The Power of Ontologies and Standardized Terminologies for Capturing Clinical Knowledge

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

THE CHALLENGE

Amazon, Google & Facebook have “Big Data” problems...

...in Oncology we have a “Small Data” Problem!

Penn Radiation Oncology
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:

\[ \text{precision in disease characterization} \uparrow \quad \text{risk stratification} \uparrow \quad \text{specificity of treatment} \]

\[ \quad = \quad \downarrow \text{patients available at any one institution to:} \]

- 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

```
"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.

REPRESENTING CLINICAL KNOWLEDGE

Common ways to represent knowledge:

<table>
<thead>
<tr>
<th>Natural Language</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Captures nuance, detail</td>
<td>• Ambiguous</td>
</tr>
<tr>
<td></td>
<td>• Infinite expressive range</td>
<td>• Imprecise</td>
</tr>
<tr>
<td></td>
<td>• Natural, familiar, established</td>
<td>• Unpredictable</td>
</tr>
<tr>
<td></td>
<td>• No-maintenance, never obsolete</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codes</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Concise, precise</td>
<td>• Rigid, tedious, cumbersome</td>
</tr>
<tr>
<td></td>
<td>• Structured, consistent</td>
<td>• Harder to capture nuance/detail</td>
</tr>
<tr>
<td></td>
<td>• Analyzable, interpretable</td>
<td>• High maintenance</td>
</tr>
</tbody>
</table>

Credit: Chris Chute, MD
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

Design considerations / sophistication

- Formal definitions of terms, or just labels?
- Redundancy – 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:**
  - morphology: third degree burn injury
  - finding site: index finger structure
  - laterality: left
  - causative agent: hot water

---

**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

---

**Example: SNOMED-CT**

Approach: Conceptual

---

**Credit:** Wikipedia

---

**Credit:** Chris Stoeckert, Ph.D.
Example: Foundational Model of Anatomy

Approach: Realist

OBO Foundry Ontologies

<table>
<thead>
<tr>
<th>RELATION TO TIME</th>
<th>CONTINUANT</th>
<th>OCCURRENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRANULARITY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORGAN AND ORGANISM</td>
<td>Organism, NCBi Taxonomy</td>
<td>Anatomical Entity (FMA, CARO)</td>
</tr>
<tr>
<td>CELL AND CELLULAR COMPONENT</td>
<td>Cell (E.U.)</td>
<td>Cellular Component (FMA, GO)</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td>Molecular Function (GO)</td>
</tr>
</tbody>
</table>

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 to make better use of terminologies in local systems and workflows

Penn Radiation Oncology