
CDC's Controlled Health Thesaurus: Common language to identify and describe information

The ability to identify and describe information is key to enabling pertinent data to be easily searched, browsed, accessed, and shared in a given domain. A controlled network of concepts, called a thesaurus, with unambiguous definitions, well-defined synonyms, and clear relationships to other terms within that network provides that necessary accurate and consistent language for public health information exchange. Whether used to metatag pieces of content on a website, classify expertise of SMEs within a public health directory, or define valid responses to a surveillance question, these concepts provide a common understanding of the identified information.

Efforts by Centers for Disease Control and Prevention

Centers for Disease Control (CDC) is developing a Controlled Health Thesaurus with an ultimate aim of using the concepts on web content to provide better access and navigation to the information. The CDC Web site at <http://www.cdc.gov> averages over 8 million visits each month and reaches citizens, public health providers, partners, congressional staff, news agencies, researchers, and clinicians. In public health emergencies, the site is a chief source of credible information as evidenced by the spike of visits to 14 million during SARS. The Thesaurus is part of a larger framework to facilitate the intuitive finding and accessing of information on the site.

CDC's Thesaurus is structured around preferred concepts, sourced whenever possible from standard vocabularies, with their associated properties (metadata) in a polyhierarchical framework. The preferred concept *bacillus anthracis*, for example, has defined associations to the concept *Anthrax* (the disease caused by the target organism) and to the concept *Anthrax vaccines* (the prevention for the disease). This organism term is "treed" in the polyhierarchy as both a soilborne bacillus organism and as a biological agent. The parent concept *Anthrax* has child concepts *Cutaneous anthrax*, *Gastrointestinal anthrax*, and *Pulmonary anthrax*. Thus the utilization of this cluster of concepts on a website or within an application enables a rich retrieval of all anthrax-related information.

Relevant Standards

The Controlled Health Thesaurus is a public health view of pertinent concepts from the **National Library of Medicine's Metathesaurus**¹. Additional concepts needed to cover the public health domain are being added and will be advanced to NLM. CDC Research



Centers for Disease Control and Prevention
www.cdc.gov/phn



revealed that the **National Library of Medicine's Medical Subject Headings (MeSH)**² provided the broadest coverage of concepts for public health. Additional concepts from other UMLS vocabularies, such as the Computerized Retrieval of Information on Scientific Projects (CRISP), Alcohol and Other Drug Thesaurus (AOD), and Library of Congress Subject Headings will be utilized as needed. CDC is leveraging the expertise of its subject matter experts and its public health partners to add additional public health concepts.

Requirements for Thesaurus structuring and metadata selection came from the **ANSI/NISO Z39.19 Guidelines for the Construction, Format, and Management of Monolingual Thesauri**.³ The standard deals with aspects of term selection and the means for establishing and displaying relationships between terms.

Further Information

<http://www.cdc.gov/PHIN>

Mamie J. Bell, MLn
CDC/National Center for Public Health Informatics
Zvt9@cdc.gov

¹ <http://www.nlm.nih.gov/research/umls/>

² <http://www.nlm.nih.gov/mesh/>

³ <http://www.niso.org/standards/resources/Z39-19.pdf>

