



# A Controlled Vocabulary Supports Knowledge Management

PHIN Conference Session #1  
Knowledge Management Lifecycle  
May 24, 2004

**SAFER • HEALTHIER • PEOPLE™**



# Overview

- ❖ The Need & Value of Common Standardized Vocabulary
- ❖ Key Features to Support Knowledge Management
- ❖ Current CDC Developments
- ❖ Public Health Scenario
- ❖ Next Steps



## The Babble Problem

*The Department of Homeland Security, charged with safeguarding the nation from terrorist strikes, is facing a growing threat: language. Different interpretations of words and phrases such as "clear the building," "cover" or "surveillance" among various government agencies could create misunderstandings and possible chaos in the aftermath of an attack....The "Lexicon Project"...advises the department to create a homeland security lexicon that includes a core set of documents, key terms with their definitions and a glossary of acronyms.*

*"Terror threat: Bureaucratic babble,"*

*Atlanta Journal-Constitution, April 23, 2004*

**SAFER • HEALTHIER • PEOPLE™**



# Eliminating the Babble: Creating Conceptual Understanding

## ❖ Surveillance Concepts

- Disease
- **Immunologic**
- Population
- Drug
- Sentinel
- Hazard
- Video
- Police
- Traffic



Preferred Term  
Definition  
Scope Note on Usage  
Semantic Type  
Synonyms  
Acronyms  
Related Concepts  
Source Authority UIDs



# Value: Accuracy and Efficiency

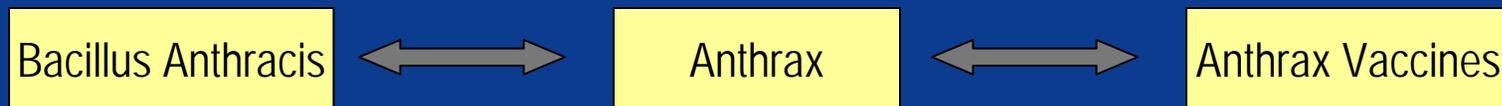
- ❖ Facilitates accurate information exchange
  - Example: Electronic Messaging for Laboratory Reporting
- ❖ Supports efficient and useful content storage and retrieval from knowledge repositories
  - Search and browse the CDC Web site for pertinent Topical Content
- ❖ Enables global search across databases in an organization (CDC) or domain (Public Health)
  - CDC Web Site
  - Public Health Directory
  - Local Databases



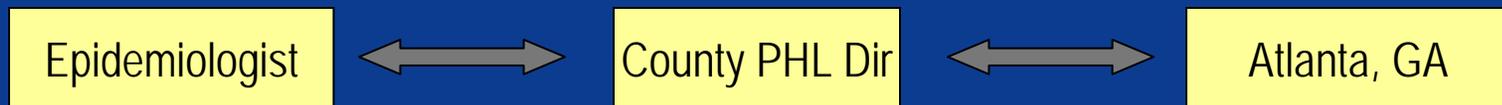
# Value: Conceptual Meaning

- ❖ Captures relationships among PH concepts

- Organism to disease to treatment



- PH role to occupation to geographical location



- ❖ Normalizes terminology by mapping acronym, synonym, misspelling, etc back to preferred term

- Acronym *BSE*, Synonym *Mad Cow Disease*, and misspelling *Bovine Spongiform Encephalopathy* are mapped to Preferred Term *Bovine Spongiform Encephalopathy*



# Value: Currency

- ❖ Rapidly captures changing concepts in PH
  - Emerging Infectious Disease
    - Example: New terminology for *Sudden Acute Respiratory Syndrome* and acronym *SARS* added and linked to *Coronavirus*
  - New Finding
    - Example: ATSDR Study identifies dust from a drywall joint compound a lung hazard
  - Update or Alert
    - Example: IMOVAX® Rabies Vaccine Recall



# Support for KM

- ❖ Provides granular metatagging of resources: content, databases and applications, people
- ❖ Allows the location, gathering, monitoring & retrieval relevant of information
- ❖ Takes advantage of existing standard authorities
- ❖ Enables content aggregation and intelligent search of multiple information sources
- ❖ Supports the repurposing of content
- ❖ Supports the personalization of results



# Challenges of Vocabulary Development for Public Health

- ❖ Public health domain has unique requirements:
  - Healthcare language is complex, yet robust
  - Public health workforce is broad, operating in varied roles in technologically diverse environments
  - PH concept deficits in available standard vocabularies are known (up to 50% gap)
- ❖ Largest proportion of gap concepts fall in the areas of:
  - Occupational health and safety
  - Environmental health
  - Injury prevention
  - Safety
  - Organisms / organism-related diseases
  - Bioterrorism and preparedness response



# CDC Thesaurus Development: Initial Build

- ❖ Inventory and Analyze CDC Domain
  - CDC Vocabularies Inventory
  - Search Logs
  - Indexes & Glossaries
  - Website Concept Analysis
  - Interviews with Stakeholders
- ❖ Research Standard Vocabularies
  - UMLS (Unified Medical Language System)
    - 100+ vocabularies (including MeSH, ICD-9, SNOMED, etc)
    - 975,000 concepts / 2.4 million concept names
  - Industry standards such as NAICS, SOC, USGS
- ❖ Select Standard Vocabularies that offer the widest CDC domain coverage



# Example Concept: Anthrax

The screenshot displays a software interface for managing concepts, likely from a biomedical ontology. It is divided into several panels:

- Hierarchy Tree :: Defined View (classic):** Shows a tree structure of concepts. The path for 'Anthrax' is: Biological Sciences > Environment and Public Health > Bioterrorism > Biological Agents > Bacillus anthracis > Biological Aerosol.
- Concept Walker:** Shows a different view of the hierarchy, highlighting 'Bacillaceae Infections' > Gram-Positive Bacterial Infections > Bacterial Infections > Infectious Diseases > Diseases and Conditions.
- Tree Editor:** Shows the detailed properties for the selected concept 'Anthrax'.
  - Primitive
  - Kind: CDC\_kind
  - Code: A0001296
  - Id: 1573
  - Bacillaceae Infections
  - Soilborne Diseases
  - all Associated\_Organism: Bacillus anthracis
  - all Associated\_Prevention: Anthrax Vaccines
  - Class: Topics(C26)
  - Date\_Created: 12/22/2002
  - Date\_Revised: 2/20/2003
  - Preferred\_Term: Anthrax
  - Production: 1
  - Scope\_Note: An acute infection caused by the spore-forming bacteria BACILLUS
  - Semantic\_Type: Disease or Syndrome(T047)
  - Sourcecode-CUI: C0003175
  - Sourcecode-MeSH: D000881
  - Synonym-MS: Anthrax
  - Synonym-MS: anthrax
  - Synonym-MS: anthax
  - Synonym-MS: anthraz
  - Synonym-MS: antrax
  - Synonym-S\_PF: Anthrax Infection
  - Synonym-S\_PF: Bacillus anthracis infection
  - Synonym-S\_PF: Infection by Anthrax
  - Synonym-S\_VWP: Infections, Anthrax
  - Synonym-S\_VWP: Infections, Bacillus anthracis
- Search:** A search bar with the text 'Bacillus\*' and a 'Search' button. Below it, 'Maximum Results' is set to 'ALL' and there is a 'Filter Properties' button.



# CDC Thesaurus Development: Expansion to Fill the PH Gap

- ❖ Quality Assurance
  - Review and modify top nodes of thesaurus hierarchy to ensure available “branches” for PH concepts
- ❖ Build out the gap
  - Review CDC Website Topics (the site map)
  - Review Search Logs
  - Tap CDC SME’s knowledge on Topics
    - Clinical knowledge
    - Programs and Campaigns
    - Consumer terminology
- ❖ Allow Term Recommendation by PHIN Partners



# Metadata Framework

- ❖ A Vocabulary is utilized in a Knowledge Repository, such as the CDC web portal, within a Metadata framework of attributes for describing and classifying information.
- ❖ Metadata Attributes for Public Health might include:

Metadata Field	Example Standard Terms in Vocabulary
Epidemiologic Method	Contact Tracing or Health Survey
Setting	Lumber Yard or Swimming Pool
Occupation	Coal Miner or Public Health Nurse
Geographical Location	DeKalb County, Georgia or Lake Lanier
Population Group	Homosexual Males or Female Workers



# Metadata Development & Implementation

- ❖ Analyzed CDC Website and consulted Agency staff to identify core metadata fields that could be used by all content, from Press Releases to Consumer Brochures to Laboratory Manuals
- ❖ Use of this core metadata consistently across CDC content ensures relevant and comprehensive searching across the knowledge portal
- ❖ Additional metadata fields can supplement the core fields to provide flexibility within applications
- ❖ Common metadata used by an organization (CDC) or organizations in a domain (PHIN) supports information sharing



# Vocabulary Utilization

- ❖ Web-based Portal Personalization
  - Enables PHIN Partner member to self-select information delivery, subscriptions, news and updates by area of expertise, interest, location, etc.
- ❖ Public Health Directory for PHIN
  - Identifies people resources by Role, Occupation, Title, Specialty, Location, etc.
  - Assessable by search or browse of categories
- ❖ Public Health Inquiries (Call Response System)
  - Knowledge bases linked with common vocabulary
- ❖ Global Search of all of above



# Scenario: Internal to CDC

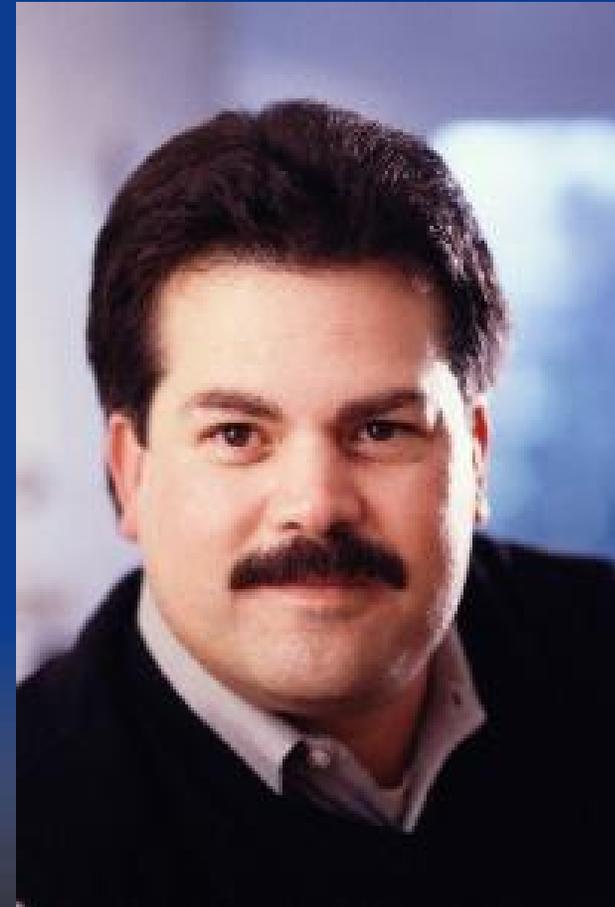


- ❖ Jennifer
- ❖ CDC Call Center Response Staff
- ❖ PH Nurse, specializing in Infectious Disease
- ❖ Responding to PH inquiry:
  - Prevention and Incidence of E. Coli Infection in swimmers in a home pool in Youngstown, Ohio
  - Performs search on website
  - Consults other internal databases with one search
  - Consults PH Directory for local PH ID specialist in Youngstown
- ❖ Results are relevant and comprehensive:
  - All applications use common vocabulary for search
  - All content tagged with vocabulary concepts



# Scenario: CDC to Partner & Partner to Partner

- ❖ Jennifer shares information with Allen, the local PH ID specialist in Youngstown
- ❖ Allen stores info in local database for future use, tagging it with common vocabulary concepts
  - Setting: swimming pool
  - Infectious Disease: Escheria Coli
  - Geography: Youngstown, Ohio
  - Aspects:
    - Prevention
    - Incidence
- ❖ This same information stored in a state database that uses the same PH Vocabulary becomes accessible by other Ohio local health departments, and thus shareable Partner to Partner





# Next Steps

- ❖ For CDC Web Thesaurus:
  - Accessible by PHIN Partners via a web browser
  - Downloadable in whole or part for use in local applications through a PHIN Vocabulary Services Provisioning System
- ❖ For PH Vocabulary developed by CDC:
  - Term submission to individual standard vocabularies such as SNOMED & MeSH
  - Vocabulary value set submission to vocabulary aggregator such as the UMLS Metathesaurus



Mamie J. Bell, MLn  
CDC/IRMO

[Mamie.bell@cdc.gov](mailto:Mamie.bell@cdc.gov)

**SAFER • HEALTHIER • PEOPLE™**