Data Content Standards

Keith W. Boone
GE Healthcare
Data Content Standards

- Key concepts
- Features of terminology systems
- Terminologies in use today
- Tools and technologies
Objectives

- By the end of this session, you should be able to:
  - Define key concepts related to data standards
  - Describe features of various terminology systems
  - List several common terminologies used in health care IT and explain for what they are used
Key Concepts

- Concept
  - An idea represented or identified by a code

- Code
  - A unique string identifying an idea

- Coding system
  - A collection of codes associated with concepts

- Vocabulary/terminology
  - A collection of words and phrases; in the health care IT context, these are associated with codes

- Ontology
  - An exhaustive organization of knowledge about particular entities and their relationships within a particular domain
Key Concepts

- **Primary term/preferred term**
  - The word or phrase best representing a concept in a coding system
  - Usually the most common or more accurate

- **Alternate term**
  - A synonymous phrase representing a concept in a coding system
  - Less common or accurate than the preferred term, may link to multiples
  - For example, *MI* for *myocardial infarction* and *mental illness*
Key Concepts

- **Hypernym**
  - A more generic term or abstract concept
  - *Ball* is a hypernym of *football*

- **Hyponym**
  - A more specific term
  - *Soccer ball* is a hyponym of *ball*
Key Concepts

- **Subsumption**
  - The code for a hypernym contains all and codes for all (subsumes all) of its children

- **Pre-coordination**
  - A pre-coordinated concept is one in which all features of the concept are identified by a single code
  - For example, LOINC codes are all pre-coordinated

- **Post-coordination**
  - Terminologies supporting post-coordination allow for multiple codes from the coding system to be used to define concepts not explicitly listed in that system
Key Concepts

- **Value set**
  - A collection of codes from one or more coding systems representing a fixed set of ideas which concretely defines a concept domain

- **Concept domain**
  - The definition of a set of codes representing a particular collection of related ideas
  - An abstraction used in HL7 Version 3 to describe the purpose or function of a value set or coding system and to constrain the codes used in communication
Key Concepts

- **Hierarchy**
  - Referring to the structure of a coding system where each concept has at most a single hypernym or higher-level concept

- **Polyhierarchy**
  - Referring to the structure of a coding system where each concept can be related to more than one higher-level concept

- **Cross-walk**
  - A set of equivalence relationships between codes in different coding systems
  - Usually supporting mapping on one direction only

- **Terminology service**
  - An application providing programmatic access to codes and terms from one or more coding systems
Best Practices

- Versioning (configuration control)
  - Code systems must be maintained, and changes must be traceable, to a given version of the code system used

- Codes are *never* reused
  - If in 2009 code X meant A, and in 2010 it meant B, systems assuming X meant B would interpret the code incorrectly
  - Best practice is that codes are never reused to prevent this from occurring
  - This practice has not always been applied to ICD-9-CM
Best Practices

- Terms and phrases have a definition
  - Short phrases and single terms don’t always have an agreed-upon definition
  - Consider *diabetes*
    - Does it include Type I and Type II?
    - What about gestationally induced?
    - What about other pancreatic disorders that are treated similarly?
Best Practices

- Codes for other not present
  - NOS: not otherwise specified
    - Used in some coding systems to create a code for a general case where specific details required for further classifying the idea being coded have not been provided, and so a more detailed code cannot be given
  - NEC: not elsewhere classified
    - Used in some coding systems to create a code for cases where specific details to further classify the idea being coded have been provided, but a more detailed code is not available
Best Practices

- Explore what happens to the meaning of these concepts when a new code is added in Year 2

- The meanings that the concepts encode change (NEC), or the statistical populations change (NOS), and yet the code will have stayed the same
Relationships

- **Hierarchical coding systems** support IS-A relationships through hypernomy
  - For example: 410.9 Myocardial Infarction of Unspecified Site IS-A 410 Myocardial Infarction

- **Ontological coding systems** support IS-A relationships through explicit relationship links and may support other kinds of relationships
  - Whole-part relationships (drug and ingredients, body parts)
  - Directionality/bilaterality (left/right)
  - Severity, onset, chronicity ...

- There may be a fixed set of relationships (for example, RxNORM)

- Set of relationships may itself be coded (for example, SNOMED-CT)
Some coding systems embed the position of the term in the code hierarchy

- ICD-9-CM
  - Chapters
  - 3-digit code
  - 1-2 digit suffix
  - Common patterns

- Health care provider taxonomy
  - Provider type (~2 digits)
  - Classification (~2 digits)
  - Specialty (5 digits)
Vocabulary-Based Interfaces

- Additional vocabulary
  - SNOMED CT used preferred and alternate terms
  - ICD-9-CM has an extensive index of synonyms
  - LOINC contains numerous alternate names

- Access to codes through synonyms requires a search index

- Many applications support search

- Terminology services provide programmatic access to search