CDC has developed the Operational Policy "**Policy on Public Health Research and Nonresearch Data Management and Access"**

1. **Reason for Update**: Revisions needed to reflect government directives requiring all science agencies of the government to develop and execute specific plans regarding data.

2. **Summary of Policy**: The purpose of this policy is to ensure that the Centers for Disease Control and Prevention (CDC) and the Agency for Toxic Substances and Disease Registry (ATSDR) manage public health data and provide access to such data for public use.

   This policy requires CDC to manage and facilitate public access to its funded public health data and remain compliant with government initiatives promoting open government and transparency. This policy addresses public health data management through all stages of the process and making data accessible.

3. **Related Issuances:** None

4. **Responsible Organization:** Office of Associate Director for Science

5. **Material Superseded:** CDC/ATSDR Policy on Releasing and Sharing Data, 9/7/2005

6. **Recertification:** This document is scheduled for recertification on or before the last working day of January 2021.

7. **Points of Contact:** Dr. Mary Ari, Office of the Associate Director for Science, 404-639-2492 or Becca Pope, Management Analysis and Services Office, 770-488-4803.

8. To go directly to the Policy, enter the following URL into the location line of your browser**:** http://masoapplications.cdc.gov/Policy/Doc/policy385.pdf

/s/ Sherri A. Berger, MSPH
Chief Operating Officer

Category: General Administration
Policy #: CDC-GA-2005-14
Date of Issue: 04/16/2003, Updated, 09/07/2005, Updated, 1/26/2016
Proponent: Office of the Associate Director for Science
Applicable Locations: All domestic and international locations
Applicable Staff: All Staff, including CDC Employees, Contractors, Commission Corps Officers, and Affiliates

POLICY ON PUBLIC HEALTH RESEARCH AND NONRESEARCH DATA MANAGEMENT AND ACCESS

Sections:

1. PURPOSE AND SCOPE

The purpose of this policy is to ensure that the Centers for Disease Control and Prevention (CDC)* and the Agency for Toxic Substances and Disease Registry (ATSDR), in line with their mission and in compliance with applicable laws, regulations, directives, and guidelines, manage public health data and provide access to such data for public-use. This policy applies to public health research and nonresearch data [1]** collected or generated using CDC funds. The goal of the policy is to ensure public access to federally funded public health data.

This policy supersedes the *CDC/ATSDR Policy on Releasing and Sharing Data (09/07/2005)* [2] and guidance based on it. This policy applies to new collections of public health data proposed after this policy becomes effective. Only new awards, as of the date of issue of this policy and all continuations thereafter, are subject to this policy. Previously established public health data collection systems that continue to collect data after this policy becomes effective should become compliant within three years of the effective date of this policy, or when the data system undergoes substantial revision—whichever comes first. However, this should not involve costly retrofitting of legacy systems. This policy will also ensure that CDC is in compliance with the following; Office of Management and Budget (OMB) memorandum titled “Open Data Policy–Managing Information as an Asset” (OMB M-13-13) [3]; Executive Order 13642 titled “Making Open and Machine Readable the New Default for Government Information” [4]; and the Office of Science and Technology Policy (OSTP) memorandum titled “Increasing Access to the Results of Federally Funded Scientific Research” (OSTP Memo) [5].

For purposes of this policy, “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 nonresearch activities. CDC research and nonresearch activities are described in the policy

*References to CDC also include Agency for Toxic Substances Disease Registry (ATSDR) throughout this document.
**Public health research and nonresearch data is referred to as public health data throughout this document

"Distinguishing Public Health Research and Public Health Nonresearch" [1]. Public health data could be quantitative, qualitative, imaging, or genomic output (for example, genome sequencing, arrays, gene expression, etc.). 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.

Public health data covered by the policy are those:

- Collected or generated by CDC
- Collected or generated by other agencies or organizations funded or co-funded by CDC (for example, through mechanisms such as grants, cooperative agreements, contracts, or other funding mechanisms). When CDC funds another federal agency, an interagency agreement should indicate who would be responsible for the data
- Reported to CDC by another entity (for example, by state health departments) that become a part of a CDC data collection system (for example, CDC surveillance systems)

Public health data covered by the policy, but which release or sharing may be limited includes those:

- Protected from disclosure by applicable laws and regulations (for example, the Privacy Act [6,7], the Trade Secrets Act [8], Section 308(d) of the Public Health Service Act [9])
- Deemed not shareable due to the potential of dual-use research of concern [10]

Public health data not covered by this policy includes those:

- Collected and generated by other organizations but that are shared with CDC
- Provided to CDC by a license agreement with restrictions on the use and sharing of the data
- Provided to CDC by another federal agency (for example the Centers for Medicare and Medicaid Services) under restricted terms of use and sharing of the data

This policy applies to all CDC staff, including CDC employees (part and full-time equivalent), contractors, Commission Corps Officers, and affiliates (such as interns, fellows, guest researchers, and locally employed staff). It applies to Centers, Institute, and Offices (CIOs) that manage intramural public health data and/or support extramural collection or generation of public health data and those external entities that collect public health data using CDC funds. This policy applies to staff at any domestic or foreign CDC location.

2. BACKGROUND

CDC is the nation's principal disease prevention and health promotion organization. In support of its mission, CDC collects, generates, stores, uses, and routinely provides access to public health data. Public health and scientific advancement are best served when public health data are released to, or shared with, other public health agencies, academic researchers, private researchers (if appropriate) and other partners in an open, timely, and appropriate way. CDC also recognizes the critical importance of:

- Maintaining standards of data quality
- Upholding individual and institutional privacy and confidentiality

- Protecting information based on national security concerns and law enforcement investigations and activities
- Protecting proprietary interests and business confidential information
- Protecting intellectual property rights; considering ethical matters
- Ensuring impartiality in the sharing of public health data

This policy concerns all types of public health data. For this reason, stakeholders, such as states, local public health agencies, public health organizations, and educational institutions have an interest in the implementation of the policy.

3. POLICY

Sections:

A. Overview

This policy ensures that CDC will manage and facilitate public access to its funded public health data. To that end, CDC seeks to make accessible public health data it has collected and generated, subject to limits imposed by law, ethical considerations, resources, technology, data quality, and must ensure the complete protection of data from physical and electronic risks to privacy and confidentiality. This policy addresses public health data management through all stages of the process and making data accessible. Before any data are made accessible, all phases of data collection, transmission, editing, processing, analysis, and storage must be evaluated for quality [11,12,13].

The requirements of the policy apply to CDC intramural and extramural public health data.

B. Requirements

1) Data Management Plan

A Data Management Plan (DMP) is required for each intramural and extramural collection of public health data covered by this policy. The DMP should be developed during the project planning phase prior to the initiation of collecting or generating public health data. The DMP should include:

- Descriptions of the data to be produced in the proposed project
- How access will be provided to the data (including provisions for protection of privacy, confidentiality, security, intellectual property, or other rights)
- Use of data standards that ensure all released data have appropriate documentation that describes the method of collection, what the data represents, and potential limitations for use
- Plans for archival and long-term preservation of the data, or explaining why long-term preservation and access cannot be justified

Principal investigators, investigators, project leads/project officers, and other points of contact, herein after referred to as “investigators”, must provide DMP with project proposals and the quality of the DMP should be evaluated by programs. CIOs should include review of DMPs in intramural project approval processes to ensure all elements are addressed, and ensure that updates are made as needed throughout the lifecycle of the data. A parallel process should be instituted for extramural projects. CDC will be able to monitor what data sets exist from the information provided in the DMP.

2) Public Health Data Security

The level of data protection should be commensurate with the risk and magnitude of the harm that would result from the loss, misuse, unauthorized access to, modification, or destruction of such information [14]. For nonpublic-use data sets that contain personally identifiable information (PII), CIOs should consider the level of security required when choosing where to store data. CIOs should maintain public health data, regardless of access level, in a secure environment to ensure the integrity of the data and prevent loss. Data maintained in paper, electronic, other formats, or on portable devices need to be physically secured when transported and stored. Additionally, the Federal Information Security Management Act (FISMA) requires each agency to provide information security for the information and information systems that support the operations and assets of the agency, including those managed on behalf of the agency [15]. CDC’s policy on protection of information resources, establishes the requirements and responsibilities for the protection of the physical, financial, and information resources, and other tangible and intangible assets of CDC [16].

3) Public Health Data Access

CDC should ensure that the public health data it collects, generates, and provides funding for are made accessible in a timely manner. To do so, and to minimize disclosure risk, CIOs must institute data review processes prior to making data accessible. Under some circumstances, alternative clearance procedures are used to disseminate public health data (for example, several CIOs must disseminate health, medical, and safety information in real time to protect the public’s health against urgent and emerging threats). CDC reserves the right to waive information quality standards temporarily for CIOs addressing urgent situations in accordance with the latitude described in both the OMB- and CDC-specific guidelines [11,13].

Authors should make data underlying papers accessible at time of publication of the paper. In general, whether publication emanates from work or not, the final version of a data set intended for release or sharing should be made accessible within 30 months after the end of data collection. This timeline applies to intramural and extramural data. CIOs are encouraged to make data available as soon as possible and no later than the deadline stated above. Data must be made accessible in a nonproprietary format. Since surveillance data collection may be ongoing in nature, CIOs should provide in line with surveillance purposes, and to the extent allowed by law and consistent with this policy, access to the surveillance data that it collects or have custody of, within a year of the end of each collection cycle. Procedures for releasing public-use data including the re-release of data provided by jurisdictions should be consistent with national standards. CIOs may handle requests for data during a public health emergency on a case-by-case basis.

a. Making the determination to make data accessible

While CDC is committed to maximizing data accessiblity, the costs in preparing data and the associated documentation for release or sharing can be considerable. When CIOs are making decisions about whether to make individual data sets accessible, and decisions about whether and when to retire data sets, programs must balance the relative value of long-term preservation and access with the associated cost and administrative burden. CIOs should determine the data retention period on a case-by-case basis depending on the value of keeping the data and the adherence to objectives, responsibilities, standards, guidelines, and instructions to meet federal records management regulations, laws, and best practices for the management of records [17]. Decisions about what data to make accessible will take into consideration resource constraints, value of the data and the cost of making them accessible, the nature of the data as it relates to issues of privacy, confidentiality, national and homeland security, dual-use research potential, trade secrets, proprietary information, and requirements of other pertinent statutes (for example Freedom of Information Act) and data held by grantees. There needs to be an adequate justification for not making data accessible and this justification must be documented in the DMP.

b. Mechanisms to make public health data accessible

Public health data can be made accessible in two ways:

- Accessible for public-use without restrictions (data release)
- Accessible to particular parties with restrictions (data sharing)

Data release: Dissemination of data either for public-use or through an *ad hoc* request so that the program or data steward or designee no longer controls the data. Data release is the usual and preferred method for making data available.

Data sharing: Granting certain individuals or organizations access to data that cannot be released publicly (for example, data that contains individually identifiable or potentially identifiable information and data releases limited by law). Data release restrictions can be imposed because of legal constraints or because releasing the data would risk disclosing proprietary or confidential information or compromising national security or law enforcement interests. Consistent with applicable law and existing CDC data security requirements, access to confidential or PII by staff outside the applicable program should:

- Be limited to those authorized based on an expressed and justifiable public health need
- Not compromise or impede public health program activities
- Not affect the public perception of confidentiality of the data collection activity
- Be approved by a designated CIO or program official

Data sharing is further subdivided into the following two categories:

i. Data sharing through a special data-use agreement. These data cannot be released publicly but need not always be under CDC's control. Typically, this applies to public health data that to truly de-identify will limit data to so few high-level variables that the data would be of little or no scientific value in any resulting analyses. In such cases, data may be shared upon request with a special data-use agreement that governs their use and further release. Before making their

data available, however, CIOs must evaluate requests for permission to use confidential or private information to ensure these data are appropriate for the use sought, will be used consistent with any applicable legal restrictions on the data, and will be used for an appropriate purpose.

ii. Data sharing under controlled conditions. These data need to remain under CDC custody at all times and are usually analyzed by users at data enclaves located at CDC facilities. Access to public health data with sensitive topics, rare outcomes, or on vulnerable populations may fall in this category.

Decisions about the mechanism used to provide access to data must be documented in the DMP.

If access to the data set has not already been provided to the public at the time of a manuscript publication, authors are required to make the data underlying the conclusions of peer-reviewed scientific publications freely available in a publicly accessible repository at initial publication and in an open, machine-readable, and nonproprietary format. At a minimum, release of this data set should consist of a machine-readable and nonproprietary version of the data tables shown in the paper. When appropriate, a limited data set publication may be followed with a more complete data release within 30 months after the end of data collection.

4) Obligations of Extramural Recipients of CDC Funds

Obligations, with respect to this policy, extend to recipients of CDC funds for research and nonresearch activities herein after referred to as awardees. These also extend to all financial awards (for example, grants, cooperative agreements, contracts, and other funding mechanisms). Applicants for project funding that involve collection or generation of data are required to develop a DMP. This replaces the data sharing plan described under “Resource Sharing Plan” in the research FOA template (includes data, AR-25) [18]. CDC requires that the cost of public health data sharing be included in grants, cooperative agreements, and contracts. The cost of archiving and long-term preservation of public health data may also be included as part of the total budget requested for first-time or continuation awards.

The DMP must describe the data to be collected or generated in the proposed project; standards to be used for collected or generated data; mechanisms for providing access to and sharing of the data (including provisions for the protection of privacy, confidentiality, security, intellectual property, or other rights); use of data standards that ensure all released data have appropriate documentation that describes the method of collection, what the data represents, and potential limitations for use; and plans for archival and long-term preservation of the data, or explaining why long-term preservation and access are not justified.

Awardees should ensure the quality of data they make accessible. All awardees whose terms of award do not include submitting data to CDC should seek to deposit a de-identified data set, accompanying data dictionary, and other documentation relevant to use of the data set in a sustainable repository for archival and long-term preservation. Awardees will be required to inform the appropriate CDC point-of-contact identified by the CIO, via an update to their DMP, the location of where the data are stored. Data underlying scientific publication should be made available coincident with publication of

the paper, unless the data set is already available via a release or sharing mechanism. At a minimum, release of the data set should consist of a machine-readable version of the data tables shown in the paper. Awardees who fail to release public health data in a timely fashion will be subject to procedures normally used to address lack of compliance (for example, reduction in funding, restriction of funds, or award termination) consistent with applicable regulations and policies [19] or other authorities as appropriate. Future awards will depend on compliance with this requirement. The final version of a data set intended for release or sharing (using the mechanism outlined above) should be made available within 30 months after the end of data collection. Awardees are required to make data available as soon as possible and expected to plan for data preparation within the funding period to make data available no later than the deadline stated above. Data must be made accessible in a nonproprietary format.

Funding opportunity announcement (FOA) templates, request for proposals (RFPs), and solicitation terms and conditions should include language that requires submission of a DMP. The Office of Financial Resources (OFR) will revise existing templates with appropriate language. The DMP should be evaluated as appropriate during the application, study proposal, or project review process. Language in the grant, cooperative agreement, or contract notice of award (NoA) should include a request for updates on the DMP, including information about where data are stored.

CDC may enter into memoranda of understanding (MOU), interagency agreements (IAAs), or other agreements with entities, including foreign entities, which involve data collection, sharing, and use. CIOs must ensure that all such agreements are consistent with any continuing funding mechanism related to these data, this data policy, and with any program-specific implementations of this data policy.

C. Compliance and Evaluation

CIOs will report to OADS on a biannual basis, their implementation of efficient processes for providing metadata to ensure discoverability of data, in compliance with requirements. These metadata will include information on intramural and extramural data, and will make it possible for CDC and the public to know what public health data sets exist, the repositories where data reside, and how those data are made accessible.

OFR, in consultation with OADS and CIOs, will develop systems and procedures to ensure compliance with submission of a DMP with each application/award (during review of proposals for award, at the time of issuance of a NoA, and during the submission of progress reports). Future awards will depend on compliance with this requirement.

For awardees, providing a DMP is a required part of their award. Reviewers of grants, cooperative agreements, and contracts must ascertain that submitted proposals for CDC funds include costs of sharing public health data. Programs, in coordination with the appropriate OFR official, should monitor each funded entity's progress and compliance. Expectations will be outlined in each FOA, NoA, or Request for Applications or Proposals (RFAs or RFPs), and contract to inform the potential awardee of the requirements for reporting data collection progress and for submitting a DMP based on the funding opportunity's specifications. Awardee compliance will be monitored through periodic and final progress reports or other performance progress reports and review of the metadata catalog. Programs should review and determine if the progress reports have met the stated objectives of the DMP. Awardees who fail to release public health data in a timely fashion will be subject to procedures normally used to address lack

of compliance (for example, reduction in funding, restriction of funds, or grant termination) consistent with applicable regulations and policies [19] or other authorities as appropriate. Future awards will depend on compliance with this requirement.

For intramural data collections, investigators should submit a DMP with project proposals and programs should review the DMP during existing project approval processes within CIOs. CIOs will review the metadata catalog periodically to track compliance. Staff who do not comply with making data accessible (releasing or sharing) as appropriate may have restrictions imposed on clearance of future publications.

4. RESPONSIBILITIES

A. CDC's Office of the Director

The CDC Office of the Director (OD), specifically the proponent of this policy, OADS, will be responsible for the following:

- Facilitate CIO implementation of the CDC data plan [20]
- Advise and assist the CIOs in the implementation of policy
- Revise policy and related resources as needed
- Provide counsel to CIOs as they develop guidance and engage partners
- Advise and assist CIOs' identification of tools and repositories within or outside of the CDC that are suitable for management and long-term preservation of public health data and recommend the use of such to CIOs
- Facilitate the identification, in collaboration with CDCs Enterprise Information Technology Portfolio Office, and CIO subject matter experts, the best practices for managing and making information on accessible data available to the public and enhance discoverability
- Ensure, working with CIOs and OFR, that the language of FOAs, RFAs/RFPs, and NoAs conveys the respective obligations of applicants and awardees for developing and updating the DMP throughout the life cycle of data; for storing data in a suitable and sustainable repository; and for making data accessible
- Provide training to staff who have access to public health data on requirements of the policy, processes, and procedures
- Provide clear reporting requirements for the CIOs to follow
- Compile reports from the CIOs on compliance with this policy
- Report on implementation progress to HHS/OSTP/OMB

B. CIOs

- Develop systems and processes for efficient and timely transmission of metadata on data to OADS
- Identify an oversight and governance mechanism to ensure compliance
- Ensure responsible data stewardship practices
- Develop guidance as indicated in this policy or as needed and document specifics of the CIO implementation plan
- Ensure that the data policy is explained to staff who collect, generate, or access public health data
- Ensure DMPs are developed for intramural and extramural projects and are reviewed for quality

- Ensure data, determined valuable to preserve, are stored in suitable and sustainable repositories, and determine how long data should be archived to comply with applicable laws, policy, and guidance
- Determine and justify whether to release or share each data set
- Ensure data to be made accessible undergo appropriate preparation including documentation, validation of quality, and review processes prior to releasing or sharing
- Ensure technical processes sustain data integrity and processes exist for data quality issue resolution
- Ensure access to restricted data is authorized and controlled and that there is a process for review of requests for public health data and approval of data use agreements
- Ensure public health data are released when the manuscript is published (or at a minimum, ensure aggregate data or tables are in a machine-readable, open and nonproprietary format and, if feasible, followed later by the complete data set)
- Maintain program public health data portfolios and monitor intramural scientists and awardees for compliance with the policy and implement procedures to address lack of compliance
- Report progress on implementation to OADS

C. Grants Management Officers/Contracting Officers
Will be responsible for the following:

- Ensure project proposals include a DMP
- Ensure the DMP is reviewed during proposal review
- Ensure there is a method that certifies data will be deposited in a repository and final reports are submitted
- Monitor compliance of awardees to this policy and implement procedures normally used to address lack of compliance

D. Extramural Research Program Office (ERPO) Directors and Equivalent (for no research awards) will:

- Work with OFR and CIO project officers to ensure language and requirement of this data policy are included in FOAs, RFAs/RFPs, and NoAs
- Develop standardized review criteria language for evaluating DMPs for quality and completeness
- Ensure DMPs are evaluated prior to funding recommendations or at other specified time depending on award
- Monitor awardees' compliance with this policy using standardized review criteria

E. Intramural and Extramural Investigators
Principal investigators will be responsible for the following:

- Create a DMP when the project is initiated and update as appropriate throughout the life cycle of data
- Prepare public health data for access (including accompanying documentation) and obtain reviews and approvals
- Ensure release of public health data when the manuscript is published (or at a minimum, ensure aggregate data or tables are in a machine-readable, open and nonproprietary format and, if feasible, followed later by the complete data set)

- Respond to requests for restricted public health data sets and ensure responses follow appropriate processes, documentation, and approval
- Other responsibilities as deemed necessary by the CIO to facilitate compliance with this policy

5. REFERENCES

A. Applicable Laws, Rules, Policies, and Guidance

1) CDC/ATSDR Distinguishing Public Health Research and Public Health Nonresearch, CDC-SA-2010-02.

2) CDC/ATSDR Policy on Releasing and Sharing Data.

3) Open Data Policy–Managing Information as an Asset OMB Memorandum M-13-13.

4) Making Open and Machine Readable the New Default for Government Information Executive Order 13642

5) Increasing Access to the Results of Federally Funded Scientific Research,

6) Privacy Act of 1974, 5 USC §552a, as amended.

7) CDC/ATSDR Privacy Act, CDC-GA-2000-01.

8) U.S. Code § 1905 – Disclosure of Confidential Information Generally (also known as Trade Secrets Act).

9) U.S. Code § 242m – General Provisions Respecting Effectiveness, Efficiency, and Quality of Health Services (also known as Assurance of Confidentiality, Public Health Service Act Section 308(d))

10) CDC/ATSDR Oversight and Clearance of Dual-Use Research of Concern, CDC-SM-2007-01.

11) Office of Management and Budget (OMB), Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies.

12) HHS Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated to the Public.

13) CDC/ATSDR Guidelines for Ensuring the Quality of Information Disseminated to the Public.

14) Office of Management and Budget (OMB) Circular A-130, Management of Federal Information Resources.

15) Federal Information Security Management Act (FISMA).

16) CDC/ATSDR Protection of Information Resources, CDC-IS-2002-06.

17) CDC/ATSDR Records Management, CDC-GA-2005-07.

18) AR-25 Release and Sharing of Data.

19) 45 CFR 75. Federal Awarding Agency Regulatory Implementation of Office of management and Budgets Uniform Administrative Requirements, Cost principles, and Audit requirements for Federal Awards

20) CDC Plan for Increasing Access to Scientific Publications and Digital Scientific Data Generated with CDC Funding.

B. Additional Resources

1) Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

2) 45 CFR Part 160.

3) 45 CFR Part 164.

4) Freedom of Information Act (FOIA), 5 USC §552, as amended.

5) Office of Management and Budget (OMB), Guidance on Agency Survey and Statistical Information Collections.

6) Common Rule (45 CFR 46).

7) CDC/ATSDR Human Research Protections, CDC-SA-2010-01

8) Project Open Data Metadata Schema v1.1

9) Federal Records Act of 1950, as amended, 44 U.S.C. Chapter 21, Chapter 29, Chapter 31, Chapter 33

10) National Research Council. Constance F. Citro, Margaret E. Martin, and Miron L. Straf, Editors. Principles and practices of a federal statistical agency 4th ed. Washington: National Academy Press; 2009.

11) Federal Committee on Statistical Methodology. Statistical Policy Working Paper 22 (Second version, 2005) Report on Statistical Disclosure Limitation Methodology. Washington: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical and Science Policy Office.

12) 42 U.S. Code § 241 – Research and Investigations Generally (also known as Certificate of Confidentiality, Public Health Service Act Section 301(d)).

13) National Institutes of Health Data Sharing Policy.

14) National Institutes of Health Genomic Data Sharing Policy.

15) Small Business Innovation Act or Small Business Innovation Development Act of 1982 (P.L. 97-219).

16) 35 U.S. Code Chapter 18 – Patent Rights in Inventions Made with Federal Assistance (also known as the Patent and Trademark Law Amendments Act).

17) 37 CFR Part 401 Rights to Inventions Made by Nonprofit Organizations and Small Business Firms under Government Grants, Contracts, and Cooperative Agreements.

18) The Health Data Initiative.

19) CDC Wonder.

20) Data.cdc.gov.

21) Paperwork Reduction Act.

22) Federal Acquisition Regulations, Subpart 52.227-14, Rights in Data.

23) HHS Open Government Plan.

24) Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).

25) CDC/ATSDR Public Access to CDC Funded Publications, CDC-GA-2013-01.

26) Office of Management and Budget (OMB) OMB Circular A-119, as amended, Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities.

27) Section 508 of the Rehabilitation Act (29 USC § 794d).

28) CDC Guidance on Scientific Integrity.

29) Data Seal of Approval.

30) Trustworthy Repositories Audit & Certification: Criteria and Checklist.

31) NIST Special Publication 800-53 Revision 4 Security and Privacy Controls for Federal Information Systems and Organizations.

6. ACRONYMS

ATSDR	Agency for Toxic Substances and Disease Registry
CDC	Centers for Disease Control and Prevention
CIO	Center, Institute, and Office
DMP	Data Management Plan
FISMA	Federal Information Security Management Act
FOA	Funding Opportunity Announcement
FTE	Full-Time Equivalent
HHS	Department of Health and Human Services
IAA	Interagency Agreement
IRB	Institutional Review Board
MOU	Memorandum of Understanding
NoA	Notice of Award
OADS	Office of Associate Director for Science
OD	Office of the Director
OFR	Office of Financial Resources
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
PII	Personally Identifiable Information
RFA	Request for Acquisition
RFP	Request for Proposal

7. DEFINITIONS

A. Specialized Words and Phrases

CDC staff: CDC employees, fellows, visiting scientists, and others (for example, contractors) involved in designing, collecting, analyzing, reporting, or interpreting data for or on behalf of CDC.

Confidentiality: Protection of information about institutions and/or individuals of research or nonresearch projects that involve the collection or maintenance of sensitive identifiable or potentially identifiable information.

Data accessibility: The capability to make data available to provide the use sought. Data accessibility (release or sharing) to the widest possible audience.

Data custodian: Person in physical possession of data or the system that houses the data.

Data custodians are responsible for the safe custody, transport, storage of the data and implementation of business rules.

Data discoverability: The capability to create, enhance, or provide awareness of data and its attribute.

Data management: The development and execution of architectures, policies, practices and procedures that properly manage the full life cycle of data.

Data management plan: Description of the data being produced in the proposed study, any standards to be used for collected data and metadata, mechanisms for providing access to and sharing of data (including provisions for protection of privacy, confidentiality, security, intellectual property, or other rights), provisions for reuse and redistribution, and plans for archiving and long-term preservation of data (or explaining why long-term preservation and access are not feasible).

Data release: Dissemination of data either for public-use or through an *ad hoc* request so that the data custodian or designee no longer controls the data.

Data set: For the purposes of this policy, the term "data set" refers to a collection of data presented in tabular or non-tabular form.

Data sharing: Granting certain individuals or organizations access to data that cannot be released publicly (for example, data that contain personally identifiable information or potentially identifiable data).

Data standards: Documented agreements on representations, formats, and definitions of common data. Data standards provide a method to codify valid, meaningful, comprehensive, and actionable ways information is captured.

Data steward: The CDC staff responsible for explaining CDC's data policy to staff and users, evaluating and obtaining approval of requests for access to data, and monitoring compliance with CDC policy.

Digital scientific data: Digitally recorded factual material commonly accepted in the scientific community as necessary to validate research findings (includes data sets used to support scholarly publications but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer-review reports, communications with colleagues, or physical objects, such as laboratory specimens).

Machine-readable: Data (or metadata) that is in a format understandable by a computer.

Metadata: Structured information that describes, explains, locates, or otherwise makes retrieving, using, or managing an information resource easier. For any particular datum, the metadata may describe how the datum is represented, ranges of acceptable values, its label, and its relationship to other data. Metadata also may provide other relevant information, such as the responsible steward, associated laws and regulations, and the access and management policy. A list of common core metadata fields required for federal data can be found at https://project-open-data.cio.gov/v1.1/schema/.

Personally identifiable information: Also called PII, this information can be used to distinguish or trace an individual's identity either alone or when combined with other personal or identifying information linked or linkable to a specific individual. The definition of PII is not anchored to any single category of information or technology.

Privacy: Protection of the disclosure of personally identifiable information when records are maintained by a federal agency in a system of records.

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

Public-use data: Data available to anyone.

Restricted data: Data sets that cannot be distributed to the general public because of, for example, participant confidentiality concerns, third-party licensing or use agreements, or national security considerations.

B. Terms

Aggregate data: Data combined from several measurements, observations, or smaller units. When data are aggregated, groups of observations are replaced with summary statistics based on those observations.

Contract: A mechanism to acquire property, goods, or services for the direct benefit or use of the United States government.

Cooperative agreement: A type of funding mechanism used when CDC anticipates substantial involvement, beyond normal oversight and monitoring activities, during the performance period. This involvement may include collaboration or participation by designated CDC staff in specified activities and, as appropriate, at particular points during performance.

Data: The recorded factual material commonly accepted in the scientific community as necessary to validate findings.

Data collection: The process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

Data collection system: System designed for the collection of data. Replaces the traditional paper-based data collection methodology to streamline data collection.

Data dictionary: Information about data, such as meaning, relationships to other data, origin, usage, and format.

Data enclave: A controlled, secure environment in which eligible individuals can perform analyses using restricted data resources.

Data file: A collection of information logically grouped into a single entity and referenced by a unique name, such as a filename.

Data quality: The accuracy and completeness of the data in a database.

Data repository: Commonly refers to a storage location often that ensures security and preservation of data.

Data table: In relational databases and flat file databases, a table is a set of data elements (values) using a model of vertical columns (which are identified by their name) and horizontal rows, the cell being the unit where a row and column intersect.

Embargo: A period during which access to data is not allowed to certain types of users. This is either to protect the revenue of the publisher or (more generally) to protect the interests of other parties (for example, partner organizations).

Extramural: Funded by grants, cooperative agreements, or contracts that are awarded by CDC to outside eligible entities. Outside institutions, non-profit organizations, or governmental agencies, etc. use CDC funding to pay for projects and resources, including the salaries of extramural scientists employed by such institutions; thus, scientists/staff conducting extramural projects are not federal employees.

Entity: An item inside or outside an information and communication technology system, such as a person, an organization, a device, a subsystem, or a group of such items that has recognizably distinct existence.

Genomics: The study of all the DNA (the genome) in an individual (human or microbe), and how parts of the genome interact with each other and the environment.

Grant: A mechanism to provide federal resources in exchange for services that help fulfill CDC's mission and are a direct benefit to the public.

Impartiality: A principle of justice holding that decisions should be based on objective criteria, rather than on the basis of bias, prejudice, or preferring the benefit to one person over another for improper reasons.

Intellectual property rights: Certain exclusive rights granted to owners of creations of the mind, such as discoveries, literature, and technology, determining who can copy, distribute, adapt, use, or profit from it. Common types of intellectual property rights are copyright and patents.

Interoperability: This is the ability for different operating and software systems, applications and services to communicate and exchange data in an accurate, effective, and consistent manner.

Intramural: Supported by CDC and conducted by CDC staff (employees, contractors, visiting scientists, fellows, and students) in its own facilities or its components.

Long-term preservation of data: Formal endeavor to ensure that digital information of continuing value remains accessible and usable. It involves planning, resource allocation, and application of preservation methods and technologies, and it combines policies, strategies, and actions to ensure access to reformatted and "born-digital" content, regardless of the challenges of media failure and technological change.

Memorandum of understanding (MOU): A document that describes the general principles of an agreement between parties, but does not amount to a substantive contract.

Metadata catalog: A collection of descriptions of data sets; each description is a metadata record.

Nonproprietary formats: Format which does not have restrictions on its use and over which no one (for example, a company) claims substantial intellectual property rights' restrictions. Preservation experts recommend using nonproprietary formats (for example, for the long-term preservation of data.

Surveillance: The ongoing systematic collection, analysis, and interpretation of health data, essential to the planning, implementation, and evaluation of public health practice, closely integrated to the dissemination of these data to those who need to know and linked to prevention and control.