

## Plan Overview

---

*A Data Management Plan created using DMPTool*

**Title:** AR or HAI Data Management Plan

**Creator:** Scott Fridkin

**Affiliation:** Emory University (emory.edu)

**Principal Investigator:** Scott Fridkin

**Data Manager:** Scott Fridkin

**Funder:** National Science Foundation (nsf.gov)

**Funding opportunity number:** CDC AR

**Template:** NSF-EAR: Earth Sciences

**Last modified:** 11-01-2016

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

---

## AR or HAI Data Management Plan

### Description

Simple description of kinds of data elements, amount of data, what the data can be used for (150 words or less), years of data included and collection cycles to expect updates

### *Project Type* (surveillance, research, administration)

Indicate the primary purpose for which the data are being collected (e.g., surveillance, research, program administration, program evaluation, investigation, or other)

### *Standards used*

Brief descriptions of standards that would be used in creating a dataset (e.g., how were variables categorized, suppression of values, level of aggregation (case, facility, region, national). Ideally provide link to full documentation or summarize key aspects here.

### *Data Representation/Limitations*

Include brief description of representativeness and limitations of use (accuracy, interpretation, validation efforts)

### *Quality Check*

Describe if data set had a specific review of data for quality, and what that was (how are you confident that data available reflects what was used in the publication)

### *Spatial*

List the range of spatial applicability of a data set. Could include a spatial region or a named place. List the geographic entities (variables) for which data are available (e.g., City, Country, Region, State, Country)

### *Population Represented*

Open text summarizing population

### *Spatial*

The range of spatial applicability of a data set. Could include a spatial region or a named place. List the geographic entities (variables) for which data are available (e.g., City, Country, Region, State, Country)

### *Population Represented*

Open text summarizing population

### *Current Project Status.*

1=planning, collection not started; 2=collection ongoing; 3=collection complete-cleaning; 4=collection complete data available

### *Project Start Year*

Enter the 4 digit year corresponding to the fiscal year the project was first awarded. If this is a recurring award enter the first year of the current cycle of funding

*Timing:* Proposed timing of availability is:

- By one year of end of collection cycle
- 30 months after end of a time limited data collection (not from time of cleaning, analysis, publication)
- Data tables only at time of publication
- Data tables at other time

*Temporal:* The range of temporal applicability of project (i.e., a start and end date of applicability for the data). Indicate the data years (or other periods) for which data are available; the date data collection started and is expected to end

*Tags:* keywords to make dataset discoverable, can be used by technical and non-technical users. consider MeSH terms

*Last Update:* date to indicate approval date of the most recent version of this DMP if updated

*Data Dictionary:* indicate Y/N if data dictionary is available, and if yes provide a URL

*Data File Format:* (e.g., Excel, SAS)

*Physical Location:* location of dataset (physically)

FOA: Link to any FOA that this DMP is linked to

*Exclusion.* Summarize if data is justifiable excluded from compliance with CDC Policy CDC-GA-2005-14 (Policy on Public Health REsearch and Nonresearch Data Management and Access): Check best single exclusion categories:

- a. Not part of CDC system
- b. Under license agreement / restricted DUA (prohibiting publication)
- c. Dual-Use Research concerns
- d. Protected under 308(d)
- e. Protected under 301(d)

*Public Access Determination.* If not excluded, label: 1-Justified non-compliance or 2- Compliant

If Justified non-compliance or restricted access levels as described below, summarize privacy, security, cost, accessibility, and impact on grantees, perceived value of data access as considerations in justification.

(255 character limit)

- Provide reasons for not providing data to the public, the benefit of providing the information does not justify the cost in terms of privacy, security, cost of provision. If the benefits are perceived as low, justify why using examples ideally or other available data.
- For 'restricted public' access level, please describe justification for having restrictions (e.g., requires IRB approval, limited access at Research Data Center, limited access to summary data)

### *Public Access Level*

The degree to which this data collected as part of this project could be made publicly-available, *regardless of whether it has been made available is:*

- Public (Data set or summary data tables could be made publicly available to all without restrictions or suppressions.)
- Restricted public (available with restrictions and/or limitations), either restricted use (data suppression, binding Data Use Agreements), restricted access (controlled access only-RDC), or summary data only
- Non-public (Data set is not available to members of the public).

*Timing.* Proposed timing of availability is:

- By one year of end of collection cycle
- 30 months after end of a time limited data collection (not from time of cleaning, analysis, publication)
- Data tables only at time of publication
- Data tables at other time

### *Access URL*

URL providing indirect access to the data set. (Must include protocol, data dictionary (variable names and valid values), data collection instrument and other relevant information. Please provide the URL for the documentation associated with this project. The information is to include protocol, data dictionary (variable names, description, attributes, valid values), data collection instrument and other relevant information.

### *Download URL*

URL providing direct access to a downloadable file of the data set. Provide the URL for access to a downloadable file of the dataset.

### *License/Other agreements*

provide the license or non-license (i.e. Public Domain) status with which the data set has been published. See Open Licenses for more information. Government works are by default in the U.S. Public Domain. If no other open license applies, the following URL should be used for the license field:

<http://www.usa.gov/publicdomain/label/1.0/>

### *Publisher*

The publishing entity and optionally their parent organization(s) (e.g., CDC, grantee). Please provide the name of the data owner who will be publishing the data associated with this DMP. If varied, can there be a location from which specific datasets publishers can be identified

### *Contact Name and Email*

Contact person's name and email address (Name, address, e-mail and phone). Please provide the name, email address and phone number of the DMP /data steward for this project.

*Data integrity.* List provisions for integrity (e.g., protection, confidentiality, intellectual property)

*Security Assessment and Authorization* (Formerly Certification and Accreditation). If CDC system - The review is handled by the OCISO/Center Privacy Officer. For more information about the process see <http://intranet.cdc.gov/ocio/docs/information-systems-security/CertificationAndAccreditation.pdf>.

*Preservation Expiration date.* Enter the 4 digit year that the dataset will be available until.

All datasets are required to have an expiration date. After this date data will be archived and available upon request

*Archival plans*

Describe justification for plans for archive (length of time to maintain, rational for duration)

.

---