Penn Medicine

Session: Ontologies Supporting Diagnostic Imaging, Therapy and Science

An Ontology, a Terminology, and a Data Dictionary Walk Into a Bar

Peter E. Gabriel, MD, MSE

AAPM | COMP Annual Meeting
July 12, 2020

"The world's most valuable resource is no longer oil, but data"

The Economist
May, 2017

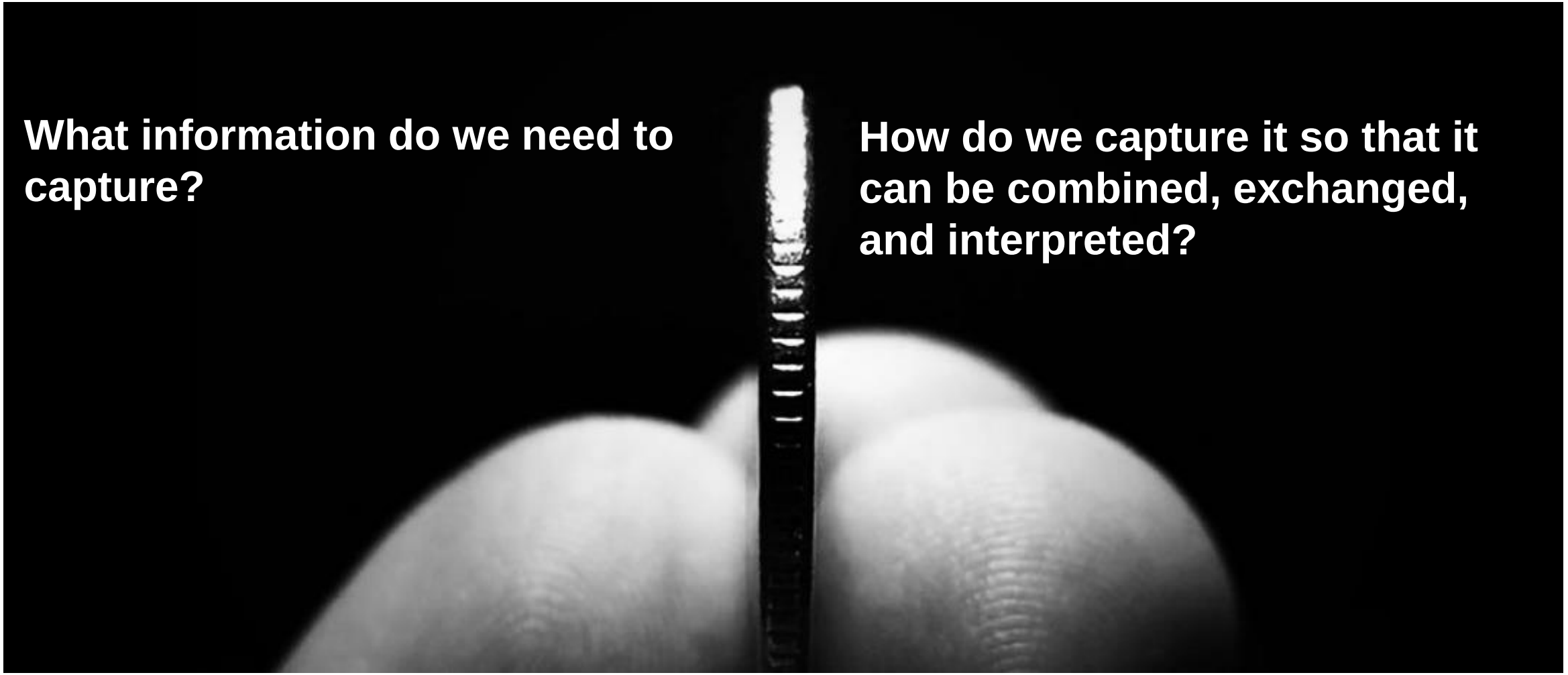

“We need an ontology…”

Multiple Oncology Data Standardization Efforts

- AAPM
  - Data Sciences Committee
    - Big Data Subcommittee
      - Ontology Working Group
  - TG 263: Standardizing Nomenclatures in Radiation Oncology[1]
- ASTRO
  - Minimum Data Elements[2]
- ACoS Commission on Cancer
  - Synoptic Radiation Treatment Summary[3]
- MITRE / ASCO / ASTRO / CMS / FDA / etc.
  - Minimum Common Oncology Data Elements (mCODE)[4]
- DICOM-RT
- IHE-RO
- ROIS and Treatment Planning System Vendors
- Others?

1. Mayo CS, et al. *Int J Radiat Oncol Biol Phys.* 2018 Mar;100(4):1057-1066
2. Hayman JA, et al. *Pract Radiat Oncol.* 2019 Nov;9(6):395-401
3. Christodouleas JP, et al. Pract Radiat Oncol. 2020 Jan 24;S1879-8500(20)30002-3.
4. https://mcodeinitiative.org/

Two Problems – Two Sides of the Same Coin



Standardizing Data Elements

- What are the entities about which we need to capture facts or attributes?
  - E.g. patient, diagnosis, treatment course, treatment session, etc.
- What are the facts or attributes we intend to capture?
  - E.g. date of birth, diagnosis code, course start and end dates, etc.
- What are the cardinal relationships between the entities?
  - E.g. a patient can have multiple diagnoses and multiple treatment courses; a course can have multiple sessions, etc.

- These details can be specified in a data dictionary (aka data model, schema)
  - Can be more or less detailed/technical (conceptual → logical → physical)

Radiation course summary	
Treatment indication	{free or structured text}
Course start date	{YYYY-MM-DD}
Course end date	{YYYY-MM-DD}
Concurrent systemic treatment?	{Yes/No} *{optional free text comments}*
RT course discontinued early?	{Yes/No} *{optional free text comments}*
Patient experience	{free or structured text}
Follow-up plan	{free or structured text}
Comment	{optional free text}

Anatomic target summary

Anatomic target	**Modality**	**Cumulative dose (cGy)**	**Delivered prescriptions**
{free or structured text}	{free or structured text}	{#}	{#}
[+]			
Comment	{optional free text}		

Delivered prescription summary

	1	**2**	**3**	**4**	**5**	[+]
Start day (date (session))	{YYYY-MM-DD} ({#})					
Dose per Fx (cGy)	{#}					
No. of Fx	{#}					
Total dose (cGy)	{#}					
Technique	{free or structured text}					
[+]						
Comment	{optional free text}					

Data element	**Definition**	**Detail**
Treatment course data elements		
Diagnosis	Identify disease(s) relevant to treatment	ICD-10
Modality	Radiation type - Records the list of all modalities used during treatment course (Check all that apply)	Reference Table 2 for detail
Technique	Treatment delivery method - Records the list of all techniques used during treatment course (Check all that apply)	Reference Table 2 for detail
Number of fractions planned	Records the total number of treatments prescribed in a treatment	
Number of fractions delivered	Records the total number of treatments delivered in a treatment course	
Start date of treatment	Indicates the date on which the patient commences course of delivered radiation treatment	MMDDYYYY
End date of treatment	Indicates the date on which the patient ends or completes a course of delivered radiation treatment	MMDDYYYY
Prescribed dose-level data elements (Note: Multiple dose levels are possible for a given treatment. The following elements are completed for each dose level.)		
Anatomic site of each prescribed dose level	Indicates the primary anatomic site(s) targets for each dose level	Reference the Standards for Oncology Registry Entry (supplementary material, available online at https://doi.org/10.1016/j.prro.2019.07.017)
Total dose planned for each prescribed dose level	Dose prescribed to each dose level	cGy
Total dose delivered for each prescribed dose level	Dose delivered to each dose level	cGy

Standardizing Meaning

"Semantic Interoperability"

- the ability for systems to exchange data with *unambiguous, shared meaning*

Example – what <u>are</u> radiation "modality" and "technique??" Some possibilities:

- 3D conformal radiotherapy (3DCRT)
- 3DCRT with electronic tissue compensation
- 3DCRT with merged subfields
- Brachytherapy
- CyberKnife
- External beam radiotherapy (EBRT)
- Gamma knife
- Halcyon
- High dose rate brachytherapy (HDR)
- High density total skin electron therapy (HDTSE)
- Intensity modulated radiation therapy (IMRT)
- Interstitial brachytherapy
- Intracavitary brachytherapy
- Intraluminal radiation therapy
- Low dose rate brachytherapy (LDR)
- Neutron therapy
- Orthovoltage
- Photon radiotherapy
- Proton double scattering
- Proton pencil beam scanning (PBS)
- Proton therapy
- Proton uniform scanning
- Pulsed dose rate brachytherapy
- Static arc therapy
- Stereotactic body radiotherapy (SBRT)
- Stereotactic radiosurgery (SRS)
- Tomotherapy
- Total body irradiation
- Volume modulated arc therapy (VMAT)

Knowledge Representation

Reality

“Representational Artifacts”

- Classification
- Codeset
- Controlled vocabulary
- Dictionary
- Metathesaurus
- Ontology
- Standardized nomenclature
- Taxonomy
- Terminology
- Thesaurus

Terminology

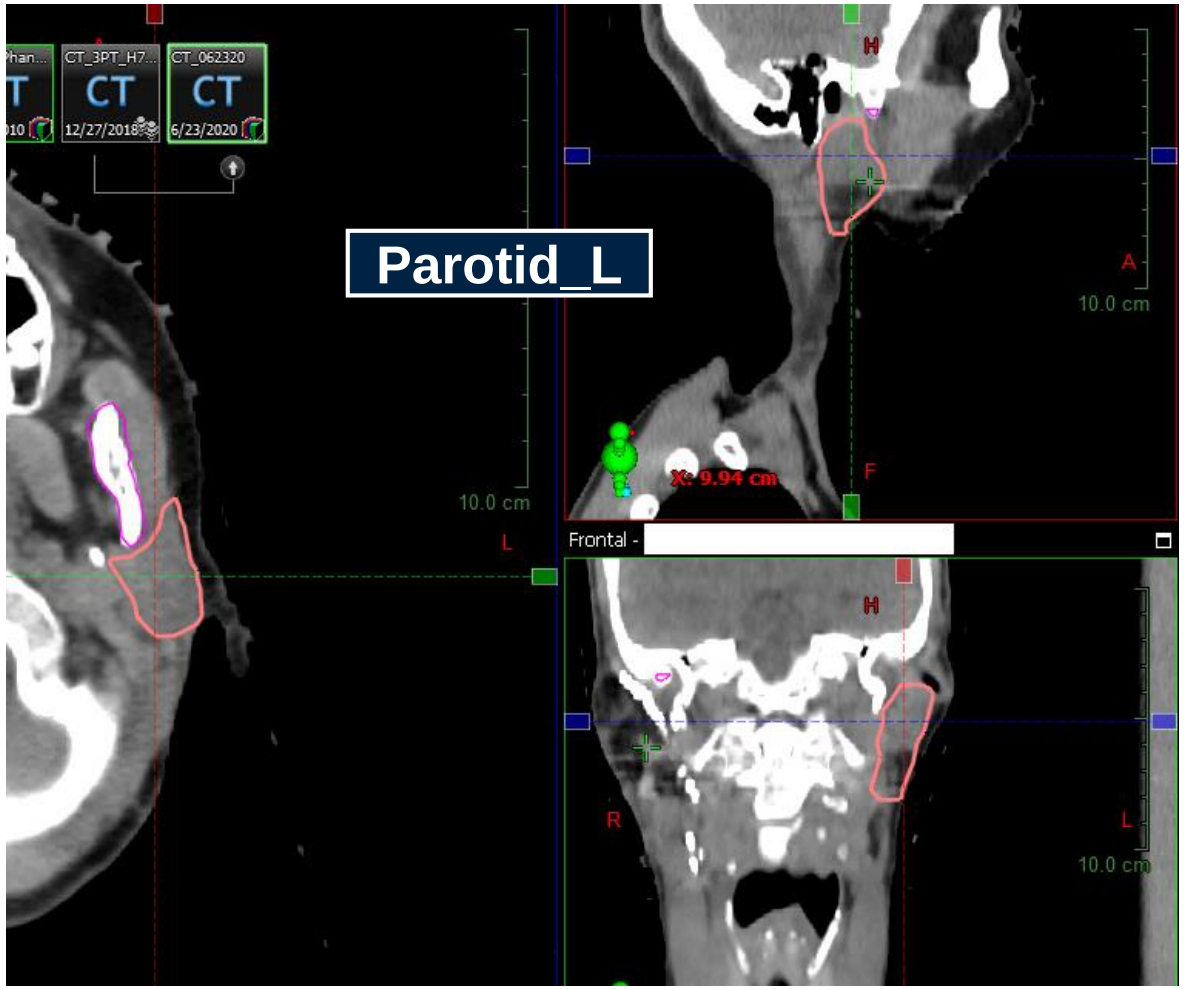
"Now! *That* should clear up a few things around here!"

- Characteristics:
  - List of unique *terms* (nouns or noun phrases +/- codes) used to represent things that exist in reality (entities)
  - One term per distinct entity (can include synonyms designated as such)
- Benefits:
  - Helps refer to things consistently
- Lacks:
  - Explicit definitions of the entities referred to
  - Specification of relationships between terms

Terminology (continued)

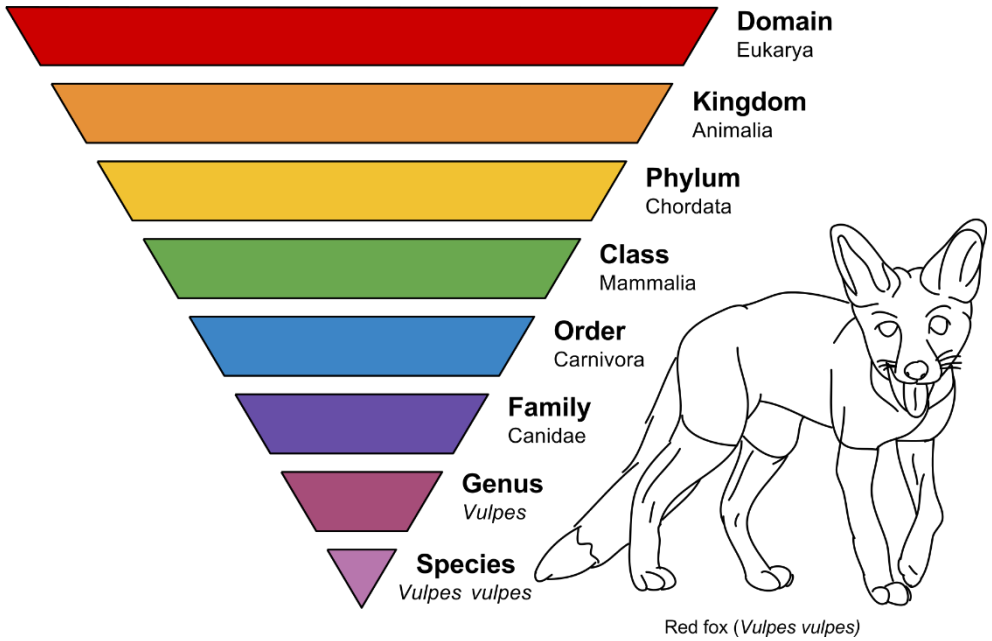
- Examples:
  - ICD-10-CM diagnosis codes*
  - AAPM Task Group 263 recommended target and structure names

- Best uses:
  - As metadata (annotating other data)
  - As permissible values for a structured data element *when complex reasoning on the captured data will not be necessary*



* Includes some hierarchical classification

Taxonomy (terminology+)



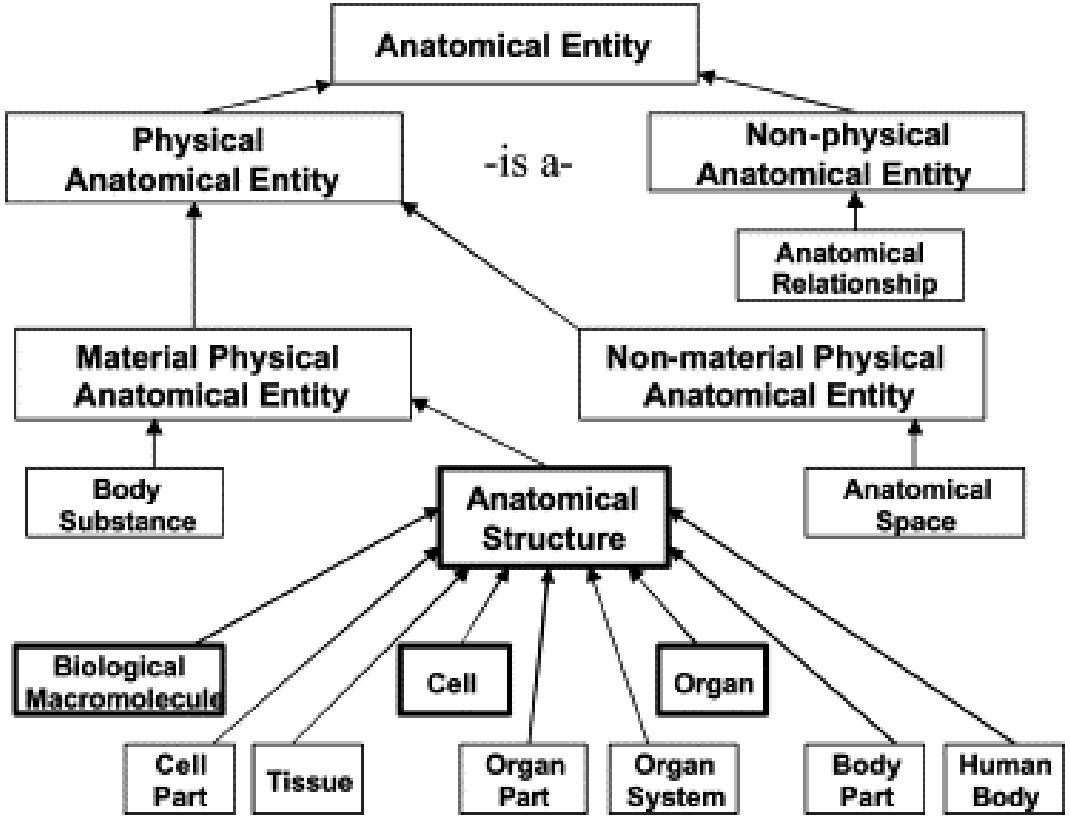
Example: Biological taxonomy

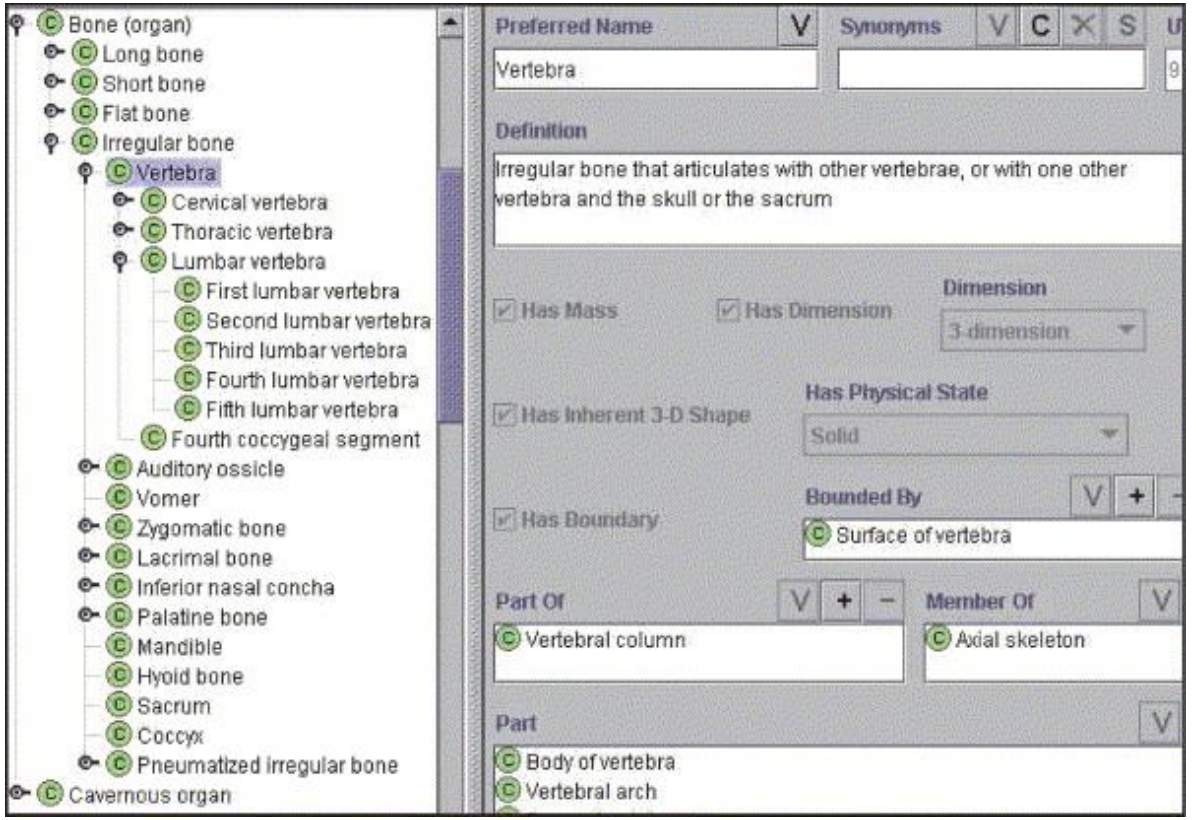
- Characteristics:
  - A terminology with *hierarchical classification* – i.e. “is-a” or “subclass” relationships expressed
  - Usually does not allow polyhierarchy
- Benefits:
  - Can enable “query expansion”
- Lacks:
  - Explicit definitions of the entities referred to
  - Other relationships between terms

Ontology (taxonomy+ / terminology++)

- Characteristics:
  - Includes a taxonomy as its proper part
  - Includes explicit, formal definitions of terms that distinguish them from each other
  - Defines other types of relationships between terms beyond the “is-a” subclass relationship
- Benefits:
  - Can be used to represent a tremendous amount of “knowledge” about a domain
  - Can enable a computer to make inferences and discover facts beyond those that are directly supplied in the data
    - Example of inference: Raw data: Flipper is a dolphin
      Ontology: All dolphins are mammals
      Computer can infer: Flipper is a mammal

Ontology Example: Foundational Model of Anatomy





Putting it all Together

- Data dictionaries:
  - Define the data elements (i.e. facts about an entity) that we want to collect
    - Keeping these to a minimum set (a la MCODE or ASTRO minimum data elements) is helpful!
  - Can also help define the way entities relate to each other via the *data model* (or schema)
- Terminologies:
  - Help to standardize the meaning and ensure the interoperability / interpretability of the information captured
  - Are best used to constrain the list of permissible values for a data element, or to annotate other pieces of data as metadata
  - Depending on level of sophistication (terminology < taxonomy < ontology), can add benefits like:
    - precise and unambiguous definitions of potentially confusing entities (e.g. “modality”)
    - the potential for “query expansion”
    - the potential for automated inferencing to expand the “facts” that are captured

Returning to the Bar…

- Bartender: “What can I get you?”
- Terminology: “A beer.”
- Taxonomy: “A Maker’s Mark. Which is_a bourbon. Which is_a whiskey…”
- Ontology: “An alcoholic beverage resulting from the planned process of fermentation which has_inputs grapes and yeast; the grapes of which have_variety “cabernet sauvignon,” the resulting product of which has_commercial_brand Robert Mondavi.
- Data Dictionary: “Can I see the menu?”


Penn Medicine