The Power of Ontologies and Standardized Terminologies for Capturing Clinical Knowledge

Peter E. Gabriel, MD, MSE
AAPM Annual Meeting | July 15, 2015

Penn Radiation Oncology
Penn Medicine

THE CHALLENGE

Amazon, Google & Facebook have "Big Data" problems…

…in Oncology we have a "Small Data" Problem!

The “Small Data” Problem

1. Structured clinical data is lacking; much valuable clinical info is trapped in free text notes and reports

2. Oncology is changing rapidly:

↑ precision in disease characterization and risk stratification

\+ ↑ specificity of treatment

= ↓ patients available at any one institution to:
- test research hypotheses
- develop predictive models

Combining EHR data across institutions is hard

Syntactic Interoperability = the ability for systems to exchange data

Semantic Interoperability = the ability for systems to exchange data **with unambiguous, shared meaning**

An example of syntactic interoperability:

The **DICOM-RT** standard tells us where information goes, but not what it means

>ROI Number	(3006,0022)	1C	Identification number of the ROI. The value of ROI Number (3006,0022) shall be unique within the Structure Set in which it is created. Required if Structure Set ROI Sequence (3006,0020) is sent.
>Referenced Frame of Reference UID	(3006,0024)	1C	Uniquely identifies Frame of Reference in which ROI is defined, specified by Frame of Reference UID (0020,0052) in Referenced Frame of Reference Sequence (3006,0010). Required if Structure Set ROI Sequence (3006,0020) is sent.
>ROI Name	(3006,0026)	2C	User-defined name for ROI. Required if Structure Set ROI Sequence (3006,0020) is sent.
>ROI Description	(3006,0028)	3	User-defined description for ROI.
>ROI Volume	(3006,002C)	3	Volume of ROI (cubic centimeters).

“Total_Lung” vs. “LUNG_TOTAL” vs. “LUNG_TOT-GTV” vs. “LUNG_L” vs. “RT_LUNG” vs. ….

An example of semantic interoperability:

The **RxNorm** standard can allow a computer to understand how these three concepts are related:

Tylenol 325mg PO tabs
Acetaminophen 500MG oral tablets
NyQuil 2tbsp Oral Liquid

Standards for representing clinical information with explicit meaning are the key to semantic interoperability.

Penn Radiation Oncology 7

REPRESENTING CLINICAL KNOWLEDGE

Common ways to represent knowledge:

	Pros	Cons
Natural Language	• Captures nuance, detail • Infinite expressive range • Natural, familiar, established • No maintenance, never obsolete	• Ambiguous • Imprecise • Unpredictable
Codes	• Concise, precise • Structured, consistent • Analyzable, interpretable	• Rigid, tedious, cumbersome • Harder to capture nuance/detail • High maintenance

Credit: Chris Chute, MD

Penn Radiation Oncology 9

Some different names for coding systems, or, the "terminology of terminologies:"

Coding
Terminology
Controlled Vocabulary
Nomenclature
Thesaurus
Taxonomy
Classification
Grouping
Ontology

"Terminologies" in simplest form are:

- Finite, enumerated sets of terms intended to convey information unambiguously
- Each "term" is a symbolic representation of a single concept; usually code + name/description

A familiar example:

International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM):

162.0 = Malignant neoplasm of trachea
162.2 = Malignant neoplasm of main bronchus
162.3 = Malignant neoplasm of upper lobe, bronchus or lung
…etc.

Design considerations / sophistication

- Formal definitions of terms, or just labels?
- Redunancy – how are synonyms handled?
- Relationships between terms:
  - Explicit or implicit?
  - Hierarchical? Polyhierarchy or strict hierarchy?
  - Other non-hierarchical relationships?
- Are axioms supported?
  - E.g. the "has laterality" relationship does not apply to unpaired organs

Design considerations (cont.)

- Granularity and “pre-coordinated” vs. “post-coordinated” concepts

Example:

Pre-coordinated: third degree burn of left index finger caused by hot water

Post-coordinated: burn of skin
- morphology: third degree burn injury
- finding site: index finger structure
- laterality: left
- causative agent: hot water

Credit: Wikipedia

Ontologies are the most sophisticated and powerful type of terminology

– Formal definitions
– Explicit relationships
– Logical axioms that support inference

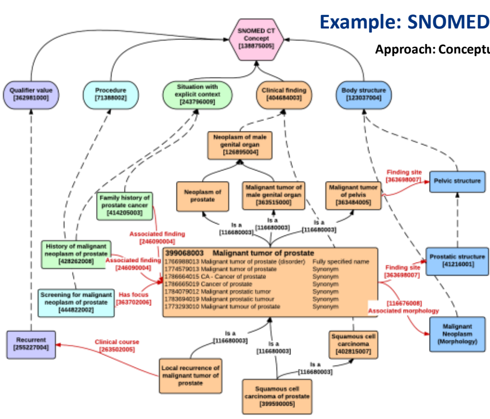
Different Approaches:

Conceptual	• model according to concepts of interest • no inherent constraints, only those defined
Realism	• model based on what exists in reality • constraints should not diverge from real world
Upper Level	• ‘thing’ • break down terms along disjoint axes

Credit: Chris Stoeckert, Ph.D.

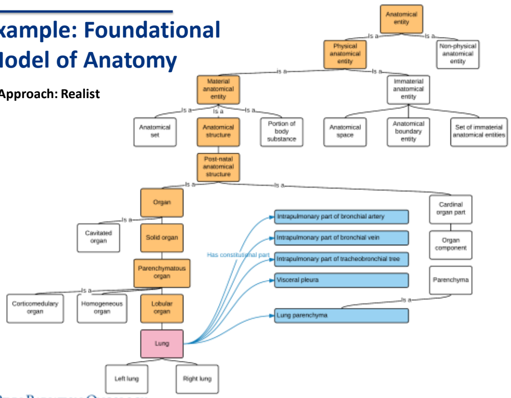
Example: SNOMED-CT

Approach: Conceptual



Example: Foundational Model of Anatomy

Approach: Realist



OBO Foundry Ontologies

RELATION TO TIME	CONTINUANT						OCCURRENT	
GRANULARITY	INDEPENDENT CONTINUANT			DEPENDENT CONTINUANT		INFORMATION ARTIFACT		
ORGAN AND ORGANISM	Organism NCBI Taxonomy	Anatomical Entity (FMA, CARO)	Environment Ontology (ENVO)	Organ Function (FMP, CPRO)	Phenotypic Quality (PATO)	IAO Software, Algorithms, …	Biological Process (GO)	OBI
CELL AND CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)		Cellular Function (GO)		Sequence Data, EHR Data …		
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)			Molecular Function (GO)		Images, Image Data, Flow Cytometry Data, …	Molecular Process (GO)	OBI: Imaging

Credit: Barry Smith. Ph.D.

A Practical Use Case

What chemotherapy drugs has a patient received?

- Penn Medicine Data Warehouse:
  - 363,484 unique medication records from inpatient and outpatient EHR dictionaries - which are chemo agents??
  - All records were mapped to concepts in an external terminology / thesaurus (3M™ Healthcare Data Dictionary)
  - A query starting from high-level concept "Antineoplastic Drugs" traverses many is_member / has_member and is_ingredient / has_ingredient relationships to find:
    - 2,049 EHR medication records representing…
    - 120 unique chemotherapeutic drug ingredients
  - Can now easily find which of 120 agents a RadOnc patient has received
  - Can share data using RxNorm or other standard concepts

This is hard work! **But this is very powerful!**

Take-Home Points

- The potential for knowledge-based planning depends on our ability to meaningfully capture and combine structured clinical data from heterogeneous sources
- Using sophisticated, standardized terminologies to represent clinical data is the key to making it usable
- Excellent terminologies and ontologies exist, but need to be extended to cover the RadOnc domain more fully
- Grass-roots effort is needed make better use of terminologies in local systems and workflows