



SNOMED Clinical Terms[®] Basics



SNOMED CT Basics

- Components and structure
- Post-coordination, Expressions & Description logic
- Terminology / Information model interaction
- Primary and Secondary uses (re-use) of data
- Extensions and Reference sets
- History tracking and Cross mapping



Building blocks

- **Concepts**
 - The anchors for meaning
- **Descriptions**
 - Terms (strings of readable characters) used to express the meanings of the concepts in human language
- **Relationships**
 - Concept-to-concept links used to express information in computer-processable language
 - First purpose: formal logical meanings
 - Other purposes: tracking retired concepts, representing “facts” that may vary, and supporting post-coordination by suggesting valid qualifiers



Codes, concepts, classes, instances

- **Code:**
 - In general, any sequence of characters used to represent something in a coding system
 - SNOMED Clinical Terms Identifier (SCTID):
 - A sequence of 6 to 18 digits that identifies a component
- **Concept:**
 - In general, an idea which has meaning. Through its meaning, a person can identify specific instances of the concept
- **Class:**
 - An abstract category of things sharing common features
- **Instance:**
 - A particular real member of a class

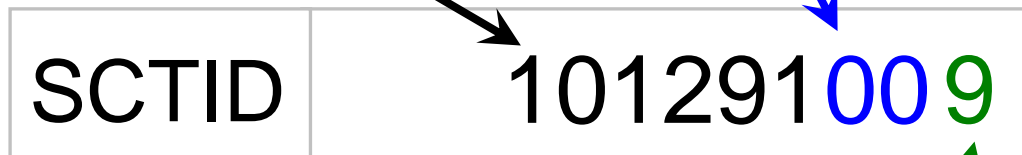


SCT Identifiers

Item identifier

Partition identifier

- 00 = Core namespace, concept
- 01 = Core namespace, description
- 02 = Core namespace, relationship



Check-digit



Concepts vs classes

- From a formal ontological standpoint
 - Concepts are in people's heads
 - Classes are abstract entities in the ontology
- Sometimes we speak of a SNOMED “concept” but really mean the code representing the class, instances of which can be identified by someone who understands the concept.



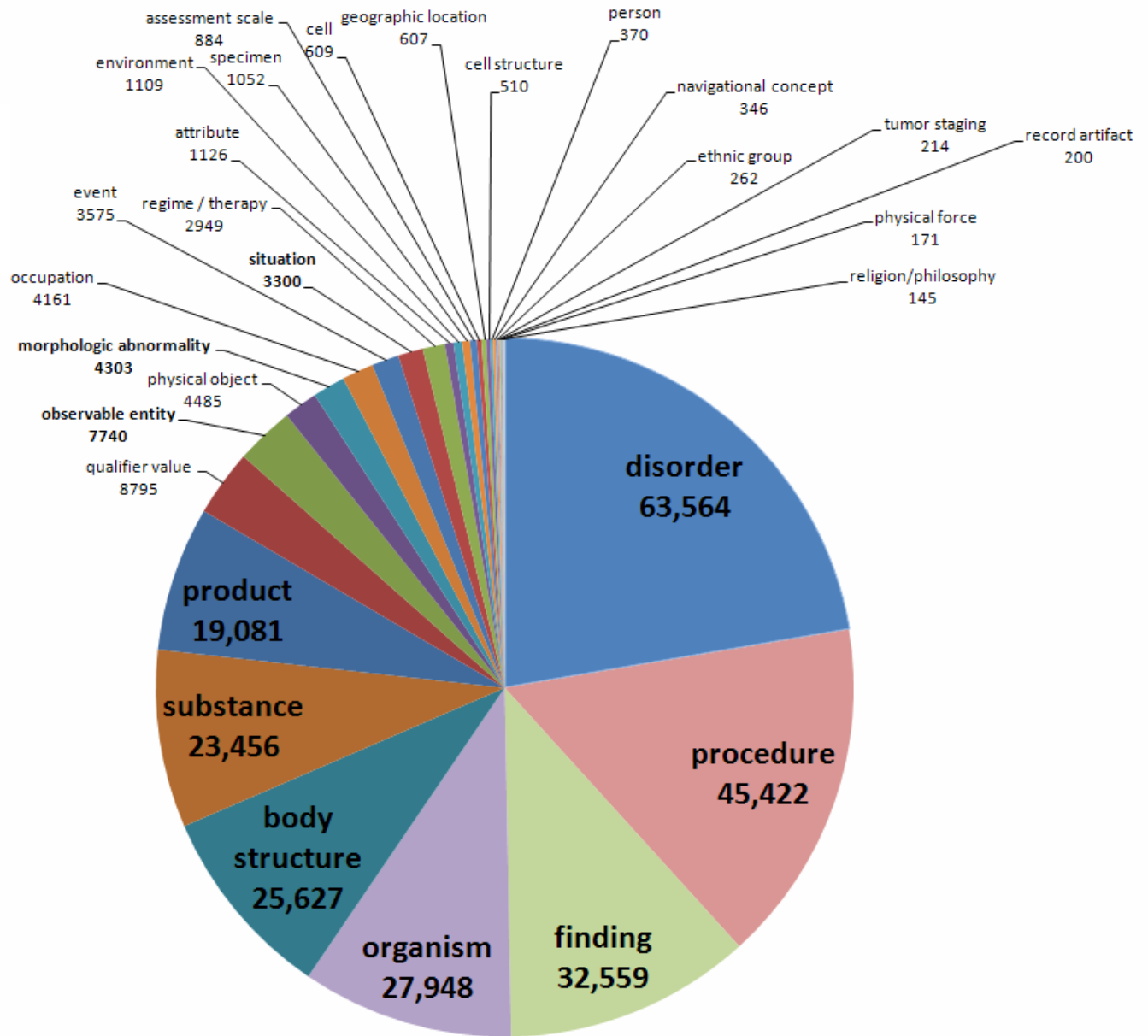
Types of concepts – by FSN tag

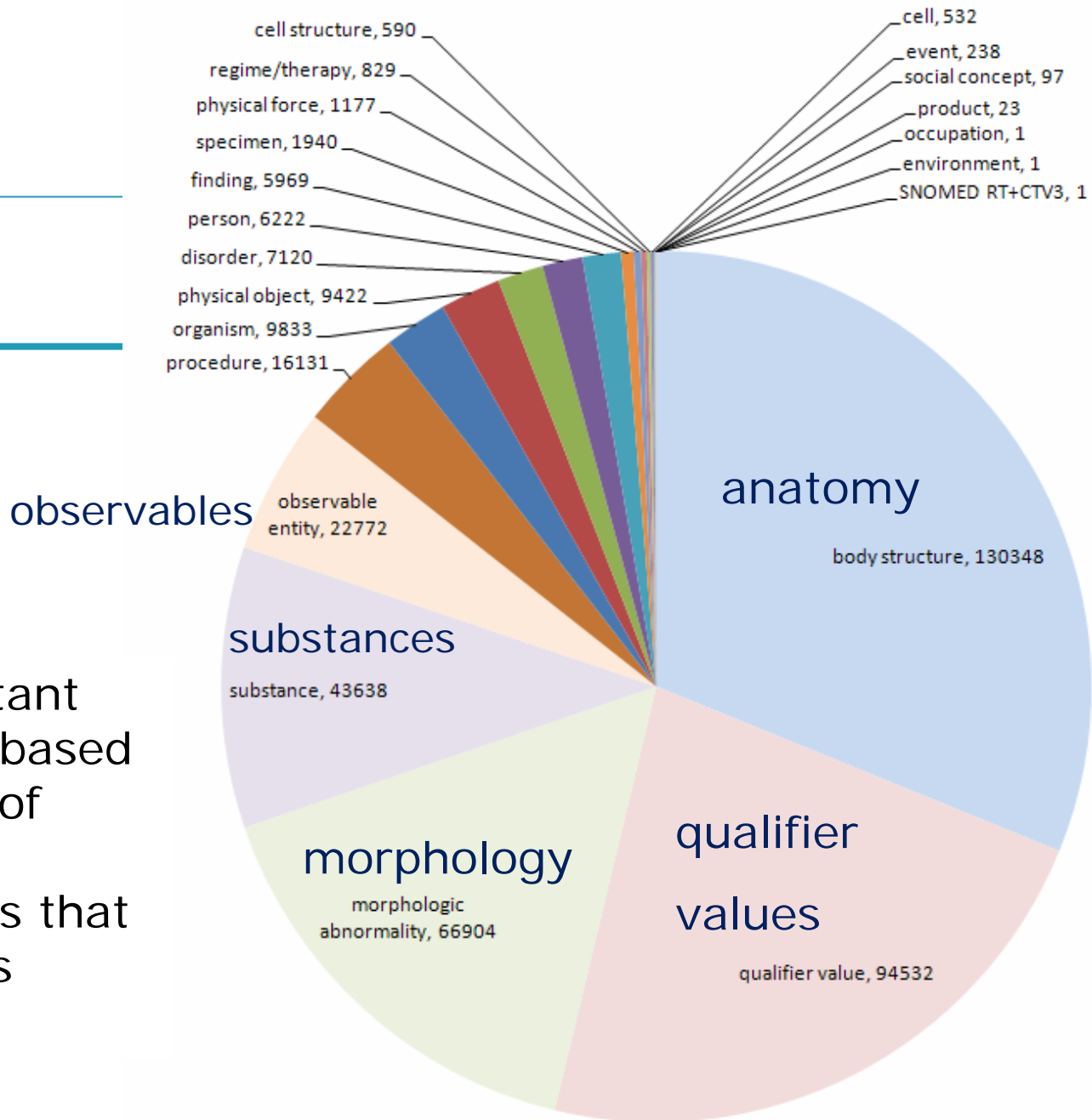
Current FSN tags (41 in January 2008 release):

TAG NAME :	Active
disorder	63564
procedure	45422
finding	32559
organism	27948
body structure	25627
substance	23456
product	19081
qualifier value	8795
observable entity	7740
physical object	4485
morphologic abnormality	4303
occupation	4161
event	3575
situation	3300
regime/therapy	2949
attribute	1126
environment	1109
specimen	1052
assessment scale	884
cell	609

TAG NAME:	Active
geographic location	607
cell structure	510
person	370
navigational concept	346
ethnic group	262
tumor staging	214
record artifact	200
physical force	171
religion/philosophy	145
namespace concept	79
administrative concept	28
social concept	26
life style	21
racial group	19
staging scale	15
link assertion	8
inactive concept	7
environment / location	1
linkage concept	1
special concept	1
SNOMED RT+CTV3	1

Active concepts
 Jan 2008 Release
 By FSN Tag





Most important hierarchies based on number of defining relationships that use them as values



“Special” hierarchies

- These may not be the source or the value of a defining relationship:
 - Linkage concept
 - Attribute
 - Link assertion
 - Special concept
 - Inactive concept
 - Namespace concept
 - Navigational concept

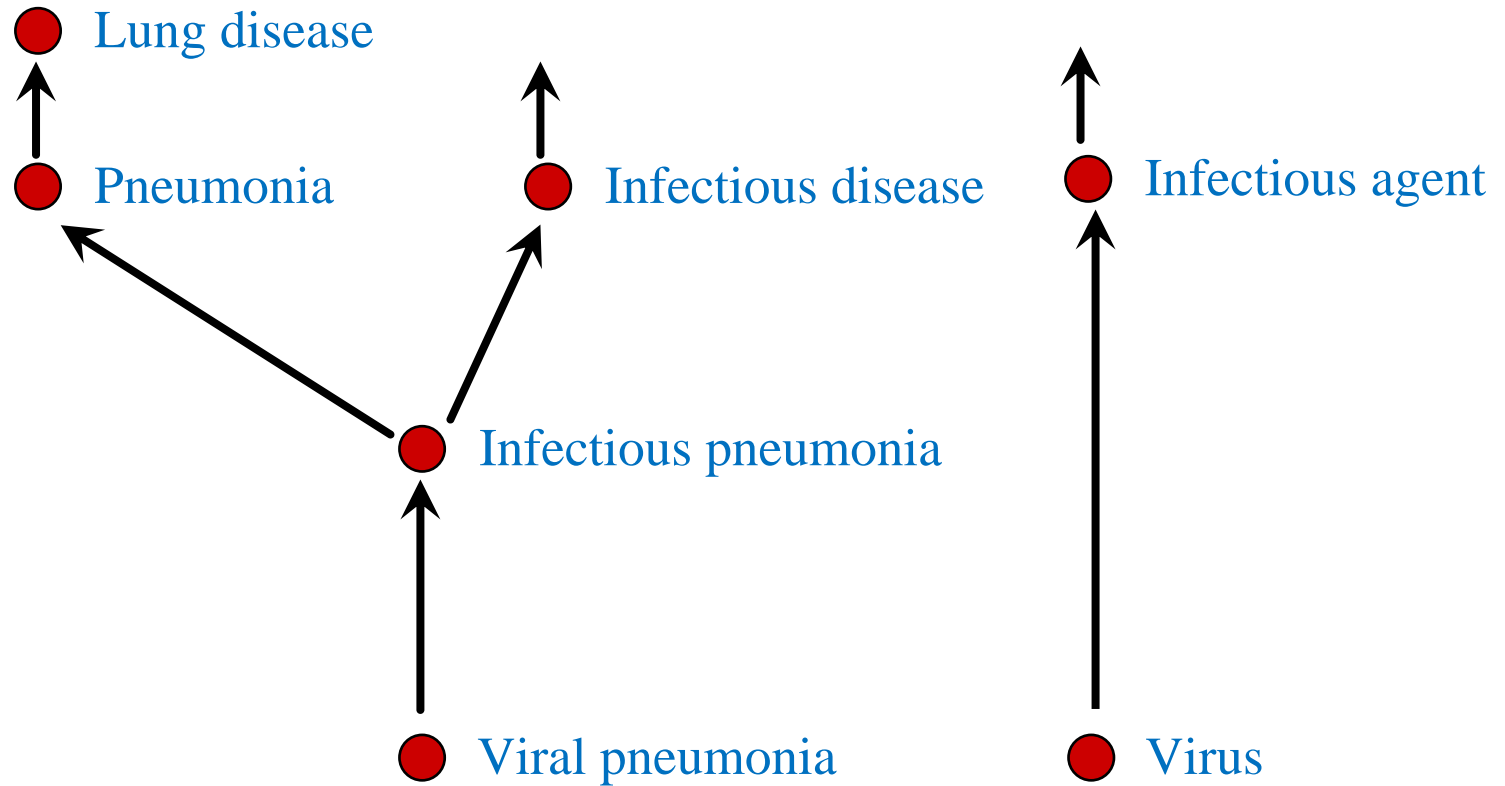


How are the codes organized?

1. **Directed acyclic graph**
 - Logical subsumption relationships, with a single root
2. **Attributes with values**
 - Necessarily true “existential restrictions”
3. **Description logic definitions of each concept code**
 - Structured combinations of isa’s and attribute-value relationships



Relationships: isa examples





Attribute-value relationships

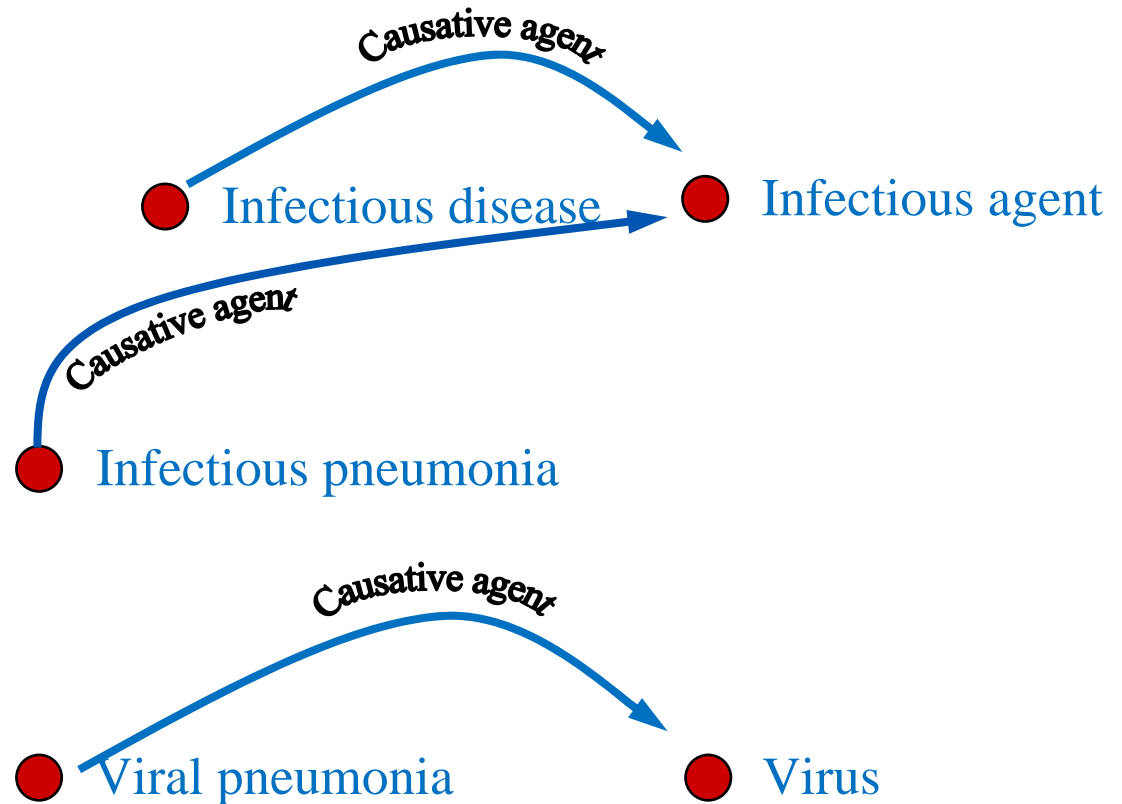
- **attribute – value**
 - Logical “existential restriction”
 - A rel B means that for every instance of A, there is at least one relationship “rel” with a value that is an instance of B



Attribute example: causative agent

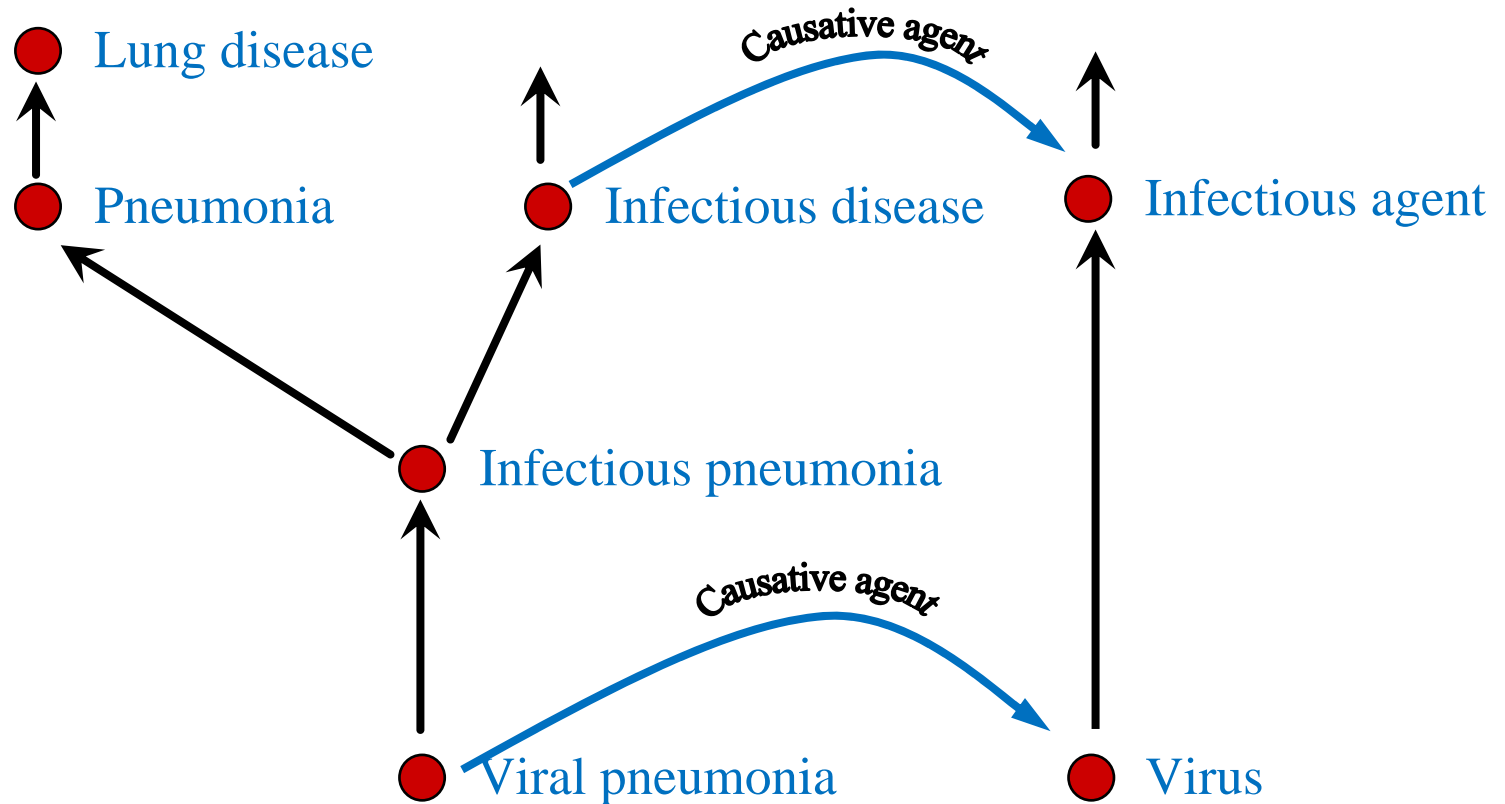
● Lung disease

● Pneumonia





Combining isa and attribute relationships





What is description logic?

- **Mathematical viewpoint:**
 - A family of logics characterized by
 - Formal set-theoretic semantics
 - Proofs of correctness and completeness of computation
 - Proofs of algorithmic complexity (PSpace, NP-complete, NExpTime, etc)
- **Knowledge representation viewpoint:**
 - A set of constructs for representing terminological knowledge (that which is always true of a meaning)
 - Algorithms and their implementations for performing:
 - Subsumption (testing pairs of expressions to see whether one is a subtype of the other & vice versa)
 - Classification (structuring a set of expressions according to their subsumption relationships)



A SNOMED example

- Headache is-a ache: finding-site = head structure
 - (and headache is marked as “defined” in concepts table).
- The class “headache” is sufficiently defined as the set of instances of the class “ache” which also have at least one finding-site relationship to an instance of the class “head structure”.
- And all instances of class “ache” with some finding-site relationship to an instance of “head structure” are instances of “headache”.
- Now, is that what you mean when you say “headache”?



SNOMED CT Basics

- Components and structure
- **Post-coordination, Expressions & Description logic**
- Terminology / Information model interaction
- Primary and Secondary uses (re-use) of data
- Extensions and Reference sets
- History tracking and Cross mapping



SNOMED CT Expressions

- **SNOMED CT coded information consists of structured (composed) collections of concept codes**
 - These are called expressions
 - The meaning of an expression depends on the situation in which it is used

Example

- **The SNOMED CT code for “fracture of femur” represents the meaning of “a break in a femur”**
 - Depending on where it is used in a patient record, the code may mean
 - The patient has a fractured femur
 - The patient’s main diagnosis is a fracture of the femur
 - The patient has a past history of fractured femur
 - The patient is suspected of having a fractured femur etc
 - In a query it may be one of several criteria for retrieving the records of patients with particular types of injury
 - In an index to the clinical literature it might indicate a paper that is relevant to this condition



Expressions

- Single code (coordinated pre-release -> “pre-coordination”)
- Combination of codes (coordinated by user post-release -> “post-”)
 - Compositional grammar specifies a way to create combinations
- Expression diagrams:
 - Illustrate the compositional grammar.
 - Focus concepts and values: yellow. Attributes: blue



- focus: attribute = value



Refinement and qualification: Two common ways to derive post-coordinated expressions

- **Refinement**
 - Replacing value C with a more specific value C1 within an existing (defining) “some R C” relationship in the definition, giving “some R C1”

 - Example
 - Fracture of femur
 - Defined as: finding-site = bone structure of femur
 - May be refined to: finding-site = structure of neck of femur
 - Yielding the new meaning: Fracture of neck of femur



Refinement and qualification: Two common ways to derive post-coordinated expressions

- **Qualification (also called “subtype qualification”)**
 - Replacing value C with a more specific value C1 within a qualifier “some R C” relationship (found in the qualifying relationships in the relationships table), giving “some R C1”
 - **Example**
 - **Bronchitis**
 - Qualifier exists as: clinical-course = courses (any course value)
 - May be qualified to: clinical-course = acute (sudden onset AND/OR short duration)
 - **Yields the meaning: Acute bronchitis**
- **End results of refinement or qualification are post-coordinated expressions with an identical logical structure**



Primitive & sufficiently-defined concepts

- A concept is “sufficiently defined”
 - If its definition is sufficient to distinguish it from all its supertype concepts
- A concept is “primitive”
 - If it is not “sufficiently defined”



Primitive & sufficiently defined concepts

- **Head injury**
 - Is a = Disease
 - Associated morphology = Traumatic abnormality
 - Finding site = Head structure
 - **Sufficiently Defined**
- **Aching pain**
 - Is a = Pain
 - **Primitive**
- **Headache**
 - Is a = Aching pain
 - Finding site = Head structure
 - **Sufficiently Defined**



Different views of relationships

- **Stated view**
 - The view that SNOMED CT modelers edit
 - Includes only the defining relationships that an author has explicitly stated to be true
 - To be distributed in KRSS and/or OWL syntax
- **Inferred view**
 - The view distributed in the distribution file
 - Generated by auto-classification
 - Includes relationships inferred from the stated view
 - Excludes redundant relationships
- **Normalized view**
 - The view best suited to comparing expressions
 - Reduces all values to their proximal primitive subtypes

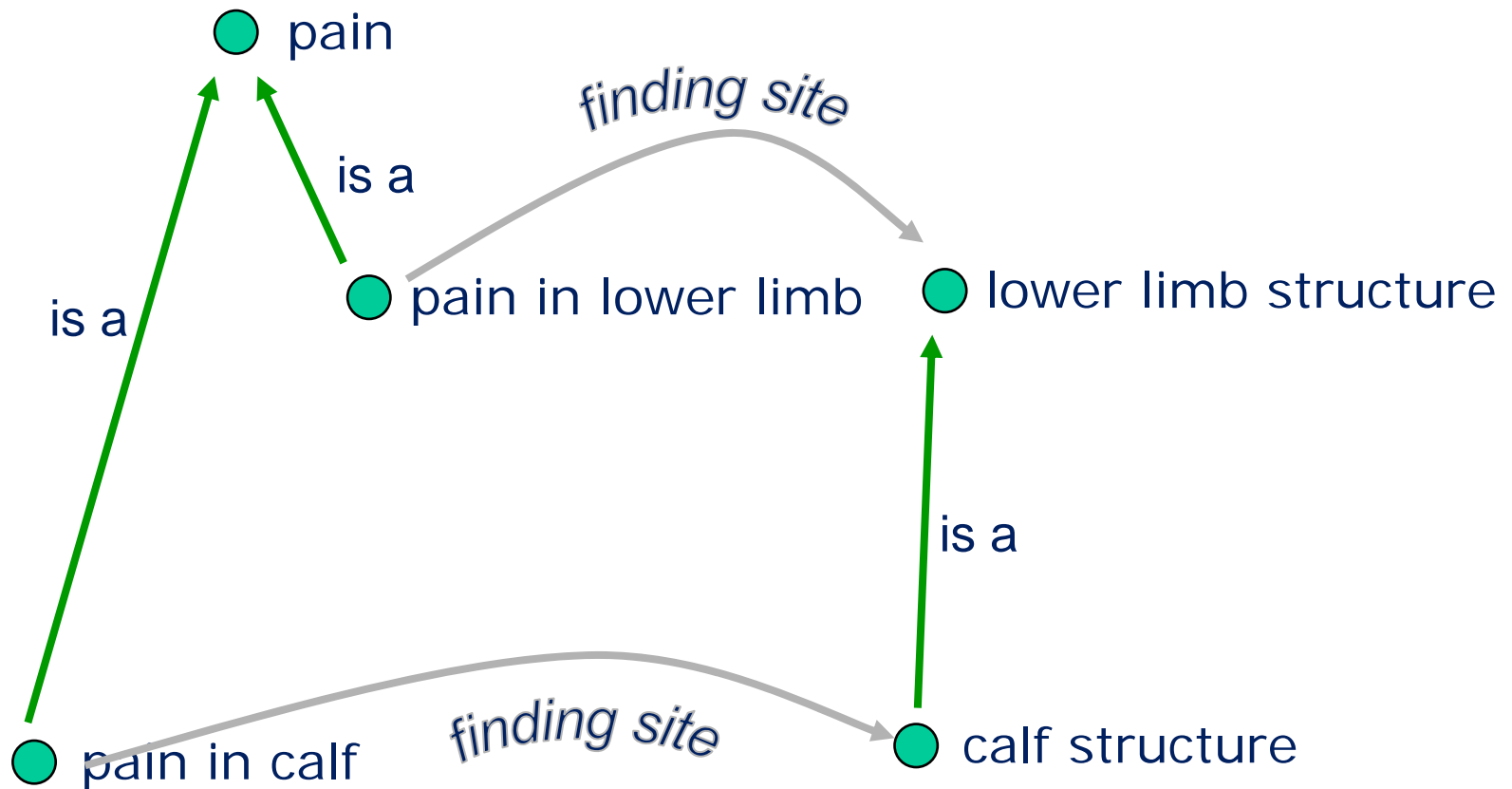


Auto-classification

- Many relationships are inferred by auto-classification rather than authored directly
- Auto-classification
 - Takes definitions “stated” by SNOMED authors and uses them to “infer” other relationships
 - Removes redundant (less specific) defining relationships
 - Creates a logically consistent parsimonious set of relationships
- Review the results of classification
 - Although logically consistent ... it may not be “correct” due to errors in “stated” definitions
 - Human errors that might otherwise be overlooked are often highlighted by auto-classification
 - Auto-classification is repeated frequently during authoring and the results are then rechecked

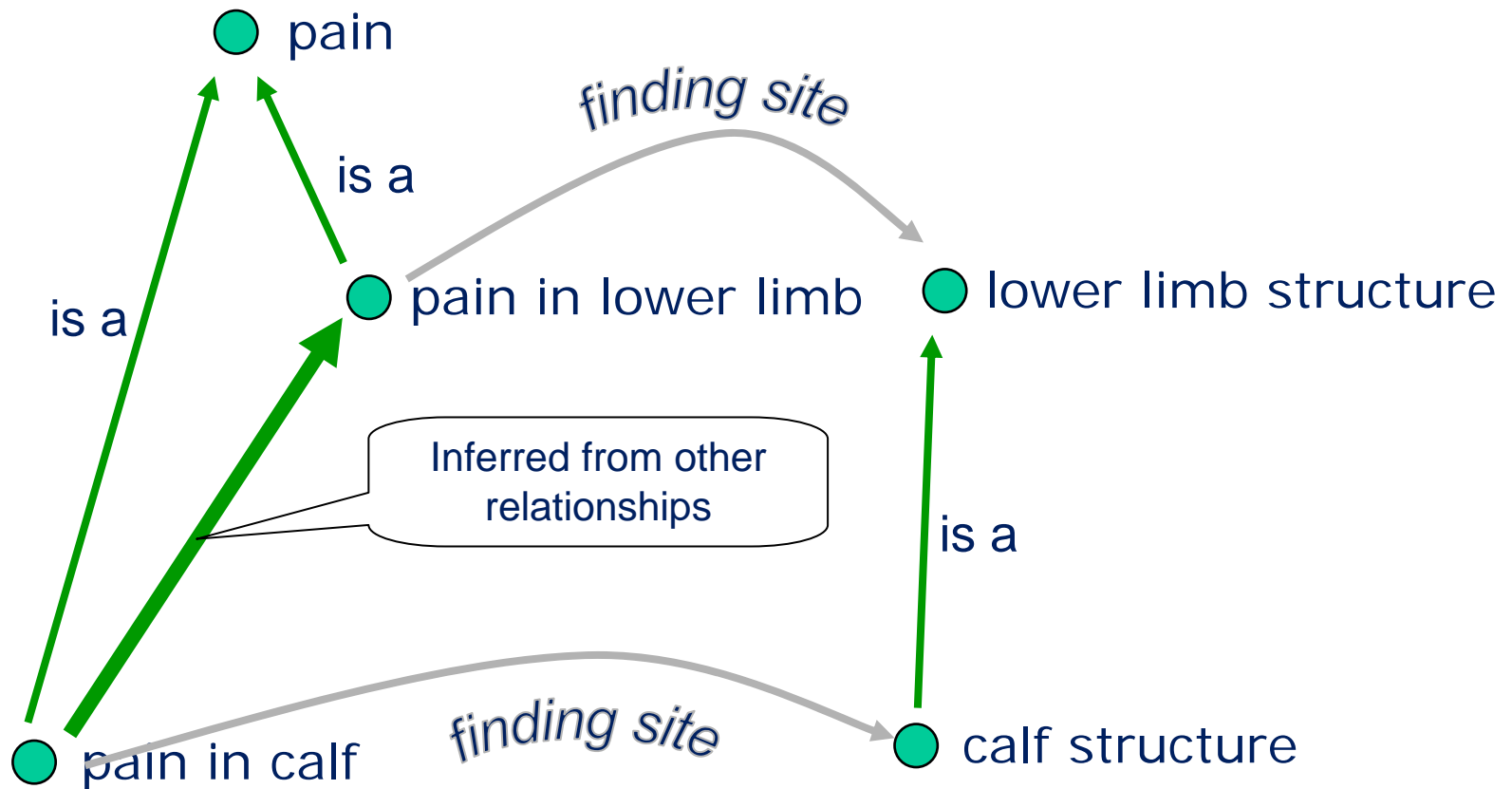


An example of a stated view



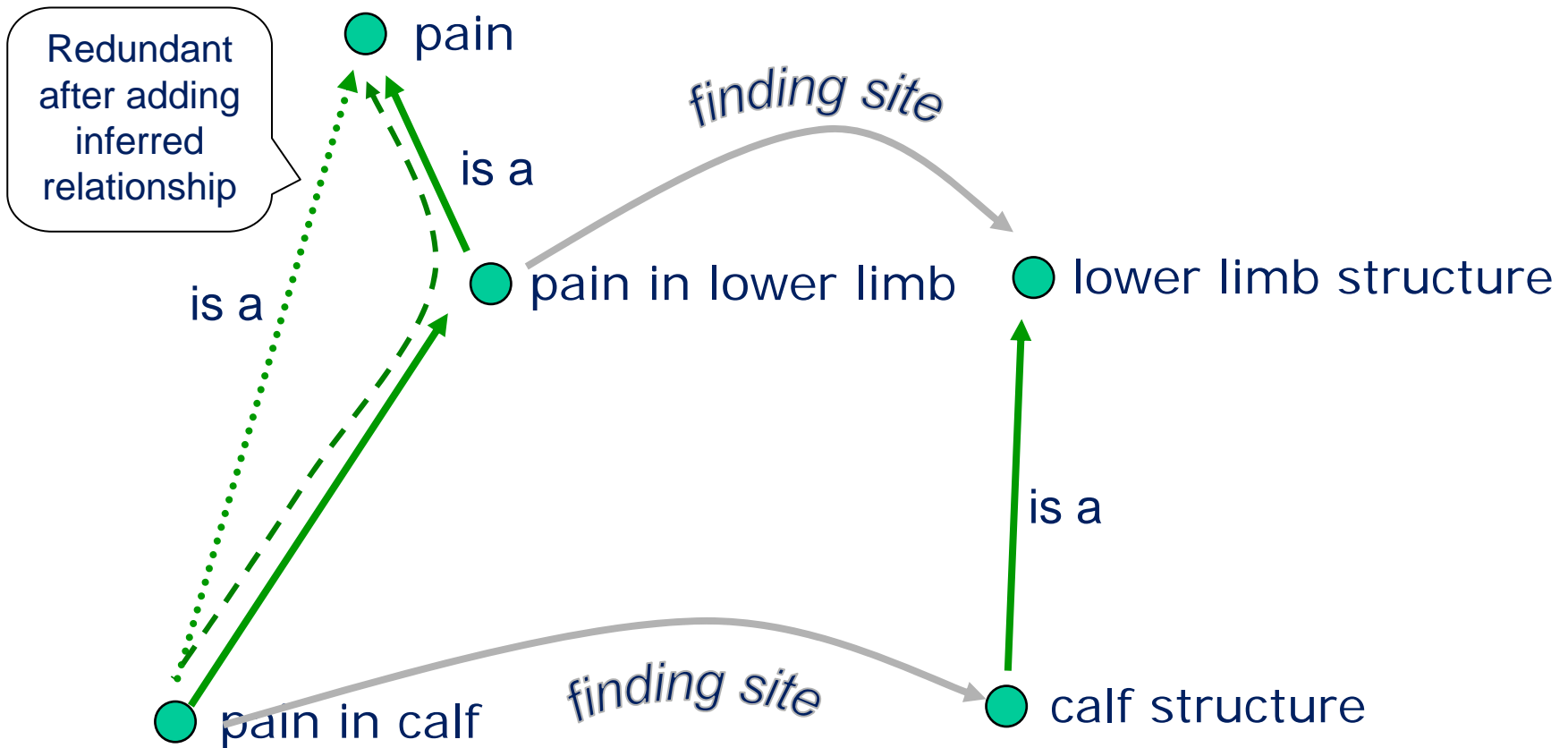


Classifier adds new is-a relationships





Redundant “direct flights” are removed for the distributed form of SNOMED CT





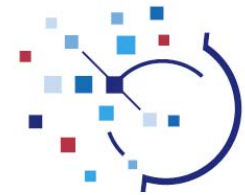
Advantages of post-coordination

- **Scope coverage**
 - Coverage of scope to an adequate level of specificity does not require every possible concept to exist
- **Terminology size**
 - Reduces the need for “combinatorial explosion” in concept numbers to cover every eventuality
- **Data entry**
 - Allows concepts to be constructed rather than searching through hundreds of similar terms for precisely the correct one
- **Retrieval**
 - Provides greater consistency between sets of similar variants for different concepts
 - For example ability to specify “severity” for any disorder or “laterality” for any bilateral body structure



Disadvantages of post-coordination

- **Human readability**
 - Extreme post-coordination can lead to loss of natural terms
 - “Appendectomy” might be expressed as “Procedure, with method excision and procedure site appendix”
- **Data entry**
 - Users may need to construct concepts by selecting multiple options rather than simply typing or choosing the term they want
- **Storage**
 - Expressions that have an unspecified number of refinements are less easy to store and manipulate than simple codes
- **Retrieval**
 - Several post-coordinated expressions may mean the same
 - Queries need to consider expressions rather than simple lists of codes



Pros and cons of post-coordination

- Both pre and post-coordination have benefits
- Any successful terminology should
 - Enable post-coordination to add flexibility
 - Include pre-coordinated concepts that express concepts with commonly used terms
 - Avoid attempting to identify an absolute boundary of what may or may not be pre-coordinated
 - Deal with the issues raised by alternative representations
 - The key issue is detection of equivalence and subsumption of alternative expressions



Which one does SNOMED allow?

- SNOMED CT allows *both* enumeration and composition
- The degree of enumeration depends on IHTSDO (for the international release) and National Release Centres (for National release extensions) and other extension developers
- Rules for composition are becoming more formal – see the machine readable concept model group for details



Equivalence

- “Laparoscopic repair of inguinal hernia”
=
- “Repair of inguinal hernia” + Endoscopic approach

Recognizing equivalence

requires more structure than simple concatenation

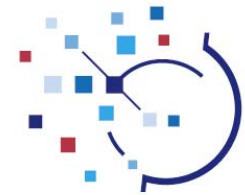


Equivalence

- Laparoscopic repair of inguinal hernia:

Method	= Repair
Site	= Inguinal canal
Object	= Hernia
Approach	= Endoscopic

- Repair of inguinal hernia: *Approach* *Endoscopic* *Qualifier*
Method = Repair
Site = Inguinal canal
Object = Hernia



Compositional grammar (1)

- Simplest expression is a single conceptid
 - For example
 - 71620000
- Optionally conceptid may be followed by a term enclosed in pipe delimiters
 - For example
 - 71620000|fracture of femur|
- Concepts can be combined with a plus sign that means logical “and” (conjunction)
 - For example
 - 31978002|fracture of tibia| +75591007|fracture of fibula|



Compositional grammar (2)

- Refinements can be added after a colon

For example

125605004: 363698007=29627003

- Refinements can be nested in parentheses

For example

53057004|hand pain|:

363698007|finding site| =(76505004|thumb structure|:

272741003|laterality| =7771000|left|)

- Refinements can be grouped in braces

For example

71388002|procedure|:

{260686004|method| =129304002|excision - action|,

363704007|procedure site| =66754008|appendix structure|}

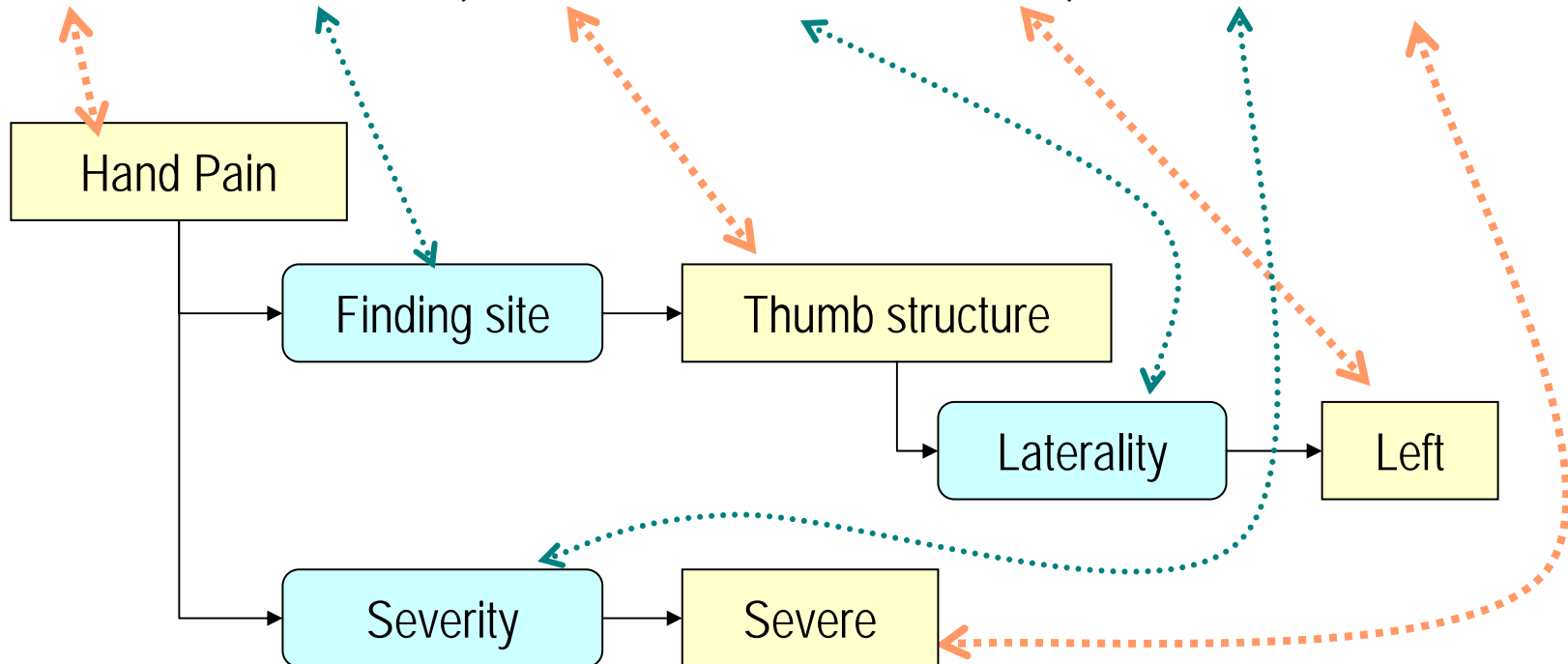
Note: the comma also means logical “and” in this expression



Expression example: “Severe pain the left thumb”

Compositional grammar mapped to diagrammatic representation

53057004:363698007=(76505004:272741003=7771000),272141005=24484000



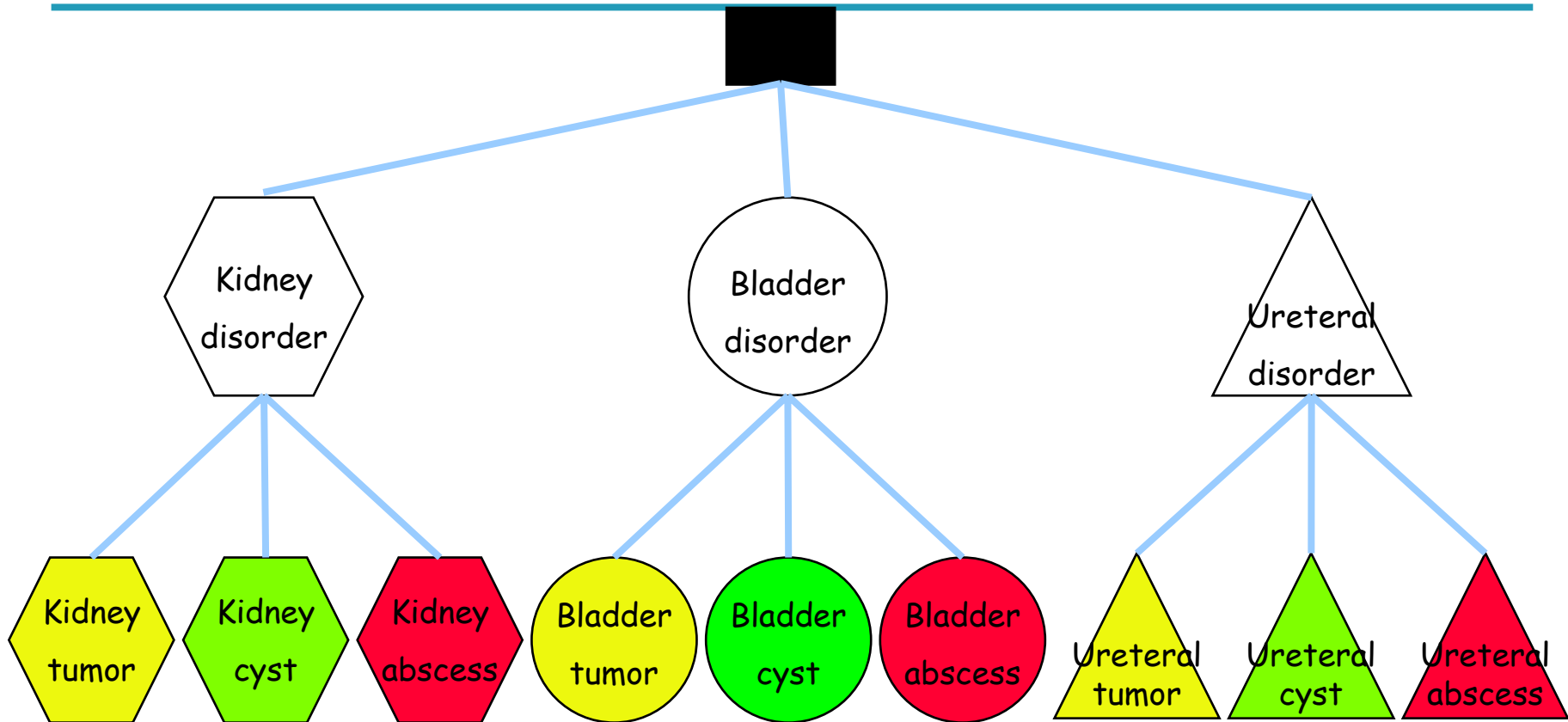


Alternative ways to represent expressions

- **SNOMED Compositional Grammar** is a recommended common form for review of expressions, and possibly for messages
- Internally within an application any form that captures the **SNOMED CT abstract logical model** for an expression is acceptable
 - SNOMED CT or a proprietary compositional grammar
 - XML in accord with (old) HL7 CD data type or proprietary schema
 - Relational database schema with table joins
 - Similar to **SNOMED CT Relationships table**
 - Use of internal unique identifiers pointing to an expression reference table

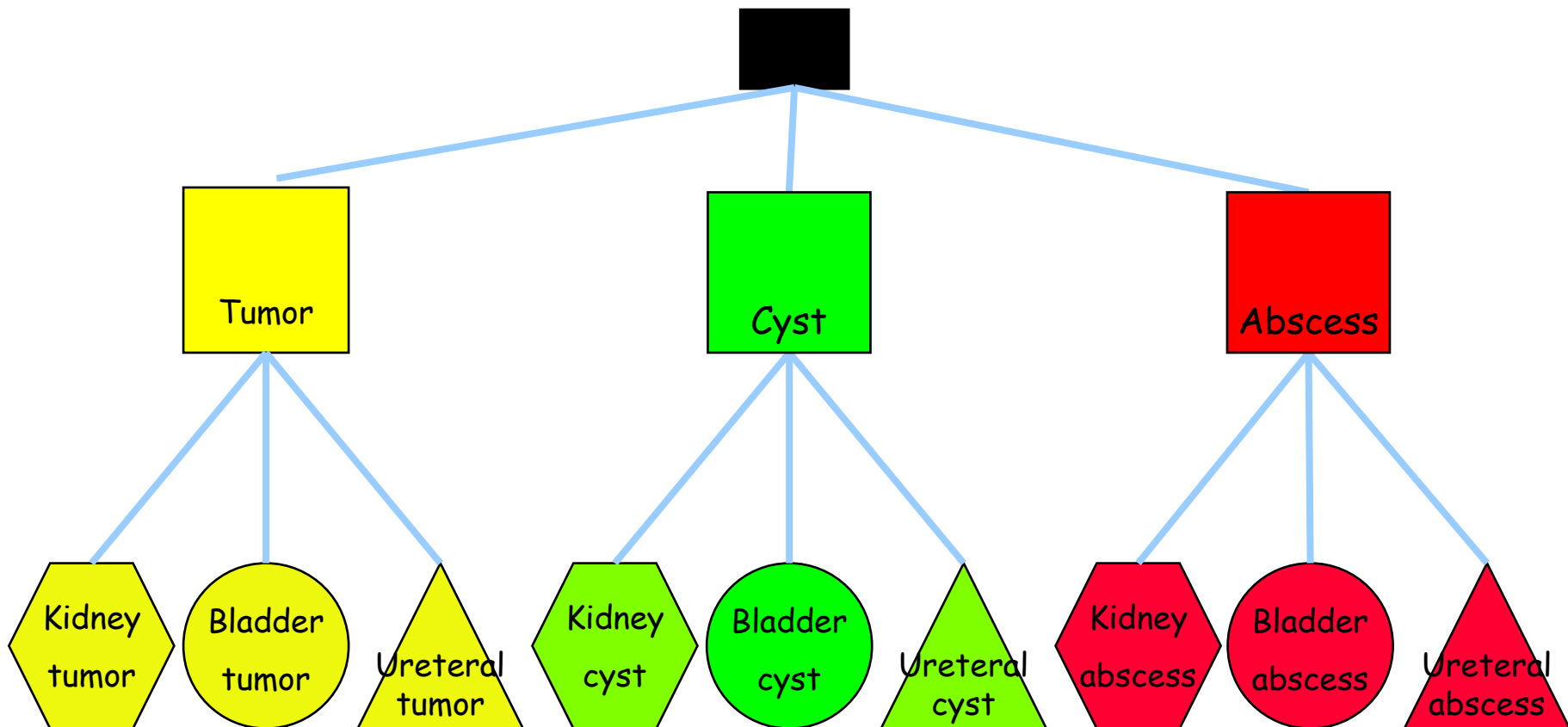


Single hierarchy: Retrieval by site



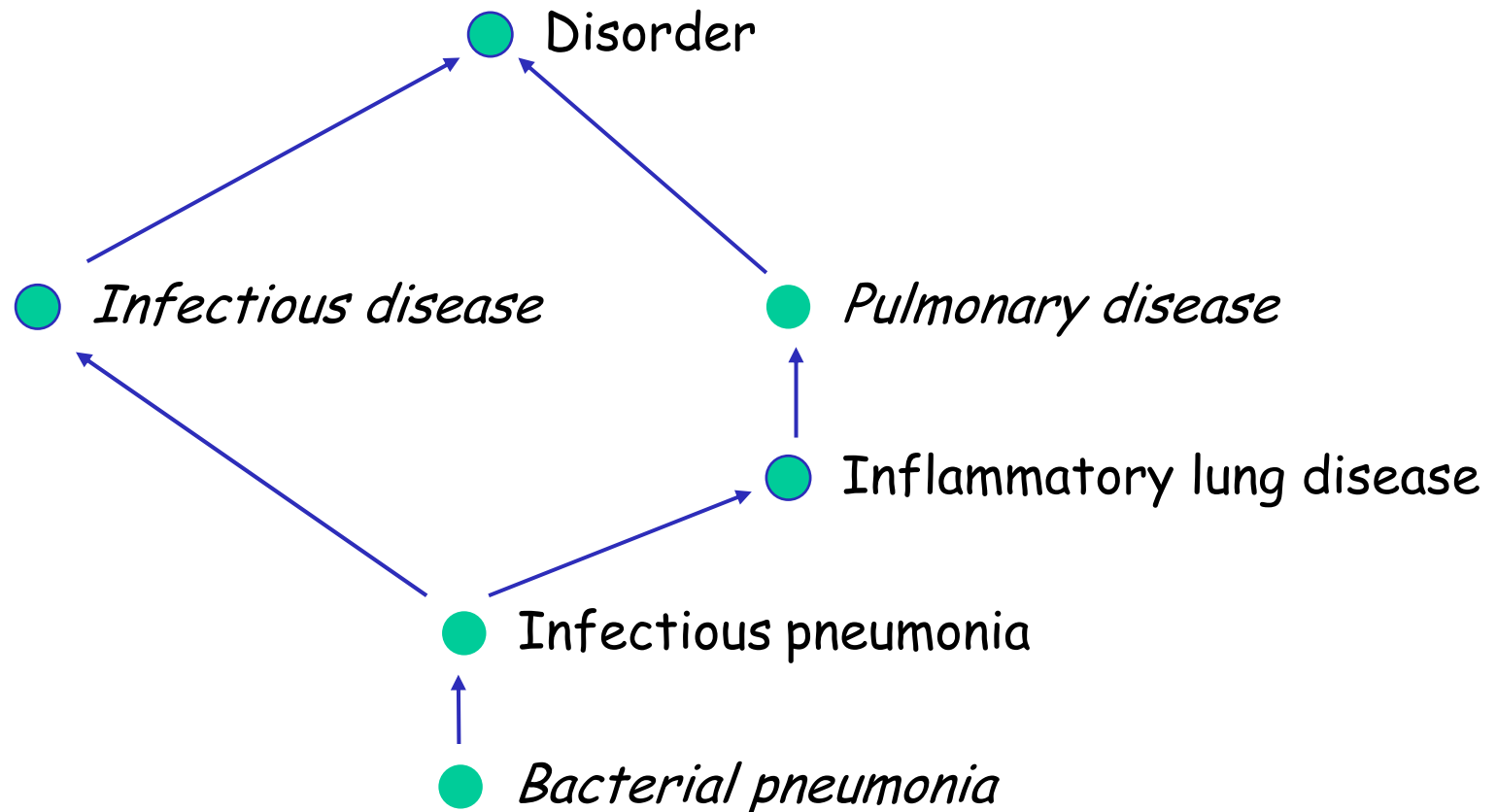


Single Hierarchy: Retrieval by morphology





Multiple Parents: More complete retrieval





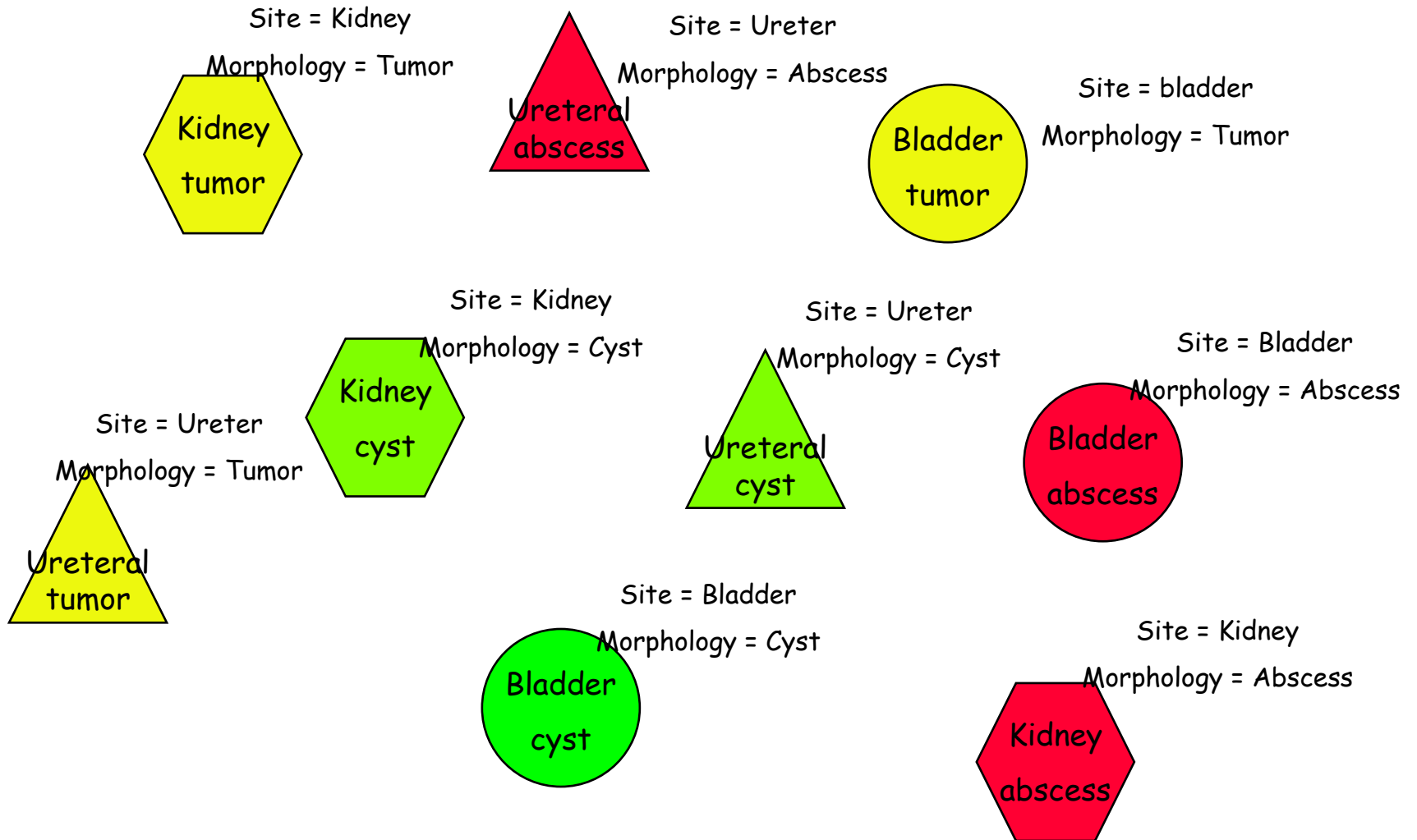
Multiple attributes

- Attribute = Site
- Values = Kidney
Bladder
Ureter

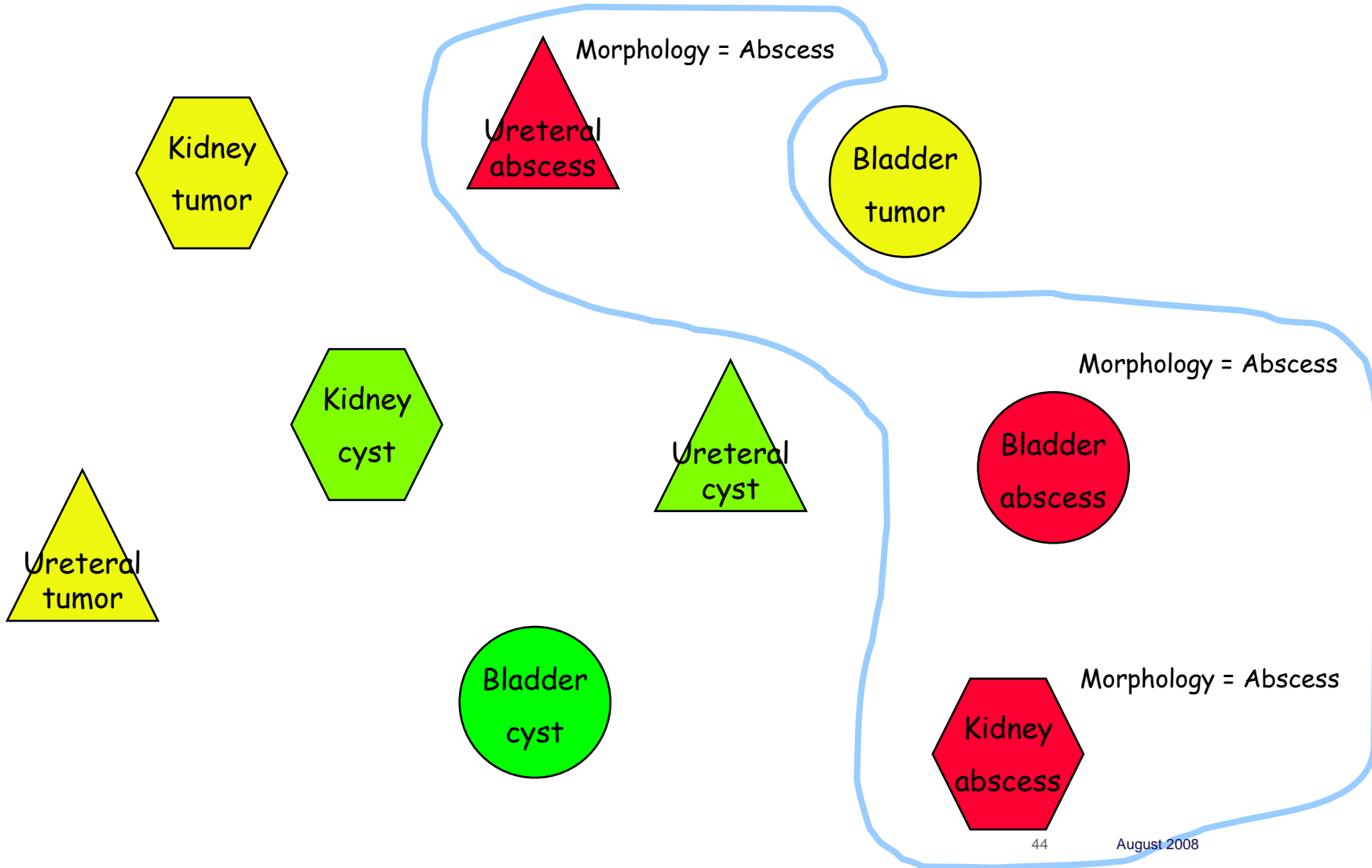
- Attribute = Morphology
- Values = Cyst
Tumor
Abscess

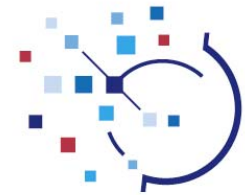
- Attribute = Etiologic agent
- Values = Infectious agents (hierarchy)
Substances (hierarchy)

Post-coordination, Expressions & Description Logic



Which ones are abscesses?





Expression transformations

- **Expressions can be transformed by applying rules that**
 - Impose formal structures from associated SNOMED CT definitions (e.g. nesting and/or grouping of attributes)
 - Remove redundancy where a refinement contains an attribute which duplicates a part of the definition of the focus concept
 - Add redundancy to assist processing at the expense of storage
 - Add explicit representation of assumed contextual information
 - Fully “normalize” the expression to support subtype and equivalence testing during selective retrieval
- **Details of the transformation rules**
 - See “Transforming expressions to normal forms” document



Expression transformation & forms used for storage and communication

- Recognition of the possibility of transformation affects decisions about representation of expression
 - Should information be transformed to a common “normal form” before storage or communication?
- or*
- Should information be stored in a “close-to-user” form that closely represents the information as actually entered?
- To preserve clinical integrity the close-to-user form is preferred as the primary representation for storage and communication
 - The reasons for this are explained on subsequent slides ...



Advantages of close-to-user form

- For clinical integrity and reliability the form stored should be as close to that entered as possible
 - The “normal form” of a given input may change as the definitions of referenced concepts are enhanced in new SNOMED CT releases
 - Close-to-user form preserves integrity across releases
 - Normal form is usually more verbose which is not ideal for storage or communication
- Formal rules specified by SNOMED allow a “close-to-user” expression to be transformed to the “normal form”
 - Normal form is vital for retrieval but for effective retrieval the normal form for all expressions must be based on the same set of SNOMED CT released definitions



SNOMED CT Basics

- Components and structure
- Post-coordination, Expressions & Description logic
- **Terminology / Information model interaction**
- Primary and Secondary uses (re-use) of data
- Extensions and Reference sets
- History tracking and Cross mapping



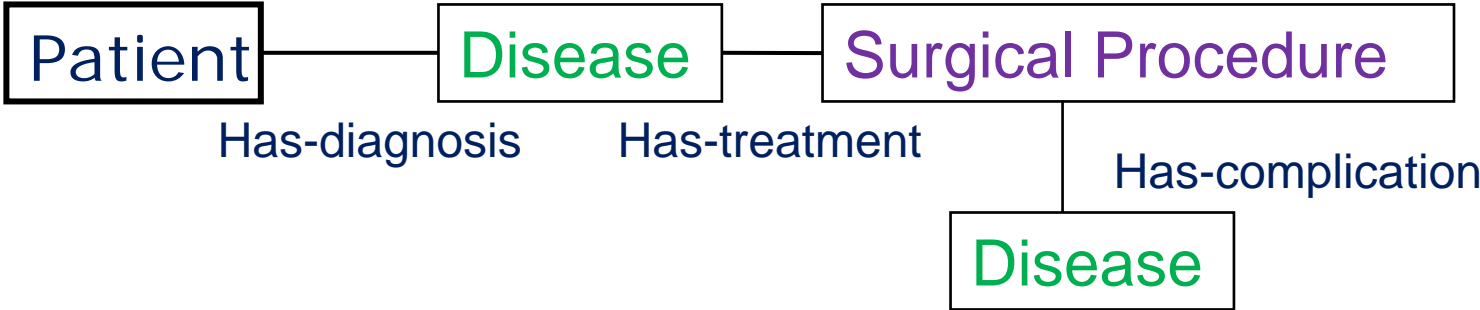
Where do the codes and expressions go in a record?

- **Statements in EHRs**
 - Electronic health record is made up of a series of statements
- **Codes and expressions are the values for fields/slots in the information model**
 - Codes and expressions from the terminology fill in some or all of the statement body
 - Information model determines the fields/slots available
- **Coordination required to avoid gaps & overlaps between**
 - terminology model
 - Information model

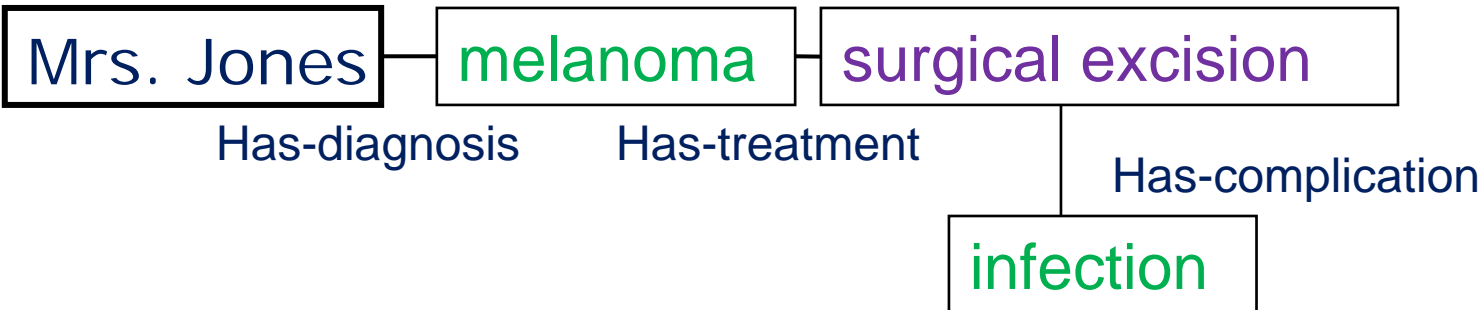


Record Structure vs Patient Data

Record Structure

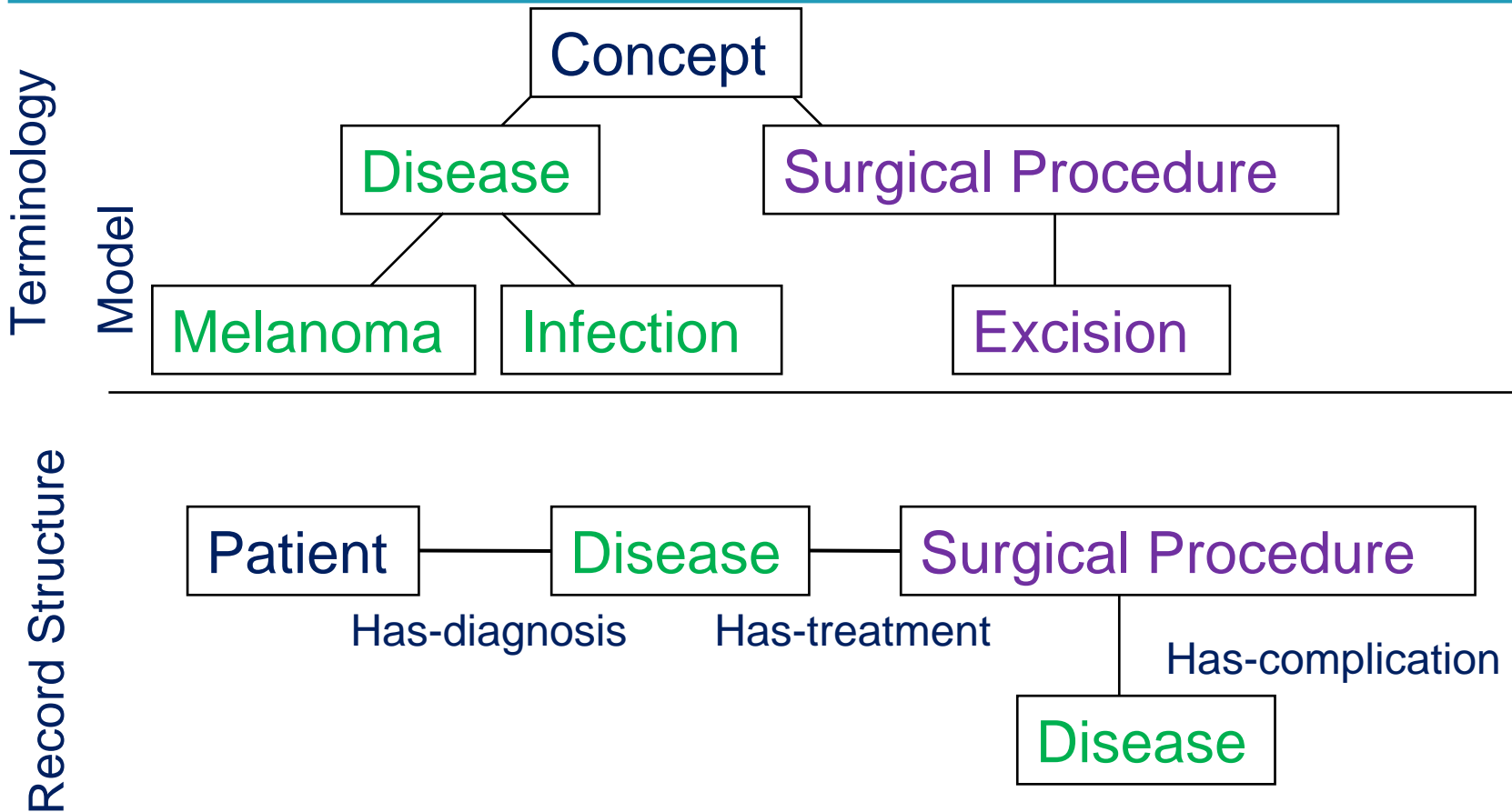


Patient Data





Terminology Model vs. Record Structure





Need for information model

- Clinical statements require an information model
- The simplest information model is just
 - “put your information here: _____”
- This is absurd, especially for data that ordinarily goes into fields such as :
 - Name, ID, Date of visit



Structuring individual observations

Date	Observation
12-Nov-2008	Free text

You can't make effective use of free text for decision support and other secondary purposes

Date	Observation
12-Nov-2008	Coded value

This extreme is absurd – requires everything to be in the coded terminology

Date	Observation type	Value
12-Nov-2008	Code for “question”	Code for “answer”

How do we decide what goes in the “question” and what goes in the “answer”???



Balance, overlaps, gaps

- Record the fact that “malignant mesothelial cells were found in a pleural fluid aspirate”:

Field or question	Terminology value
Pleural fluid finding	Malignant mesothelial cells
Site of malignant mesothelial cells	Pleural fluid
Lab test result	Malignant mesothelial cells in pleural fluid
Type of mesothelial cells in pleural fluid	Malignant
Type of malignant cells in pleural fluid	Mesothelial



Clinical statements

- Are the basis for a common view of patient record structure.
- The electronic medical record can be viewed as a collection of statements
- A faithful record of what clinicians have heard, seen, thought, and done
 - Not a collection of facts
- Other requirements for a medical record, follow naturally from this view
 - e.g. that it be attributable and permanent

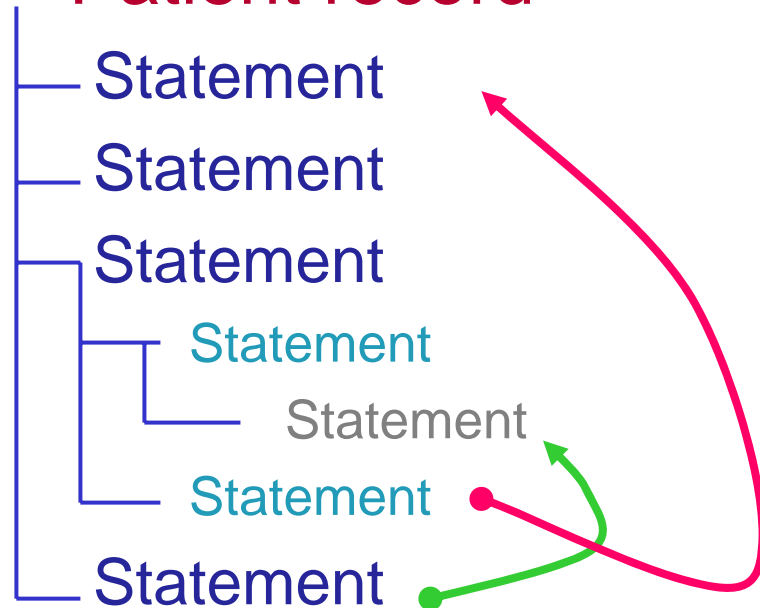
*Rector ,Nowlan & Kay (1991) Foundations for an electronic medical record.
Methods Inf Med. 30:179-186, 1991*



Statement based view of a patient record

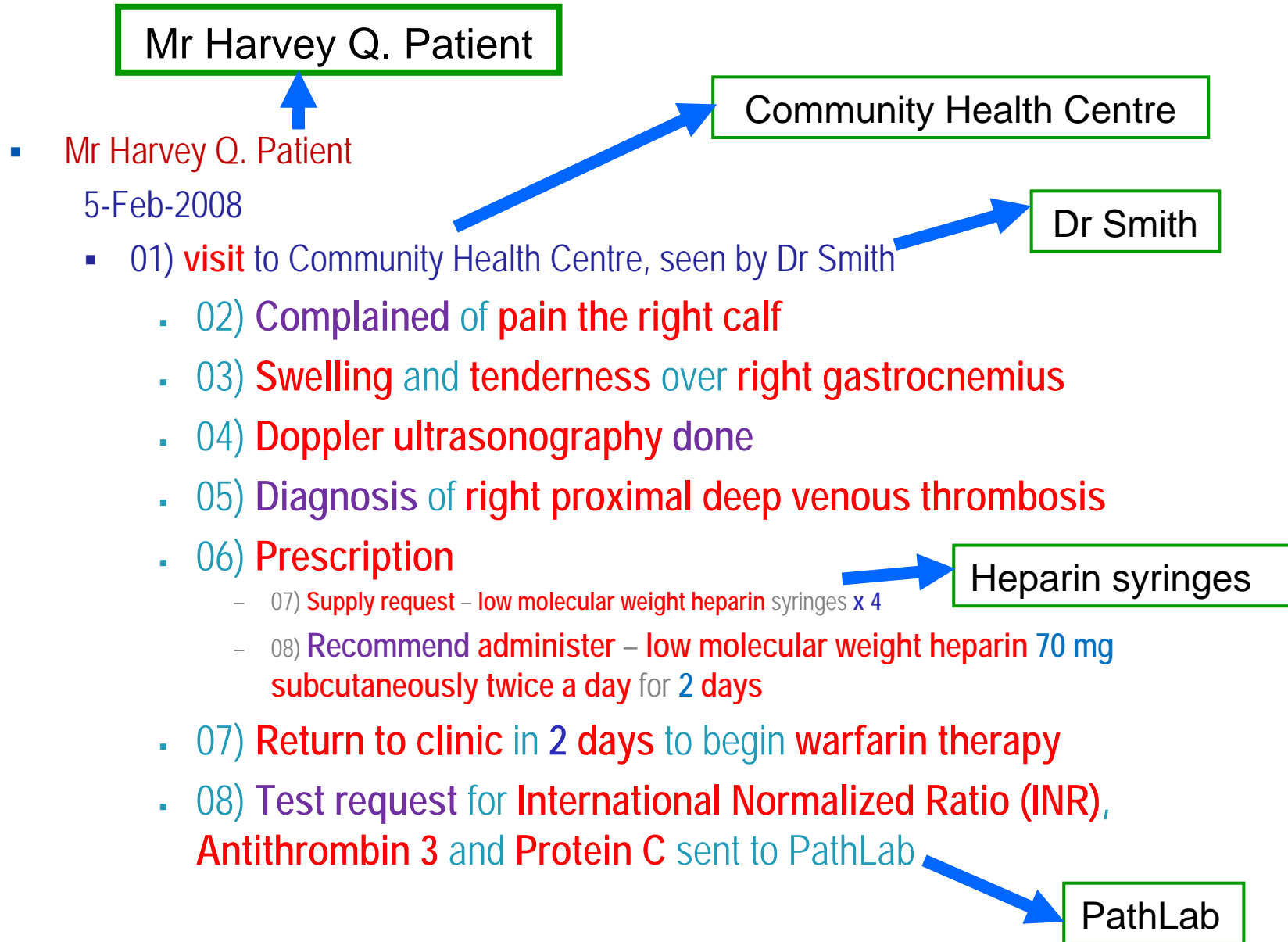
Overall high-level view

■ Patient record



- Common repeating elements represent statements
- Subsidiary statements may be nested
- Statements may be related to one another
- All statements have a common basic structure with a minimum set of specializations

Identifying concepts, context, values





SNOMED CT Basics

- Components and structure
- Post-coordination, Expressions & Description logic
- Terminology / Information model interaction
- **Primary and Secondary uses (re-use) of data**
- Extensions and Reference sets
- History tracking and Cross mapping



First rule of coding

- Yesterday's data should be usable tomorrow
- Corollary: If no-one is going to re-use the data, then no-one needs to code it.



Why do clinicians record patient data?

- To aid their memory
- To legally document what they saw & did (and sometimes why)
- To communicate to other members of a team
- To support and justify reimbursement
- To satisfy requirements of protocols & systems
 - Research protocols
 - Minimum data sets
 - Professional guidelines
 - (? And some incidental constraints imposed by software)



Secondary uses

- **Secondary uses of clinical data are any uses other than the primary purpose(s) for which the data is recorded**
 - ICD-10 coding for reimbursement can be derived from the dictated discharge summary, where the primary purpose may be documentation +/- communication, (not reimbursement).
 - Communicable disease reports to the health department can be derived from routine lab culture reports, where the primary purpose is communication to the ordering physician, (not epidemic detection).



Examples where the ideal is beginning to work

- **Microbiology laboratories**
 - positive *Salmonella* culture
 - Reports go to the physician(s) caring for the patient
 - Reports also go to local/state reportable disease registries



Data re-use

- **Requires the ability to:**
 - Query databases containing coded clinical data
 - Systematically retrieve patients based on general criteria
 - Aggregate data in ways *not directly encoded*
 - Consider the following type of rule:
 - If the patient has had an MI but has no CHF, AV block, asthma, peripheral vascular disease, or Type 1 diabetes mellitus, and is not taking a beta blocker, you need to consider adding beta blocker therapy ...



Dilemmas requiring attention

- The value of secondary data accrues (mainly) to parties other than those who collect it
- The value of secondary data depends on its quality, while the quality of data is directly proportional to the care with which it is collected



SNOMED CT Basics

- Components and structure
- Post-coordination, Expressions & Description logic
- Terminology / Information model interaction
- Primary and Secondary uses (re-use) of data
- **Extensions and Reference sets**
- History tracking and Cross mapping



Extensions and namespaces

- **Local extension is a fact of life**
 - If so-called “local” identifiers (codes) are added to patient records, it is unlikely they will remain solely within an organisation, since patients and their records are inherently mobile
- **The extension namespace mechanism is intended to allow organisations other than IHTSDO to create and distribute valid identifiers**
 - **Main benefits:**
 - prevention of collisions
 - Software systems can be built to ignore certain namespaces, or to allow special handling of others
 - Proper use of namespaces is the responsibility of the namespace holder



Extension namespaces

- **Contained within part of the code (SCTID)**
 - A namespace is controlled by an organization other than IHTSDO
- **Extensions allow addition of content to National Releases that is not required in the international release**
 - leave granted under the Mental Health Act 1983 (England and Wales)
- **Extensions also allow addition of local/regional content:**
 - Multnomah County (Oregon) jail cell number



Identifiers for extensions

Extension item identifier

Partition identifier

- 10 = Extension, concept
- 11 = Extension, description
- 12 = Extension, relationship



Namespace identifier

Check-digit



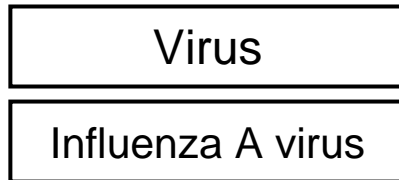
Reference Sets (RefSets)

- Formerly called “subsets”
- Define groups of SNOMED components to be used for a particular purpose
- Types of RefSets
 - Simple
 - Group
 - Tagged
 - Language
 - Navigation
 - Aggregation
 - Prioritized

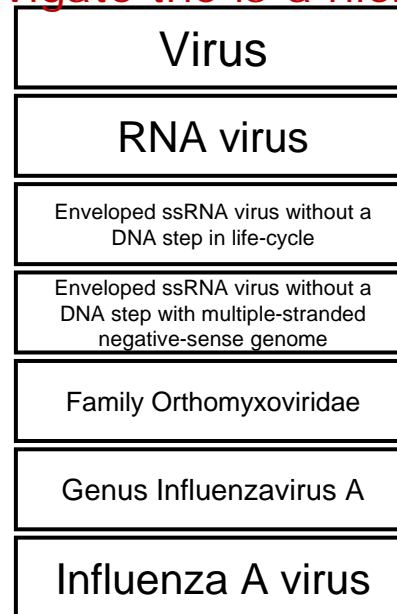


Navigational RefSet example

How a GP might like to navigate to Influenza A virus from "virus":



Six levels deep if you try to navigate the is-a hierarchy:





Making SNOMED usable

- Requires design and selection of usable components
 - Examples:
 - Problem list reference set
 - Primary care reference set
 - Non-human reference set (to be removed from browser for practitioners of human medicine)
- Requires hiding some of the complexity from the users
 - Search, adding qualifiers, doing queries, etc.
- Requires software that enables the users to accomplish their goals



SNOMED CT Basics

- Components and structure
- Post-coordination, Expressions & Description logic
- Terminology / Information model interaction
- Primary and Secondary uses (re-use) of data
- Extensions and Reference sets
- **History tracking and Cross mapping**



Use case for history tracking

- IHTSDO issues a new release containing
 - New concepts
 - Changed relationships
 - Changed concept status (concepts are inactivated)
 - Different subsets, cross maps, and other components
- Temporary local identifiers (for content submitted for inclusion in the international release) have been replaced by official international release identifiers.
- ICD-10-AM & other classifications are released, with new content that requires updating the mappings
- Application content needs to be updated



History representation

- **Component History Table**
 - Identifies each change in the status of a *Component* and the *Version* in which the change was made, along with the *Change type* and *Status* after the change and (sometimes) a *reason* for the change
- **Relationships Table**
 - Relates inactive concepts to active concepts in the terminology. The relationship types that support history are SAME AS, MAYBE A, REPLACED BY, WAS A, MOVED TO and MOVED FROM.



Principles of change

- **Graceful evolution of content and structure**
 - Avoid radical change
 - Clearly identify changes and why they occurred
- **Concept persistence**
 - Meaning of a concept must not be changed or deleted
 - Concept may be inactivated if its meaning is found to be redundant, ambiguous, or otherwise incorrect
- **Address redundancy**
 - Recognize redundancy
 - Provide a mechanism for resolution



Managing changes in systems

- Systematic planning for change management is essential to efficient updating
- Some systems that require “rip and replace” are never updated, with implications for safety
- The best software is designed to ensure modularity of :
 - Terminology and terminology services
 - Patient data repository and query services
 - User Interface and UI services



General change policies

- **Always retain data in the form it was recorded**
 - Data over-writing has serious quality issues and (in some realms) also medico-legal implications
- **Update terminology regularly as necessary to keep pace with improved terminology quality and advanced knowledge**
 - Terminology can be updated either incrementally or as a whole
 - *Incremental changes* generally require more effort to design & set up and less effort to do it each time.
 - *Complete replacement* is less difficult to design, requires more effort to actually do it.



Assumptions about cross maps

- **Clinical care is documented using a terminology**
- **Reimbursement and statistical coding use:**
 - Classifications (e.g. ICD10)
 - Groupers (e.g. HRGs, DRGs)
 - Administrative or financial codes
- **Mapping is the way to link and coordinate**
 - Reduce but not eliminate need for separate coding
- **All maps are directional (from a source to a target)**
- **Producing a crossmap is not trivial**
 - Different uses implies different levels of detail & grouping
 - Mapping is rarely one-to-one and may be many-to-many
- **Using a crossmap can be straightforward,**
 - if it is well designed, and supported with software



Direction of a crossmap

- **Every mapping has a direction**
 - The direction depends on the use case
 - Use case 1: code clinical statements in SNOMED, derive ICD codes afterwards:
 - Direction is from a SNOMED CT concept to one or more target codes, taking account of inclusion and exclusion criteria, etc
 - Use case 2: legacy data already coded in ICD, now want to do queries on new and legacy data using SNOMED CT
 - Direction is from ICD to SNOMED CT
 - General and “disjunctive” ICD codes may not have a corresponding code in SNOMED
 - E.g. “Other superficial injuries of abdomen, lower back and pelvis”
 - » Best you can do is “superficial injury of trunk”



Scope of mapping to ICD

- What SNOMED Concepts should be mapped?
- Hierarchies:
 - Clinical findings
 - Situations with explicit context
- Concept status
 - Current concepts (status 0)
 - *Not* legacy ('status 6')
- Excludes codes in the non-human subset
- Any concept meeting these conditions, but has no corresponding ICD code, is explicitly noted to be unmappable



Example crossmap

46635009 Diabetes mellitus type 1

maps via the ICD-9-CM crossmap to

250.01 Diabetes mellitus without mention of complication,
type I, not stated as uncontrolled

- Note that “without mention of” and “not stated as” are meaningless in a terminology context.
- They become meaningful in a classification / coding context, and depend on the contents of a pre-existing clinical record.