Metadata Sample Content
Applies to Version 3 of the MCT

THIS DOCUMENT WAS DEVELOPED FOR THE ENVIRONMENTAL PUBLIC HEALTH TRACKING NETWORK TO PROVIDE GUIDANCE ON HOW TO CREATE METADATA USING TRACKING'S METADATA CREATION TOOL. EACH DATASET SUBMITTED TO THE TRACKING NETWORK NEEDS TO HAVE A METADATA RECORD DESCRIBING THE DATASET. IF YOU HAVE ANY QUESTIONS, CONTACT EPHTMETADATA@CDC.GOV OR TRACKINGSUPPORT@CDC.GOV.

*Means this element is required

I.   IDENTIFICATION TAB
A. CITATION PAGE

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* PUBLICATION DATE	What is the official public date of release of these data?

	The 'date of release' refers to the date the data are sent to CDC. Since the MD record has to be created prior to sending data to CDC, please enter the date you anticipate uploading to their server.
*TITLE	By what official name is this data set known when referenced by the data steward? For datasets that are not supplied to the CDC for the national EPHT portal, use the name provided by the data steward. For EPHT national datasets the title is standardized and provided by CDC. In the Distribution Section you can enter the data steward title in the RESOURCE DESCRIPTION field.
URL	If these data, as described by this metadata document, are available online, what is the grantee web address (e.g., grantee portal, data steward site where the data is being held, etc.) that would take the requester directly to the data? This should be a URL that will take a user directly to the download or view location. Do not reference SharePoint – this site is not available to the public.

B. DESCRIPTION PAGE

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* ABSTRACT	(Generally the What, Where, and Who) Provide a brief summary of the content of the data found in this dataset. The summary should include a statement about the time frame covered by the data, the affected population covered by the data (e.g., for DW the data covers Public Water Supplies), geographic coverage (e.g. entire state by county). What does a record describe? To where are the data referring?

	About whom does the data refer? (Children with elevated lead?, asthmatics? Etc.) Please try to limit the abstract to 500 words.
* PURPOSE	(Generally the Why) Why was this dataset compiled and who is the target audience for these data? What are people supposed to learn from these data (e.g. data compiled to show the trend in incidence of carbon monoxide hospitalization throughout New Jersey between 2000 and 2005). Legislation that mandates the collection of these data can also be listed or referenced here.
SUPPLEMENTAL INFORMATION	(Caveat Section) May be a good place to state a data update policy (e.g. Since data are routinely updated and corrected throughout the year, it is recommended that data for all years be requested when requesting data and not just data for the most recent year.). In addition, may be a good place to denote how missing data are identified and how special situations were handled (e.g. how are non-detects identified or coded in the dataset). The date a dataset is considered "final" by the data steward should be included in this field. This is also the location to add additional information clarifying the date entered in the 'Currentness' field. There can be a statement here as to how to properly acknowledge the originator/source of the data. Note: This can be your state's or department's official indemnification statement on the use of data. The data usability information goes into **Liability** (what it can/cannot be used for)
* UPDATE FREQ.	How often are the publically available data updated on your grantee portal? Choices from this pull-down menu include: Continually, Daily, Weekly, Monthly, Annually, Unknown, As Needed, Irregular, None Planned.
* OUT OF STATE HOSPITALIZATIONS OR ED VISITS	Are hospitalizations or ED visits by state residents to hospitals/EDs in other states included in the dataset?

	• Yes • No • Other If other, explain.
* HOSPITAL AMI TRANSFERS	Are hospital transfers excluded from the dataset? Note that exclusion of transfers is only required for AMI hospitalizations. • Yes • No • Other If other, explain.
* BIRTH DEFECTS SURVEILLANCE METHOD	What was the surveillance method used to ascertain birth defect cases? • Active • Passive with follow-up • Passive • Other If other, explain.
* BIRTH DEFECTS RACE AND ETHNICITY	Were maternal race and ethnicity collected and reported separately by the primary data steward? • Yes • No • Race and ethnicity not collected • Other If other, explain.

* BIRTH DEFECTS CODING	What medical coding standard was used to classify birth defects? • ICD-9 CM • ICD-9, CDC coding based on British Pediatric Association • ICD-10 CM
* BIRTH DEFECTS REPORTING	Were birth defects data collected and reported in every county within the state for all 12 birth defects? • Yes • No, all counties collected and reported but not for all birth defects • No, all birth defects were collected and report but not in every county • No, data were not collected and reported for all 12 birth defects and those that were collected were not collected in every county Provide more details
* BIRTH DEFECTS OUTCOMES	What pregnancy outcomes were included in ascertaining cases? • Live births only • Live births and fetal deaths • Live births and pregnancy terminations Provide more details

C. TIME PERIOD INFO PAGE

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* DATE TYPE	Options from the pull-down menu:

	- Single date - Multiple dates within a year or series of non-consecutive years(generally does not apply to EPHT at this point in time) - Range of dates (i.e., data spanning a year or a multi-year period as is the case for almost all EPHT submissions) Most EPHT datasets are a range of dates unless there are gaps between years (e.g., 2000-2008)
SINGLE DATE	What is the one date covered by this entire dataset (e.g. November 1, 2008)
MULTIPLE DATES	
Date 1	What is the first date covered by these data in this dataset (November 1, 2008)
Date 2	November 15, 2008
Date 3	November 30, 2008
* RANGE OF DATES	This field is used to denote the ENTIRE date range that can be found in this dataset. So, if you are providing data covering the years 2000 though 2008, the range would be indicated as: January 1, 2000 – December 31, 2008 even if data are only collected on a monthly or bi-annual basis.
* CURRENTNESS	Since many datasets are updated with corrected information, provide the time frame for which the data are believed to be accurate (e.g. these data, spanning the years 2000-2007, are known to be complete and accurate as of November 1, 2009). The two choices from this pull-down menu are: **Time Period End Date** or **Publication Date**. Use **Time Period End Date** for original source data. Use **Publication Date** for data that is secondary or has undergone processing by the data steward.

D. GEOCODING

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
*STANDARD	Were data geocoded using the Tracking Network Geocoding Standards? • Yes • No, not to standard • No, data were not geocoded If no, please describe geocoding process user including entity responsible for geocoding.
* SOFTWARE	What geocoding software was used? Include software name (Texas A&M Geocoder, ESRI, ArcGIS, etc.), version, and quarter.
* DATABASE	What underlying databases were used by the geocoding software? (e911, tiger files, etc.)
* RECORD COUNT	How many individual records were included in the input dataset?
* RURAL ROUTE COUNT	How many rural routes were included in the input dataset?
* PO BOX COUNT	How many PO Boxes were included in the input dataset?
* PO BOX PROCESS	Did you remove PO Boxes before geocoding the data? • Yes

	• No • Unknown If you did not remove PO Boxes, please describe the process used to geocode PO Boxes to a census tract.

E. KEYWORDS PAGE

MCT FIELD	Common Language Question Equivalent Interpretive Statement
* THEME	The theme field denotes the overall topic of the data (e.g. health, environmental, ICD 9, SNOMED etc.) Note: CWG Teams are developing best themes for each content area – this will be incorporated into the tool as it becomes available. Guidance for selecting the theme will be provided by your EPHT representative If no specific theme is being referenced, select 'NONE' from the choices provided.
* THEME KEYWORD	If you are using a specific theme, ascertain that the keyword chosen is a valid keyword for that theme via a keyword dictionary. If no specific theme is being used, determine what is/are the most logical word/words that one would use when searching for these data on your website or when searching the internet. *CWG Teams are developing best themes for each content area – this will be incorporated into the tool as it becomes available (via links).
* THEME	The theme field denotes the overall topic of the data (e.g. health, environmental, ICD 9, SNOMED etc.) Note: CWG Teams are developing best themes for each content area – this will be incorporated into the tool as it becomes available. Guidance for selecting the theme will be provided by your EPHT representative If no specific theme is being referenced, select 'NONE' from the choices provided.

* THEME KEYWORD	If you are using a specific theme, ascertain that the keyword chosen is a valid keyword for that theme via a keyword dictionary. If no specific theme is being used, determine what is/are the most logical word/words that one would use when searching for these data on your website or when searching the internet. *CWG Teams are developing best themes for each content area – this will be incorporated into the tool as it becomes available (via links).
* THEME	The theme field denotes the overall topic of the data (e.g. health, environmental, ICD 9, SNOMED etc.) Note: CWG Teams are developing best themes for each content area – this will be incorporated into the tool as it becomes available. Guidance for selecting the theme will be provided by your EPHT representative If no specific theme is being referenced, select 'NONE' from the choices provided.
* THEME KEYWORD	If you are using a specific theme, ascertain that the keyword chosen is a valid keyword for that theme via a keyword dictionary. If no specific theme is being used, determine what is/are the most logical word/words that one would use when searching for these data on your website or when searching the internet. *CWG Teams are developing best themes for each content area – this will be incorporated into the tool as it becomes available (via links).
* CATEGORY	Category has been included as a theme to describe broadly the category of your data, and you need to include category as one of your themes. There are currently three broad categories which describe EPHT data: Environmental Hazard, Environmental Exposure, or Health Effect. There are, however, more than three categories listed on the menu of the MCT. At the present time, creators of metadata should choose one of the following category identifiers to describe their dataset: Environment, Environmental Hazard, Exposure, Health, Health Effects. For datasets that could fall into multiple categories the question should be posed to the

	CWG Team responsible for data.
* PLACES	By what geographic coding system are these data denoted (e.g. FIPS, GNIS, other)? The choices from the pull-down menu include: FIPS 5-2 (state), FIPS 6-4 (county), ISO 3166-1 (country), ISO 3166-2 (country subdivision). Current NCDMs use FIPS state or FIPS county.
* PLACES KEYWORD	Note your state name, abbreviation, and FIPS code (e.g., Washington, WA, 53).

F. SECURITY PAGE

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* CLASSIFICATION	Are these data available to the public or are they restricted to those individuals who meet a certain "handling" criteria (e.g. choices from the pull-down menu are "Unclassified", "Restricted", "Top Secret", "Secret", "Confidential", "Sensitive" or "None"). "Unclassified" is the handling restriction for data available on the public portal.
* SECURITY HANDLING DESCRIPTION	Describe, if applicable, the manner in which these data are to be stored once received by the requester (e.g. store dataset on secure server). Default entry into this free text field should be "None".
* ACCESS CONSTRAINTS	If these are publicly available data, write 'NONE' in this space. If a certain criteria must be met before obtaining these data, describe the criteria in detail here (e.g. only researchers who have an approved protocol from an IRB may access these data). This field should include both legal (liability) information and non-legal access constraints. Information that pertains to dissemination of the data should be included in the Distribution tab.

	For all datasets available on the Public Portal this field should be "None".
* USE CONSTRAINTS	There should be a statement here stating that the user must understand the metadata content before attempting to understand, interpret and use the data on the portal. Additional questions to be addressed in this field include: How should this dataset be used? And not used? Can it be linked to other datasets? Can these data be used for commercial purposes? Can these data be used to form a basis for additional health studies or some remediation actions? What are the constraints for data interpretation? This is also the field where messaging information should be included.

II.   DATA QUALITY TAB

MCT FIELD	Common Language Question Equivalent or Interpretive Statement

This section describes how the data were manipulated from their raw format to the current state in which they have been made available. Each step in the manipulation process should have its own processing date and description. Consider steps such as conversion from one program to another, data cleaning, data aggregating, geocoding, deduplication, addition of data fields, additional computations, etc.). You can list multiple process steps. Examples of process steps shown below. These fields will vary by grantee. These fields describe how and when you manipulated the data prior to uploading it to the Portal.

Do not reference EPHT SharePoint site.

* PROCESS DATE	Date that the processing described below was completed – NOT when the data were sent to the portal. (e.g., November 12, 2009)
* PROCESS DESCRIPTION	Process that occurred. (e.g., Data downloaded from server and record deduplication was performed using XXXX criteria.)
* COMPLETENESS REPORT	This field can contain information as to what happened during processing to create the dataset into its available form. Address the following points: How many records lost to deduplication, geocoding errors, incomplete record information, geographic boundary changes, etc? Identify any data that has been omitted from the dataset that you might logically expect to see and the reason for exclusion. This field can also contain information as to what percentage of data are missing, the accuracy of the data, the version of a specific coding system that was used (e.g. ICD-9 vs. ICD-10 or data based on 2000 census population vs. intercensal values). You can list the test(s) used to check for data inconsistencies here. If this metadata record is being created for data available on a GRANTEE PORTAL, add information regarding suppression and aggregation

III. ENTITY AND ATTRIBUTES TAB

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* DETAILED CITATION	This field can be thought of as a user guide and data dictionary to each data element. It may contain a complete data dictionary listing, a data key or a link to a data dictionary and supporting documents. Do not link to EPHT SharePoint in this field.

IV. DISTRIBUTION TAB

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
LIABILITY	This should contain a statement of liability against improper usage and endorsements of any kind. Generally speaking, each grantee should have a liability statement that indicates a department or data steward is not responsible if an individual misinterprets the presented information, links data in a manner other than that mentioned in the 'Use Constraints' field, somehow repackages and redistributes the available data, typographical errors are found, if data errors are found, etc.

V. METADATA TAB

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* DATE CREATED	On what date was this metadata document finalized?
* STANDARD NAME	EPHTN Tracking Network Profile Version 1.2 or FGDC Content Standard for Geospatial Metadata can be chosen from the pull-down menu.

VI. CONTACTS TAB

A. MATRIX PAGE (Note: only MCT users see the Matrix Page. When this form is sent to data stewards you can highlight the fields they need to complete.

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* CONTACT 1 NAME	Corporate Contact Title of Data Steward
* CONTACT 1 TYPE	Ex: NJDHSS Family Health Services
CONTACT 2 NAME	Corporate Contact Title of Metadata Creator
CONTACT 2 TYPE	Ex: NJ EPHT Metadata Coordinator
CONTACT 3 NAME	Corporate Contact Title of Data Distributor
CONTACT 3 TYPE	Distributor of the data
CONTACT 4 NAME	
CONTACT 4 TYPE	

B. Contact Fields (for originators, creator, distributor) – give us the contact info for:
   i. Metadata contact
      1. To whom the MCN/approval email goes to;
      2. Usually is a grantee contact
   ii. Originator/Creator
      1. The contact to send questions to about the data;
      2. The contact that created the MD record;
   Distributor contact
      3. The contact to obtain the data (e.g., distributor).

MCT FIELD	Common Language Question Equivalent or Interpretive Statement
* PERSON	Corporate Contact Title of Group responsible for providing dataset to EPHT program
* ORGANIZATION	
* TITLE	

USERID	
HOURS	Hours during which they can be contacted
INSTRUCTIONS	Special instructions (e.g. leave a voicemail and your call will be returned within X days)
* PHONE NO. 1	
PHONE NO. 2	
* FAX	
* E-MAIL	Note: Corporate Contact E-mail Address ONLY. No Personal E-mail addresses.
TDD/TTY	
* STREET ADDRESS	
* CITY	
STATE	
COUNTRY	U.S.A.
* ZIP	