

CDC'S NATIONAL ENVIRONMENTAL PUBLIC HEALTH TRACKING PROGRAM

TECHNICAL NETWORK IMPLEMENTATION PLAN

National Center for Environmental Health
Division of Environmental Hazards and Health Effects



TNIP 2010

National Tracking Network Implementation Plan 2010

Environmental Public Health Tracking Network

CONTENTS

- 1. Introduction.....4
 - 1.1. Overview of the Technical Infrastructure of the Tracking Network and State/local Tracking Networks 5
 - 1.2. Overview of the Content of Tracking Network.....5
 - 1.3. Document Organization.....7
- 2. Nationwide Tracking Network Technical Infrastructure9
 - 2.1. Portals.....9
 - 2.2. Transport Protocols (Gateways)17
 - 2.3. Data Repositories.....18
- 3. National Tracking Network Content.....19
 - 3.1. Nationally Consistent Data and Measures (NCDM)19
 - 3.2. Other Tracking Network Data22
 - 3.3. Metadata23
 - 3.4. Tracking Network Tools24
 - 3.5. Messages25
- 4. National Tracking Network Processes27
 - 4.1. Technical Infrastructure Implementation.....28
 - 4.2. Tracking Data Identification and Organization28
 - 4.3. Metadata Creation, Submission, and Validation29
 - 4.4. Data File Creation30
 - 4.5. Data Submission and Validation31
 - 4.6. Data Processing (Tracking Program)32
 - 4.7. Data Publishing on Portals.....33
- 5. Appendix A: Overview of National Tracking Network Workgroups35
- 6. Appendix B: Organization of National Tracking Network Workgroups37
 - Standards and Network Development Workgroup (SND)37
 - Content Workgroup (CWG)38
 - Program Marketing and Outreach Workgroup (PMO)39
 - Geospatial Workgroup.....40
- 7. Appendix C: Tracking Network Required Metadata Fields.....42
- 8. Appendix D: Data Transport Processes46
 - Tracking Transport Overview: How to send Nationally Consistent Data and Measures to the National Tracking Portal46
- 9. Appendix E: Glossary of Terms51

1. INTRODUCTION

Environmental Public Health Tracking is a concerted effort on the part of many people, agencies, and organizations to further access to environmental, exposure, and health data. This National Tracking Network Implementation Plan 2010 (TNIP 2010) developed by the Centers for Disease Control and Prevention's (CDC) Environmental Public Health Tracking Program (Tracking Program) serves as documentation to support the development and evolution of the National Tracking Network (Tracking Network) (www.cdc.gov/ephtracking). It provides necessary information for participating in and contributing to the evolution of the Tracking Network. It includes descriptions of the technical and content components of the Tracking Network. It also provides the tools and resources that an organization needs to participate in the Tracking Network. This TNIP 2010 document is intended for use by

- **Network Architects**—who conceive the evolving architecture and functioning of the Tracking Network.
- **Network Builders**—who require a framework within which to put in place the architects' vision, Network Builders include programmers who develop system specifications and write code.
- **Network Contributors**—who make content available on the Tracking Network. Network Contributors may include grantees, national partners, data owners, and data stewards.

The TNIP 2010 is not intended to be a guide to how to use the Tracking Network. It is also not intended to document all the activities that have gone in to building the Tracking Network. Other documents provide details on the origins, vision, and goals of the Tracking Program and the strategies for and use of the Tracking Network. Such documents can be found at <http://www.cdc.gov/nceh/tracking/publications.htm>. They include the National Network Implementation Plan (NNIP) and the previous version of the Technical Network Implementation Plan (TNIP) v 1.1.

Terminology

National Tracking Network (Tracking Network): the National components of the Tracking Network developed by CDC

State/local tracking networks: the portals developed by funded grantees

Nationwide Tracking Network: combination of the above

1.1. OVERVIEW OF THE TECHNICAL INFRASTRUCTURE OF THE TRACKING NETWORK AND STATE/LOCAL TRACKING NETWORKS

The Tracking Network and state/local tracking networks are made up of a collection of Web-based portals, gateways, and repositories. These components exist in various forms. The forms are based on different technologies available at the Tracking Program and at state/local (grantee) sites. Such forms are dependent on site-specific information technology policies and practices.

Both the national and the state/local public portals are developed with the various kinds of users in mind.. The portals provide the public with the ability to access Nationally Consistent Data and Measures (NCDM) (see the [NCDM Guidance Document](#)), metadata, other data of interest, and tools with which to visualize and analyze the publicly accessible datasets (see next section and Section 3). National and state/local secure portals provide access to resources that are more detailed (spatially and/or temporally). The secure portals may have confidentiality constraints, or they may be only working prototypes for eventual public access. For security purposes, access to the secure portals requires the user to complete a registration process and is generally intended for scientists and researchers.

Gateways are the means to transport data securely between the National Tracking Network and state/local tracking networks. The gateways are used by the Tracking Program, grantees, and partners who have installed the required gateway software made available by CDC. The Tracking Program maintains data repositories by which to store data (and eventually tools and other resources) that can be accessed via the gateways and made available on the state/local portals.

1.2. OVERVIEW OF THE CONTENT OF TRACKING NETWORK

TNIP 2010

The content on the Tracking Network, showing health, exposure, and environmental conditions, is derived from a variety of sources. The Content Workgroup (CWG), relying on teams to conduct its work, is responsible for identifying and making recommendations regarding potential NCDM. The NCDM recommendations are forwarded to the Tracking Program for consideration. In some cases, the Tracking Program works with national partners (e.g., the United States Environmental Protection Agency) to identify relevant national level datasets that could become content for the Tracking Network. Grantees may also identify and provide specific datasets, such as hospitalizations, that become part of the Tracking Network. In some instances, grantees identify content areas they would like to make accessible on their portals to users of the state/local tracking networks within their jurisdictions. Content on the National Tracking Network includes the following:

- Data about health outcomes, environmental hazards, and exposures. These are the NCDM. Each of these content areas includes public health messages.
- Additional data and tools such as demographics, modeling results, and visualization tools.
- Metadata records for datasets uploaded to the Tracking Network. These records provide a means for data discovery. They describe the details of dataset creation prior to the dataset's submission to the Tracking Network. The records also identify unique characteristics for a given dataset, describe use or data availability constraints, provide data owner contact information, and assist users with understanding and using the data.
- Various tools available on the national secure portal. These tools may include externally developed software, such as spatial analysis and data linkage tools, and internally developed software, such as the Metadata Creation Tool.

NCDM Categories on the Tracking Network

- Air Quality
- Asthma
- Birth Defects
- Cancer
- Carbon Monoxide Poisoning
- Childhood Lead Poisoning and Age of Housing
- Community Drinking Water
- Myocardial Infarctions (Heart Attacks)
- Vital Statistics/Reproductive Health Outcomes

Examples of Other Data on the Tracking Network

- Population Characteristics
- Well Water Testing Results

1.3. DOCUMENT ORGANIZATION

The components of the Tracking Network are described in the next sections. Section 2 describes in detail the technical infrastructure—the basic framework of the system. Section 3 provides more details on content. Section 4 depicts in detail the processes and resources available by which users can contribute data to the Tracking Network. Throughout the TNIP 2010, sidebars call out details of grantee activities, organizational structures, and best practices. The sidebars provide background, context, and examples that aid in understanding the Tracking Network and its relationship to state/local tracking networks. The colors shown below are used throughout the remainder of this document to code different sections and components of the Tracking Network.

Section 2:
Nationwide
Tracking Network
Technical
Infrastructure

- Portals
- Transport protocols (gateways)
- Repositories

Section 3:
National Tracking
Network Content

- NCDM
- OtherTracking Network data
- Metadata
- Tracking Network tools
- Messages

Section 4:
National Tracking
Network
Processes

- Prepare infrastructure for submission
- Identify and organize needed tracking data
- Create and submit metadata and receive validation
- Create data files
- Submit data and receive validation
- Process data (Tracking Program)
- Publish data on portals

2. NATIONWIDE TRACKING NETWORK TECHNICAL INFRASTRUCTURE

Since 2003, the Nationwide Tracking Network has developed, primarily as a result of technology changes and increasing knowledge and expertise within the Tracking Program and the grantee community. The fundamental infrastructure, however, has not changed significantly from the original conceptions discussed in 2003–2004.

Figure 1 depicts the relationships among the basic components of the Nationwide Tracking Network. These basic components and their relationships are discussed in detail in the following sections.

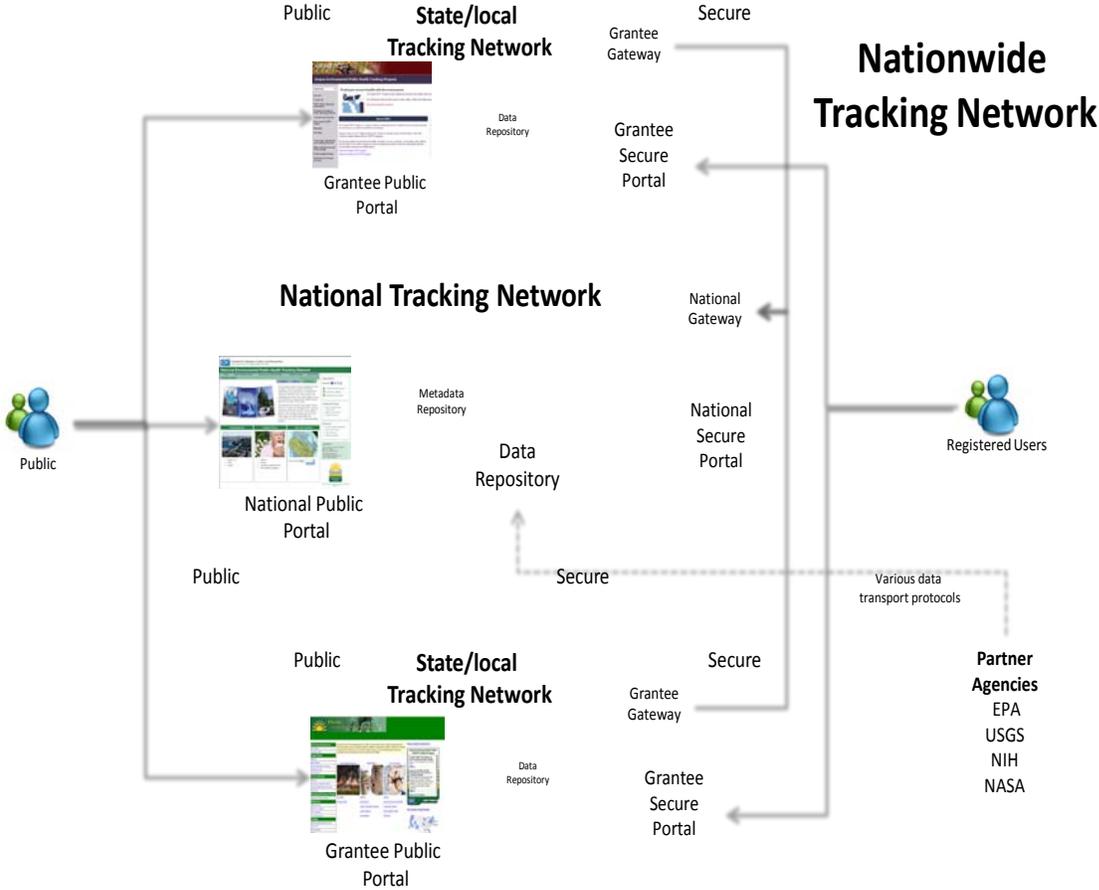


Figure 1: Nationwide Tracking Network Technical Infrastructure

2.1.1. PORTALS

The Nationwide Tracking Network contains a national portal maintained by the Tracking Program and a series of individual portals maintained by Tracking Program-funded grantees. There are two types of portals: public and secure. Public portals are fully accessible on the Internet. They provide the primary means of access for most users to view general information and non-sensitive data on the Nationwide Tracking Network. Secure portals, in contrast, require that a user be pre-authorized to use the network before that user can access more detailed data available. This distinction is clarified in the following sections.

2.1.1. PUBLIC PORTALS

Access to the Nationwide Tracking Network for most users is via the public portals. All information made available to the public, including Nationally Consistent Data and Measures (NCDM), messages, and metadata, are available via the public portals. A public portal requires only an Internet connection and a Web browser.

National Public Portal

The national public portal (Figure 2) provides users the ability to do the following:

- Access public data, NCDM, and other compiled data.
- Search and view metadata (data that describe and define other data) for Tracking Network data assets (NCDM and other);
- Use metadata as a pointer to find information on regulations and procedures for accessing secure data.
- Browse and view other relevant Tracking Network data sources, including EPA environmental data, the Census, WONDER, and other CDC data;
- Perform analysis and visualization functions on the NCDM and other data sources, including
 - Execution of pre-configured and user-defined queries on selected data
 - Generation of descriptive statistics, graphing, and mapping.
- Browse and use relevant categorized links to other information sites, including the following:

- Grantee and partner portals (e.g., EPA) and Websites for other data and other information
- Other CDC Websites
- Access additional, descriptive content designed to assist in interpretation of NCDM and other data, such as
 - Fact sheets and other public health information materials
 - FAQ documents
 - Online reference information on selected topics

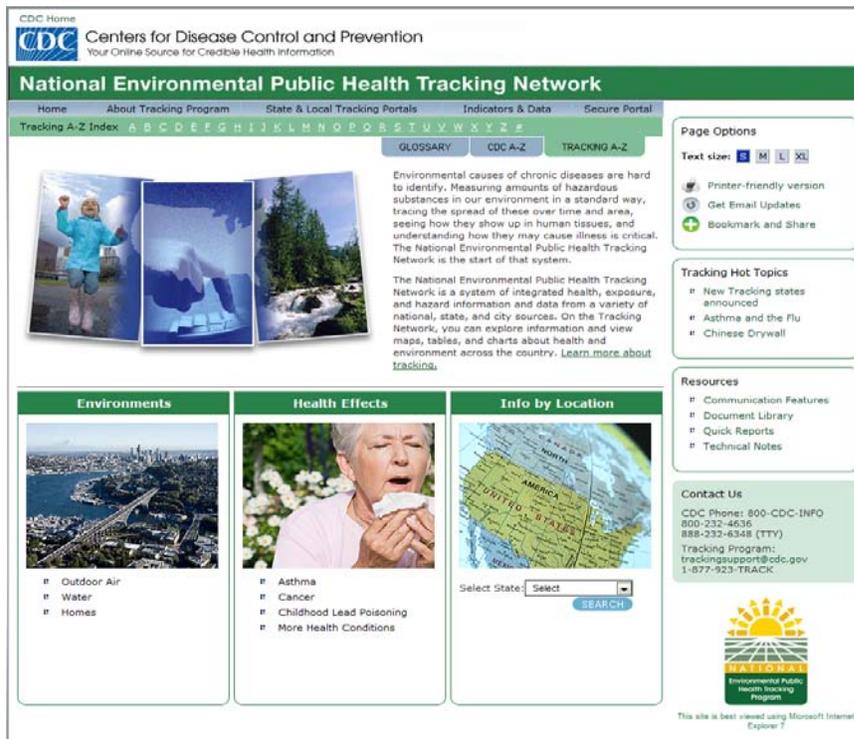


Figure 2: Tracking Network National Public Portal

Grantee Public Portals

Grantee public portals serve state or local users. These portals make available environmental health data that may be of specific interest within grantee communities. Grantee portals use a variety of technologies and interfaces to make data accessible, but they follow standards established by the Tracking Program, including the following:

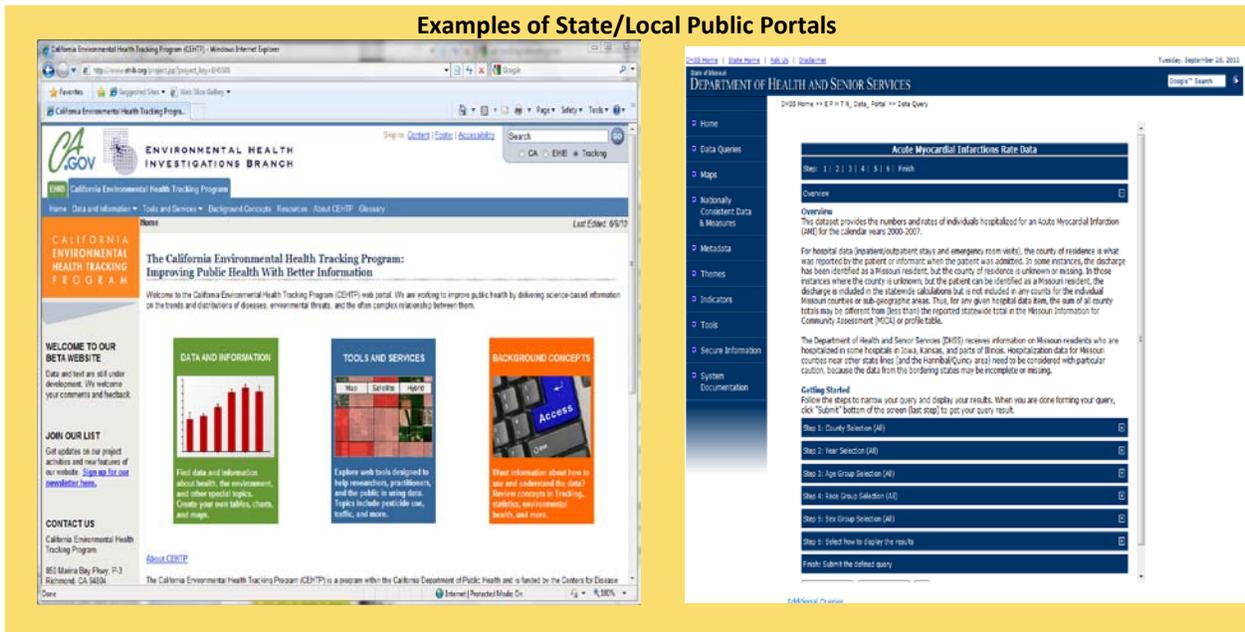
- Standardized Portals
 - Provide a link to the National Tracking Network public portal from individual grantee public portals.
 - Provide access to environmental, exposure, and health effect data.
 - Comply with Section 508 standards (a requirement that electronic and information technology developed by or purchased by the Federal Agencies be accessible by people with disabilities).
- Offer analysis, visualization, and reporting
 - Maintain visual consistency among pages on grantee portals.
 - Support the means to generate visual displays of data, including maps, charts, and tables.
 - Provide the ability to display both counts and rates on a map.
 - Use standard color pallets from proven research that shows the effectiveness of the use of colors
 - Provide the capability to combine multiple data sources in a single graph, table, or chart.
 - Ensure protection for people included in the data as necessary through offering small-number protection—using data aggregation, smoothing, etc.
 - Provide users with the ability to interact with data tools and displays.
 - Provide users with the ability to see alternate views of the information.
 - Ensure that analysis functionality provides for “minimally aggregated” data, as in counts and rates.
 - Provide users with the capability to execute queries to obtain data.

State/local Grantees

California
 Colorado
 Connecticut
 Florida
 Iowa
 Kansas
 Louisiana
 Massachusetts
 Maryland
 Maine
 Minnesota
 Missouri
 New Hampshire
 New Jersey
 New Mexico
 New York
 New York City
 Oregon
 Pennsylvania
 South Carolina
 Utah
 Vermont
 Washington
 Wisconsin

- Metadata
 - Allow for the discovery of data by searchable metadata by use of the standardized metadata template for all data available on the grantee portal.
- Data Content and Discovery
 - Provide data results and information on grantee public portals within three clicks of the search initiation.
 - Provide a trail by which users can return to the starting or entry point.
 - Allow for export of appropriate data to a common format, such as comma delimited.
 - Provide a structured flow that links related information and data sources.
 - Organize contents of pages to make it easy for users to find detailed information.
 - Provide the ability for users to combine and display multiple data query results.
 - Establish categories to make it easy for users to do data browsing and discovery.
 - Provide clear labeling for categorical information.
 - Provide a set of Frequently Asked Questions (FAQ).
 - Provide clear definitions of terms.
 - Provide on-line documentation.
 - Provide a phone number or an e-mail address that users may use to get additional information or clarification.

These standards are intentionally broad-based. They are developed in this way because grantee public portals are also required to comply with standards that are specific to their agencies or jurisdictions. Grantees have used a variety of portal platforms and technologies in the development of their public portals. As a result, look, feel, and



functionality vary across portals. However, grantee public portals are all developed in such a way that the national public portal currently links to them (<http://ephtracking.cdc.gov/showStateTracking.action>).

2.1.2. SECURE PORTALS

National Secure Portal

The national secure portal is a component of the Tracking Network. For users who require more detailed data and have registered for access, the national secure portal is another primary means to access the Tracking

CDC Secure Access Management Services (SAMS)

Only registered users can gain access to the national secure portal. The registration process ensures that individuals requesting access have an appropriate professional need to use the data and resources within the secure portal. Users who are approved will receive a username and password to access the system. The SAMS portal is a Website designed to provide centralized access to public health information and computer applications operated by the CDC.

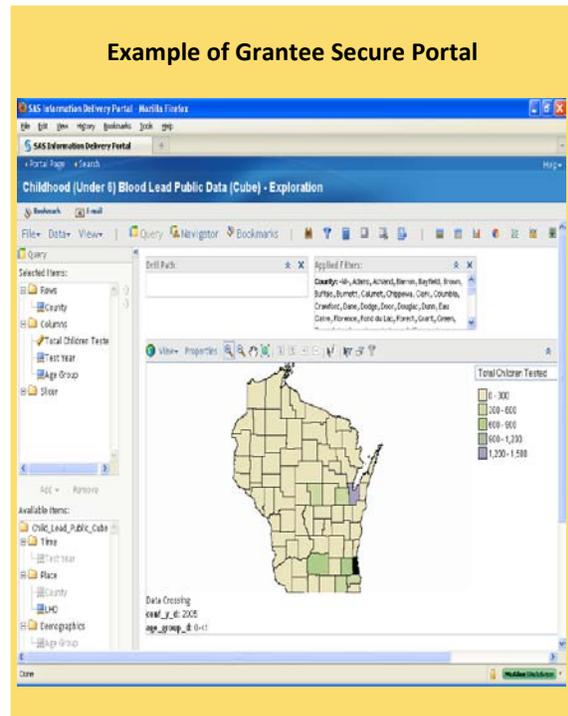
Network. The national secure portal is designed by and will support environmental public health practitioners and researchers. It complements and uses the system of state/local portals, gateways, clients, and partner interfaces. The primary function of the national secure portal is to directly support access to NCDM and other data. It also provides a secure, online workspace for grantees, partners, researchers, and other public health practitioners. The national secure portal consists of a secure Website that can be accessed only by registered users. In addition, it provides an ongoing process for the identification, organization, and hosting of relevant Tracking Network information, data, and tools. Using a Web browser and security credentials, registered users are able (or in the future will be able) to access

- Compiled NCDM stored in the national repositories
- Data discovery and description services (metadata)
- Restricted access data (with appropriate permissions)
- Robust online analysis, visualization, and reporting tools
- A repository of analytical tools and methods
- Collaboration and knowledge management functions (document sharing, discussion, contact management), including
 - Support for Tracking Network Workgroups, and
 - Online reference information for selected topics (NCDM and others).

Grantee Secure Portals

Grantees are in the process of developing secure portals. These portals will have the ability to provide access to data for registered users and partners. Some are already functioning. As is the case with the state/local public portals, the secure portal technology is based on technology requirements in place at state/local sites. The Tracking Program has provided the following standards for development of grantee secure portals:

- Network
 - Comply with Section 508 standards.
 - Provide access to environmental, exposure, and health effect data.
- Analysis Visualization and Reporting
 - Support the means to generate, visual displays of data, including maps, charts, and tables.
 - Provide the ability to display both counts and rates on a map.
 - Use standard color pallets from proven research that shows the effectiveness of the use of colors
 - Provide the capability to combine multiple appropriate data sources in a single graph, table, or chart.
 - Ensure that analysis abilities provide for minimally aggregated data, as in counts and rates.
- Data Content and Discovery
 - Provide the ability for users to make queries to obtain data
 - Provide a set of Frequently Asked Questions (FAQ).
 - Provide clear definitions of terms.



- Provide on-line documentation.
- Security
 - Provide for various levels of data access depending on the user's role on the secure portal.
 - On secure portals, ensure that access to data is secure.
 - Provide a secure login by which users can be verified by the system.
 - Provide access to varying levels of data and services according to the user's authorized role.

These standards will change as the Tracking Program collects feedback from users of the portals and identifies typical uses and the functions desired by users.

2.2. TRANSPORT PROTOCOLS (GATEWAYS)

Gateways provide the means for Nationwide Tracking Network participants to exchange data with the Tracking Program. They are used to transport

- Nationally Consistent Data and Measures (NCDM); and
- Other data to be published or used by the Tracking Network.

A gateway is not a single technology. Rather, it is a number of ways by which people can exchange data securely. Grantees have developed gateways by using the prescribed technology standards to exchange data with the national gateway. Such gateways are used for machine-to-machine transactions, where a server at one site pulls or pushes data from/to another. Currently, the Tracking Network supports two data transport conventions for gateways:

- Public Health Information Network Messaging System (PHIN MS)
- Secure Data Network Secure File Upload (SDN SFU)¹

Public Health Information Network Messaging System (PHIN MS)

¹ SDN SFU will be replaced by the newer Secure Access Management System (SAMS)

The CDC Public Health Information Network (PHIN) is a national initiative. Its purpose is to improve the ability of public health agencies and officials to use and exchange information electronically. It does so by promoting the use of standards and by defining functional and technical requirements for information exchange. The PHIN Messaging System (PHIN MS) transport protocol is a generic, standards-based, interoperable, and extensible message transport system. PHIN MS is platform-independent. It is loosely coupled with systems that produce outgoing messages or consume incoming messages. PHIN MS is ideally suited for frequent periodic exchanges of data (e.g., monthly exchanges of structured pre-defined data), but it is not ideal for one-time exchanges of small quantities of data.

Secure Data Network Secure File Upload (SDN SFU)

SDN SFU provides a convenient, Web-based system used to send data to the Tracking Program. Grantees who are unable to install PHIN MS at their agencies can use SDN SFU. For the present, users can continue to use Secure File Upload to send data to the Tracking Program. However, the SDN security layer will be replaced by the Secure Access Management System (SAMS) over time.

2.3. DATA REPOSITORIES

The Tracking Program and grantees keep their separate data repositories. The Tracking Program has not prescribed standards for the technology or the architecture for grantee data repositories. Rather, grantees choose solutions that fit within their institutional architecture and policies. The data stored in these repositories are made available on the Tracking Network by the use of the gateways and/or by the display of them on the portals. The Tracking Program does not specify standards for how grantees store or manage their internal databases. Rather, the Tracking Program has specified a data exchange standard (the use of XML Schema—see Section 4.4.) In this way, grantees have the ability to maintain their own databases and data structures, but the data being exchanged with the Tracking Program have a standardized structure and format. This feature makes it possible for the Tracking Program to collect data from multiple sources and database systems and place the data in an integrated central repository.

3. NATIONAL TRACKING NETWORK CONTENT

The Tracking Network provides a location for users to access the tracking-specific data and measures developed by the Tracking Program. The Tracking Network also provides links to other datasets by which users can explore environmental hazard and public health outcomes. Data discovery and search for these data are driven by metadata records associated with each dataset. As the Tracking Network continues to evolve, it will make additional tools available on the secure portal for data linkages and advanced epidemiological analyses. The Network includes messages that describe the Network content and that support environmental health tracking. The following sections describe Tracking Network content and how data, measures, tools, and messages are developed and approved for the Network.

3.1. NATIONALLY CONSISTENT DATA AND MEASURES (NCDM)

The NCDM make up the primary information source of the Tracking Network. NCDM stem from the efforts of grantees and CDC to compile, on a national basis, datasets that are derived from many sources but are standardized to provide consistent reporting across the nation.

Initial decisions about the health and environmental data to include on the Network were based on many factors. These factors included giving priority to the needs of state and local health departments as well as priority to the needs of CDC. Also considered was whether data were available for a particular content area. New priority areas have been and will continue to be identified over time.

Content Workgroup Teams

NCDM are developed by Content Workgroup (CWG) Teams and recommended by the Principal Investigators (PI) for consideration by the CDC Tracking Program for possible inclusion on the Tracking Network. This process is outlined in Appendix B. Ten teams have been established, as shown below. Eight of these have developed recommendations to the Tracking Program:

- Air Quality
- Birth Defects

- Carbon Monoxide Poisoning
- Cancer
- Childhood Lead Poisoning
- Climate Change*
- Drinking Water Quality
- Hospitalization Data for Asthma and Myocardial Infarction (Heart Attacks)
- Pesticides*
- Vital Statistics /Reproductive Health Outcomes

(*The Climate Change and Pesticides teams have not completed recommendations as of the date of TNIP 2010)

The Tracking Network made Initial decisions about the health and environmental data to include on the basis of the priorities discussed on the previous page.

To help identify the underlying data for development of the NCDM on the Tracking Network, teams assess the quality and completeness of existing content-specific data sets. They then determine how best to construct NCDM that can be tracked over time and across geography (e.g., incidence rates, average annual concentrations of pollutant). As part of this process, the teams strongly consider the epidemiologic significance of the data. They consider the relevance and importance of the data for the public and for surveillance by researchers, known and suspected etiologic agents, and how to frame meaningful, responsible public health messages. Teams also consider the suitability of datasets for linkage with other datasets. They also examine various approaches for deriving or collecting nationally consistent data that are compatible with Tracking Network standards.

CWG team members include experts from various partner agencies who are knowledgeable about existing data and data stewards. Some of the participating organizations are as follows:

- US Environmental Protection Agency (EPA);
- US Geological Survey (USGS);
- National Institutes of Health (NIH);
- American Association of Poison Control Centers (AAPCC);
- National Birth Defects Prevention Network (NBDPN);
- National Association of Health Data Organizations (NAHDO);
- National Association for Public Health Statistics and Information Systems (NAPHSIS) and;
- North American Association of Central Cancer Registries (NAACCR).

Content workgroup teams filled in Indicator and measure templates describing proposed NCDMs. These documents provided specific information about the derivation of the measure, measurement units (e.g., days, number of children tested), geographic scope and scale, rationale for inclusion on the Tracking Network, and limitations of the measure.

CONTENT DOMAIN: HOSPITALIZATIONS INDICATOR: HOSPITALIZATIONS FOR ACUTE MYOCARDIAL INFARCTION	
Type of EPHT Indicator	Health Outcome
Measures	<ol style="list-style-type: none"> 1. Annual number of hospitalizations for Acute Myocardial Infarction (AMI) 2. Average Daily Number of hospitalizations for AMI by month 3. Annual age-specific rate of hospitalizations for AMI 4. Annual unadjusted (crude) rate of hospitalizations for AMI (ages ≥35) 5. Annual age-adjusted rate of hospitalizations for AMI, (ages ≥35) <p>When supported by sufficient data volume, the measures may also be reported stratified by age, sex, race, & ethnicity.</p>
Derivation of Measures	<p>Numerator: Resident hospitalizations for AMI, ICD-9-CM: 410.00 – 410.92 by gender and total for state and by county</p> <p>Denominator: Midyear resident population, by gender, for state and by county</p> <p>Adjustment: Age-adjustment by the direct method to Year 2000 US Standard population</p>
Unit	Hospital admission (categorized by discharge diagnosis)
Geographic Scope	State and National (tracking network states)
Geographic Scale	Residents of jurisdiction – State, County (ZIP code available for all measures once postcensal population data source identified)
Time Period	Hospital admissions between January 1 to December 31, inclusive, for each year, 2000-2005; annually thereafter
Time Scale	Daily, monthly, and annual (as appropriate for the measure)
Rationale	<p>There currently is no single AMI surveillance system in place in the US, nor does this exist for coronary heart disease (CHD) in general. Mortality is the sole descriptor for national data for AMI. Estimates of incidence and prevalence of AMI and CHD are largely based on survey samples (e.g., National Health and Nutrition Examination Survey) or large cohort studies such as the Atherosclerosis Risk in Communities (ARIC) study.</p> <p>In 2007 the American Heart Association estimated 565,000 new attacks and 300,000 recurrent attacks of myocardial infarction</p>

CWG co-chairs referred group NCDM recommendations to the Tracking Program for review and possible adoption. In some cases, CDC revised the recommended data and measures standards to allow them to be used on the Tracking Network.

Data-Sharing Agreements

Access to the data frequently requires development of partnership agreements. A data-sharing agreement is a formal contract between two parties (e.g., a grantee and state/local data steward or CDC and a national data steward) that clearly documents the data to be shared.

The characteristics and the content of data-sharing agreements vary by the needs of each data steward.

However, the agreements generally include the intended use of the data, constraints on use of the data, required procedures for data confidentiality and security, and the methods of data-sharing (i.e., having a secure connection over the Internet).

Certain documentation may be used in support of the development of data-sharing agreements. This documentation includes state/local cooperative agreements with the Tracking Program to provide core data and measures to the Tracking Network and make them available on state/local Tracking Networks. It may also include [CDC's Data Re-Release Plan](#). The Data Re-Release Plan describes the principles and procedures for the re-release of data on the Tracking Network through the national public and secure portals, in accordance with applicable federal and state laws.

Example NCDM on the Tracking Network

3.2. OTHER TRACKING NETWORK DATA

The Tracking Program uses other types of data in the Tracking Network. An example is the use of Census data on the number of children living in poverty. In addition, the Tracking Program staff is actively exploring options for including other data in the future. Examples include asthma prevalence data from the CDC Asthma Call-Back Survey, remote sensing data, and climate data from the National Oceanic and Atmospheric Administration.

3.3. METADATA

The term *metadata* is used for data that describe and define other data. Creation and maintenance of metadata are essential for the success of the Nationwide Tracking Network. Descriptive metadata describe the content, quality, and context of data resources on the Tracking Network. Such metadata exist for the purpose of making it easier to identify, discover, and use Tracking Network data. Metadata records may reference other resources, such as data dictionaries. Metadata provide the “what, why, when, who, where, and how” for a data resource. Metadata are important to the Tracking Network for two main reasons:

- They allow Tracking Network users to locate resources through a variety of means, including keywords, geographic boundaries, date, and time.
- They support the ability of a Tracking Network user to determine the content of a resource, why it was created, how it was created, limitations of a resource, access and use restrictions, data quality, and contact information. Such information helps a user decide if a data resource is appropriate for the intended purpose.

Grantees create metadata records in conjunction with local data stewards, the Tracking Program, and other national data partners. Data are available on the Tracking Network only if the data are accompanied by Federal Geographic Data Committee (FGDC) and Tracking Network-compliant metadata. The Tracking Network requires that a standard file format for NCDM metadata accompany all data.

The *Tracking Network Metadata Profile* is a subset of the full FGDC Content Standard. It provides the minimum set of descriptive metadata elements required for making data resources available on the Tracking Network (see Appendix C). Metadata Subgroup members, data stewards, and partners

Metadata Example	
Data Facts	
Description: Ozone data	
Owner: U.S. Environmental Protection Agency	
Original Home: EP& website more	
Years available	2000-2006
Measurement Method	Ozone monitors
	Number of Observations
	What's Reported
Ozone measurements	
Yearly average of hourly readings more	parts per million (ppm)
Other Information	
Location	county
Time frame	calendar
Level of geographic detail: County	
Total number of counties in state: 24 (including Baltimore City)	
Counties included in data: 20 more	
Explanation: Ozone monitors are not placed in every county more	
Why these data are important	
Ozone is an airborne substance that can be harmful to health. High levels of ozone can cause several problems. more	
Keywords: ozone, monitor, warm season, climatology, meteorology, atmosphere	
Questions	
Can I download this data from the MTN? Yes more	
Is more detailed data available from the Original Home? Yes more	
How often does the MTN get this data? Once a year more	
Is this data checked to make sure it's valid? Yes more	
Contacts	

developed the profile. It contains the minimum elements required by FGDC and several optional elements that the group believed was essential for documenting Tracking Network data. Use of this standard profile allows Tracking Network metadata records to be indexed on other data discovery sites, such as www.data.gov. In addition, reliance on the FGDC metadata standard supports the capture of both spatial and non-spatial information that may be present in Tracking Network datasets. Currently, the FGDC standard's name is the Content Standard for Digital Geospatial Metadata (CSDGM), Version 2, (or FGDC-STD-001-1998). The FGDC standard is now being modified to be compatible with international standards. When this modification is complete, the Tracking Network will follow suit.

3.4. TRACKING NETWORK TOOLS

The Tracking Network provides access to tools created specifically for tracking. It also provides external tools and software that may be useful for epidemiological studies. The Tracking Network provides access to the Metadata Creation Tool. The Tracking Program is evaluating additional spatial and linkage analysis tools for incorporation.

3.4.1. METADATA CREATION TOOL (MCT)

A Metadata Creation Tool (MCT) resides on the national secure portal. It can also be installed locally by grantees. The MCT supports the ability to populate the Tracking Network metadata template (Appendix C). Users can access the MCT by logging on to the national secure portal via SAMS. The Tracking Program updates the MCT regularly on the basis of user input and the receipt of other information.

3.4.2 RAPID INQUIRY FACILITY

The Rapid Inquiry Facility (RIF) is an automated geographic information system (GIS) tool used for epidemiologic analysis and mapping. RIF is an extension of the Environmental Systems Research Institute's (Esri) ArcGIS software product. RIF was developed by the Small Area Health Statistics Unit (SAHSU) at the Department of Epidemiology and Public Health, Imperial College London. In collaboration with the Tracking Program, the Imperial College London has transformed the RIF to US conditions. Imperial College London has redesigned RIF to include increased functionality and flexibility. The RIF is now database-independent. It allows input of exposure modeling data. Users

can export data for further analysis in other (statistical) software packages, such as WinBUGS and SaTScan. Users can also link directly with these two software packages, and RIF automatically imports and displays results.

3.5. MESSAGES

The content of the national public portal is supported by different types of messages. These messages help describe what can be found in the data, how the data might be used, and the constraints on the use. Three main types of messages appear on all Nationwide Tracking Network portals (national and state/local, public and secure): 1) general messages, 2) content-specific messages, and 3) data quality/metadata messages.

General messages may include

- Background information: general information about the Tracking Program and the Tracking Network.
- Global public messages: What is known and not known about how the environment affects health and what tracking data can reveal.
- Risk messages: What the risks are, who is at risk, factors that increase or decrease risk, and how one knows he/she is at risk.
- Self-efficacy and prevention: How one can protect one's self and family and where more information can be found.

To provide guidance and consistency across portals, the Tracking Program developed a set of general messages to be used on the national public portal. Grantees can adapt the national messages to best fit the format of their sites. State/local personnel can obtain resources for message development by contacting Tracking Program

Sample General Messages

Background: *The goal of environmental public health tracking is to protect communities by providing information to federal, state, and local agencies. (CDC)*

Global: *Unless we have evidence that people were exposed to a chemical, there is no way to examine whether it is related to a health effect. (New York State)*

Risk: *Smoking is the greatest risk factor associated with bladder cancer. Persons who smoke have more than twice the risk for bladder cancer than non-smokers. Research indicates that smoking causes approximately 20–30% of bladder cancers among women and 50–65% among men..(CDC)*

Self-Efficacy and Prevention: *Carpet holds many asthma triggers. It is best to remove all carpeting in the bedroom. If this is not possible, vacuum weekly when the person with asthma is not in the room. Use a vacuum with special allergy-proof vacuum bags and HEPA filters, or use a central vacuum. (New Jersey)*

Communications staff. Additionally, the Tracking Program is developing guidance on consistent terminology for use across all portals. Such consistent terminology will contribute to Nationwide Tracking Network branding.

Content-specific messages are unique to each health or environmental topic. Such messages appear as the user drills down into specific content areas. On the National Public Portal, these messages occur outside the query function to provide users with access to information within one or two clicks from the home page. In general, they address the following questions:

- Why is this content area an environmental public health issue?
- What is being tracked in this content area?
- What is known about the content area's relationship to health or the environment?
- Why is this indicator used to make a statement about this content area?
- Why is this measure used to support this indicator?

A set of general messages pertaining to each content area is available for use by all grantees. These general messages provide introductory information on the landing page for each NCDM.

Data quality messages address specific aspects of the data. These messages may be derived from or supported by the metadata. Messages may describe specific data limitations (e.g., data are available only for certain dates or under certain conditions) or background information on data collection. The system generates these messages as appropriate and as needed by grantees and the Tracking Program for specific data displays on portals.

4. NATIONAL TRACKING NETWORK PROCESSES

Several processes support the identification, creation, submission, processing and publication of data on the Tracking Network. The following diagram (Figure 3) gives a high-level overview of the processes involved from identification to publication of data. The processes described are focused on grantee submission of data to the Tracking Network. Data processing and submission steps for data not provided by grantees are different for every national partner of the Tracking Program. This document describes the steps in this figure in more detail on the following pages.

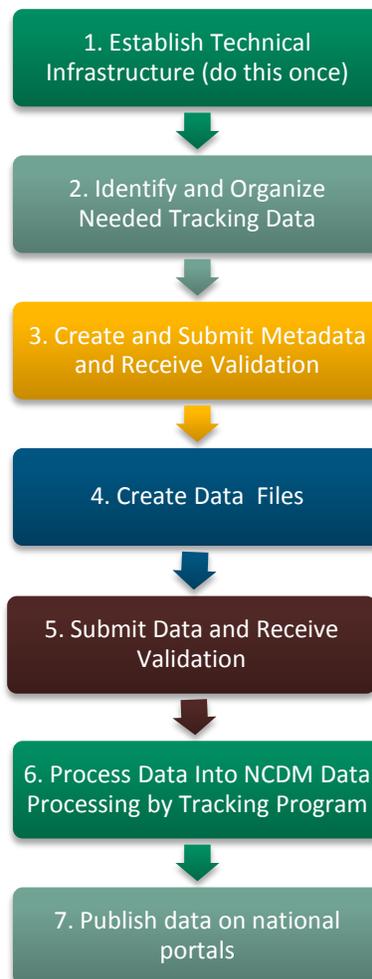


Figure 3: Overview of Tracking Network Processes for Data on the National Portals

4.1. TECHNICAL INFRASTRUCTURE IMPLEMENTATION

The Tracking Program has identified a series of technology standards for grantees to use. These standards help make it possible to create and securely exchange data on the Tracking Network. The *Tracking Network Transport Overview* provides grantees with information on protocols and procedures. Figure 4 below depicts the steps that are described in more detail in Appendix D. Grantees must first determine the transport protocol they will use. They must then identify the approach that will support the creation of metadata. The proper choice results in the production of a file that complies with the *Tracking Network Metadata Profile* previously described. The choice will also ensure that a grantee has the right security credentials. Finally, grantees validate their metadata and then validate the XML implementations of their metadata. The Tracking Program does not tell grantees what specific XML validation software to use. In fact, grantees use a variety of free and commercially available software according to their needs and policies.

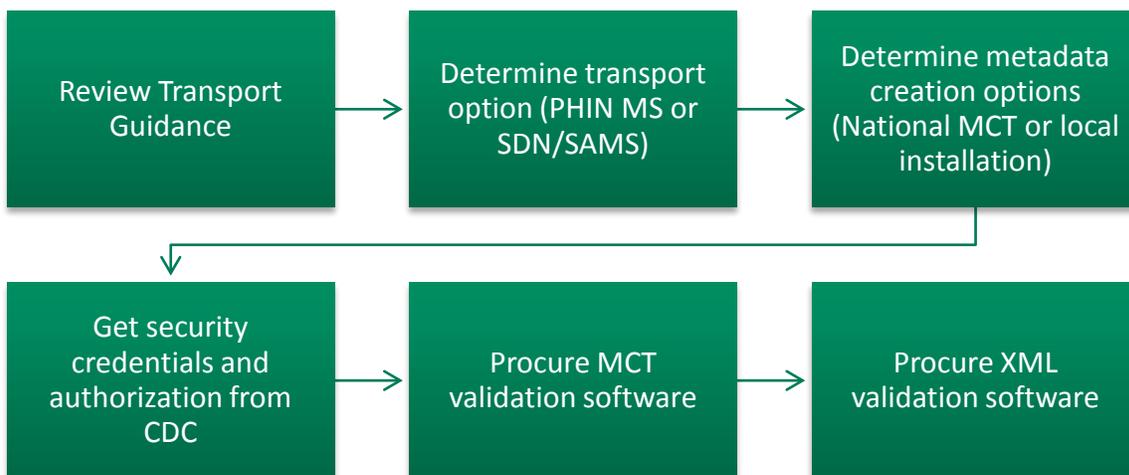


Figure 4: Establish Technical Infrastructure

4.2. TRACKING DATA IDENTIFICATION AND ORGANIZATION

The Tracking Program issues a call to grantees to submit data to the Tracking Network (Figure 5) by a specific time frame. Grantees refer to the CDC Standards for Nationally Consistent Data and Measures

within the Environmental Public Health Tracking Network to identify the data sources/sets they need by which to create and submit required NCDM. To acquire data from state, local, or other data agencies, grantees must identify the data stewards of those agencies and negotiate data-sharing agreements. The data-sharing agreements specify how the data will be used on the Tracking Network and the means to be used to ensure confidentiality. Once grantees establish the data-sharing agreements, grantees are able to acquire data from data stewards for processing.

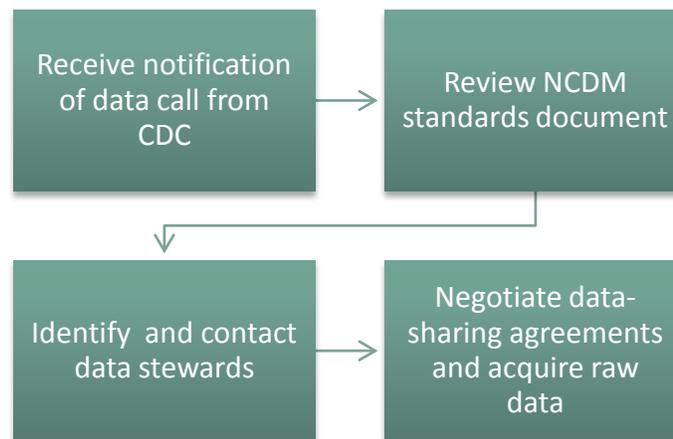


Figure 5: Identify and Organize Tracking Network Data

4.3. METADATA CREATION, SUBMISSION, AND VALIDATION

Every NCDM submitted to the Tracking Program by a grantee must be associated with a metadata record. This connection is established via a Metadata Control Number (MCN). One way for a grantee to create metadata records is by use of the MCT. Another way is by use of any tool that will create the standard FGDC-compliant XML file output that can be accepted by the MCT. The record that the tool generates is submitted to the Tracking Program for approval. The submission is through the MCT or through an upload of the XML file to the MCT. Upon validation of the metadata by the Tracking Program, the grantee submitter will receive a MCN to be associated with the NCDM to be submitted.

There is a two-step approval process for each metadata record. The first approval of the record occurs at the grantee site by the staff member designated as the “Grantee Approver.” This person also serves as the contact in case the Tracking Program has questions or needs record revisions during the second round of approval undertaken by the Tracking Program staff. The Tracking Program reviewers are responsible for identifying all of the required elements within the metadata record and for ensuring that information is in the correct field. Once the Tracking Program staff completes this second round of the approval process, the Tracking Program notifies the person who submitted the record and the Grantee Approver that the metadata record has been approved. The Tracking Program will send the grantee an MCN for the record. If the Tracking Program rejects the record, no MCN will be sent. The Tracking Program will provide a later explanation of the changes required for the record to be approved (see Figure 6).

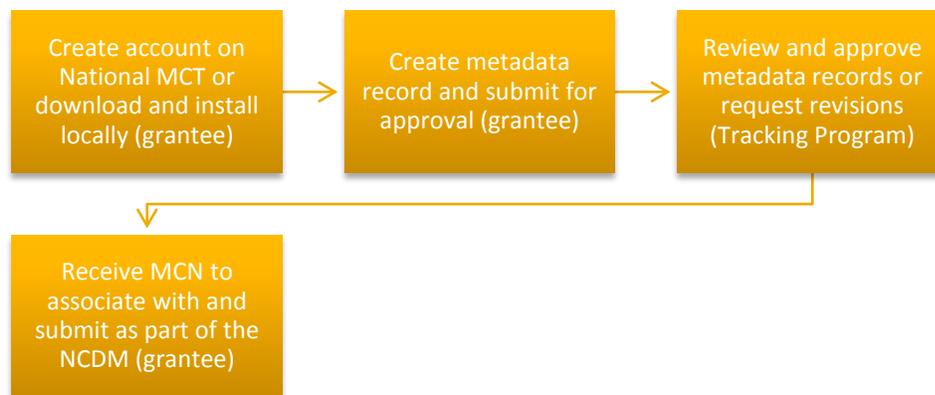


Figure 6: Metadata Creation, Submission, and Validation

4.4. DATA FILE CREATION

Data submission is a formal exchange between gateways on the Tracking Network. The National Tracking Program has formalized this exchange through the development of standard templates (XML schemas). These templates make the process easier and quicker. They also encourage grantees to use the same standards. The schemas are custom-made for various data sets. The schemas make it easy for grantees to submit data by reuse of common elements. The design also makes it easy for the National Tracking Program to make future changes in one schema without making serious changes to the functions of other schemas.

To create datasets for submission (Figure 7), grantees review the [published](#) standards for NCDM within the Environmental Public Health Tracking Network. They also review available dataset recommendations from CWG teams and XML Schema variable definitions in order to identify required fields for submission. Grantees acquire the latest XML Schema made available for download by the Tracking Program. The schema will reflect the latest updates to the requirements for NCDM. Grantees can use ready-made code to process their datasets into the format prescribed by the XML Schema.



Figure 7: Data File Creation

4.5. DATA SUBMISSION AND VALIDATION

As previously discussed, the submission of data to the national portal requires the submitter to have an MCN that is generated at the time of the approval of metadata (Figure 8). The MCN is then tied to the NCDM for which the metadata were created. The MCN is included with the XML file to be submitted to the Tracking Program. Once the Tracking Program receives the data, it sends a confirmation of receipt to the grantee. The Tracking Program validates the content of the data, and it may require the grantee to resubmit the data if it finds errors in the data.

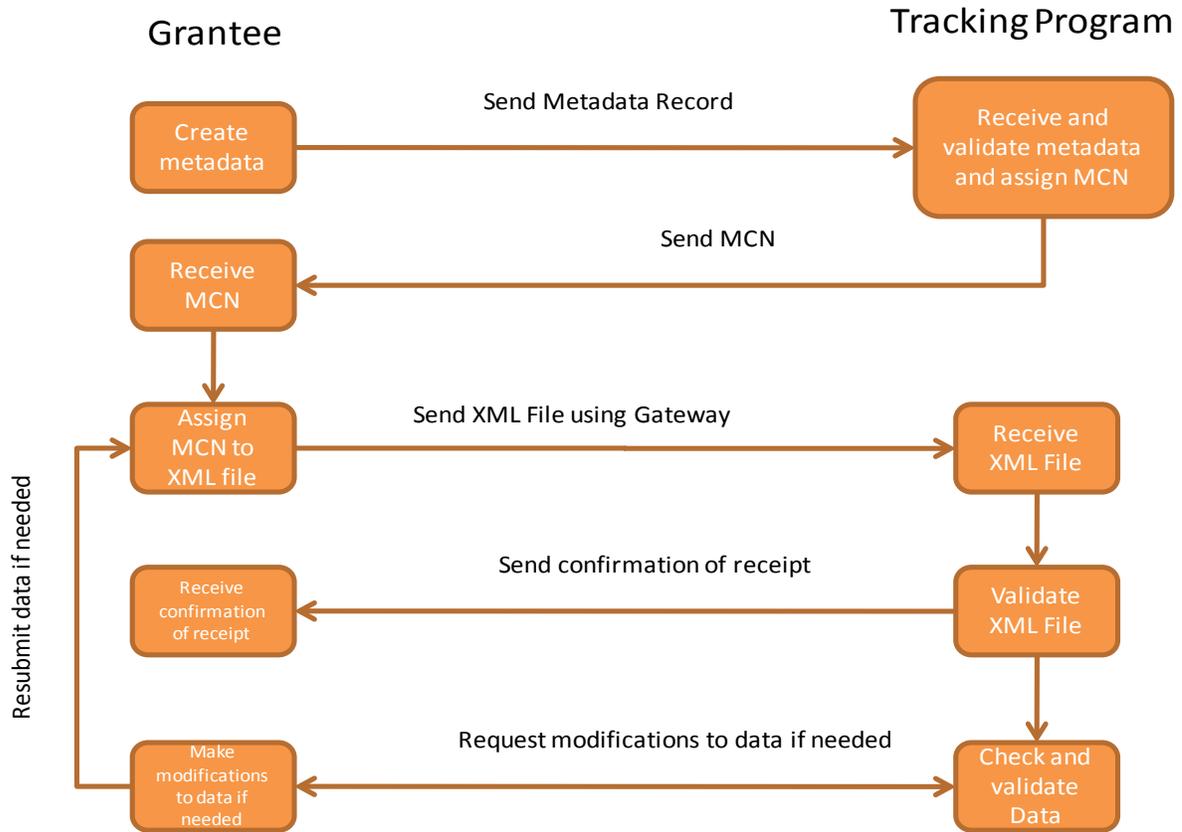


Figure 8: Data Submission and Validation

4.6. DATA PROCESSING (TRACKING PROGRAM)

Once the Tracking Program receives the NCDM or other data file via the national gateway, the Tracking Program staff begins the process of validating and compiling the data. The main steps are detailed in Figure 9 below.

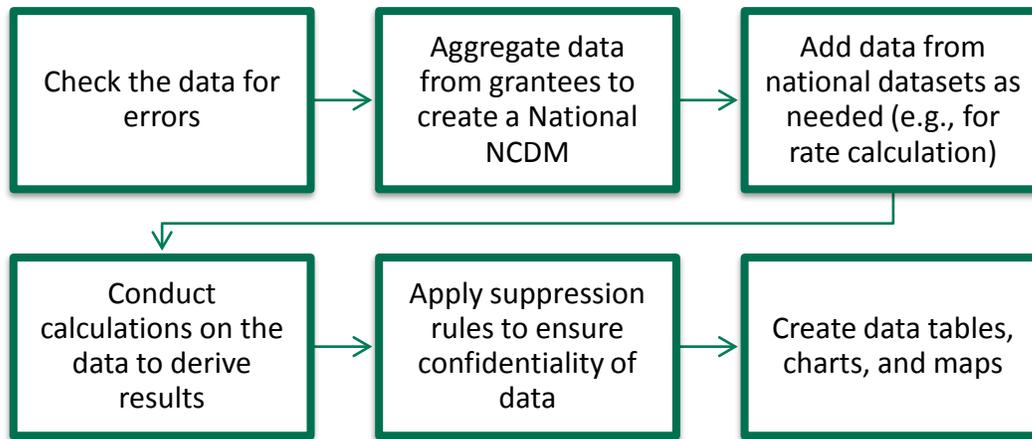


Figure 9: Tracking Program Data Processing and Validation

4.7. DATA PUBLISHING ON PORTALS

Once grantees submit data to the Tracking Program, they become part of a data management process. This process results in data that are available for display in various formats on the Tracking Network (Figure 10).

Before such data are published, however, a key step is to develop messages that help describe the data limitations and sources. After the development and review of these messages, the data content and messages are published together on the national portal.

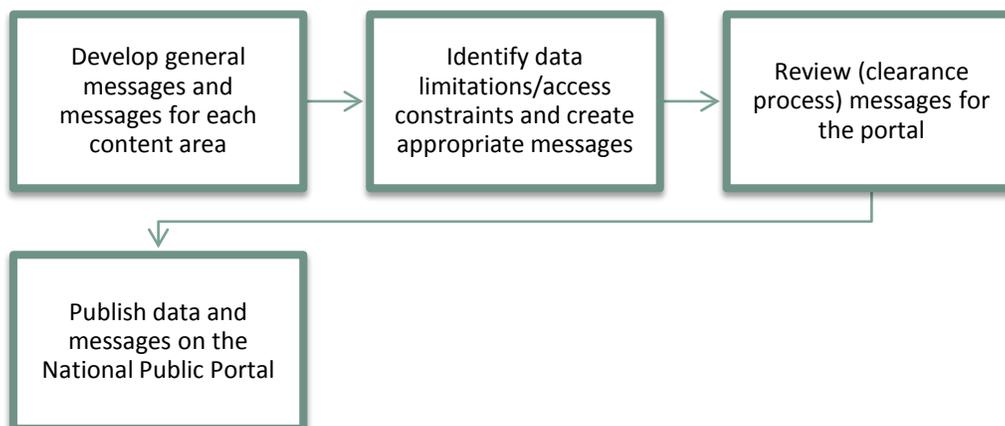


Figure 10: Publish Data on Portals

APPENDICES

5. APPENDIX A: OVERVIEW OF NATIONAL TRACKING NETWORK

WORKGROUPS

The National Environmental Public Health Tracking Program (Tracking Program) at the Centers for Disease Control and Prevention (CDC) has led the effort to develop the Nation-wide Environmental Public Health Tracking Network. It has done so in partnership with funded state/local governments (grantees) and academic organizations, federal agencies, and non-governmental organizations. The workgroups described here have helped guide the National Tracking Network (Tracking Network) to its current state.

Principal Investigators Forum

The Principal Investigators (PI) Forum consists of Tracking Program staff and the PIs from each of the current state and local grantee organizations and academic partners. The PI Forum convenes monthly or as needed to discuss broad logistic and strategic issues regarding the development of the Tracking Program and Network.

Content Workgroup

The Content Workgroup (CWG) recommends Nationally Consistent Data and Measures (NCDM) for addressing priority information needs for the Network. The CWG consists of a Steering Group (SG) of Principal Investigators, three co-chairs (National Tracking Program Staff and two grantees), and various teams. The teams review and make recommendations on specific types of environmental and health data/measures. Teams consist of subject matter experts, National Tracking Program staff, and a number of grantee representatives. The teams and the SG meet as needed to accomplish their work.

Standards and Network Development Workgroup

The Standards and Development (SND) Workgroup is led by two co-chairs, one from the Tracking Program and the second from a grantee organization. The SND Workgroup has technical staff from the Tracking Program, grantee

organizations, academic partners, and federal partners. The SND examines issues related to the technical infrastructure and standards of the Tracking Network and the state/local Tracking Networks. The SND forms technical subgroups as needed. Such subgroups have worked on network architecture, security, and challenges regarding metadata and geospatial data. The technical subgroups develop and submit recommendations to the SND Workgroup, which reviews and then forwards the recommendations to the Tracking Program for consideration.

Geospatial Workgroup

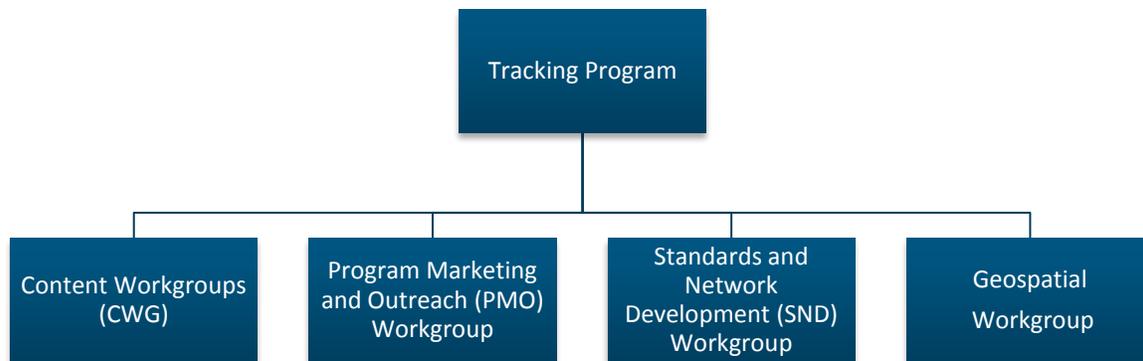
The Geospatial Workgroup was formed to address many of the interests related to representing, displaying, and linking data geographically. It is co-chaired by the Tracking Program staff and a member of a grantee organization. Three teams initiate work to explore spatial tools and methods. Another aim is to explore approaches to maintaining confidentiality and reducing unstable disease rates due to small numbers. These teams also work to optimize consistent display and visualization of data and support the ability to link data geographically.

Program Marketing and Outreach Workgroup

The Program Marketing and Outreach (PMO) Workgroup is led by two co-chairs, a member of the Tracking Program staff and member of a grantee organization. The PMO consists of staff members from the Tracking Program, grantee organizations, academic partners, and non-governmental organization partners. PMO provides recommendations for conducting outreach and marketing for the Tracking Program and state/local Tracking Programs. Activities of the PMO include developing materials that communicate consistent tracking messages to assist grantees in their state/local communication and conducting outreach with key audiences and users. The PMO also identifies new audiences for outreach and helps other Tracking Program workgroups with communication needs.

6. APPENDIX B: ORGANIZATION OF NATIONAL TRACKING NETWORK WORKGROUPS

Figure 11 shows the relationship among the Workgroups supporting development of the Tracking Network and state/local Tracking Networks. The subgroups and teams that support these groups are described on the following



pages.

Figure 11: Tracking Program Workgroups

STANDARDS AND NETWORK DEVELOPMENT WORKGROUP (SND)

The Standards and Network Development (SND) Workgroup is made up of members of the National Tracking Program and grantee technical staff. The SND Workgroup recommends technical network development standards to the Tracking Program. These standards are to promote the functioning and use of the Tracking Network. The SND Workgroup has a number of subgroups that work on different aspects of the Tracking Network. As needed, the subgroups create ad hoc teams that work on specific tasks for short periods of time. Figure 12 depicts the current SND subgroups and teams.

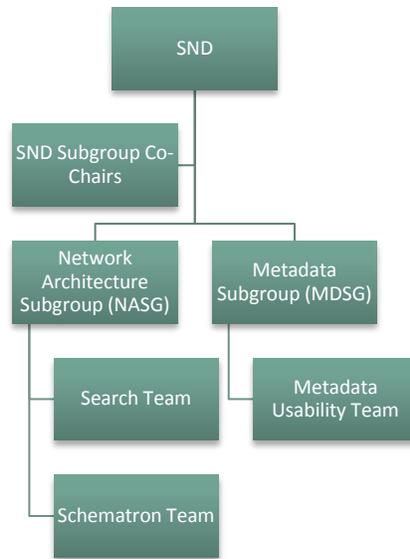


Figure 12: Standards and Network Development Subgroups and Teams

CONTENT WORKGROUP (CWG)

The Content Workgroup is made up of members of the National Tracking Program and grantee staff with epidemiologic and subject matter expertise in various content areas. The mission of the CWG is to make recommendations to CDC on the development of NCDM that address priority information needs for the Tracking Network. The structure of the CWG consists of a Steering Group (SG) of Principal Investigators and various Content Teams (see Figure 13 below). Team members consist of experts with knowledge of existing data. These experts are from various partner agencies, data stewards, and other organizations, including CDC. Other organizations represented are the Environmental Protection Agency (EPA); the US Geological Survey (USGS); the National Institutes of Health (NIH); and non-governmental organizations such as the American Association of Poison Control Centers (AAPCC), the National Birth Defects Prevention Network (NBDPN), the National Association of Health Data Organizations (NAHDO), the National Association for Public Health Statistics and Information Systems (NAPHSIS), and the North American Association of Central Cancer Registries (NAACCR).

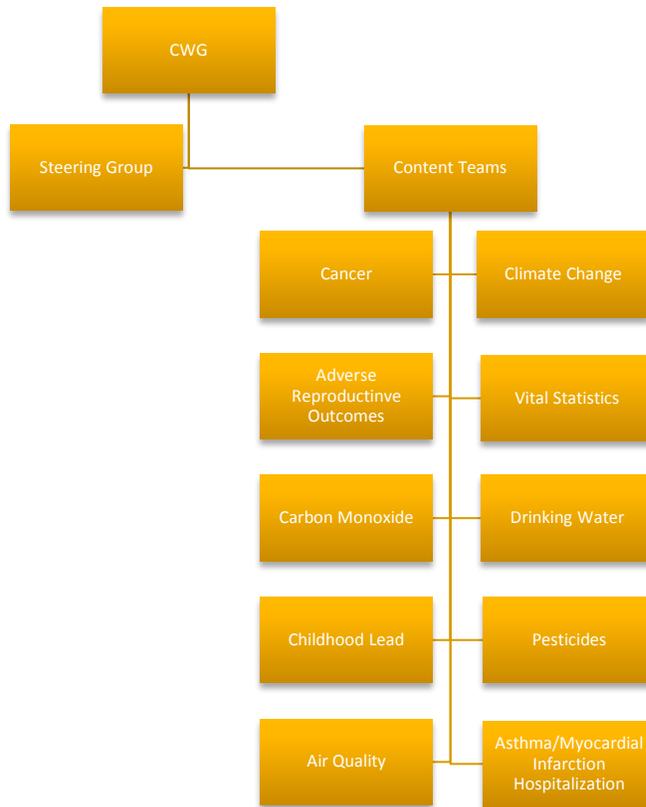


Figure 13: CWG Structure and Teams

To help identify existing data available for the development of NCDM on the Tracking Network, the Content Teams assess the quality and completeness of existing content-specific datasets. The teams determine how best to construct an NCDM that can be tracked over time and geography (e.g., incidence rates, average annual concentrations of pollutant). Teams strongly consider the epidemiologic significance of the data; the relevance and importance of the data for surveillance for the public and researchers; the known and suspected causes; and how to frame meaningful, responsible public health messages. Teams also consider the suitability of datasets for linkage with other datasets. They also examine various approaches to deriving or collecting nationally consistent data that are compatible with Tracking Network standards.

PROGRAM MARKETING AND OUTREACH WORKGROUP (PMO)

The Program Marketing and Outreach workgroup (PMO) has the purpose of helping the Environmental Health Tracking Branch and its grantees to develop, put in place, and evaluate communication and outreach activities.

PMO activities include developing appropriate education and outreach activities and materials that support the goals, objectives, and timely promotion of the Tracking Network.

The PMO is specifically responsible for identifying new local or national audiences. It is responsible for undertaking research about any newly identified audience as a part of developing a recommendation for an outreach method to use toward that audience. PMO develops materials that provide tracking messages to assist grantees in their state/local communication. It also supports the communication needs of other Tracking Workgroups. The structure of the PMO workgroup and its teams changes periodically to address the present and emerging needs of the Tracking Program. In its history, the PMO has undergone three major reorganizations, with the current structure focused on audience-specific outreach. Four initial audiences were chosen for focus in 2009, as shown in Figure 14. Each audience team worked on outreach, communication, and other activities pertaining to its specific audience.

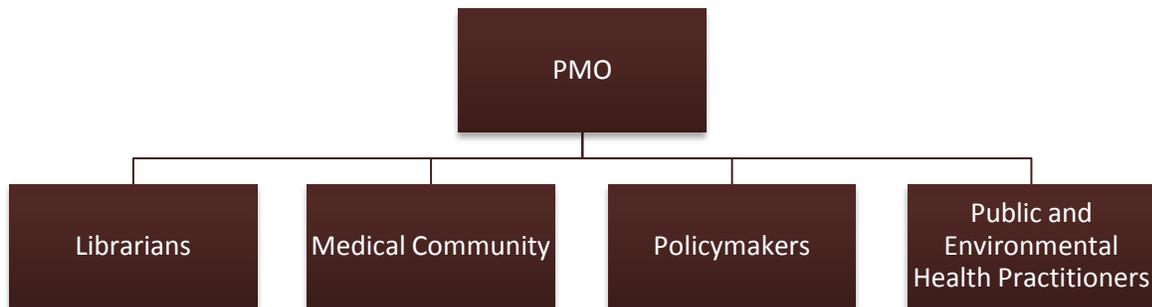


Figure 14: 2010 PMO Structure and Audience Teams

GEOSPATIAL WORKGROUP

The Geospatial Workgroup came into formal existence in early 2010. The group’s main objective is to investigate issues of geography and the display of geographic information. The workgroup will research and put priorities on the issues. It will then create and manage teams that are focused on creating and communicating best practices—

the practices that have proven to work best for creating and displaying geographic information. Workgroup membership is open to Tracking Branch staff, other CDC staff, grantees, other federal partners, and national partner organizations.

Three teams are currently in the Geospatial Workgroup. The Community Mapping Team's objectives are to identify and recommend the best ways to display useful data at different geographic resolutions while still preserving confidentiality. The Visualization Guidance Team's objectives are to identify and recommend best practices for creating visual consistency within the Tracking Network, including the national and grantee sites. The Data Linkage Team's objectives are to identify and recommend best practices for linking dissimilar data across geography and time.



Figure 15: 2010 Geospatial Workgroup Structure

7. APPENDIX C: TRACKING NETWORK REQUIRED METADATA FIELDS

I. IDENTIFICATION TAB

Field	Definition
CITATION PAGE	
* CATEGORY	Environmental Hazard, Environmental Exposure, or Health Effect
* PUBLICATION DATE	Public date of release of the data
* TITLE	Official name of the data set
* NATIVE DATASET ENVIRONMENT	Format of raw data
DESCRIPTION PAGE	
* ABSTRACT	Summary of the data (who, what, where)
* PURPOSE	Why do the data exist and for whom?
* PROGRESS	Whether completed, planned, in progress
* UPDATE FREQ.	Updated Continually, Daily, Weekly, Monthly, Annually, Unknown, As Needed, Irregularly, None Planned.
TIME PERIOD PAGE	
* DATE TYPE	Single date, multiple dates within a year, range of dates
* SINGLE DATE	Date covered by entire dataset
* MULTIPLE DATES	First, others
* RANGE OF DATES	Entire data range covered
* CURRENTNESS	Time Period End Date, Publication Date
GEOGRAPHIC AREA PAGE	
* WEST COORDINATE	Geographic corners of state, county, etc.
* EAST COORDINATE	
* NORTH COORDINATE	

* SOUTH COORDINATE	
KEYWORDS PAGE	
* THEME	Overall topic of the data
* THEME KEYWORD	Keywords describing the theme
* PLACES	Geographic coding system (FIPS 5-2 (state), FIPS 6-4 (county), ISO 3166-1 (country), ISO 3166-2 (country subdivision).
* PLACES KEYWORD	State name, abbreviation, and FIPS code (e.g., Washington, WA, 53).
SECURITY PAGE	
* SECURITY CLASSIFICATION SYSTEM	State or agency security classification system (free text field)
* CLASSIFICATION	Unclassified, restricted, top secret, secret, confidential, sensitive, none
* SECURITY HANDLING DESCRIPTION	Date storage requirements, if any
* ACCESS CONSTRAINTS	Criteria for access, if any; if publically available, write "none"
* USE CONSTRAINTS	Constraints to use (e.g., linking to other data, commercial use)

II. DATA QUALITY TAB

Field	Definition
* PROCESS DATE	Date that the processing described below was completed
* PROCESS DESCRIPTION	Process that occurred. (e.g., data downloaded from server, de-duplication)
* LOGISTICAL CONSISTENCY REPORT	For GIS data (e.g., polygons closed, missing polygons). If not GIS = none
* COMPLETENESS REPORT	Text field describing any issues (e.g., records lost during de-duplication, omissions, percent data missing)

III. ENTITY AND ATTRIBUTES TAB

Field	Definition
* OVERVIEW	Overview of the primary attribute fields included in the dataset
* DETAILED CITATION	User guide/data dictionary for each data element

IV. DISTRIBUTION TAB (NOT REQUIRED)

Field	Definition
RESOURCE DESCRIPTION	
LIABILITY	
CUSTOM ORDER PROCESS	

V. METADATA TAB

Field	Definition
* DATE CREATED	Date metadata were finalized
* STANDARD NAME	Tracking Network Profile Version 1.2 or FGDC Content Standard
* ACCESS CONSTRAINTS	Constraints to accessing or reviewing this metadata file
* USE CONSTRAINTS	Limitations for use of this metadata file

VI. CONTACTS TAB

Field	Definition
MATRIX PAGE	
* CONTACT 1 NAME	Corporate Contact Title of Data Steward
* CONTACT 1 TYPE	Agency
CONTACTS PAGE (Complete the following for originator, creator, distributor)	
* PERSON	
* ORGANIZATION	
* TITLE	
* PHONE NO. 1	
* FAX	
* E-MAIL	

* STREET ADDRESS	
* CITY	
* ZIP	

* = Required Field

8. APPENDIX D: DATA TRANSPORT PROCESSES

TRACKING TRANSPORT OVERVIEW: HOW TO SEND NATIONALLY CONSISTENT DATA AND MEASURES TO THE NATIONAL TRACKING PORTAL

INTRODUCTION

Grantees are required to send NCDM to the Tracking Program for compilation into national datasets. A number of tasks are involved in this process and needed software tools may vary in preparation for the data for submission.

Grantees are expected to assign a person who is accountable to the Tracking Program for addressing technical issues involved with sending NCDM. That individual is expected to be able to establish relationships with data providers in order to resolve quality control issues as well as with their Tracking Program and IT staff to resolve transport problems. Grantees must be able to generate and validate XML from their data sources before sending the files to the Tracking Program.

DATA SUBMISSION PROCESS

The flow of the process for sending data is outlined in Figure 16. A description of the terms used in the diagram is provided in the table that follows.

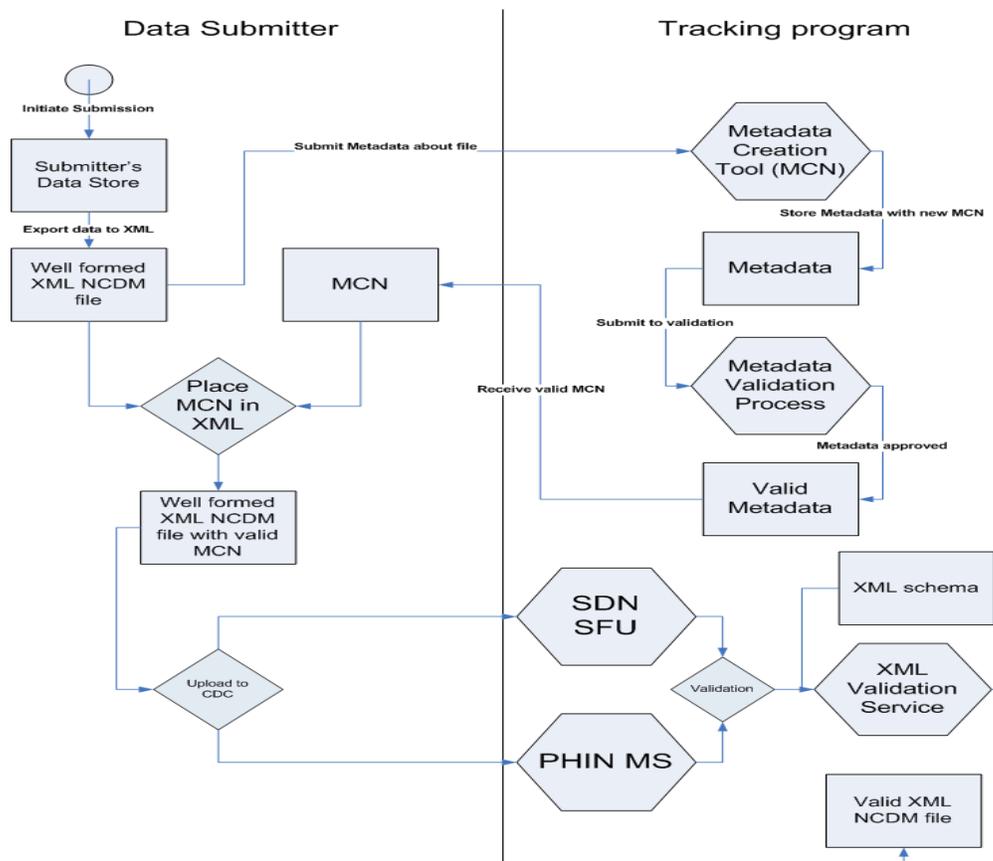


Figure 16: Data Submission Process

Part	Function
Submitter's data store	Where the data provider keeps the data in whatever format the provider chooses
Well formed XML file	An export of NCDM data into the format designated by Tracking's schema file for the NCDM
Metadata creation tool	Custom tool used to create FGDC-compliant metadata; alternatively, grantees can use comparable software

Metadata	FGDC metadata captured for each data submission to the Tracking network
Metadata validation process	Tracking Program's content review of submitted metadata
Valid metadata	Metadata that has been accepted by the Tracking Program
MCN	Metadata Control Number, used to reference each data submission
Well formed XML + MCN	An NCDM submission file with containing proper MCN
SDN SFU	Option 2 of upload choices
PHIN MS	Option 1 of upload choices
XML validation service	Automated validation takes place after NCDM file upload
XML schema	Each NCDM has a schema used to validate a file format
Valid XML NCDM file	An NCDM file that has passed schema validation and is accepted into the Tracking Data warehouse.

PHIN MS VS SDN SFU

Most steps in preparing for NCDM submission are contained in the Procedure section below. Grantees have two options for submitting data to the Tracking Program. The choice of options for each grantee is likely to be dependent on the grantee's existing infrastructure and information technology requirements.

Complete information on the system requirements and instructions for downloading, installing, and using PHIN MS information is available in the [PHIN MS V2.8.01 Installation Guide](#).

SDN SFU does not require the installation of any software. Grantees can use a compatible Web-browser to upload files. Before they can use SDN SFU, grantees must register for Secure Access Management System (SAMS) authorization, as discussed earlier in this document. Additional information on SAMS can be found in the [SAMS User Guide V1.0](#).

PROCEDURE

The steps in the table below outline the activities required for preparing to submit NCDM submissions:

Step	Action
1	Identify contacts
1a	Identify Science person (who creates the data submission)
1b	Identify Technical person (who submits the data)
2	Review Transport Guide
3	Determine PHIN MS or SDN SFU
3a	If PHIN MS, get software and install client
4	Review Parts CDC NCDM Standards, CWG dataset recommendations, and CDC XML Schema
5	Determine how to create Metadata
5a	If CDC ID is needed for user keyfob, submit required papers (This process can take several weeks)
5b	If it is needed, get the MCT software to install
6	Get SDN certificate (Please see section SDN Certificate below)

7 Procure XML validation software
--

The steps in the table below are required for the submission of NCDM to the Tracking Program.

Step	Action
1	Retrieve latest XML schema for submission of NCDM
2	Create XML file containing required data for submission
2a	Ensure that the XML xsi:schemaLocation value is updated to reflect the location of your stored XML schema
3	Assure that Science contact e-mail is contained in the XML
4	Create metadata about submission (using MCT or other appropriate software)
5	Submit metadata
6	Receive MCN to be inserted into XML file
7	Validate XML file, using validation software and schema from step 1
8	ZIP the XML file
9	Submit XML file via PHIN MS or SDN SFU
9a	If SFU is used, confirmation of receipt will be given via Web interface
9b	If PHIN MS is used, confirmation of receipt will be given via PHIN MS client
10	Confirm receipt of valid file e-mail to Science contact identified in XML
11	If errors are identified, correct and resubmit.

9. APPENDIX E: GLOSSARY OF TERMS

Term	Definition
NCDM	Nationally consistent data and measure, approved by CDC Tracking
PHIN MS	Public Health Information Network Messaging System, primary transport tool for sending data to CDC
SDN SFU	Secure Data Network's Secure File Upload, secondary transport tool for sending data to CDC
Metadata	Descriptive data of NCDMs, required before submission of data
MCT	Metadata Creation Tool, produced by Tracking to assist grantees in generating and maintaining metadata
MCN	Metadata Control Number
XML	Extensible Markup Language
Schema	A way to define the structure, content, and, to some extent, the semantics of XML documents
SDN certificate	Secure data networks digital certificate, used to gain access to PHIN MS or SFU
JRD	Jurisdiction: unique ID given as a parameter

	to PHIN MS & SFU
DF	Data feed type: identified as a parameter to PHIN MS & SFU
NASG	Network architecture subgroup under the SND
SND	Standards and Network Development Workgroup focused on IT issues
SharePoint	Our communications site at www.ephtn.org maintained by Ross and Associates
Transport Guide	A guide detailing the steps and procedures needed to submit data to the Tracking Network. This guide is available by request from the Tracking Program.