposed Collection Start Date (YYYY-MM-DD):

◯ ◯ To be determined

AD2.1.04 Proposed Collection End Date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned / Continual without end

AD2.1.05 List expected Language(s) of the data (English, Spanish, Others):

AD2.1.06 List expected or proposed Tags or Keywords:

Release

AD2.1.07 Expected initial public data release date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned

AD2.1.08 Expected type(s) of data released (ctrl+click for multiple selections):

Complete Raw Dataset
Research Replication Dataset
Aggregated Data
Microdata
Public-use Dataset
To be determined
Other

AD2.1.08.a If you selected "Other", please specify:

Frequency

AD2.1.09 Expected data update frequency, if known (ctrl+click for multiple selections):

One-time collection, no updated expected
Baseline and follow-up(s)
Ongoing Surveillance
Continually Updated
Updated Monthly
Updated Annually
Updated Irregularly
To be determined
Other

AD2.1.09.a If you selected "Other", please specify:

AD2.1.10 Proposed timing of data availability to the Public, if known:

AD2.1.10.a If you selected "Other", please specify:

Related Software, and Code

AD2.1.11 Expected software(s) to be used in the creation of the data (ctrl+click for multiple selections):

Electronic Health Records (EHR) interface
Fyre
Microsoft Excel
Microsoft Word
PDF
Plain Text Editor (.txt)
Qualtrics
REDCap (Research Electronic Data Capture)
RStudio
SAS
SPSS Statistics
To be determined
Other

AD2.1.11.a If you selected "Other", please specify:

AD2.1.12 Site name or URL to the external site where the script or code will be located or shared (if known):

AD2.1.13 Expected scripting language(s) to be used (ctrl+click for multiple selections):

Python
R
SAS
SPSS
To be determined
Other

AD2.1.13.a If you selected "Other", please specify:

Proceed to next page.

Spatiality

Include the geographic area that is the focus of the research. Be as specific as possible.

AD2.1.14 Spatiality Details:

Country/ Tribal Nation	State/ Province	County/ Region	City/ Town	Specific Place Name/ Other	Geographic Coordinates

Appendix A: Additional Section 1 Dataset Description for Projects with more than one Dataset

AD3. Description of Additional Dataset 3

In this section, provide a description of the data to be collected, generated, obtained, or transfered in the proposed project. You may not have answers to some or many of the prompts in this section at proposal time. Use the "To be determined" response as needed in this section.

Additional Dataset 3

AD3.1.01 Working Dataset Title, if known:

AD3.1.02 Please provide an initial Dataset Description:

AD3.1.03 Proposed Collection Start Date (YYYY-MM-DD):

◯ ◯ To be determined

AD3.1.04 Proposed Collection End Date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned / Continual without end

AD3.1.05 List expected Language(s) of the data (English, Spanish, Others):

AD3.1.06 List expected or proposed Tags or Keywords:

Release

AD3.1.07 Expected initial public data release date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned

AD3.1.08 Expected type(s) of data released (ctrl+click for multiple selections):

Complete Raw Dataset
Research Replication Dataset
Aggregated Data
Microdata
Public-use Dataset
To be determined
Other

AD3.1.08.a If you selected "Other", please specify:

Frequency

AD3.1.09 Expected data update frequency, if known (ctrl+click for multiple):

One-time collection, no updated expected
Baseline and follow-up(s)
Ongoing Surveillance
Continually Updated
Updated Monthly
Updated Annually
Updated Irregularly
To be determined
Other

AD3.1.09.a If you selected "Other", please specify:

AD3.1.10 Proposed timing of data availability to the Public, if known:

AD3.1.10.a If you selected "Other", please specify:

Related Software, and Code

AD3.1.11 Expected software(s) to be used in the creation of the data (ctrl+click for multiple selections):

Electronic Health Records (EHR) interface
Fyre
Microsoft Excel
Microsoft Word
PDF
Plain Text Editor (.txt)
Qualtrics
REDCap (Research Electronic Data Capture)
RStudio
SAS
SPSS Statistics
To be determined
Other

AD3.1.11.a If you selected "Other", please specify:

AD3.1.12 Site name or URL to the external site where the script or code will be located or shared (if known):

AD3.1.13 Expected scripting language(s) to be used (ctrl+click for multiple selections):

Python
R
SAS
SPSS
To be determined
Other

AD3.1.13.a If you selected "Other", please specify:

Proceed to next page.

Spatiality

Include the geographic area that is the focus of the research. Be as specific as possible.

AD3.1.14 Spatiality Details:

Country/ Tribal Nation	State/ Province	County/ Region	City/ Town	Specific Place Name/ Other	Geographic Coordinates

Appendix A: Additional Section 1 Dataset Description for Projects with more than one Dataset

AD4. Description of Additional Dataset 4

In this section, provide a description of the data to be collected, generated, obtained, or transfered in the proposed project. You may not have answers to some or many of the prompts in this section at proposal time. Use the "To be determined" response as needed in this section.

Additional Dataset 4

AD4.1.01 Working Dataset Title, if known:

AD4.1.02 Please provide an initial Dataset Description:

AD4.1.03 Proposed Collection Start Date (YYYY-MM-DD):

◯ ◯ To be determined

AD4.1.04 Proposed Collection End Date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned / Continual without end

AD4.1.05 List expected Language(s) of the data (English, Spanish, Others):

AD4.1.06 List expected or proposed Tags or Keywords:

Release

AD4.1.07 Expected initial public data release date (YYYY-MM-DD):

◯ ◯ To be determined ◯ None Planned

AD4.1.08 Expected type(s) of data released (ctrl+click for multiple selections):

Complete Raw Dataset
Research Replication Dataset
Aggregated Data
Microdata
Public-use Dataset
To be determined
Other

AD4.1.08.a If you selected "Other", please specify:

Frequency

AD4.1.09 Expected data update frequency, if known (ctrl+click for multiple selections):

One-time collection, no updated expected
Baseline and follow-up(s)
Ongoing Surveillance
Continually Updated
Updated Monthly
Updated Annually
Updated Irregularly
To be determined
Other

AD4.1.09.a If you selected "Other", please specify:

AD4.1.10 Proposed timing of data availability to the Public, if known:

AD4.1.10.a If you selected "Other", please specify:

Related Software, and Code

AD4.1.11 Expected software(s) to be used in the creation of the data (ctrl+click for multiple selections):

Electronic Health Records (EHR) interface
Fyre
Microsoft Excel
Microsoft Word
PDF
Plain Text Editor (.txt)
Qualtrics
REDCap (Research Electronic Data Capture)
RStudio
SAS
SPSS Statistics
To be determined
Other

AD4.1.11.a If you selected "Other", please specify:

AD4.1.12 Site name or URL to the external site where the script or code will be located or shared (if known):

AD4.1.13 Expected scripting language(s) to be used (ctrl+click for multiple selections):

Python
R
SAS
SPSS
To be determined
Other

AD4.1.13.a If you selected "Other", please specify:

Proceed to next page.

Spatiality

Include the geographic area that is the focus of the research. Be as specific as possible.

AD4.1.14 Spatiality Details:

Country/ Tribal Nation	State/ Province	County/ Region	City/ Town	Specific Place Name/ Other	Geographic Coordinates

Appendix B: Links to Standards and Metadata Schema listed in Section 2 drop-down menus

2.04 Metadata schema(s) to be used to describe the data:

1. Biotechnology — Requirements for data formatting and description in the life sciences (ISO 20691:2022) [https://www.iso.org/standard/68848.html]
2. Brain Imaging Data Structure (BIDS) [https://bids.neuroimaging.io/]
3. Core Information for Metabolomics Reporting (CIMR) [https://rdamsc.bath.ac.uk/msc/m130]
4. Core Scientific Metadata Model (CSMD) [https://rdamsc.bath.ac.uk/msc/m8]
5. DCAT-AP-NL [https://geonovum.github.io/DCAT-AP-NL30/]
6. Fast Healthcare Interoperability Resources (FHIR) [https://rdamsc.bath.ac.uk/msc/m103]
7. Genome Metadata [https://www.bv-brc.org/docs/system_documentation/data.html#genome-metadata]
8. Genomics Informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records 90ISO/TS 20428:2024) [https://www.iso.org/standard/82205.html]
9. Health-RI [https://github.com/Health-RI/health-ri-metadata]
10. Health Data CATalogue – Application Profile (HealthDCAT-AP) [https://healthdcat-ap.github.io/]
11. Health informatics – Clinical knowledge resources (ISO 13119:2022) [https://www.iso.org/standard/78392.html]
12. Minimum Information about Any Sequence (MIxS) [https://rdamsc.bath.ac.uk/msc/m108]
13. Minimum Information for Biological and Biomedical Investigations (MIBBI) [https://rdamsc.bath.ac.uk/msc/m72]
14. National Research Data Infrastructure for Personal Health Data (NFDI4Health) [https://repository.publisso.de/resource/frl:6450625]
15. NIH Common Metadata Elements for Cataloging Biomedical Datasets [https://doi.org/10.6084/m9.figshare.1496573.v1]
16. NIH Protocol Data Element Definition for ClinicalTrials.gov [https://clinicaltrials.gov/policy/protocol-definitions]
17. Open Microscopy Environment TIFF (OME-TIFF) [https://rdamsc.bath.ac.uk/msc/m73]
18. Open Microscopy Environment XML (OME-XML) [https://rdamsc.bath.ac.uk/msc/m29]
19. Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework (PDBx/mmCIF) [https://rdamsc.bath.ac.uk/msc/m30]
20. Recommended Metadata for Biological Images (REMBI) [https://rdamsc.bath.ac.uk/msc/m112]
21. Geographic information — Metadata (ISO 19115) [https://www.iso.org/standard/53798.html]
22. Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (FGDC/CSDGM) [https://rdamsc.bath.ac.uk/msc/m17]

2.05 Standards for persistent identifiers (PIDs):

1. Digital Object Identifiers (DOI) [https://www.doi.org/]
2. Archival Resource Key (ARK) [https://arks.org/about/]
3. Handle System (handles) [https://datatracker.ietf.org/doc/html/rfc3650]
4. ORCID (Open Researcher and Contributor Identifier) [https://orcid.org/]
5. Research Organization Registry (ROR) [https://ror.org/]
6. System Award Management (SAM) identifier [https://sam.gov/entity-registration]

2.06.a Standards used for creating Data Dictionaries:

1. FAIRCodeBook-Elixir [https://faircookbook.elixir-europe.org/content/recipes/interoperability/creating-data-dictionary.html]
2. Open Science Framework [https://help.osf.io/article/217-how-to-make-a-data-dictionary]
3. Smithsonian Data Management Best Practices: Describing You Data: Data Dictionaries [https://library.si.edu/sites/default/files/tutorial/pdf/datadictionaries20180226.pdf]
4. United States Geological Survey [https://www.usgs.gov/data-management/data-dictionaries]
5. University of Pennsylvania [https://repository.upenn.edu/entities/publication/0430ccdd-cbd8-4404-9f54-11cb81d5b3b1]

2.06.b Standards used for creating README files:

1. ETS Bibliotech [https://etsmtl.libguides.com/c.php?g=734817&p=5318309]
2. National Transportation Library [https://transportation.libguides.com/researchdatamanagement/datapackages]

2.06.c Standards used for creating Codebooks:

1. ICPSR [https://www.icpsr.umich.edu/web/ICPSR/cms/1983]