Chapter 12
SNOMED CT

SNOMED CT is the most comprehensive, multilingual clinical healthcare terminology in the world. It is used in electronic health record systems for clinical documentation and reporting. It can be used to retrieve, and analyze clinical data.

When considering SNOMED CT and its use, we need to remember that on its own SNOMED CT does very little. The value of SNOMED CT can only be realized when it is built into software and systems that are designed around it.

Kent Spackman, the leader of the team that developed SNOMED CT, has postulated two golden rules (Spackman 2005):

- The first rule of data quality is that the quality of data collected is directly proportional to the care with which options are presented to the user.
- The first rule of coding is that yesterday's data should be usable today.

The rule of data quality recognizes the heterogeneity of medical care and implies that we cannot adopt a one-size-fits-all approach to the detailed design of clinical applications.

The rule of coding recognizes that clinical data need to be treated as being permanent. We have to be able to use yesterday's data today and today's data tomorrow and for the indefinite future.

SNOMED CT has been designed and built for this. SNOMED CT is a comprehensive clinical terminology for recording the health and care of individual patients in a way that the information can be indexed and retrieved for reuse both at the point of care and subsequently for management, surveillance, and research. It is able to provide the clinical content and expressivity required for precise clinical documentation.

SNOMED CT was founded on four basic principles that have guided development activities and will continue to guide the future directions of SNOMED.

1. Development efforts must encompass broad, inclusive involvement of diverse clinical groups and medical informatics experts.
2. The clinical content must be quality-focused and adhere to strict editorial policies.

T. Benson, *Principles of Health Interoperability HL7 and SNOMED*, HI,
DOI 10.1007/978-1-84882-803-2_12,

3. The quality improvement process must be open to public scrutiny and vendor input to ensure that the terminology is truly useful within healthcare applications.
4. There must be minimal barriers to adoption and use.

SNOMED CT is the most comprehensive, multilingual clinical terminology in the world. In January 2009, it contained over 310,000 active concepts, 990,000 English descriptions, and 1.38 million relationships. There is no paper version. The sheer size of SNOMED CT is a significant issue in developing, using, and maintaining it.

In comparison, ICD-10 has 10,760 classes (excluding Chapter XX, external causes) and comes in three large volumes.

SNOMED CT is both a coding scheme, identifying concepts and terms, and a multidimensional classification, enabling concepts to be related to each other, grouped, and analyzed according to different criteria.

Numeric codes (the SNOMED CT identifier – SCTID) identify every instance of the three core building blocks: concepts, descriptions, and relationships. Each *concept* represents a single specific meaning; each *description* associates a single term with a concept (any concept may have any number of descriptions or names); and each *relationship* represents a logical relationship between two concepts.

SNOMED CT is used only in computer systems – it cannot be used manually, partly because it is so large, and more importantly because it works in a different way than earlier coding schemes such as ICD or the Read Codes. The relationship structure in SNOMED CT relies on a large number (well over a million) of explicitly defined relationships to work. The mechanism used is more complex than a code-dependant hierarchy used in earlier schemes, but is enormously more powerful, flexible, and future-proof. It allows any concept to be classified or qualified in any number of ways.

SNOMED CT provides an extensible foundation for expressing clinical data in local systems, for interoperability, and for use in data warehouses.

SNOMED CT is designed for clinical documentation and reporting. The terminology is made up from concepts, terms, and relationships to provide a way of representing clinical information across the broad scope of health care to support analysis and clinical decision support.

The content of SNOMED CT is organized into a number of hierarchies, including clinical finding, procedure, observable entity, body structure, organism, substance, pharmaceutical/biological product, specimen, physical force, event, environment/geographical locations, social context, staging and scales, etc.

12.1 SNOMED CT Documentation

The SNOMED CT User Guide, Technical Reference Guide, and Technical Implementation Guide are three key documents describing SNOMED CT in detail.

12.1.1 SNOMED CT User Guide

The User Guide is intended for clinical personnel, business directors, software product managers, and project leaders; information technology experience, though not necessary, can be helpful (IHTSDO 2009a). The User Guide is intended to explain SNOMED CT's capabilities and uses from a content perspective. It explains the content and the principles used to model the terminology.

12.1.2 SNOMED CT Technical Reference Guide (TRG)

The TRG is intended for SNOMED CT implementers, such as software developers (IHTSDO 2009b). The TRG assumes an information technology background. Clinical knowledge is not a prerequisite. The TRG contains reference material related to the current release of SNOMED CT and includes file layouts, field sizes, required values and their meanings, and high-level data diagrams. It can be used to install and use SNOMED.

12.1.3 SNOMED CT Technical Implementation Guide (TIG)

The TIG is intended for SNOMED CT implementers, such as software designers (IHTSDO 2009c). The TIG assumes information technology and software development experience. Clinical knowledge is not required, although some background is helpful to understand the application context and needs. The TIG contains guidelines and advice about the design of applications using SNOMED CT and covers topics such as terminology services, entering and storing information, and migration of legacy information.

12.2 Commom Aspects

SNOMED CT is composed of components, which include concepts, relationships, descriptions, subsets, and cross maps, each of which is identified by a SNOMED CT Identifier (SCTID) and has a validity status (Fig. 12.1).

12.2.1 The SCTID

All components are identified using a special SNOMED Clinical Terms Identifier (SCTID). The SCTID is an integer between 6 and 18 digits long. One way of thinking of the SCTID is as a 64-bit integer, although it does have an internal structure.

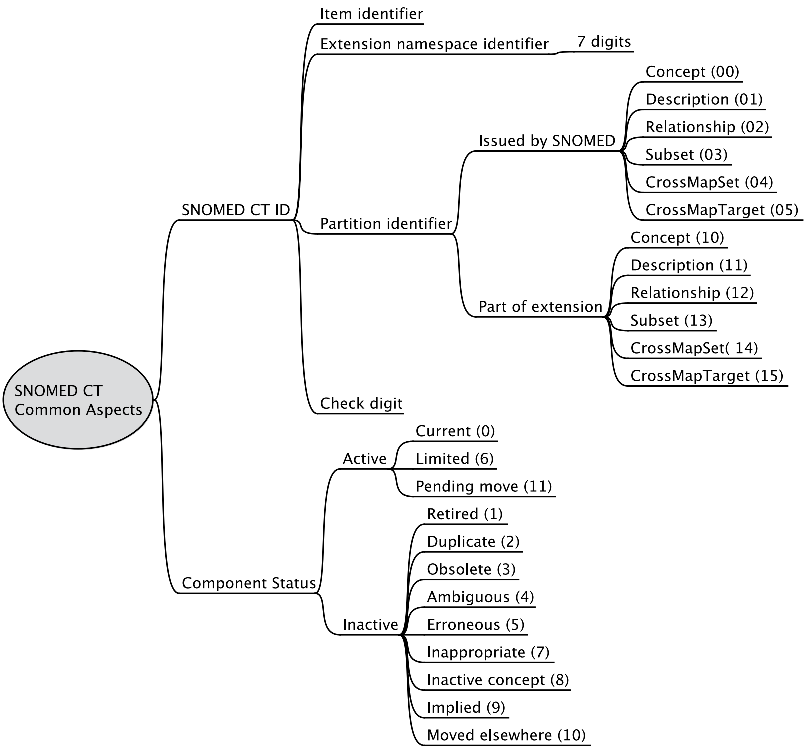

Fig. 12.1 SNOMED CT Common Aspects

The internal structure of the SCTID includes a meaningless item identifier, a check digit at the end and a partition identifier just in front of the check digit (the check digit algorithm is Verhoeff's Dihedral Group D5 Check).

The partition identifier indicates the type of component referred to by that SCTID. The SCTID may also contain a namespace-identifier, which is used in local extensions.

The partition identifier is a two-digit number. If the first digit of the partition identifier is a zero (0) then this component is part of the International Release; if it is a 1, then the component is part of an extension set. The second of the two digits in the partition identifier indicates which of the partitions of SNOMED CT the SCTID is identifying, where:

- Concept (0)
- Description (1)
- Relationship (2)
- Subset (3)
- Cross Map Set (4)
- Cross Map Target (5)

Extensions are additions to SNOMED CT, usually specific to one country or organization, and are identified using a seven-digit namespace-identifier.

12.2.2 Status

An important principle of SNOMED CT is that of permanence. Once a component such as a concept or description has been created it is never deleted, but may be given an inactive status, which also indicates the reason why it is inactive. The status field has the following values, with code values in parenthesis:

- Active: Current (0), Limited (6), Pending move (11)
- Inactive: Retired (1), Duplicate (2), Obsolete (3), Ambiguous (4), Erroneous (5), Inappropriate (7), Inactive concept (8), Implied (9), Moved elsewhere (10)

12.3 Concept

SNOMED CT is concept-oriented (Fig. 12.2). A concept is just a clinical idea to which a unique ConceptID (which is a SCTID) has been assigned in SNOMED CT. Concepts are in people's heads; codes (ConceptIDs) are in the terminology and they refer to real things in the real world. A concept is a clinical meaning that never changes.

In addition to the ConceptID, each concept has a unique fully specified name (FSN), which is human-readable. Each concept is also linked to a set of terms (descriptions), which name the concept in a human-readable way.

Concepts are formally defined in terms of their relationships with other concepts, based on the principles of reference terminology. These defining relationships may be either subtype relationships (also called IS_A) or attribute relationships. For example, the concept "appendicectomy" IS_A "procedure" and also has attributes "method" = "excision" and "procedure site" = "appendix."

SNOMED CT is organized into a set of hierarchies in which concepts are related by IS_A relationships to their more general parent concepts directly above them in a hierarchy. General concepts are at the top of the hierarchy; at each level down the hierarchy, concepts become increasingly specialized. Unlike a pure tree-structure, any SNOMED CT concept can be the subtype of more than one concept in the same hierarchy – concepts can have more than one parent.

SNOMED CT has 19 Top-level hierarchies (the number changes from time to time as the system evolves), which descend from a single Root concept (the SNOMED CT Concept). Some of the hierarchies also have well-defined sub-hierarchies. For example, the Clinical Finding top-level hierarchy has a sub-hierarchy for Disease (or disorder); the Organism hierarchy has sub-hierarchies for Animals, Plants, and Microorganisms (Fig. 12.3).

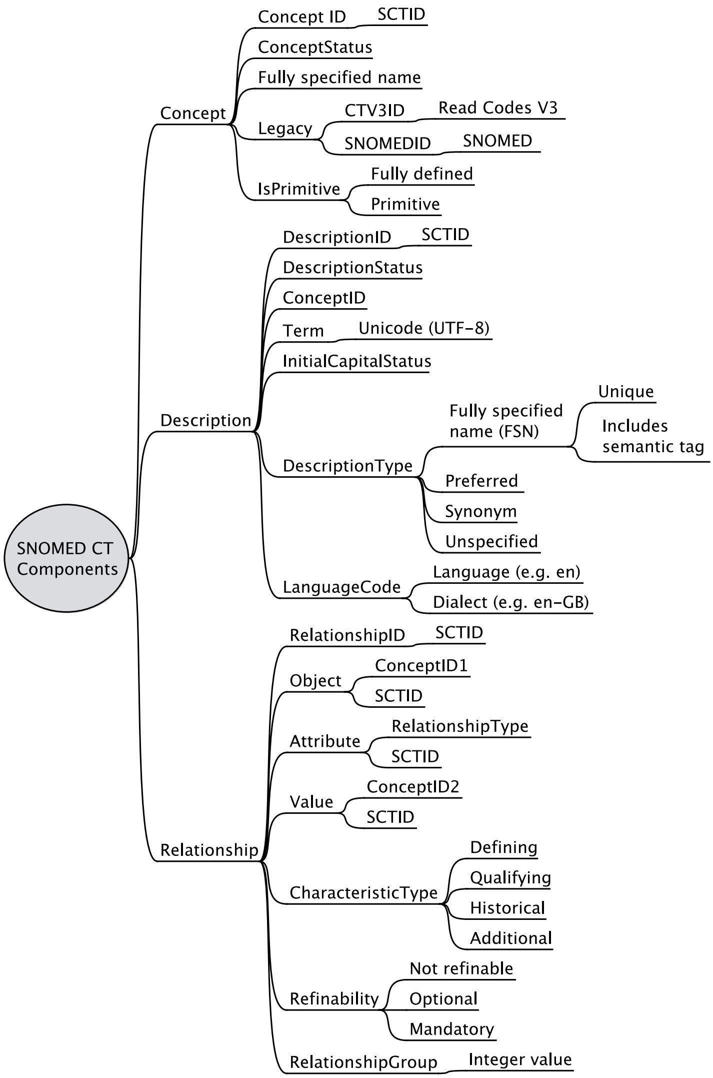

Fig. 12.2 SNOMED CT components

The SNOMED CT hierarchies fall in three main groups: object hierarchies, which mainly comprise concepts which are likely to be qualified; value hierarchies, which are mainly concepts which act as values, when qualifying or defining an object concept in a relationship as an object-attribute-value triple; and finally a miscellaneous group.

Legacy references are provided to the primary sources of SNOMED CT, the NHS Clinical Terms Version 3 (CTV3) and the original version of SNOMED.

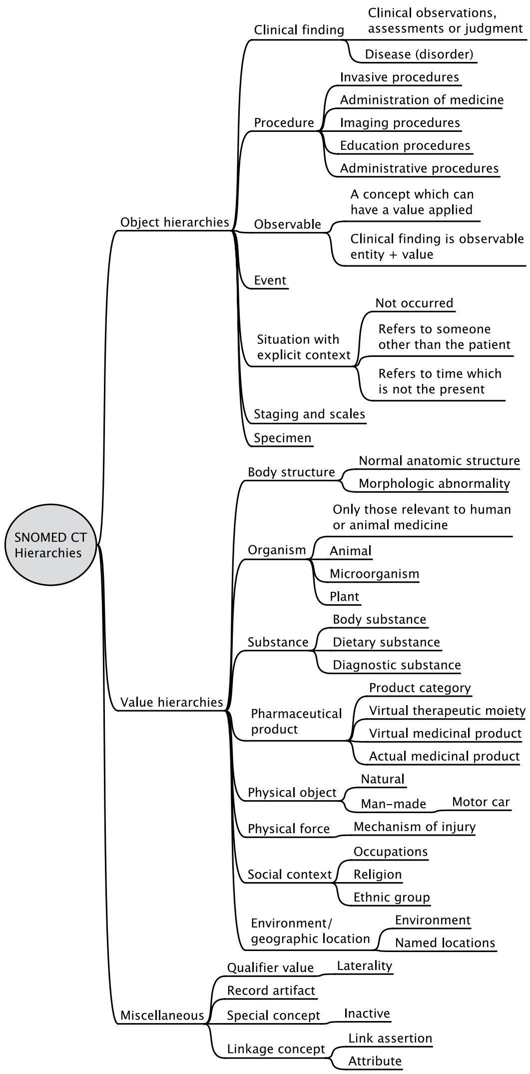

Fig. 12.3 SNOMED CT hierarchies

An indicator is also provided to indicate whether the concept is fully defined (using relationships) or is a primitive concept. A concept is "primitive" if it is not "fully defined," which is to say that its modeling (attributes and parents) does not fully express its meaning. Fully defined concepts are differentiated from their parent and sibling concepts by virtue of their attribute relationships.

12.4 Description

Each description links a human-readable term (a sequence of readable characters) with a concept. It has an associated unique numeric DescriptionID, which is a SCTID.

Every concept has at least two descriptions: the fully specified name (FSN) for that concept and a preferred term, sometimes called the display term, for that concept in the language or dialect being used.

Terms are encoded using Unicode (UTF-8), which supports all languages.

SNOMED CT supports multiple dialects and languages. For example, British English (en-GB) and US English (en-US) are different dialects of English in which many medical terms have different spellings; English, French, and Spanish are different languages.

The fully specified name (FSN) is a phrase that names a concept in a way that is both unique and unambiguous. Each FSN contains a suffix (hierarchy tag) that indicates its primary hierarchy, e.g., myocardial infarction (disorder).

The preferred term is the common phrase or word used by clinicians to name a concept and is used as default display term for that concept. Each concept has one preferred term in any dialect. The preferred term is often the FSN without its suffix (hierarchy tag), e.g., myocardial infarction.

Each concept may have multiple synonyms and translations e.g., heart attack or cardiac infarct. A synonym in SNOMED CT must be a clinically acceptable alternative to the preferred term as a way of expressing a concept. A list of synonyms for a concept shows the various ways a concept may be described, rather like a thesaurus.

SNOMED CT terms can also be homonyms in which the same term is used for different concepts. The preferred terms and synonyms are not necessarily unique within a language or dialect. For example, the FSN "cold sensation quality (qualifier value)" has a preferred term of "cold," but "cold" is also a synonym of "common cold (disorder)."

The case of the first letter of the term can be marked as significant as in pH or Alzheimer's.

12.5 Relationship

Relationships are at the heart of SNOMED CT and are the distinguishing feature of any reference terminology. More than 1.3 million relationships have been defined in SNOMED CT and this number is likely to continue to grow.

Each relationship is defined as an object-attribute-value triple. The object is the source concept – the one that has the relationship, identified by a concept identifier (ConceptID1). The attribute specified the type of relationship (RelationshipType) and is also a SNOMED CT concept. The value is the target (ConceptID2).

All relationships may be written using a notation known as description logic, such as:

Concept: attribute = value

The attributes and values which may be used to define or qualify concepts are set out in the SNOMED CT Concept Model (see below).

There are four categories of relationship (CharacteristicType).

(1) Defining relationships define the meaning of concepts. These are pre-coordinated concepts.
(2) Qualifying relationships do not define concepts but may be used in clinical systems to modify the meaning of terms using post-coordination.
(3) Historical relationships provide a link from retired or inactive concepts to the current replacement concept.
(4) Additional relationships allow nondefinitional information to be distributed.

12.5.1 Defining Relationships

Defining relationships are used to define concepts by its relationships with other concepts. Defining relationships are either supertypes (parents) or defining attributes. Defining attributes are specified as attribute–value pairs, where each attribute and value is itself a concept. Only relationships that are always true are used as defining relationships.

A concept is fully defined if its defining relationships are sufficient to distinguish it from all its supertype and sibling concepts. Primitive concepts do not have the unique relationships needed to distinguish them from their parent or sibling concepts.

A sufficient definition consists of a set of defining relationships (and relationship groups) which taken together imply a particular meaning. A necessary definition consists of a set of defining relationships (and relationship groups) that express all the attributes that are necessarily true about a concept. Other information, which is usually but not necessarily true, is outside the scope of SNOMED CT.

A lot of SNOMED CT is not yet fully defined.

Every active SNOMED CT concept (except the SNOMED CT root concept) has at least one supertype. Supertype relationships are used in subsumption testing, which allows users to identify whether a patient with a specific condition has a more general condition that subsumes the specific one. It allows one to answer questions such as: "Is angina pectoris a type of heart disease?"

This is useful in information retrieval, because most research, audit, and decision-support applications usually assume that a supertype includes all its subtypes (children and descendents). For example, a project may need to identify all patients with diabetes, which implicitly assumes that all types of diabetes will be included.

12.5.2 Qualifying Relationships

Qualifying attributes are used in post-coordinated expressions. They are optional nondefining relationships that may be applied by a user or implementer.

The possible values an implementer can select in assigning a qualifying characteristic to a concept are constrained by the SNOMED CT Concept Model.

12.6 Concept Model

The SNOMED CT Concept Model is the complete set of rules that govern the ways in which concepts are permitted to be modeled using relationships to other concepts.

The most common rules assert which attributes and values can be applied to a particular type of concept. For example, the Concept Model asserts that a concept that is a subtype of "clinical finding" can be related using the attribute "finding site" to a concept that is a subtype of either "anatomical structure" or "acquired body structure."

Other rules indicate whether particular relationships should be grouped together and whether the same attribute can be applied more than once to the same concept. These more complex rules are not discussed further here.

The Concept Model has around 50 attributes that may be combined with the various subtype hierarchies in a complex web of relationships. A machine-readable concept model (MRCM) is being developed to make this easier to use.

Each attribute has a domain and a range. The domain is the set of concepts (objects) to which this attribute may be applied; the range is the set of values which are allowed.

The Concept Model is too large and complex to describe in full here. See the SNOMED CT User Guide.

12.7 Clinical Findings

Clinical findings in SNOMED CT represent the result of clinical observations, assessments, or judgments and include both normal and abnormal clinical states; this covers a very broad range of concepts, with a similar range as the HL7 Observation (Fig. 12.4).

The default context for a clinical finding is that the finding has actually occurred, it relates to the subject of record (the patient), currently or at a stated past time.

The disease (or disorder) sub-hierarchy of clinical findings covers abnormal clinical states only.

Clinical finding allows the attributes including: finding site, associated morphology, associated with (including after, due to, and caused by), severity, and clinical course.

Finding Site specifies the body site affected by a condition and has values from either the body structure hierarchy. For example – appendicitis: finding site = appendix.

Associated Morphology specifies morphologic changes seen at the tissue or cellular level that are characteristic features of a disease and has values from the

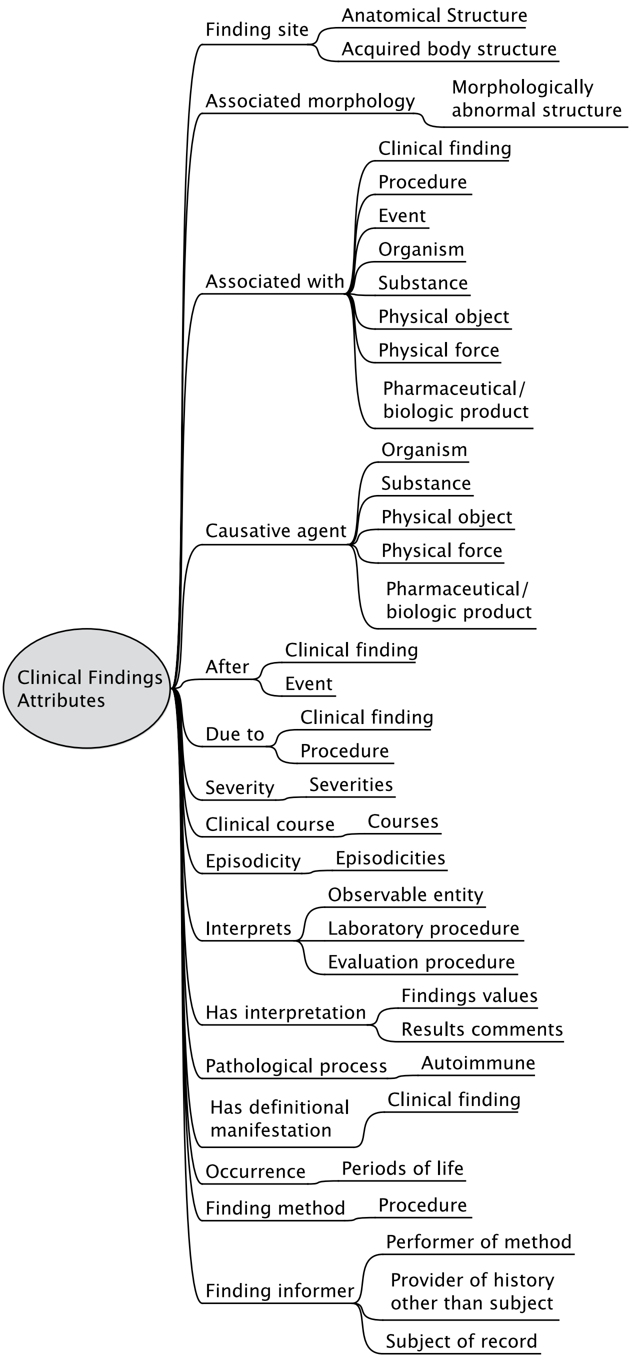

Fig. 12.4 Clinical Findings' Attributes

morphologic abnormality hierarchy. For example – Appendicitis: associated morphology = inflammation.

Associated With asserts that a clinical finding is associated with another clinical finding, procedure, pharmaceutical product, substance, organism, physical object, physical force, or event in some unspecified way. It has three subtypes: After, Due To, and Causative Agent. After is used when a clinical finding occurs after another clinical finding or procedure, emphasizing the sequence of events. Due To relates a clinical finding to its cause, which may be another clinical finding or an event. Caused By identifies the causative agent of a disease such as an organism, substance, pharmaceutical product, physical object, or force.

Severity is used to represent the severity level of a clinical finding.

Clinical Course represents the course and/or onset of a disease, such as acute or chronic.

Other attributes of clinical findings include episodicity, interprets, has interpretation, pathological process, has definitional manifestation, occurrence, finding method, and finding informer.

12.8 Procedure

In SNOMED CT the definition of a procedure is broader than the definition in the HL7 RIM. In SNOMED a procedure is broadly defined as including any type of action done intentionally as part of the process of delivering health care, including history taking, physical examination, testing, imaging, surgical procedures, disease-specific training and education, counseling, and administrative procedures (in HL7, the term procedure is limited in the main to surgical procedures).

The Procedure Model introduces the idea of direct and indirect attributes. A procedure may have a direct site or an indirect site. "Procedure site – direct" is used when a procedure is on the structure itself (e.g., excision of appendix), while "procedure site – indirect" is used when the procedure is on some other material such as a thrombus or removing a foreign body from the site.

The same idea is applied to morphology. An operation to remove a neoplasm uses the "direct morphology" attribute to link the procedure to neoplasm. While a procedure to remove stitches from a wound has "indirect morphology" link to the wound.

Procedure device has four subtypes: direct device (the subject of the procedure), indirect device (which is rarely used), using device (e.g., core biopsy needle), and using access device (e.g., endoscope).

Other procedure attributes include: approach, direct substance, priority, has focus, has intent, recipient category, revision status, route of administration, surgical approach, using substance, and using energy (Fig. 12.5).

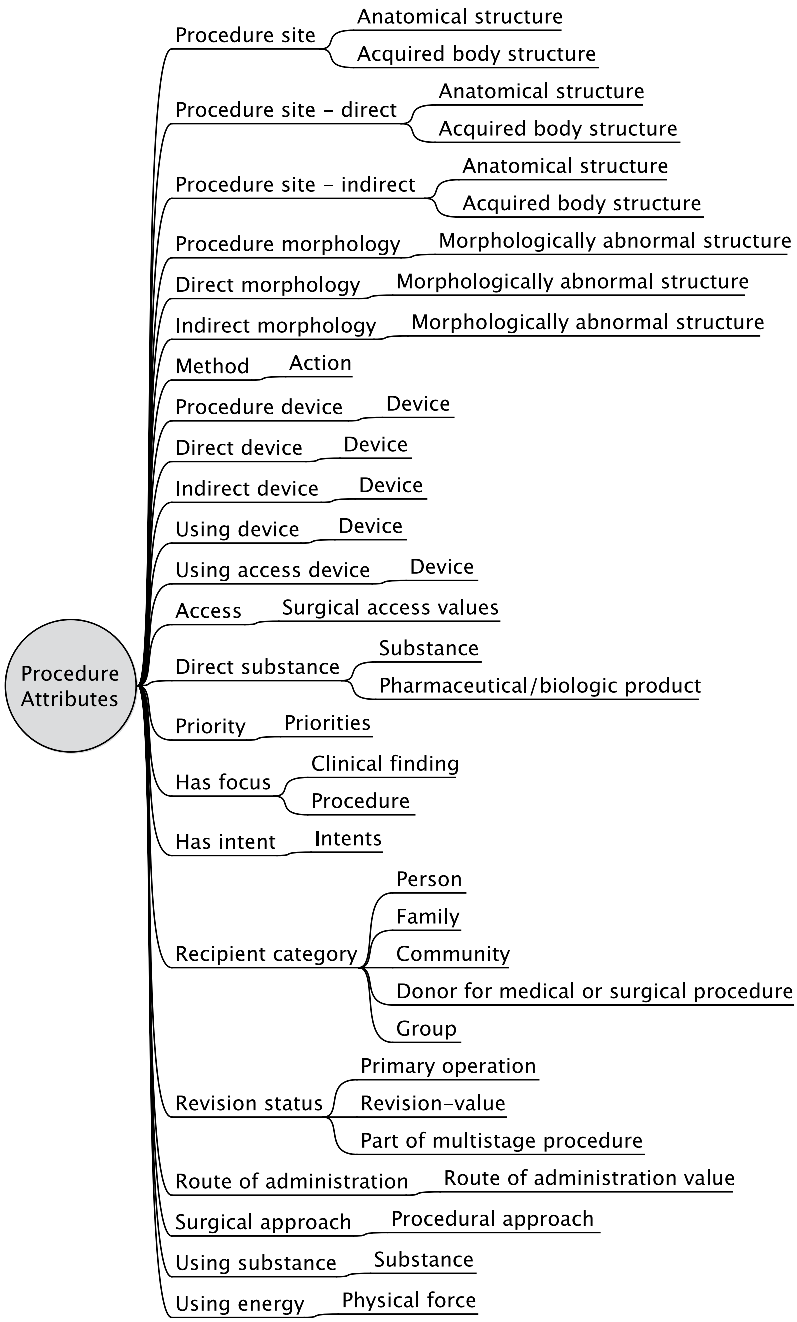

Fig. 12.5 SNOMED CT Procedure Attributes

12.9 Other Attributes

12.9.1 Observable

Observable entities (usually referred to as observables) are qualities or properties which can have values applied to them. When given a value, observables provide a specific finding or assertion about health-related information. Examples include the names of lab tests, physical exam tests, and dates of significant events.

A clinical finding can usually be modeled as an observable entity plus a value. One use is as headings on a template. For example, the concept "gender" is an observable entity, while the concept "female gender" is a clinical finding.

12.9.2 Event

Events are occurrences that happen which are not healthcare procedures or interventions, such as travel, earthquake, and death.

12.9.3 Situation with Explicit Context

Statements with explicit context are those that express something about who is the subject of the record (the patient or someone else), when the event took place (past, present, or future), whether a finding was present, absent, or unknown, and if a procedure was done, not done, or planned.

12.9.4 Staging and Scales

This hierarchy contains assessment scales and tumor-staging systems.

12.9.5 Specimen

Specimens are entities that are obtained, usually from a patient, for examination or analysis. They can be defined by attributes such as:

- Body structure from which they are obtained
- Procedure used to collect the specimen
- Source from which it was collected
- Substance of which it is comprised

12.9.6 Body Structure

Body Structure includes normal as well as abnormal anatomical structures. Normal anatomical structures can be used to specify the body site involved by a disease or procedure.

Body Structure has just one attribute, laterality (e.g., Left, Right, Left, and Right).

12.9.7 Organism

The organism hierarchy is limited to organisms of significance in human and animal medicine, including causes of diseases. Sub-hierarchies of organism include: animal, microorganism, and plant.

12.9.8 Pharmaceutical/Biologic Product

The SNOMED CT core distinguishes between medicinal substances, such as active ingredients and manufactured products.

Medicinal products have three distinct types:

- Virtual Therapeutic Moiety (VTM) is product name only (e.g., aspirin). The VMP is linked to the active ingredient substance.
- Virtual Medicinal Product (VMP) includes strength and form as used on a drug prescription (e.g., aspirin 75 mg tablet).
- Actual Medicinal Product (AMP) is a single unit dose of a marketed medicinal product and includes specific manufacturer and proprietary brand names. AMPs are often country-specific and are therefore found in national extensions, such as the NHS dictionary of medicines and devices (dm + d).

12.9.9 Physical Object

Physical Objects may be natural and man-made (e.g., a motor car).

12.9.10 Physical Force

Physical Force is used to represent forces which play a role in causing injuries.

12.9.11 Social Context

The Social Context includes social conditions and circumstances relevant to health care such as employment, education, housing, care provision, family relationships and lifestyle.

12.9.12 Environments and Geographic Locations

The Environments and Geographic Locations hierarchy includes types of environments as well as named places such as countries, states, and regions.

12.9.13 Qualifier Value

Includes concepts such as laterality.

12.9.14 Record Artifact

Used to refer to parts of electronic patient records.

12.9.15 Special Concept

One sub-hierarchy of Special Concept is Inactive concept, which is the supertype for all concepts that have been retired. Another is for special navigational concepts.

12.9.16 Linkage Concept

Linkage Concepts include SNOMED CT attributes used for defining concepts and in post-coordinated expressions.

12.10 SNOMED Expressions

Clinical records are created for the purpose of providing information about events or states of affairs. For example, a complete health record is a record artifact that also may contain other record artifacts in the form of individual documents or reports, which in turn may contain more finely granular record artifacts such as sections and section headers.

A SNOMED CT expression is a collection of references to one or more concepts used to express an instance of a clinical idea. An expression may be just a single conceptID (and optional term) or a combination of two or more concepts, which is known as a post-coordinated expression. The term expression is used to indicate that it expresses an instance of a real-world phenomenon (such as a headache), represented by a referenced concept occurring in a particular patient.

An expression is said to be pre-coordinated when a single concept identifier is used to represent a clinical idea. Including commonly used concepts in a pre-coordinated form makes the terminology easier to use (Fig. 12.6).

12.11 Post-Coordination

SNOMED CT allows the use of post-coordinated expressions to represent a meaning using a combination of two or more concept identifiers. Post-coordination describes representation of a clinical meaning using a combination of two or more codes. SNOMED CT allows concepts to be represented in a post-coordinated form. One form of post-coordination involves creating a single expression consisting of several concepts related by attributes.

Post-coordinated expressions may be single-level expressions or nested to any number of levels of detail. In a nested expression each attribute value is itself an expression, which can be single-level or nested.

Nested post-coordinated expressions provide a powerful but complex means to allow SNOMED CT to describe things in great detail and to cover unexpected requirements.

There are three main forms of post-coordination:

12.11.1 Subtype Qualification

Subtype Qualification is where the concept is elaborated (that is, linked with an attribute concept) in such a way as to result in a post-coordinated expression which is equivalent to a subtype of the unelaborated concept.

For example, the concept “Asthma” can be qualified with the attribute concept “Severe” to produce an expression that is the subtype of the concept “Asthma.” Ideally, where expressions have been post-coordinated and saved in this way, at the point of retrieving the stored codes the application should be able to compute equivalence and, therefore, subsumption.

There are four types of subtype qualification: qualification; refinement of a defining attribute; addition of unsanctioned qualifiers; and addition of “nested” qualifiers.

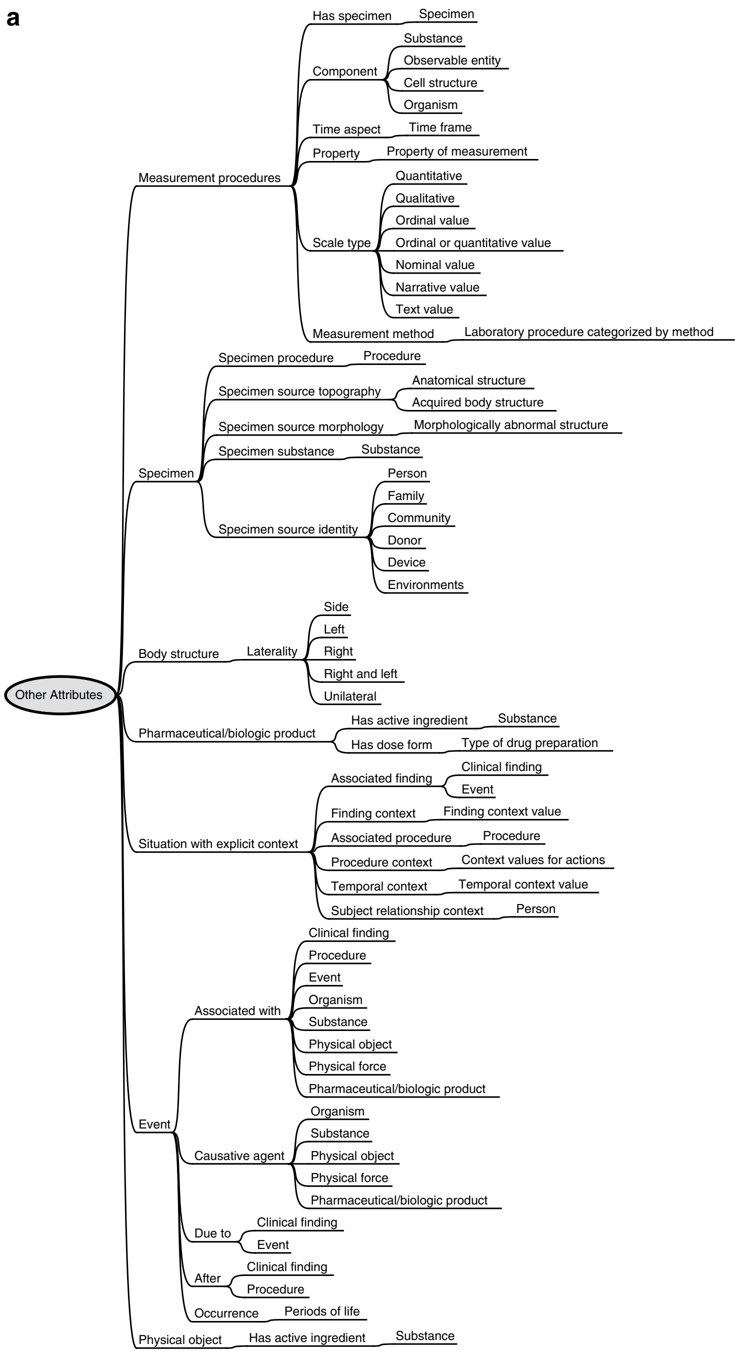

Fig. 12.6 SNOMED Expressions

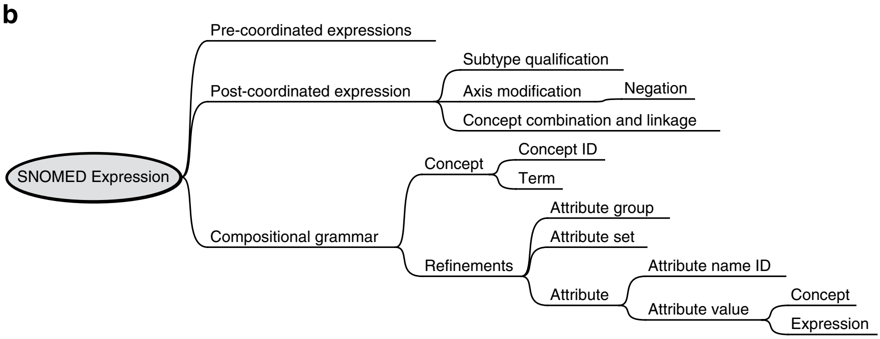

Fig. 12.6 (continued)

12.11.2 Axis Modification

Axis Modification is where the elaboration fundamentally changes the meaning of the concept, rather than simply qualifies or refines it. Such an elaboration of the concept results in it no longer being subsumed as a subtype of a parent concept. For example, if we elaborated the concept “Asthma” to associate it with a patient’s mother, the meaning of the resulting expression would have a fundamentally different meaning, and therefore different clinical implications, from the noting of the concept of “Asthma” by itself. If the clinician were running a query of all instances of asthma in their practice, they would not expect the system to return instances where asthma is linked to a family member.

Additionally, concepts that are post-coordinated with a negation concept (such as “known absent”), have their meanings fundamentally shifted. A modified concept is not a subtype of the unmodified instance of the concept. For example, “Asthma not present” is not the subtype of the context-neutral concept, “Asthma.” Instead, it is a type of “Explicit context.” Negated concepts subsume in the opposite direction to their positive counterparts.

Whereas positive expressions are subsumed by more general instances, negated expressions are subsumed by more specific negative expressions. For this reason modified concepts need to be treated differently to concepts that have been refined through subtype qualification.

12.11.3 Concept Combination and Linkage

Concepts can be combined by way of a number of relationships, such as causal or temporal relationships. For example, one may post-coordinate the concept “Laparoscopic procedure” with the concept “Cholecystectomy,” which is the canonical equivalent of the concept “Laparoscopic cholecystectomy.” Concepts can also be linked together, for example, to indicate causality. In this way, the concept "Anemia" can be linked by the attribute "Due to" to the concept "Ascorbic acid deficiency." The resulting post-coordination is equivalent to the concept "Anemia due to ascorbic acid deficiency." In this example, the concept "Anemia" is the "base" concept, with "Due to" the attribute name, and "Ascorbic acid deficiency" the attribute value.

12.11.4 Compositional Grammar

SNOMED CT expressions are usually presented using a notation known as compositional grammar (IHTSDO 2008a).

Concept identifiers (ConceptID) are shown as a sequence of digits. Other SCTIDs are not usually shown in compositional grammar.

Display names (either FSN or preferred term) are delimited by a pair of pipe ("|") symbols following the concept identifier, e.g.,

87628006 | bacterial infectious disease |

Whitespace characters, such as space, tab, and linefeed, are ignored, except within a display name.

A colon (":") is used as a refinement prefix, between the concept and its refinement. A refinement consists of one or more attributes and/or attributes groups.

The equals sign ("=") is used as an attribute value prefix between the attribute name and its value. Each attribute consists of an attribute name and an attribute value.

The following example specifies a bacterial infectious disease caused by streptococcus pneumoniae.

87628006 | bacterial infectious disease |:
246075003 | causative agent |=9861002 | streptococcus pneumoniae |

When a refinement includes more than one attribute a comma (",") is used as the attribute separator.

The value of an attribute may be represented by a nested expression rather than a single concept identifier. In this case, the nested expression is enclosed in parentheses.

The following example specifies a bacterial infectious disease affecting the left upper lobe of the lung and caused by streptococcus pneumoniae. The nested expression localizes and lateralizes the site of the disease.

87628006 | bacterial infectious disease |:
246075003 | causative agent |=9861002 | streptococcus pneumoniae |,
363698007 | finding site |=(45653009 | structure of upper lobe of lung | :
272741003 | laterality |=7771000 | left |)

The plus sign ("+") is used to combine attributes to indicate that both apply to a concept. For example:

87628006 | bacterial infectious disease |+50043002 | disorder of respiratory system |

This means a disorder that is both a bacterial disease and disorder of the respiratory systems, for example, "bacterial pneumonia."

A stated definition view is the set of relationships (and groups of relationships) that an author has stated to be defining characteristics of a concept. An inferred definition view is derived from the stated concept definition by applying a consistent set of logical rules to the definition taking account of the definitions of related concepts. The standard SNOMED CT distribution includes the relationships table which represents an inferred view of the definitions of all active concepts, restricted to the proximal supertypes for each concept.

Another of these inferred concept definition views provides a comprehensive view of all supertype ancestors and is known as transitive closure. This includes a high-level of redundancy and when applied to the full content of SNOMED CT it results in tens of millions of relationships. However, a transitive closure table provides a direct way of checking whether any one concept is a subtype of any other and provides a means of high-performance subsumption testing. "A pre-computed transitive closure table appears to out-perform other options and is robust, flexible and easy to implement" (Fig. 12.7) (IHTSDO 2008b).

12.12 Extensions

When an organization creates an extension to SNOMED CT, the new components in the extension need to be identified as part of that particular organization's extension. SNOMED CT does this by allocating an identifier to the organization (the Namespace-identifier). The organization would include its namespace-identifiers as part of the identifiers originated in its namespace. The Namespace-identifier is part of the SCTID.

If no namespace is identified in a SCTID, it is assumed that the component is part of the International Release of SNOMED CT. In these cases, SCTIDs are used in an abbreviated form, without the seven-digit namespace-identifier.

12.13 Subsets

Subsets are key to the practical application of SNOMED CT. When using a system, a user is only interested in a tiny proportion of the whole of SNOMED CT and subsets provide one way of providing this. Subsets are used to specify picking lists for specific data entry fields, to provide priority search lists containing frequently used context-specific terms and for exclusions (terms and concepts that should not be used). Subsets are one of the keys to making SNOMED CT usable.

The number of members in a subset depends on its purpose. For example, a language subset will normally contain at least one description for every concept and so may have hundreds of thousands of members. A subset containing the concepts

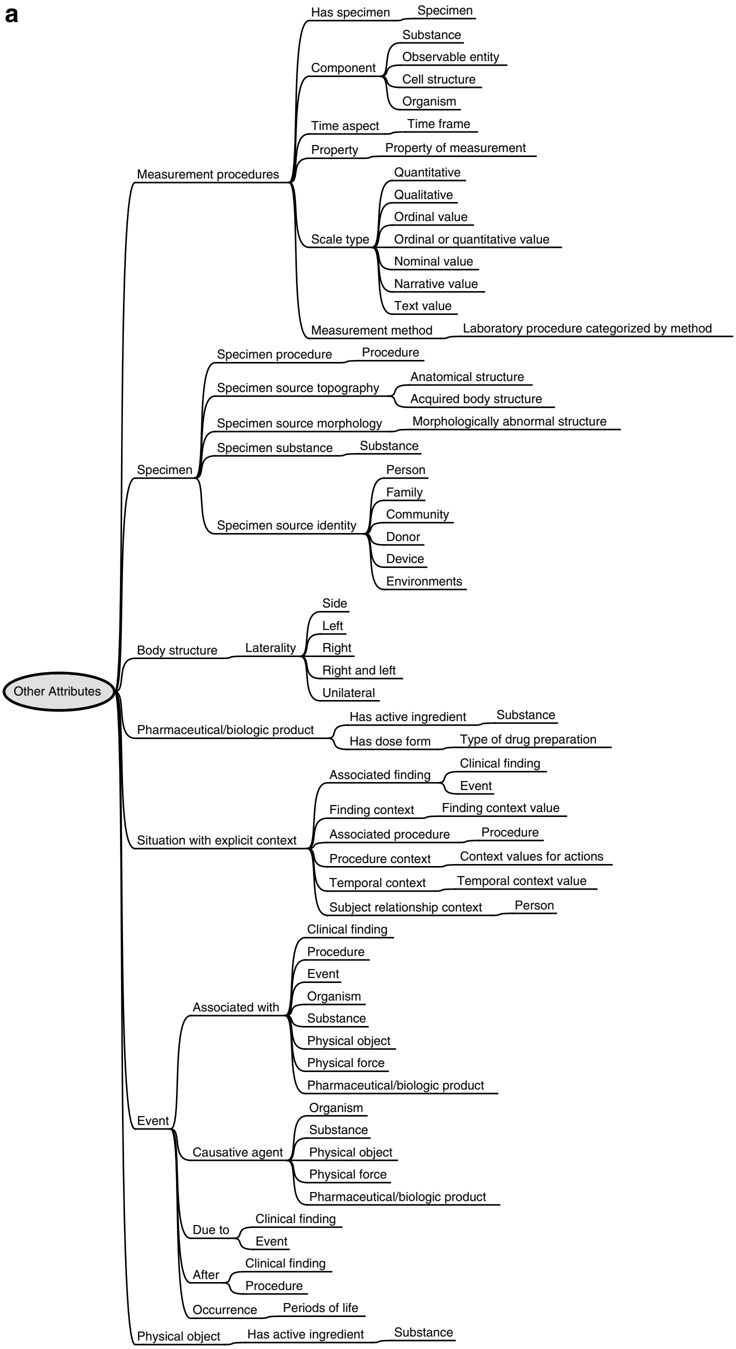

Fig. 12.7 SNOMED CT expressions

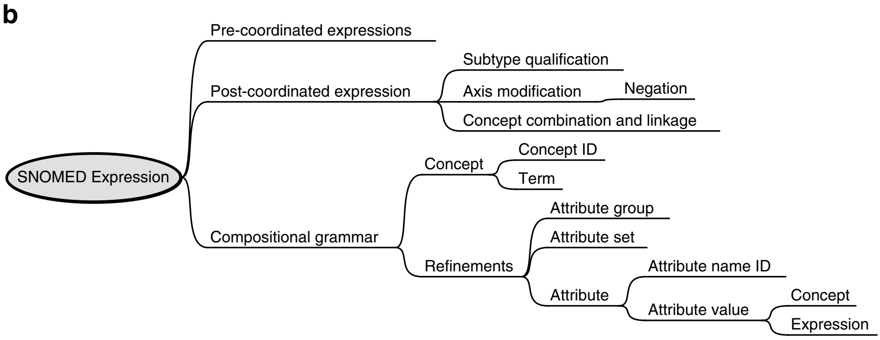

Fig. 12.7 (continued)

commonly used in a particular specialty may contain several thousand members, but the set of concepts or descriptions for a specific clinical protocol, template, or data entry field may only contain a few members (Fig. 12.8).

One way to think of a subset is as an index entry pointing to a set of pages relevant to a topic. A subset is simply a set of Concepts or Descriptions, or Relationships that are appropriate to a particular language, dialect, country, specialty, organization, user, or context. Subsets can be of any size, very large and very small. In its simplest form, a subset is a list of SNOMED identifiers (SCTIDs) each of which refers to a subset member, which is a SNOMED CT description or concept.

However, the subset mechanism provided in SNOMED CT appears complex, in part because the same mechanism is used to do several different jobs. Although subset members only have four fields, the use of three of these fields (MemberID, MemberStatus, and LinkedID) differs markedly depending on the type of subset.

There are eight types of subset

- Language Subset
- Navigation Subset
- Realm Concept Subset
- Realm Description Subset
- Realm Relationship Subset
- Context Concept Subset
- Context Description Subset
- Duplicate Terms Subset

Language subsets contain pointers to the terms (descriptions) used in a particular language or dialect; they are usually the largest. SNOMED CT can be translated into any language or dialect. Each translation uses existing concepts with new language-specific descriptions. A language subset is a set of references to the descriptions that make up that language edition and the Member status field indicates whether each description is the preferred term in that language.

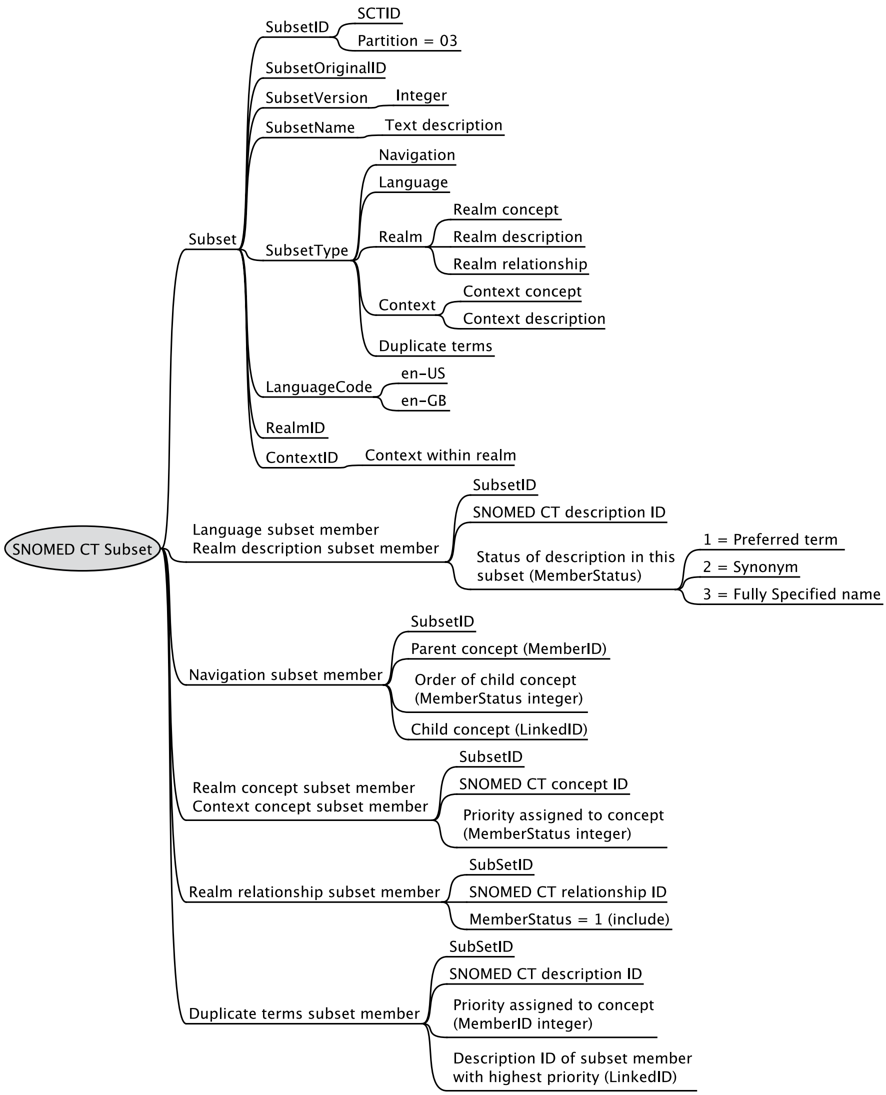

Fig. 12.8 Subsets

Navigation subsets provide an alternative tree-view of a set of terms in a specified order. Navigation hierarchies can reflect the way that people think when entering data, sometimes referred to as the model of use. Navigation hierarchies are useful for display, navigation, and data entry. These are usually handcrafted to limit the number of levels and the number of choices at each level, to list terms in a sensible order, and to ensure consistency over time. However, large numbers of handcrafted hierarchies are difficult to maintain. Each subset member includes a reference to the parent concept, a child concept, and the sequence order of that child.

Realm subsets are used for the terminology used in a specific area of expertise, preference, or authority. Examples of realms include: a specialty, a professional discipline, an organization, a country, or a specialty within a country (e.g., US dentists).

Context subsets are used for a specified part or field of a patient record, application, protocol, query message, or other communication specification.

Note that subsets are not necessarily mutually exclusive and the content of subsets may overlap.

Each subset comprises two logical tables: the Subsets Table provides metadata information with one row for each subset describing its scope, including an identifier for the whole subset, its name, version, type, and additional data such as language, realm, and context domain.

The subset members table contains one record for each subset member, which may refer to a concept, description, or relationship and assigns a member status to each.

The process of subset development and maintenance is challenging, time-consuming, and labor-intensive. It is no easier to develop and maintain a subset that has the support of a large clinical community than it is to develop any other consensus standard. Clinicians want subsets that meet their particular needs, complete and yet focused. Subset development is likely to remain a growth area for many years to come.

A typical subset development project is likely to involve the following steps:

- Establish scope and team
- Identify relevant terms from existing records and evidence base (literature)
- Extract terms and compare to SNOMED CT content
- Derive subsets, including hiding some of the complexity of SNOMED, allocating priorities, and the sequence of terms
- Validation using panels and in practice (comprehensiveness, relevance, reliability, usability)
- Implementation and deployment of software that enables users to achieve their goals
- Maintenance

A number of tools have been developed to help with the task of building, maintaining, and using subsets.

It is planned to replace the subset mechanism with a new Reference Set (RefSet) mechanism, which will better distinguish between the different purposes for which subsets are being developed.

12.14 Cross Mappings

Cross Mappings are used to reference other terminologies and classifications, such as the International Classification of Diseases (ICD). Cross mapping is needed to allow data collected for one purpose, such as clinical care, and to be used for another purpose such as reimbursement to avoid the costs and errors of having to reenter data. It is also needed when data need to be migrated to newer systems Fig. 12.9.

Ideally, computer programs will use the cross maps table to translate codes automatically, but unfortunately the rules of many coding systems, such as ICD-9 CM and ICD-10 are such that fully automated coding is not yet feasible.

The structure of the cross-mapping files allows a single SNOMED concept to be mapped to one or more target codes, with a stated level of precision.

Each cross map matches a SNOMED CT concept with another coding scheme (the target scheme).

The SNOMED CT structure to support Cross Mapping includes three tables:

- Cross Map Sets Table: Each row in this table represents a Target Scheme to which SNOMED CT is mapped (e.g., ICD-10).

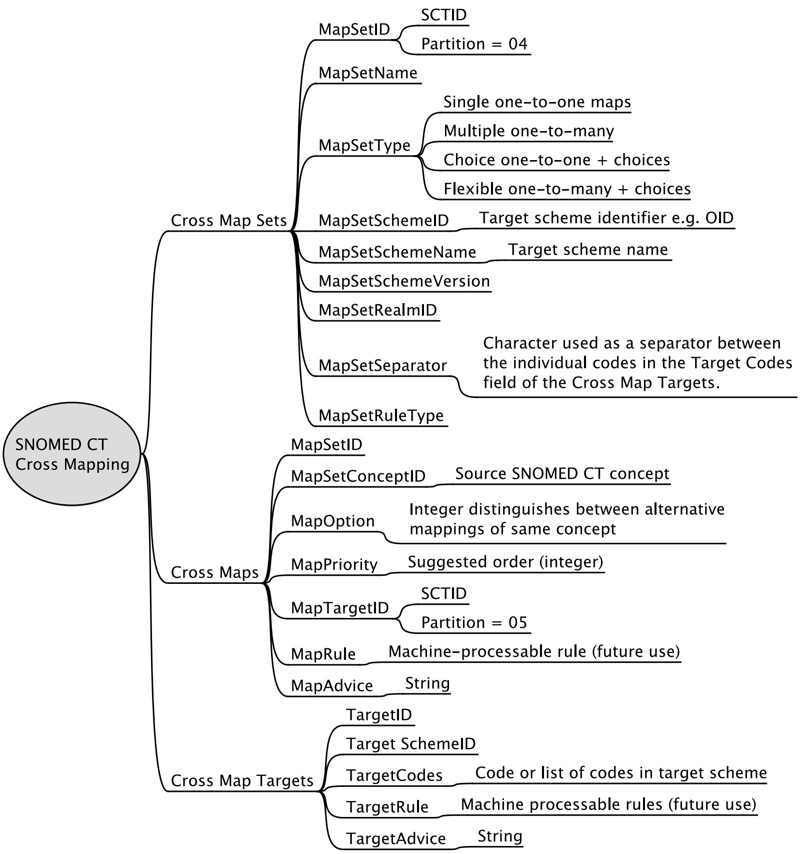

Fig. 12.9 Cross maps

- Maps Table (Cross Maps Table): Each row in this table represents one option for mapping a SNOMED CT Concept. This identifies the row in the Targets Table with the target codes for this SNOMED concept. Each concept may have one or more mappings to the Target Scheme.
- Targets Table (Cross Map Targets Table): Each row in this table represents a code or set of codes in the Target Scheme, which provides a mapping for one or more SNOMED CT Concepts.

Currently available cross maps include ICD-9CM, ICD-1,0, and the OPCS-4 classification of surgical procedures.

12.15 History Files

SNOMED CT includes component history files, which maintain a record of changes to existing components, in line with the principle of permanence.

12.16 Releases

The International Release consists of the Core, the IHTSDO's Specifications and other documents and software that the IHTSDO includes within the International Release.

The Core is the SNOMED CT content that is controlled, maintained, and distributed by the IHTSDO. The IHTSDO has the sole right to modify the Core.

The International Release is licensed by the IHTSDO, whether distributed on its own or as part of a Member's National Release (see below). Members' National Releases consist of the International Release plus National Extensions, Derivatives, and other documents and software that members choose to include within their National Releases.