

# Describing Data Resources on the Environmental Public Health Tracking Network

2007 Public Health Information Network  
Conference

Atlanta

August 29, 2007





# History of Metadata Development (2003-2007)





# Today's Presentation

---

- Review Metadata efforts undertaken for the Tracking Network since the 2005 PHIN Conference.
- Set the stage for presentation on the Metadata Content Guidance Document and the Metadata Creation Tool demonstration.



# PHIN Conference 2005

---

## Developing a Metadata Template for CDC

The Efforts of the Metadata and Data Quality  
Subgroup to Capture Data about Data

The Metadata and Data Quality Subgroup presented in 2005 on efforts to select a descriptive metadata standard for the Tracking Network.



# Selection of a Standard...

---

- Review of several metadata standards, including Dublin Core; International Organization for Standardization (ISO); and the Federal Geographic Data Committee (FGDC)
  - ◆ Talked with individuals involved with PHIN and the Environmental Protection Agency's Exchange Network
  - ◆ Reviewed data inventories from the various Tracking Grant partners
  - ◆ Evaluated common elements against the different metadata standards



# FGDC Standard Selected...

---

Why?

- Network will have both Geospatial and Non-Geospatial data
- Elements to describe data on Network
- Well established standard
- Supported by Presidential Executive Order 12906



# Testing

---

- Developed a test for the FGDC Content Standard
- Tested the usefulness of the freeware tool Tkme
- Gathered feedback from test participants
- Made recommendations to Metadata Subgroup members.



# Testing

---

## Results:

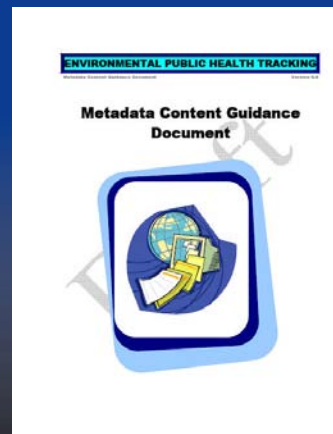
- Determine a core set of FGDC elements that will constitute the bare minimum data elements that will be required to put data on the EPHTN. Additional elements could be added if applicable.
- Develop a tool that allows the easy creation of metadata. The tool should be user-friendly and should have built in constraints.





# Metadata Profile

- Resulted in the creation of a EPHT-specific Metadata Profile based on the FGDC standard
- Creation of a Content Guidance Document



Metadata Profile Version 1.2 - Spreadsheet Version - For Internal Use Only - For more detailed information see the upcoming guidance documents.  
 Last Updated on February 6, 2007

LEGEND  
 Light Blue Shade = Optional Elements  
 Dark Blue Shade = Conditional Operator (and/or)  
 Light Grey Shade = Compound Element, No Data Added

Element	Field	Definition	Short Name	Element Type	Domain	Format	Repeating Element?	Notes
<b>Section 1: IDENTIFICATION</b>								
Citation	Citation (1-1)	Information to be used to reference the data set.	citation	Required				
Originator	Originator (3-5)	The name of an organization or individual that developed the data set. If the name of a firm or company is provided, the name must be followed by "Inc." or "Corp." respectively.	origin	Text	Minimum:Free Text			
Publication Date	Publication_Date (3-2)	The date when the data set is published or otherwise made available for release.	pubdate	Date	Minimum:YYYY-MM-DD Maximum:YYYY-MM-DD			
Title	Title (3-4)	The name by which the data set is known.	title	Text	Free Text			
URL	On-line_Linkage (URL) (3-10)	The name of an online computer resource that contains the data set. Entries should follow the Uniform Resource Locator convention of the Internet. (Completes if applicable.)	url	Text	Free Text			Optional
Description	Description (1-2)	A characterization of the data set, including its intended use and limitations.	description	Required				
Abstract	Abstract (3-3)	A brief narrative summary of the data set.	abstract	Text	Free Text			
Purpose	Purpose (3-2)	A summary of the intentions with which the data set was developed.	purpose	Text	Free Text			
Supplemental Info	Supplemental_Info (3-2)	Other descriptive information about the data set. (Completes if applicable.)	supplinfo	Text	Free Text			Optional



# Tool Requirements

---

- Created a mock-up of a tool
- Gathered tool requirements from participants
- Package those requirements for use by a developer (Metadata Tool Vision 1.1)



# Metadata Registry Schema / NYC Toolset

---

- Developed Initially from work undertaken by NYC Department of Health
- Final product based on a collaborative effort between NYC Department of Health and EPA Region 2



# Metadata Use Cases


Use Case ID #	Use Case Name	Actors
MD UC1	Create Metadata Record	<ul style="list-style-type: none"> <li>&gt; Metadata Creator</li> </ul>
MD UC2	Upload Metadata to Metadata Registry	<ul style="list-style-type: none"> <li>&gt; Metadata Creator</li> <li>&gt; Tracking Network Metadata Registry Administrator</li> </ul>
MD UC3	Update Metadata	<ul style="list-style-type: none"> <li>&gt; Metadata Creator</li> <li>&gt; Local Metadata Registry Administrator</li> <li>&gt; Tracking Network Metadata Registry Administrator</li> </ul>
MD UC4	Search Metadata	<ul style="list-style-type: none"> <li>&gt; Metadata Owner/Creator</li> <li>&gt; Tracking Network Metadata Registry Administrator</li> <li>&gt; Metadata Viewer</li> </ul>
MD UC5	Manage Metadata	<ul style="list-style-type: none"> <li>&gt; Metadata Owner</li> <li>&gt; Local Metadata Registry Administrator</li> <li>&gt; Tracking Network Metadata Registry Administrator</li> <li>&gt; Metadata Viewer</li> </ul>



# Conclusion

---

- Through out this process there has been data steward and EPHT partner involvement.
- Adaptation of existing industry standards



# What is the Tracking Network Metadata Content Guidance Document?





# Today's Presentation

---

- Evolution from a Tracking Network Metadata Profile (Profile) to a Tracking Network Metadata Content Guidance Document (Guidance Document)
- Components of the Guidance Document
- Expected Results from the Guidance Document

# Evolution from a Profile to a Guidance Document







# Profile

---

- Asked data stewards to create metadata using the entire FGDC Standard and provide feedback
- Feedback resulted in the creation of a limited set of elements (the Profile)
- The Profile contains 57 of the 195 total elements in the FGDC Standard
- Not all are mandatory but recommended if applicable



# Example of the Profile

Metadata Profile Version 1.2 - Spreadsheet Version - For Internal Use Only - For more detailed information see the upcoming guidance documents.

Last Updated on February 8, 2007

## LEGEND

Light Blue Shade = Optional Elements

Dark Blue Shade = Conditional Operator (and/or)

Light Grey Shade = Compound Element, No Data Added

<i>Element</i>	<i>Field</i>	<i>Definition</i>	<i>Short Name</i>	<i>Element Type</i>	<i>Domain</i>	<i>Format</i>	<i>Repeating Element?</i>	<i>Notes</i>
<b>Section 1: IDENTIFICATION</b>								
<b>Citation</b>	<b>Citation (1.1)</b>	<b>Information to be used to reference the data set.</b>	<b>citation</b>	<b>Compound</b>				
<b>Originator</b>	Originator (8.1)	The name of an organization or individual that developed the data set. If the name of editors or compilers are provided, the name must be followed by "(ed.*)" or "(comp.*)" respectively.	origin	Text	"Unknown"; Free Text			
<b>Publication Date</b>	Publication_Date (8.2)	The date when the data set is published or otherwise made available for release.	pub-date	Date	"Unknown"; "Unpublished Material"; Free Date	YYYY for year; YYYYMM for month of year; YYYYMMDD for day of the year		
<b>Title</b>	Title (8.4)	The name by which the data set is known.	title	Text	Free Text			
<b>URL</b>	On-line_Linkage (URL) (8.10)	The name of an online computer resource that contains the data set. Entries should follow the Uniform Resource Locator convention of the Internet. (Complete if applicable).	online	Text	Free Text			Optional
<b>Description</b>	<b>Description (1.2)</b>	<b>A characterization of the data set, including its intended use and limitations.</b>	<b>descript</b>	<b>Compound</b>				
<b>Abstract</b>	Abstract (1.2.1)	A brief narrative summary of the data set.	abstract	Text	Free Text			
<b>Purpose</b>	Purpose (1.2.2)	A summary of the intentions with which the data set was developed.	purpose	Text	Free Text			
<b>Supplemental Info</b>	Supplemental_Info (1.2.3)	Other descriptive information about the data set. (Complete if applicable).	supplinf	Text	Free Text			Optional



# Example of the Profile

Metadata Profile Version 1.2 - Spreadsheet Version - For Internal Use Only - For more detailed information see the upcoming guidance documents.

Last Updated on February 8, 2007

## LEGEND

Light Blue Shade = Optional Elements

Dark Blue Shade = Conditional Operator (and/or)

Light Grey Shade = Compound Element, No Data Added

<i>Element</i>	<i>Field</i>	<i>Definition</i>	<i>Short Name</i>	<i>Element Type</i>	<i>Domain</i>	<i>Format</i>	<i>Repeating Element?</i>	<i>Notes</i>
<b>Section 2: DATA QUALITY INFORMATION</b>								
Logical Consistency Report	Logics_Conistency_Report (2.2)	An explanation of the fidelity of relationships in the data set and tests used.	logic	Text	Free Text			
Completeness Report	Completeness_Report (2.3)	Information about omissions, selection criteria, generalization, definitions used, and other rules used to derive the data set.	complete	Text	Free Text			
<b>Lineage</b>	<b>Lineage (2.5)</b>	<b>Information about the events, parameters, and source data which constructed the data set, and information about the responsible parties.</b>	<b>lineage</b>	<b>Compound</b>				
<b>Process Step</b>	<b>Process_Step (2.5.2)</b>	<b>Information about a single event.</b>	<b>procstep</b>	<b>Compound</b>				
Process Description	Process_Description (2.5.2.1)	An explanation of the event and related parameters or tolerances.	procdesc	Text	Free Text		YES; Unlimited	This element can be repeated infinitely and must have a corresponding element 2.5.2.3 (seen below).
Process Date	Process_Date (2.5.2.3)	The date when the event was completed.	procdate	Date	Free Date		YES; Unlimited	This element can be repeated infinitely and must have a corresponding element 2.5.2.1 (seen above).

## Section 5: ENTITY AND ATTRIBUTES



# Example of the Profile

Metadata Profile Version 1.2 - Spreadsheet Version - For Internal Use Only - For more detailed information see the upcoming guidance documents.

Last Updated on February 8, 2007

## LEGEND

Light Blue Shade = Optional Elements

Dark Blue Shade = Conditional Operator (and/or)

Light Grey Shade = Compound Element, No Data Added

<i>Element</i>	<i>Field</i>	<i>Definition</i>	<i>Short Name</i>	<i>Element Type</i>	<i>Domain</i>	<i>Format</i>	<i>Repeating Element?</i>	<i>Notes</i>
<b>Section 5: ENTITY AND ATTRIBUTES</b>								
<b>Overview Description</b>	<b>Overview_Description (5.2)</b>	<b>Description of the entities, attributes, attribute values, and related characteristics encoded.</b>	<b>overview</b>	<b>Compound</b>				
<b>Entity and Attribute Overview</b>	Entity_and_Attribute_Overview (5.2.1)	Detailed summary of the information contained in a data set.	eaovovr	Text	Free Text			
<b>Entity and Attribute Detail Citation</b>	Entity_and_Attribute_Detail_Citation (5.2.2)	Reference used to the complete description of the entity types, attributes, and attribute values for the data set.	eadctcit	Text	Free Text			
<b>Section 6: DISTRIBUTION INFORMATION</b>								
<b>Distribution Information</b>	<b>Distribution_Information (6)</b>	<b>Information about the distributor of the option for obtaining the data</b>	<b>dirtinfo</b>	<b>Compound</b>			<b>YES; Unlimited in a group</b>	
<b>Distributor</b>	<b>Distributor (6.1)</b>	<b>The part from whom the data set may be obtained</b>	<b>dirtirk</b>	<b>Compound</b>	<b>See Section 10.</b>			
<b>Resource Description</b>	<b>Resource_Description (6.2)</b>	<b>The identifier by which the distributor knows the data set.</b>	<b>resdesc</b>	Text	Free Text			
<b>Distribution Liability</b>	<b>Distribution_Liability (6.3)</b>	<b>Statement of the liability assumed by the distributor.</b>	<b>dirtliab</b>	Text	Free Text			



# Needs for Content Guidance

---

- The Profile was not intended as a metadata entry guide
- There was a desire for clarity on how to enter information from a health perspective
- Examples of public health and environmental metadata were needed
- Metadata needs to be standardized across the network



# Developing the Guidance Document

---

- Ten members of the metadata subgroup formed a new team to develop metadata guidance
- The team reviewed/evaluated guidance developed by other organizations



- The team worked with data stewards to create metadata using the profile, document issues and provide feedback



# Developing the Guidance Document

---

- The Guidance Document was modeled on:
  - ◆ The United Kingdom Gemini standard document
  - ◆ The Environmental Protection Agency guidance document
  - ◆ The Federal Geographic Data Committee content guidance
  - ◆ Content guides developed within the Tracking Network



# Developing the Guidance Document

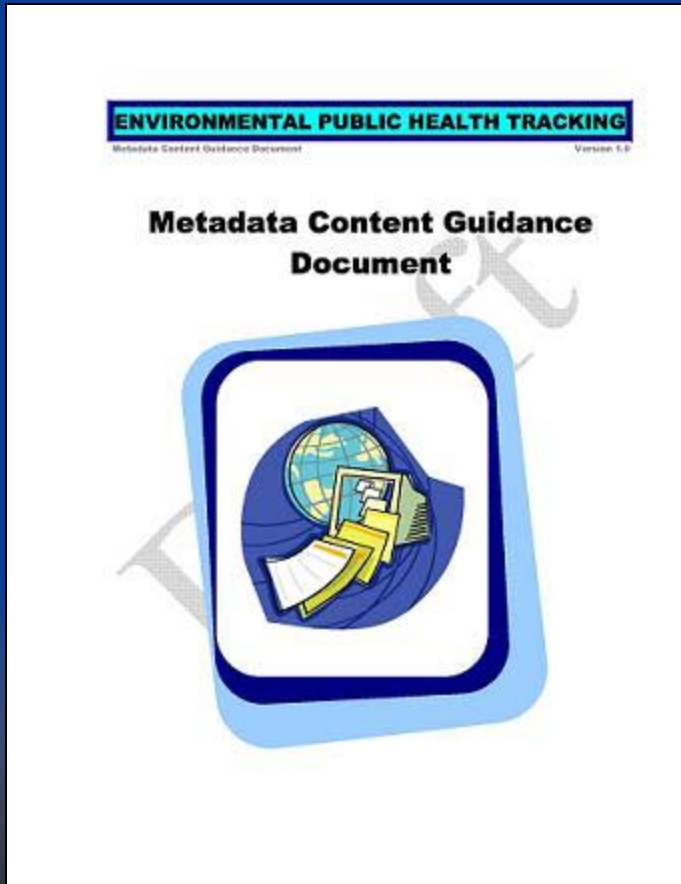
---

- Working with data stewards to create metadata using the Profile, documenting issues and provide feedback:
  - ◆ Worked with an Environmental Protection Agency air modeling team
  - ◆ Five state health departments created metadata on:
    - ★ Air monitoring data
    - ★ Asthma hospitalizations data
    - ★ Cancer data
    - ★ Drinking water contamination data





# The Result



- A easy to read document with guidance on how to enter information into each element in the metadata profile
- Examples of metadata created for public health and environmental datasets for use as guides



# Components of the Guidance Document





# Components

---

- Introduction and definition of metadata
- Structure of a FGDC metadata document
- How to deal with information not covered by the profile
- Tables for each metadata element
- Examples
- Document change request form



# Tables For Each Element

- Element Name
- Definition
- Purpose and Meaning
- Obligation
- Occurrence
- Date Type
- Domain (Controlled Terms)
- Recommendations
- Examples
- Additional Resources
- Other Comments
- Things to Note

Abstract	
Element Name	<b>Abstract</b>
Definition	A brief narrative summary of the dataset.
Purpose and Meaning	This information provides a user with a brief description of the source and contents of the dataset
Obligation	Mandatory
Occurrence	Single
Date Type	Text
Domain (Controlled Terms)	<b>Free Text:</b> There are no controlled terms for this element.
Recommendations for Filling in the Entry	<p>The abstract should contain information about:</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> subject, topic or theme of the dataset;</li> <li><input type="checkbox"/> the population included in or covered by the dataset;</li> <li><input type="checkbox"/> the spatial domain (largest spatial unit; for example state) and scale (smallest spatial unit; for example county) of the dataset; and</li> <li><input type="checkbox"/> the temporal domain (range) and scale (time unit) of the dataset.</li> </ul> <p>The abstract may also contain a brief description of the processes used to create or compile this dataset. Detailed description of the processes used should be provided in one or more Process Description elements.</p>
Examples	<p><i>Health Data:</i> This data set contains the annual case counts and standardized rates of liver cancers (ICDO-3 C220) among Misoretah residents for each county of the state of Misoretah from 1970 through 2007. Data from the Misoretah Cancer Registry were used to compile this dataset.</p> <p><i>Environmental Data:</i> Distribution point monitoring data from the Misoretah Division of Safe Drinking Water (MDSDW) were used to compile average municipal water system arsenic concentration levels for all municipalities in the state of Misoretah. The dataset provide the number of samples, the earliest and latest sampling dates, and the minimum maximum, mean, median and standard deviation of the arsenic concentration.</p>
Additional Resources	None
Other Comments	None
Things to Note	As a minimum, the name should include a theme (data subject).



# Tables For Each Element



## Recommendations for Filling in the Entry

The abstract should contain information about:

- ☐ subject, topic or theme of the dataset the population included in or covered by the dataset; the spatial domain (largest spatial unit; for example state) and scale (smallest spatial unit; for example county) of the dataset; and the temporal domain (range) and scale (time unit) of the dataset.

The abstract may also contain a brief description of the processes used to create or compile this dataset. Detailed description of the processes used should be provided in one or more Process Description elements.



# Tables For Each Element



## Examples

***Health Data:*** This data set contains the annual case counts and standardized rates of liver cancers (ICDO-3 C220) among Misoretah residents for each county of the state of Misoretah from 1970 through 2007. Data from the Misoretah Cancer Registry were used to compile this dataset.

***Environmental Data:*** Distribution point monitoring data from the Misoretah Division of Safe Drinking Water (MDSDW) were used to compile average municipal water system arsenic concentration levels for all municipalities in the state of Misoretah. The dataset provide the number of samples, the earliest and latest sampling dates, and the minimum, maximum, mean, median and standard deviation of the arsenic concentration.

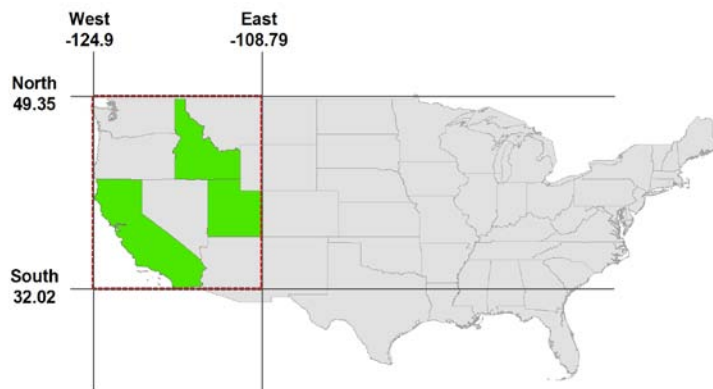


# Exceptions

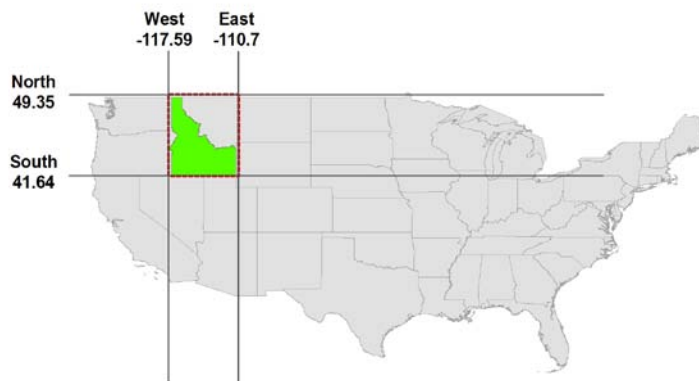
## Examples



### Regional Dataset:



### State Dataset:





# Expected Results from the Guidance Document







# Expected Results

- A document that helps health and environmental professionals new to metadata creation quickly learn how to populate a metadata document
- Standardized metadata that can be searched for the discovery of available health and environmental hazards data useful for public health surveillance and research





# Using the Guidance Document

---

- How NOT to use the Guidance Document
  - ◆ Don't read (and/or memorize) the Guidance Document – cover to cover. *This is a reference document to help “guide” entry of metadata content*
  
- How TO use the Guidance Document
  - ◆ If you have a question on what content to enter for a specific element in a metadata record:
    - ★ Go to the index and look up the name of the element
    - ★ Got to the page and read the table for that element



# Metadata Creation Tool Demonstration





# Questions

---





# Contact Information

---

- For more information on today's presentations contact Metadata Subgroup Co-Chairs
  - ◆ Jeff Patridge, [Jeff.Patridge@dhss.mo.gov](mailto:Jeff.Patridge@dhss.mo.gov)
  - ◆ Gonza Namulanda, [FOS0@cdc.gov](mailto:FOS0@cdc.gov)