U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

Centers for Disease Control
and Prevention (CDC)
Atlanta GA 30329-4027

CDC New Operational Policy

Policy Title: Data Collection and Management

Summary of Policy: To ensure that CDC appropriately collects and manages data in line with its mission and in compliance with applicable statutes, regulations, directives, and guidelines. The goal of the policy is to ensure responsible data collection and stewardship and to advance interoperability through shared data standards for public health both within CDC and with external partners.

Related Issuances: None

Responsible Organization: Office of Public Health Data, Surveillance, and Technology (OPHDST)

Material Superseded: None

Recertification:
This policy is due for recertification on or before the last working day of July 2031.

Points of Contact: Eric Wortman (ltr3@cdc.gov)

To go directly to the policy, use the following URL:

/s/ Sean Slovenski
Chief Operating Officer

NOTE: References to CDC also apply to the Agency for Toxic Substances and Disease Registry (ATSDR)
CS327630-B 10/31/2022

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

Centers for Disease Control
and Prevention (CDC)
Atlanta GA 30329-4027

Category: General Administration
Policy #: CDC-GA-2005-14
Date of Issue: 07/14/2026
Proponent: Office of Public Health Data, Surveillance, and Technology (OPHDST) **Application:** Domestic and International
Applicable Staff: CDC Employees and Non-Employees

DATA COLLECTION AND MANAGEMENT

Sections:

1. PURPOSE AND SCOPE

To prevent disease and promote health, the Centers for Disease Control and Prevention (CDC)[1] collects, generates, stores, uses, and routinely provides data for public health purposes. The purpose of the Data Collection and Management Policy is to ensure that CDC collects and manages data in line with its mission and in compliance with applicable statutes, regulations, directives, and guidelines.[2] The goals of the policy are to ensure responsible data collection and stewardship and to advance interoperability by using shared public health data standards both within CDC and with external partners.

[1] References to CDC also include Agency for Toxic Substances Disease Registry (ATSDR) throughout this document.
[2] This policy helps ensure that CDC is in compliance applicable laws, including the following; Office of Management and Budget (OMB) memorandum titled “Open Data Policy–Managing Information as an Asset” (OMB M-13-13); Executive Order 13642 titled “Making Open and Machine Readable the New Default for Government Information”; the Office of Science and Technology Policy (OSTP) memorandum titled “Ensuring Free, Immediate, and Equitable Access to Federally Funded Research” (OSTP Memo); the Foundations for Evidence Based Policy Making Act of 2018; and Executive Order 13994 titled “Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats”.

This policy applies to all CDC Staff CDC employees[3] and non-employees[4] at all locations, domestic and international, and to CDC's Centers, Institute, and Offices (CIOs) and Business Services Offices, which are hereafter collectively called "CDC programs."[5]

This policy covers data collected or generated for public health purposes, including, but not limited to, science, surveillance, research, and public health practice, except as noted below. Data may be quantitative or qualitative and can be collected or generated in many ways, such as through imaging, paper, audio, electronic transmission, observation, or analysis of samples or specimens.

Data covered by the policy includes intramural and extramural research, and non-research public health data.

Examples of data generally not covered by this policy:

- Data collected and generated solely by other agencies or organizations, and not shared or reported to CDC (data is not within CDC's custody and control), unless otherwise specified
- Data CDC generates or collects primarily for performance monitoring and evaluation (for example, grant recipient progress reports)
- Data CDC generates or collects primarily for financial or administrative purposes
- Preliminary analyses, drafts of scientific papers, plans for future research, communications between colleagues or funding recipients, or physical objects (for example, laboratory notebooks or laboratory specimens)
- Simulated data generated primarily for purposes of testing or developing a tool process, or some other similar mechanism.

NOTE: Depending on underlying agreements between CDC and the external entity generating or collecting the data listed above, this data, once actually shared with CDC may not be subject to this policy but are still subject to applicable federal law for purposes such as use, sharing, and disposition.

This policy applies to all new data collection and generation activities initiated as of this policy's effective date, including data collected or generated based on new grant and contract awards published and issued as of this policy's effective date and all continuations of such awards. *Previously established CDC-operated or - funded data collection systems that continue to collect covered data after this policy becomes effective must be compliant* ***within three years of this policy's effective date,*** *or when the data system undergoes substantial revision—whichever comes first (see Section 3C).*

---

[3] For the purposes of this policy, the term "employees" consists of members of the civil service, Commissioned Corps officers, and locally employed staff. For more information on these categories, refer to "Employee Categories (Updated March 2023)," available at: https://intranet.cdc.gov/ocio/docs/systems-tools/EmployeeCategoryHelp_March-2023.docx.

[4] For the purposes of this policy, the term "non-employees" includes individuals who provide consistent services to CDC, or maintain a regular presence on a CDC facility, or have been issued a physical or logical access credential and are funded by CDC-managed appropriations. As used in this policy, non-employees include groups of individuals such as guest researchers, contractors, Intergovernmental Personnel Act (IPA) personnel, or students. For more information on these categories, refer to "Non-Employee Categories (Updated March 2023)," available at: https://intranet.cdc.gov/ocio/docs/systems-tools/Non-EmployeeCategoryHelp_March2023.docx. http://intranet.cdc.gov/ocio/docs/systems-tools/Non-Employee_Category_Definitions.pdf

[5] More information on CDC organizational nomenclature is available at: https://sbi.cdc.gov/DOA/pdf/orgnom.pdf.

Programs must evaluate data quality during all phases of data generation, collection, transmission, editing, processing, analysis, and storage consistent with CDC's Enterprise Technology and Data Governance (ETDG), and during consultation with the Office of Public Health Data, Surveillance, and Technology (OPHDST), as needed, before making the data available.

2. BACKGROUND

CDC is the nation's principal disease prevention and health promotion organization. In support of its public health mission, CDC collects, generates, stores, uses, and routinely provides public health data. Public health and scientific advancement are best served when public health data are generated, collected, and managed in a timely and appropriate way. In addition, data modernization is essential for improving public health; therefore, this policy update considers technological advancements that aim to maximize data interoperability, accessibility, and utility, to the extent practicable and permitted by federal law.

3. POLICY

A. General Requirements

Data Generation and Collection

CDC generates, collects, and uses valuable and relevant public health data. Data generation and collection can occur through intramural or extramural activities. CDC must comply with all applicable federal laws, current HHS policy, and CDC policy, governance, and other agency requirements applicable to the proposed data activities. Before starting data activities covered by this policy, CDC programs must include the details about the proposed data collection or data generation activity into CDC's electronic clearance system to determine whether any reviews and specific approvals are required (for example, Office of Management and Budget (OMB) approval under the Paperwork Reduction Act (PRA), human subjects research determination). CDC programs cannot begin or fund data collection or data generation without appropriate CDC or other documented approvals, as applicable.

Security and Privacy

As a federal agency, CDC must ensure proper security and privacy safeguards for data it collects or generates throughout the data life cycle. Under federal law, CDC programs must limit the collection and generation of data only to those data collections and generations that are legally authorized, relevant, and reasonably necessary in furtherance of CDC's mission and goals. Similarly, CDC programs should limit the collection of personally identifiable information (PII) and other legally protected data only to the minimum needed for public health purposes, consistent with applicable federal law.

During the project planning phase, CDC programs must work with the Office of Chief Information Officer (OCIO) Cybersecurity Program Office (CSPO), including the appropriate Systems Security and Privacy Officer (SSPO), and the Office of Science (OS), as appropriate, to ensure any planned data collection, generation, storage, and use will comply with applicable privacy and security laws and guidelines, and current CDC policies and practices.

Intramural Data Activities

For intramural data collection and generation, CDC programs must develop a Data Management Plan (DMP) during project planning and submit it with the corresponding project proposals into the CDC's electronic clearance system before starting the data collection or generation activities (see Section 3B). CDC programs must follow all applicable federal laws, and CDC policies and procedures related to the proposed data activities.

Funded Extramural Data Activities

This policy also applies to recipients of CDC funds for research and non-research activities, (referred to as "recipients") through the terms and conditions of their financial awards. Financial awards include grants, cooperative agreements, contracts, and other funding mechanisms entered between CDC and the recipient. External entities must submit a DMP when applying for funding if the project involves data collection or generation covered by this policy (see Section 3B)[6]. Notice of Funding Opportunity (NOFO) announcement templates, Requests for Applications or Proposals (RFAs or RFPs), and solicitation terms and conditions that cover data activities must include language that requires submission of a DMP. The Office of Financial Resources (OFR) will ensure that existing templates contain the appropriate language.

For grants and cooperative agreements, CDC programs must include required template language and incorporate Additional Requirement (AR)-25 in the NOFO to ensure applicants are aware of CDC's DMP requirements. CDC requires applicants to develop and submit a DMP when applying for project funding if the project involves collection or generation of data. Language in the grant or cooperative agreement terms and conditions must include a request for updates on the DMP, including information about where data are stored.

For contracts, interagency agreements (IAAs), and other transaction (OT) agreements, CDC programs must include explicit language in the contract terms requiring the contractual party to develop and submit a DMP if the project involves collection or generation of data. CDC programs must work with the CDC Office of Acquisition Services (OAS) to ensure the appropriate contract terms are included.

Regardless of funding mechanism, CDC programs should encourage sharing and use of federally funded data collections, consistent with applicable federal law and the terms of any underlying funding mechanism. When the project involves data covered by this policy, CDC programs must require use of certain data terminology and content standards in the funding announcement and terms of the funding mechanism, as applicable and consistent with CDC enterprise-wide data quality standards and guidance. Generally, CDC allows the reasonable costs of data sharing to be included in grants, cooperative agreements, and contracts. The cost of archiving and long-term preservation of data collected or generated under the CDC-funded project may also be included as part of the total budget requested for first-time or continuation awards.

[6] If a DMP is not required as components of a funding mechanism, the funding award applicant must include a statement with its funding award application submission that a DMP is not needed because the project will not collect or generate public health data covered by this policy.

All recipients must ensure the quality of the data they make accessible to CDC meets CDC's enterprise-wide data quality standards and guidance. For funding mechanisms that require the sharing or release of data, the final or provisional version of a dataset must be made available as required by the terms of the award or, if the award does not specify timing, recipients must make the data available as soon as possible within the funding period, but no later than 12 months after data collection ends. Data must be made accessible in a nonproprietary format. Unless the dataset is already publicly available, data underlying scientific publication should be made available to the public at the time the paper is published, to the extent permitted by law and subject to valid restrictions. At a minimum, the released dataset should consist of a machine-readable version of the data used to produce the results in the paper.

If the funding award terms do not address submitting data to CDC, the recipient must, at minimum, deposit a de-identified dataset, accompanying data dictionary, and other documentation relevant to use of the dataset in a sustainable repository for archival and long-term preservation consistent with the award terms. Recipients must inform the CDC point-of-contact where the data are stored by updating their DMP.

Recipients who fail to share or release data on time 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 or other authorities as appropriate. Future awards may depend on compliance with this requirement.

Unfunded Data Collection or Generation

CDC may receive public health data from public health organizations, state, tribal, local and territorial (STLT) agencies, and other partners on a voluntary basis, even when CDC does not provide federal funds for data collection or generation activity. Even if unfunded, public health data in CDC's custody and control must be treated as an asset and as such must be properly managed throughout the data life cycle. When possible, this management should begin during the early planning stages before the data is collected or generated. If CDC did not fund the data collection or data generation and does not otherwise have authority over how the external entity manages that data, CDC's responsibility to properly manage the data begins once the data comes under CDC's custody and control. Proper planning for and managing data as early as possible, to the extent practical, helps CDC to strategically design systems and processes that maximize data utility and support downstream information processing and dissemination activities, consistent with applicable federal law, CDC policy, CDC data governance, agreements, and other requirements.

CDC programs collecting unfunded extramural data must comply with applicable provisions of this policy. These include, but are not limited to, obtaining required clearances and approvals before data collection or generation, when possible; developing and submitting DMPs (Section 3B): integrating metadata into the CDC enterprise-wide data catalog (Section 3B); and complying with applicable federal data privacy and security laws and guidelines, as well as CDC data security and privacy policy and practices. CDC may enter into data use agreements (DUAs) or other agreements with entities for unfunded data collection, sharing, or use only to the extent permitted by federal law and CDC policy. CDC programs must ensure these agreements are consistent with CDC data policy and data governance requirements.

B. Data Management

CDC must properly manage data in its custody and control, consistent with applicable federal laws, policies, and agreements. CDC must manage data throughout the data's life cycle in a way that supports the data's intended use.

Data Management Plans

A Data Management Plan (DMP) is a written description of the processes involved for the generation, collection, protection, use, sharing, storage, long-term preservation, and, as applicable, disposition or destruction of public health data. DMPs support the proper management of data at each stage of the data life cycle and help apply industry standard data principles. A DMP is required for all intramural and extramural data collection or generation covered by this policy and must be kept up to date to reflect the status of the data activity and CDC's current data governance requirements. Generally, DMPs contain information that CDC programs and recipients already collect or be held accountable for, so DMPs should not be burdensome to develop and maintain.

CDC programs must use the CDC DMP template to ensure agency-wide standardization and consistency with applicable federal law, policy, and other requirements. CDC project officers or project leads must evaluate the completeness of the DMP to ensure that it includes all required terms, and any additional information required by CDC data governance requirements, in sufficient detail, before submitting to CDC's electronic clearance system. After submission, reviewers will independently evaluate the quality of the DMP and, if needed, will work with the CDC project lead, officer, or assignee to refine the DMP to ensure quality and completeness before a project determination is made. CDC reviewers must evaluate and rate the quality of the DMP during the application, study proposal, and project review process as needed. Reviewers may evaluate the DMP after the completion of the activity on an ongoing basis to ensure continued accuracy and quality.

CDC programs and recipients should not make substantive changes to DMP terms unless needed to keep them accurate and aligned with changes in applicable law, policy, technology, standards, or processes or changes reflective of the data activities. When DMPs are required under a funding mechanism, CDC programs must ensure recipients comply with the DMP terms at the time of award and throughout the award period, including any continuation. CDC programs must maintain and store the DMP in accordance with CDC's current data governance processes.

Data Quality

Data quality is important throughout the data life cycle. Using established data standards to ensure data quality promotes interoperability and makes data more useful for other public health practitioners, researchers, and the public. CDC programs must evaluate data quality throughout the data lifecycle in accordance with CDC's enterprise-wide data quality standards and CDC's enterprise-wide data governance procedures. This helps ensure data meets applicable data standards to support interoperability across CDC data collection and data generation activities. CDC programs may consult with OPHDST (datapolicy@cdc.gov) as needed, to ensure compliance with any CDC enterprise-wide requirements.

CDC Data Management Platform

Data management can be improved by implementing data modernization practices and applying industry standard data principles throughout the data life cycle to support accessibility,

interoperability, utility, and efficiency. CDC programs should ensure that public health data in CDC's custody and control are integrated and maintained in a CDC-designated cloud-based data management platform. CDC programs should use the platform and associated technologies to electronically collect, analyze, and provide data consistent with applicable federal law, CDC policy, CDC data governance, agreements, and other requirements. When a CDC-designated cloud-based data management platform cannot be used to maintain the data and on-premises storage is necessary (for example, a computer hard drive or paper), CDC programs must expressly state this and explain the justification in the DMP during the initial project planning phase and obtain OPHDST approval. If use of a cloud-based data management platform becomes feasible, CDC programs must integrate and maintain the data in a CDC-designated cloud-based data management platform, update the respective DMP accordingly, and maintain any hard drives or paper records consistent with applicable records retention requirements.

CDC's Enterprise-wide Data Catalog

Cataloging metadata helps CDC staff find and access important data more quickly. CDC programs and applicable recipients must identify and provide the metadata for each dataset to the appropriate CDC enterprise-wide data catalog[7]. CDC programs must ensure the metadata for intramural data and applicable extramural data are added to the CDC enterprise-wide data catalog within ninety (90) days of when the data collection is complete. This helps ensure metadata is captured and available in the CDC data catalog dataset for use by data practitioners and data governance. For ongoing data collection or data received or generated over time (for example, public health surveillance), data stewards should start cataloging the dataset as early as possible, consistent with CDC enterprise-wide data catalog guidance.

All data must have documentation that describes how the data were collected, what the data represents, the completeness and accuracy of the data, and potential limitations for use, including information to help prevent misinterpretation. This documentation may be included in or linked to the data metadata record. CDC programs may update this documentation, as necessary, to ensure information remains accurate.

Data Security

CDC must protect its data in a way that reflects the level of risk and potential harm that would result from the loss, misuse, unauthorized access to, modification, or destruction of the data. For datasets that contain protected data, such as personally identifiable information (PII), CDC programs must store data with the appropriate level of security in accordance with applicable federal laws and CDC data governance requirements.

CDC programs must maintain data in a secure environment to ensure privacy, integrity, confidentiality, and data loss prevention. Data maintained in paper, electronic, portable devices, or other formats need to be physically secured when transported and stored. Additionally, the Federal Information Security Management Act (FISMA) requires each agency to provide security for its information and information systems, including those managed on behalf of the agency. CDC's protection of information resources policy establishes the requirements and responsibilities for the protection of the physical, financial, and information resources, and other tangible and intangible assets of CDC.

[7] Data cataloging capabilities are shifting to the 1CDP environment.

Making Data Accessible to Internal CDC programs

Generally, CDC has custody and control over the data it generates or collects as an agency, and responsibilities to use and manage the data appropriately apply agency wide. Data access is a component of data management and is governed by CDC data security policy and practices.

CDC programs should maximize data use and reuse within the agency, consistent with applicable federal law. Unless prohibited by federal law or contract or other legal agreement, CDC programs may make data accessible to authorized staff in other CDC programs, provided those staff meet and comply with certain requirements as established by CDC policy, governance, practices, and procedures. At a minimum, these requirements may include:

- Authorizing access only to the minimum necessary identifiable data, using appropriate privacy protections consistent with applicable federal laws, CDC security policy, and agreements
- CDC staff must not attempt to identify a person in a dataset, unless permitted or required by federal law
- Limiting data access to CDC staff consistent with applicable federal law and based on an expressed and justifiable public health need for the data
- Ensuring access will not compromise or impede public health program activities
- If CDC received the data under a data use agreement or an underlying funding agreement, accessing and using the data consistent with the terms of the agreement, which must align with applicable federal law
- Completing required CDC training (for example, security, ethics, and Assurance of Confidentiality/Certificate of Confidentiality training) and meeting the appropriate level of data skills, as applicable
- Agreeing to collaborate with the CDC Component's data steward to review any proposed publications, derived data, or visualizations based on the data to ensure use is consistent with how the data was obtained and appropriate protection is afforded the data
- Understanding and agreeing to the appropriate CDC data platform's Terms of Use or Rules of Behavior, as applicable
- Adhering to other current CDC security policy and CDC governance requirements, as applicable

Generally, the requirements for internal data access depend on whether the data is:

1. **Restricted**: internal access to microdata (individual-level data) and certain aggregate-data that requires restrictions due to the sensitivity of the data and/or legal requirements and protections that cover the data.
2. **Unrestricted**: internal access to data can be provided to all CDC employees with a legitimate reason to access data.

Additional requirements may apply based on the data access tier-level, permissions applicable to the CDC staff requesting access to data, or other legal that apply to the data.

CDC will set data access permissions based on the level of protection required consistent with applicable federal law. Data access will be authorized in accordance with CDC security policies

and the National Institute of Standards and Technology (NIST) Risk Management Framework[8], including formal System Security Authorization for the transmission systems involved.

Intra-agency DUAs should not be used to make data accessible to other CDC programs, unless expressly required by a legal authority, mechanism, or other circumstance. CDC programs should refer to CDC's DUA Policy for more information.

Data Disposition

Public health data in CDC's custody and control are federal records, and CDC must manage and dispose of the data in accordance with federal retention schedules and other requirements. The DMP must include applicable data retention periods and disposition method, as appropriate. CDC programs must follow applicable federal law, CDC Records Management Policy, and CDC data governance requirements to ensure data are transferred or disposed of in accordance with governing retention periods and procedures. CDC programs may dispose of federal records only in accordance with approved federal records retention schedules and should coordinate with their CDC records management liaison and OPHDST for guidance.

C. Compliance and Evaluation

CDC must properly manage data in its custody and control consistently with applicable federal law, policy, and agreements. Data management must be maintained throughout the data's life cycle and should be aligned with its intended use.

To monitor compliance with this policy, CDC programs must maintain, at minimum:

- A list of investigators/Associate Directors for Science (ADSs)/Associate Directors for Informatics (ADIs) or assignees who oversee compliance
- A list of data assets in CDC's enterprise-wide data catalog
- The reason for any new or old data collection not being included in CDC's enterprise-wide data catalog
- A list of data steward(s) for each respective data asset
- The DMP associated with each new or ongoing data collection or data generation the data steward manages
- If the data has been published, details about what data was published, when, and to whom or where
- Confirmation that CDC data stewards reviewed any proposed publications or visualizations in advance to ensure the data is used as intended and properly protected
- If the data has been shared, details about what data was shared, when, and with whom
- Confirmation that data assets that can be made publicly available are listed at data.cdc.gov

CDC programs may use CDC designated data management systems or other CDC approved technology tools to capture the above information as a record of compliance. OPHDST may require CDC programs to report on their compliance with this policy, as necessary. OPHDST

[8] https://csrc.nist.gov/projects/risk-management/about-rmf

may use data in these data management systems or reports to assess CDC program compliance with this policy.

Previously established CDC-operated or -funded data collection systems that continue to collect covered data and are not in compliance with this policy within three years of its effective date will be subject to corrective action in accordance with OPHDST data governance processes.

For funded extramural data collection, reviewers of grants, cooperative agreements, contracts, IAAs, and OTs must ensure that proposals for CDC funds include costs of sharing data, if any. CDC programs must monitor each funded entity's progress and compliance with the terms and DMPs associated with each respective funding mechanism.

For intramural data collections, CDC program project leads or assignees must submit a DMP with project proposals. Programs should review the DMP during existing project approval processes within their respective CDC programs. CDC programs must review the metadata catalog periodically to track compliance with this policy's requirements, as applicable.

D. Relationship with Other CDC Data Policies

CDC policies address various stages of the data life cycle including data generation/collection, data management and use, data sharing and release, and data disposition. CDC programs must follow the CDC policy and governance requirements that apply to each stage of their data activities, in addition to CDC security and privacy policies.

4. RESPONSIBILITIES

A. Office of Public Health Data, Surveillance, and Technology (OPHDST)

- Manage and revise this policy and related agency-wide data governance and guidance, as needed
- Assist CDC programs with implementing this policy by providing technical assistance and guidance
- Provide training to CDC staff who collect, manage, or access data on the requirements of this policy
- Work with CDC programs and OFR to ensure that NOFOs, RFAs/RFPs, and Notice of Awards (NoAs) contain the requirements for applicants and funding recipients for developing and updating the DMP, storing data in a suitable repository, and making data accessible
- Provide clear reporting requirements for CDC programs to follow, if necessary
- Assess CDC programs' compliance with this policy
- Coordinate collaboration across CDC programs to develop, host, and improve data quality standards
- Provide Enterprise Data Catalog
- Identify best practices for managing data, making data available, and making data easier to find
- Identify suitable tools and repositories within or outside of CDC for management and preservation of data and recommend the use of such to CDC programs

B. Office of Science (OS)

- Collaborate with OPHDST to train data stewards on systems and tools that support the requirements of this policy
- Add processes to clearance tools to help CDC programs comply with CDC data policies
- Consult with OPHDST and CDC Component subject matter experts to provide best practices for improving data quality, management, accessibility, use, and making data easier to find

C. Office of Financial Resources' (OFR) Office of Grant Services (OGS) and Office of Acquisition Services (OAS)

- Develop protocols to ensure submission of a compliant DMP with each application/award, in consultation with OS and CDC programs
- Develop and implement terms for funding mechanisms
- Provide guidance to CDC programs during review of proposals, at the time of issuance of a NoA, and during the submission of progress reports
- Ensure project proposals include a DMP, when a DMP is a required element of an application
- Ensure the DMP is reviewed for completeness and quality during proposal review
- Ensure there is a method that certifies agreement by a funding recipient that data will be deposited in a repository and that final reports are submitted
- Implement procedures that may be used to address lack of compliance with requirements from this policy, as specified in terms of funding mechanisms
- Ensure appropriate data clauses are placed in contracts
- Maintain guidance for programs concerning the collection of data from grant and cooperative agreement recipients and contracts

D. CDC Component Associate Directors of Science, Associate Directors of Informatics, or Assignee

- Identify an oversight mechanism within the Component to help ensure compliance
- Ensure CDC programs are aware of and able to implement this policy
- Ensure that this policy is explained to staff who collect, generate, or access data

E. CDC Component Data Stewards

- Submit information for each applicable data collection to CDC's enterprise-wide data catalog
- Ensure responsible data stewardship practices consistent with this policy and any applicable CDC data governance requirements
- Adhere to CDC data standards and data governance requirements, as applicable, throughout the data lifecycle
- Ensure DMPs are developed for intramural and extramural projects in CDC's electronic clearance system and are reviewed for quality prior to submission to CDC's electronic clearance system
- Determine how long data should be archived to comply with applicable laws, policy, and guidance

- Ensure data to be made accessible are properly prepared before release or sharing the data, including documentation, validation of quality, and review processes

F. CDC Component Data Custodians

- Accept accountability for the operation of a system(s) in support of the overall program mission
- Coordinate with Technical Stewards to ensure the system complies with HHS and CDC security and privacy policies and federal regulations to protect the confidentiality, integrity, and availability (CIA) of systems and information
- Coordinate with the Data Stewards and System Security and Privacy Officers to develop and maintain security and privacy plans, and ensure that the system is operated in accordance with the selected and implemented controls
- Ensure technical processes maintain data integrity and that processes exist to resolve data quality issues
- Ensure access to restricted data is authorized and controlled
- Ensure there is a process for reviewing requests for data that follows data use agreement policies
- Obtain appropriate interconnection security agreements (ISAs) or memoranda of understanding (MOUs), or other similar agreements, as appropriate, prior to connecting with other systems and/or sharing sensitive data
- Coordinate with the Data Steward and the System Security and Privacy Officer to determine user access requirements, including who should be given access and what level of access they need, and grant users only the minimum access needed to perform their jobs based on a legitimate need to satisfy mission, business, or operational needs
- If there is a data breach or unauthorized disclosure of data from a CDC controlled platform, ensure that formal notice is provided to CDC's Cybersecurity Program Office (CSPO) as soon as the breach or unauthorized disclosure is discovered

G. CDC Component Project Officers, Project Leads, or Assignees

- Ensure DMPs are developed for intramural and extramural projects and are reviewed for quality prior to submission to CDC's electronic clearance system
- Ensure data are stored in suitable and sustainable repositories, as appropriate and consistent with applicable records retention laws and policies
- Ensure that a metadata record is submitted to the CDC enterprise-wide data catalog and, for publicly available datasets, to a public catalog dataset
- Monitor intramural scientists and recipients for compliance with the policy and to address lack of compliance
- Report progress on implementation to OPHDS, if required

5. REFERENCES

A. Pub. Law 115-435 – Foundations for Evidence-Based Policymaking Act of 2018
B. Open, Public, Electronic, and Necessary (OPEN) Government Data Act, H.R. 1700 (2017), https://www.congress.gov/bill/115th-congress/house-bill/1770
C. 5 U.S.C. § 552a (1995) – Records Maintained on Individuals (Privacy Act)
D. Presidential and Federal Records Act Amendments of 2014, Pub. L. No. 113-187, 128 STAT 2003 (2014), https://www.govinfo.gov/content/pkg/PLAW-113publ187/pdf/PLAW-

113publ187.pdf
E. 18 U.S.C. § 1905 (2024) – Disclosure of Confidential Information Generally
F. 42 U.S.C. § 242(m) (2024) – General Provisions Respecting Effectiveness, Efficiency, and Quality of Health Services
G. 42 U.S.C. § 241 (2016) – Research and Investigations Generally
H. 5 U.S.C. § 552 (2013) – Freedom of Information Act (FOIA)
I. 44 U.S.C. § 3501 (2001) – Paperwork Reduction Act
J. The Health Insurance Portability and Accountability Act of 1996, Pub. L. 104-191 (1996) – https://uscode.house.gov/statutes/pl/104/191.pdf https://www.congress.gov/bill/104th-congress/house-bill/3103/text
K. 45 CFR § 46 (2018) – Common Rule
L. M2 CFR Part 200, Uniform Administrative Requirements, Cost Principles, and Audit Requirements For Federal Awards
M. 45 CFR § 160 – General Administrative Requirements
N. 45 CFR § 164 – Security and Privacy
O. 48 C.F.R. 52 § 52.227-14 (2014) - Rights in Data — General Rights in Data — General
P. NIST. "Risk Management Framework," dated September 24, 2024, https://csrc.nist.gov/projects/risk-management/fisma-background
Q. OMB. Circular A-130: *Managing Information as a Strategic Resource*, dated July 28, 2016, https://www.federalregister.gov/documents/2016/07/28/2016-17872/revision-of-omb-circular-no-a-130-managing-information-as-a-strategic-resource
R. OMB. "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies," dated January 3, 2002,
S. OMB. "Statistical Programs and Standards," https://www.federalregister.gov/documents/2002/02/22/R2-59/guidelines-for-ensuring-and-maximizing-the-quality-objectivity-utility-and-integrity-of-information
T.
U. OMB. "Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)," dated October 2006, https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/omb/inforeg/proposed_cispea_guidance.pdf
V. OMB. Circular A-119: *Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities*, dated January 27, 2016, https://www.whitehouse.gov/wp-content/uploads/2020/07/revised_circular_a-119_as_of_1_22.pdf
W. 36 CFR Part 1194 § 508 (2018) − Section 508 of the Rehabilitation Act
X. Presidential Memorandum for the Heads of Executive Departments and Agencies: *Open Data Policy–Managing Information as an Asset*, May 9, 2013, https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf
Y. Executive Order 13642 (2013) – Making Open and Machine Readable the New Default for Government Information, May 9, 2013, https://www.federalregister.gov/documents/2013/05/14/2013-11533/making-open-and-machine-readable-the-new-default-for-government-information
Z. Presidential Memorandum for the Heads of Executive Departments and Agencies, "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research," dated August 25, 2022, https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf
AA. Presidential Memorandum for the Heads of Executive Departments and Agencies, "Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Government Data Access and Management Guidance," dated January 15, 2025, https://www.whitehouse.gov/wp-content/uploads/2025/01/M-25-05-Phase-2-Implementation-

of-the-Foundations-for-Evidence-Based-Policymaking-Act-of-2018-Open-Government-Data-Access-and-Management-Guidance.pdf
BB. HHS. "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated to the Public," dated October 1, 2002, https://aspe.hhs.gov/hhs-guidelines-ensuring-maximizing-disseminated-information.
CC. HHS. "Open Government Plan. "Open Government Plan," dated December 11, 2024, https://aspe.hhs.gov/hhs-guidelines-ensuring-maximizing-disseminated-information
DD. CDC-GA-2025-02: *Data Use Agreements*, dated March 30, 2026, https://cdc.sharepoint.com/sites/SBI/CDCOperationalDocuments/Forms/AllItems.aspx?id=%2Fsites%2FSBI%2FCDCOperationalDocuments%2FCDC%2DGA%2D2025%2D02%2Epdf&parent=%2Fsites%2FSBI%2FCDCOperationalDocuments
EE. CDC.CDC-SA-2010-02: *Distinguishing Public Health Research and Public Health Nonresearch*, dated March 10, 2023, https://cdc.sharepoint.com/:b:/r/sites/SBI/CDCOperationalDocuments/CDC-SA-2010-02.pdf
FF. CDC. CDC-SA-2010-01: *Protection of Human Subjects in Research and Clinical Investigations*, dated October 15, 2021, https://cdc.sharepoint.com/:b:/r/sites/SBI/CDCOperationalDocuments/CDC-SA-2010-01.pdf
GG. CDC. CDC-GA-2000-01, *Privacy Act*, dated January 26, 2022, https://cdc.sharepoint.com/:b:/r/sites/SBI/CDCOperationalDocuments/CDC-GA-2000-01.pdf
HH. CDC. CDC-IS-2002-06: *Protection of Information Resources*, dated April 29, 2016, https://cdc.sharepoint.com/:b:/r/sites/SBI/CDCOperationalDocuments/CDC-IS-2002-06.pdf
II. CDC. CDC-GA-2005-07: *Records Management Policy*, September 14, 2021, https://cdc.sharepoint.com/:b:/r/sites/SBI/CDCOperationalDocuments/CDC-GA-2005-07.pdf
JJ. CDC. CDC-SM-2007-01. *Oversight and Clearance of Dual Use Research and Dual Use Research of Concern*, November 13, 2019, https://cdc.sharepoint.com/:b:/r/sites/SBI/CDCOperationalDocuments/CDC-SM-2007-01.pdf
KK. CDC-GA-2013-01: *Public Access to CDC Funded Publications*, April 8, 2024, https://cdc.sharepoint.com/:b:/r/sites/SBI/CDCOperationalDocuments/CDC-GA-2013-01.pdf
LL. CDC. Additional Requirement 25, "Data Management and Access," https://www.cdc.gov/grants/additional-requirements/ar-25.html
MM. Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. "The Fair Guiding Principles for Scientific Data Management and Stewardship." Scientific Data 3, no. 1 (March 15, 2016). https://dash.harvard.edu/entities/publication/73120379-2cc8-6bd4-e053-0100007fdf3b
NN. Project Open Data Metadata Schema v1.1, dated November 6, 2014, https://resources.data.gov/resources/dcat-us/

6. ACRONYMS AND ABBREVIATIONS

CDC − Centers for Disease Control and Prevention
CIO − Centers, Institute, and Offices
DMP − Data Management Plan
HHS− Department of Health and Human Services
IAA − Interagency Agreement
MOU − Memorandum of Understanding
NoA − Notice of Award
NOFO − Notice of Funding Opportunity
OFR − Office of Financial Resources
OMB − Office of Management and Budget
OPHDST − Office of Public Health Data, Surveillance, and Technology
OS − Office of Science

OSTP − Office of Science and Technology Policy
PII − Personally Identifiable Information
RFA − Request for Application
RFP − Request for Proposal

7. DEFINITIONS

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.

Authorized User − A person who has been granted access to data or an information system to carry out official, authorized government duties. Authorized Users will generally be employees, contractors, and other agents specified by CDC that are engaged in public health efforts consistent with CDC authorities.

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 − The obligation to protect sensitive or identifiable/potentially identifiable information, including personally identifiable information (PII) and protected health information (PHI), from unauthorized access, use, or disclosure.

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

Cooperative Agreement − A financial assistance mechanism between an awarding agency (CDC) and a recipient in furtherance of carrying out a public purpose but which anticipates substantial involvement of the awarding agency (CDC), beyond normal oversight and monitoring activities, during the performance period.

Data − Recorded information, regardless of form or the media on which it is recorded.

Data Access − For purposes of this policy, making data available to authorized users within and between CDC programs, to the extent consistent with applicable federal law, policies, and agreements.

Data Accessibility −Data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched.

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 Custodian − The designated individual in physical possession of the data or authorized to manage the technical system that houses the data.

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

Data Disposition − Actions taken regarding data at the end of the data's life cycle and after its legal retention period has been met, as applicable.

Data Life Cycle – A series of stages that data progresses through. The data life cycle stages include data generation/collection, data orchestration, data storage, data cleansing and integration, data analysis and visualization, data dissemination, and data disposition.

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 Quality − Degree to which data is accurate, complete, timely, and consistent with all requirements and business rules, and relevant and appropriate for a specific use.

Data Release − Process for making data publicly available for use and analysis. Dissemination of data either for public-use or through an *ad hoc* request so that the data custodian or assignee no longer controls the data.

Data Repository − Commonly refers to a storage location that ensures security and preservation of data.

Dataset − A data collection of separate sets of information that a computer treats as a single unit

Data Sharing − When an organization sends data with restrictions imposed (for example, data use agreement or under other controlled conditions) due to legal constraints or because releasing the data may potentially disclose proprietary or confidential information or compromise national security or law enforcement interests.

Data Standards − An agreed upon set of rules, requirements, or characteristics for governing architectural components of data. Data standards define the approach and practices for data representation, access, and distribution, and provide a common language to promote efficiency and comparability

Data Steward −. Point of contact who can identify the appropriate individuals to populate the metadata record, respond to scientific questions about the data, coordinate access requests, and oversee data management processes and procedures (for example, program staff, technical steward, security steward, data engineers, data SMEs).

De-identified Data − Data where personally identifiable information has been removed or obscured in such a way that an individual cannot be identified, and there is no reasonable basis that the data could be used to identify an individual.

External Entity – Any organization or group that is not part of CDC, such as a private industry, academic institution, international entity, other Federal government agency, or state agency.

Extramural Data[9] **−** Data that are collected or generated by external agencies or organizations, either through funded or unfunded mechanisms such as grants, cooperative agreements, contracts, memoranda of understanding, or data use agreements.

Grant − A financial assistance mechanism between an awarding agency (CDC) and a recipient in furtherance of carrying out a public purpose.

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 Data – Data collected or generated solely by CDC staff (employees, contractors, visiting scientists, fellows, and students) in its own facilities or its components.

Long-term Preservation of Data[10] **−** 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

Machine-readable − Data (or metadata) is in a format understandable by a computer without human intervention while ensuring no sematic meaning is lost.

Memorandum of Understanding (MOU) – A non-binding 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 datasets; each description is a metadata record.

Metadata – Structured or descriptive information about data such as content, format, source, rights, accuracy, provenance, frequency, periodicity, granularity, publisher or responsible party, contact information, method of collection, and other descriptions that makes retrieving, using, or managing an information resource easier.

Nonproprietary Formats[11] **−** Format which does not have restrictions on its use and over which no one (for example, a company) claims substantial intellectual property rights' restrictions.

Personally Identifiable Information[12] **−** 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.

---

[9] Extramural Data is collected or generated by other agencies or organizations where the data collection or data generation is funded or co-funded by CDC, through mechanisms such as grants, cooperative agreements, contracts, or other funding mechanisms; or collected or generated by other agencies or organizations not using CDC funds to directly support the data collection or generation, but the data is shared or released to CDC and comes under CDC's custody and control (for example, STLT health department voluntarily providing public health surveillance data, research data, and non-research data).

[10] 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.

[11] Preservation experts recommend using nonproprietary formats (for example, for the long-term preservation of data

[12] The definition of PII is not anchored to any single category of information or technology.

Privacy – An individual's right to control the acquisition, uses, or disclosures of his or her identifiable data.

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

Surveillance − The ongoing systematic collection, analysis, and interpretation of data tracked over time to detect patterns and changes, 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 of disease.