

The Future of Public Health Vocabularies and Public Health Data Standards

Vivian A. Auld, MLIS
Senior Specialist for Health Data Standards
National Library of Medicine, NIH, HHS

Medical Library Association
Annual Meeting
May 23, 2004



Data Content Standards Include:

- Data elements, e.g., gender, presenting complaint
- Descriptions of entities, e.g., birth certificate
- Messages, e.g., lab test order
- Allowable values for data elements, which can be entire *vocabularies*
- Mappings between different value sets, e.g., between SNOMED and ICD-9-CM
- Information models that define the context in which standards are used
- Survey questions and any coded responses
- Guideline, protocol, and algorithm formats

Electronic Health Data Standards and Standard Vocabularies

- Key element of the information infrastructure
- Needed for:
 - Safe and effective health care
 - Efficient clinical and health services research
 - Timely public health and bioterrorism surveillance
- Public health requirements must be fed into the standards development process
- Public health must participate in testing and promoting the use of standards

Barriers to Standardized Health Data

- Lack of interest among system buyers
- Competing candidates and no “perfect” candidate, especially in vocabulary area
- Presumed increased costs of data creation
- No hard data on value of vocabulary standardization
 - E.g., for which data elements is there a real payoff?

Recommended Steps to Achieving U.S. Health Data Standards (1990-2003)

- ✓ Establish a mechanism for designating U.S. Standards - HIPAA, NCVHS, CHI
- ✓ Pick best available as starting point - NCVHS, CHI
- Broaden participation in standards development
- *Support development, maintenance, and low/no cost distribution
- Promote use and improvement
- Coordinate development of selected standards to achieve non-overlapping, interlocking set

Support Development, Maintenance, and Low/No Cost Distribution – *NLM-led Actions*

- **1999** – LOINC (lab tests/instrument observations) - contract support
- **2002** – RxNorm (clinical drugs) - direct development
- **2003** – SNOMED CT contract & license for U.S-wide use (*as distributed by NLM in UMLS*)
- **2004** – HL7 (messaging) contract support to
 - Align HL7 messages with CHI vocabularies
 - Create implementation guide(s) for the use of HL7 to transmit electronic health record (EHR) data

Accelerate Adoption and Use

- Promote use of standards in NIH-funded clinical research networks
 - NIH Roadmap priority: *Re-engineering the clinical research enterprise*
- Encourage manufacturers to include LOINC in device output/test kit packaging
 - Need help to identify manufacturers important to public health community
- Collaborate with other HHS agencies to support demonstration/testing

Types of Health Data

- Administrative health data
E.g., health insurance claims, billing, ordering
- Clinical data
E.g., lab test results, problems, diagnoses, history and physical
- Public health data
E.g., disease surveillance, immunization rates, environmental monitoring

Note: Health Services Research spans all three types of data.

Administrative Health Data

- Health Insurance Portability and Accountability Act of 1996 (HIPAA)
- Purpose is to:
 - Improve the efficiency and effectiveness of the health care system, by standardizing the electronic transmission of certain administrative and financial transactions
 - Protect the security and privacy of health information

Administrative Health Data (2)

- Regulations cover (effective dates):
 - Transactions and Code Sets (10/16/2003)
 - Privacy (4/14/2003)
 - Security (4/21/2005)
 - Claims Attachments (final rule expected 2004)
 - Unique Identifiers
 - Employers (7/30/2004)
 - Health Care Provider (final rule expected 2004)
 - Health Plan (final rule expected 2004)
 - Individuals (on hold indefinitely)

Clinical Data

- Enhanced ability to provide access to knowledge where clinical decisions are being made
- Generation of research data as a by-product of health care
- Efficient exchange of data between health care and public health


Clinical Data (2)

- National Committee on Vital and Health Statistics (NCVHS)
 - Recommended messaging and vocabulary standards for Patient Medical Record Information
- Consolidated Health Initiative (CHI)
 - Adopt existing clinical vocabulary & messaging standards to enable interoperability *in the federal health care enterprise*
- Public/private collaborations recommending standards
 - Markle Foundation's Connecting for Health
 - EHealth Initiative

Public Health Data

- National Electronic Disease Surveillance System (NEDSS)
 - Facilitates electronic transfer of appropriate information from clinical information systems to PH departments
- Public Health Information Network (PHIN)
 - Enables consistent exchange of response, health, and disease tracking data between PH partners
- Consolidated Health Informatics (CHI)
 - Evaluated existing vocabularies for several domains important to PH including population health, immunization, etc.
- National Library of Medicine (NLM)
 - Partnering with other federal agencies to identify vocabularies relevant to environmental health

Clinical Versus Public Health Concepts

Clinical  Public Health

- Anthrax
- Asthma
- Influenza

- Fever
- Diabetes

- Anthrax
- Asthma control
- Adult influenza vaccination schedule
- Fever >104
- Diabetes prevention

Unified Medical Language System® (UMLS®) Metathesaurus® - *a Vocabulary Database*

- **Preserves** the **meanings**, hierarchical connections, and other relationships between terms present in its source vocabularies
- **Adds** certain basic definitional **information** about each of its concepts
- **Establishes new relationships** between concepts and terms from different source vocabularies
- **Distributes many vocabularies in a common, explicit format**

UMLS Source “Vocabularies”

- Widely varying purposes, structures, properties, but all are in essence “sets of valid values” for data elements:
 - HIPAA code sets, e.g., ICD-9-CM, CPT
 - NCVHS and CHI recommended clinical standards, e.g., SNOMED CT, LOINC, RxNorm
 - Thesauri for biomedical literature, e.g., MeSH
 - Other lists of controlled terms, e.g., HL7 value sets

2004AA UMLS Metathesaurus (May 2004)

- ~1,022,000 concepts
- ~2,383,000 unique “strings”
(Eye, Eyes, eye = 3)
- ~3,030,000 source vocabulary terms
- 111 source vocabularies
- 15 different languages
 - Includes *active* SNOMED CT (English)
 - Spanish will appear in 2004AB; *inactive* in 2004AC

Expedite Enhancement, Coordination, Alignment, & Dissemination of Standards

- UMLS® Metathesaurus® – common distribution format/mechanism for CHI standards and HIPAA code sets
- 2004AA - UMLS distribution format changes to support
 - Complete “Source Transparency”
 - Easier extraction of subsets for particular purposes
 - Complete “change sets” from previous versions
 - Sophisticated, purpose-specific inter-vocabulary mapping

Electronic Health Record (EHR)

- EHRs currently in use are for closed systems
 - Intermountain Health Care, Utah
 - Regenstrief Medical Record System, Indianapolis, IN
- Need to develop standards so records can be shared across systems
 - HL7 just successfully balloted standard for EHR – first step in the process
- NLM contract with HL7 to create implementation guide(s) for the use of HL7 to transmit EHR data
- Widespread use of EHRs will facilitate collection of PH data in a timely manner and enhance tools for disease control and prevention

National Health Information Infrastructure (NHII)

- Initiative to improve the effectiveness, efficiency and overall quality of health and health care in the United States
- Comprehensive knowledge-based network of interoperable systems that would improve decision-making by making health information available when and where it is needed
- NOT a centralized database of medical records or a government regulation

Development of a PH Vocabulary

- CDC Web Redesign Controlled Health Thesaurus
 - Good example of the process needed for developing a PH Vocabulary
 - Developed to meet the needs of the CDC for their Web site redesign
 - Will need input from PH community in order to become a PH vocabulary

CDC Process

- Analyzed CDC Website for PH concepts
 - 19,000 unique concepts identified
 - 11,000 not matched to existing terminology
- Best matches:
 - MeSH - provided the highest number of matches
 - CRISP
 - Alcohol & Other Drugs Thesaurus
- SNOMED RT had 4,400 concept matches, many overlap with MeSH
- Demonstrated clear need for PH specific terminology

CDC Process (2)

- Imported a good portion of MeSH to form initial framework of vocabulary tool
- Structure adopted from MeSH only shows clinical perspective – need to adjust to show PH perspective. E.g.:
 - For obesity – define a relationship between:
 - Obesity and BMI or CDC 5-a-day program
 - For diabetes (or other diseases) – define a relationship between:
 - Diabetes and prevention programs or specific measurements
- Evaluating top 3 levels of the hierarchical structure to see if it provides a complete framework
- Creating placeholders for what needs to be added

CDC Process (3)

- Looking at concepts extracted from CDC programs and services
- Evaluating search logs of CDC Web site – what are people looking for?
- Matching site map for new CDC Web site against the Thesaurus, adding concepts as identified.
 - CDC migrating information by topic – fill in concepts as new topic pages developed
- Ongoing process

Future Steps for CDC

- Continue enhancing and refining Thesaurus
- Determine the amount of overlap with SNOMED CT
- Fall 2004
 - Begin talking with vocabulary developers about enhancing their vocabulary with concepts from CDC Thesaurus (MeSH, SNOMED CT, etc.)
 - Talk with NLM about including what remains in NLM's UMLS Metathesaurus

Future for Public Health

- Public health requirements must be fed into the development process
- Public health must participate in testing and promoting the use of all elements of the developing systems
- Involvement in the Public Health Data Standards Consortium (PHDSC) is a good place to start

Additional Information

- CDC Web Redesign Controlled Health Thesaurus
 - Mamie Bell
CDC/Information Resources Management Office
email: zvt9@cdc.gov
- PHIN Standards (www.cdc.gov/phin)
- Consolidated Health Informatics (CHI)
 - www.whitehouse.gov/omb/egov
Choose “Government to Business”
- National Committee on Vital and Health Statistics (NCVHS)
 - ncvhs.hhs.gov

Contact Information

Vivian A. Auld, MLIS

Senior Specialist for Health Data Standards

National Information Center on Health Services

Research and Health Care Technology

National Library of Medicine, NIH, HHS

8600 Rockville Pike

Bethesda, MD 20894

auld@nlm.nih.gov