Chapter 13
Using SNOMED and HL7 Together

One of the lessons of the past few years is that we cannot separate the issues of information structure from those of terminology. The idea that syntax and semantics are independent has been shown to be an oversimplification. We cannot slot any terminology into any data structure and expect it to work. Perhaps, it is surprising that for so many years, so many people thought that we could. After all, in every spoken language, such as English, the grammar (syntax) and words (terminology) have evolved together.

For reasons, which seemed good at the time, the standards development organization responsible for healthcare terminologies (such as SNOMED CT), message syntaxes (such as the HL7 V2 and V3) and information models, developed independently and in parallel. The main justification was the perceived need to interoperate with a wide range of legacy schemes and to meet the needs of different national requirements and languages. SNOMED CT was designed to be syntax-neutral, so it could work with any syntax. Reference models such as the HL7 RIM and EN 13606 set out to be terminology-neutral.

We now recognize that we need a special process called terminology binding, to specify how a specific terminology and a structural model shall be used for each use case (Markwell 2008). This is not straightforward because the flexibility built into both HL7 and SNOMED CT means that there is normally more than one way to perform the binding.

An important complication is that electronic information systems operate at two different levels, which Rector et al. (Rector et al. 2006) describe as "models of use" and "models of meaning."

13.1 Model of Use

The model of use describes how the system, such as an EHR, is actually used, and in particular the way that data are captured and displayed. The model of use represents the human interface. The same data could be captured in any number of different ways and the skill of the designer is to make this as easy and efficient as possible for each work process. Different use cases warrant different models of use.

T. Benson, *Principles of Health Interoperability HL7 and SNOMED*, HI,
DOI 10.1007/978-1-84882-803-2_13, © Springer-Verlag London Limited 2010

Data capture (the model of use) needs to be tuned to make individual work process as quick and easy as possible. Healthcare work is made up of a relatively small number of common high-volume processes, such as requesting tests, prescribing medication, and making referrals, intermingled with a much larger number of less common processes that are specific to the individual patient's condition, the location along their care pathway, and even the specialty of the clinician. The specifics of many tasks (such as patient assessment, diagnosis, and monitoring, and planning treatment, tests, and follow-up) differ greatly according to what is the matter with the patient.

One method of data capture is what we call ad hoc data entry in which the user simply types in the first few letters of the term in mind and the system responds with a list of matching terms. These terms are often organized hierarchically in a tree-structure, as exemplified by the Read Codes. The user can scan up and down the hierarchy, moving between levels of greater or lesser detail to find just the right term. The term selected is shown, providing the user with an opportunity to validate the choice. This method of data capture has been used successfully by almost all GPs in the UK for over 20 years, though there remain problems with the completeness and consistency of the data captured in this ad hoc way.

The practicality ad hoc data entry depends on keeping the size of the list of matching terms to a size that can be scanned quickly and easily. Psychological research has shown that the human eye can read about six lines of text of about 15 characters without moving the eye muscles. Anything more takes longer to read and increases the error rate.

One way to limit the number of options displayed is to provide subsets (value sets) for each task, which greatly restrict the number of options available for each task, or to display the most commonly used items first (known as velocity coding).

Structured data entry removes the requirement to type in the first few letters of the term, and the user simply points and picks with a mouse or on a touch screen, which can be very quick and easy. This works well when the work process has a narrow scope and clearly defined pathway; but the catch is, first of all, the user has to identify the specific work process. Every option is unambiguously defined in terms of the flow of questions, the terms used and the way that each item of data are recorded and coded. The computer protocol must follow the natural clinical order of the task. Items on picking lists need to be grouped in a natural order.

Another difficulty of data capture is the need to integrate data collection with the display of previous records and to integrate with clinical decision support warnings and alerts. At any one moment, the user's screen can only show a limited amount of information, although one patient's EHR may contain thousands of separate items of information. The solution is to provide multiple views into the record with almost instant switching between views.

Typical views are:

- Chronological. The most natural way to display information is in date/time accession order. Reverse order shows the most recent first. Data may be grouped by the clinically relevant date, so tests may be linked both to the date of sample, as well as to the report date.

- Summary views showing current demographics, clinical episodes or problem list. These summaries may be a starting point for further drill-down.
- Problem-orientation shows entries linked to a specific problem, along the lines of problem oriented medical records (POMR).
- Reminders and alerts for drug sensitivities, data that are missing from the records, etc.
- Clinician view allows users to remind themselves of the last entries they made for this patient.
- Notes type views provide a way to display information by type, to list separately the diagnoses, medication, operations, test results, and other classes of entry.
- Detailed flow charts and graphs can be provided for specific types of data, which need to be monitored and such as blood pressure, blood chemistry, and other numeric values.

Excellent work has been done on how best to capture and display data as part of the NHS Common User Interface project, undertaken in collaboration with Microsoft. This set out to deliver a consistent user experience across applications and devices, which would be useful, usable, and compelling to clinicians, thus encouraging increased adoption of health IT systems, improve patient safety, reduce the cost of adopting health IT systems, in terms of end-user training and support costs, and increase the productivity and effectiveness of clinicians.[1]

The work is presented in a number of documents:

- Terminology Matching[2]
- Terminology Elaboration[3]
- Display Standards for Coded Information[4]

The three main ways of entering SNOMED CT coded data are to use forms, single concept matching or text parser matching.

Forms are used where the user selects set options rather than entering text; here, an encoding interface is not needed, because the clinical codes should be embedded within the form itself.

In single concept matching, the user enters a note for a single clinical concept and selects an appropriate match returned by the SNOMED CT database and elaborates this as required.

---

[1] See www.mscui.net

[2] NHS CUI Design Guide Workstream – Design Guide Entry – Terminology – Matching, 2007. www.cuisecure.nhs.uk/CAPS/Clinical%20Noting%20and%20Terminology/Terminology%20-%20Matching.pdf

[3] NHS CUI Design Guide Workstream – Design Guide Entry – Terminology – Elaboration, 2007. www.cuisecure.nhs.uk/CAPS/Clinical%20Noting%20and%20Terminology/Terminology%20-%20Elaboration.pdf

[4] NHS CUI Design Guide Workstream – Design Guide Entry – Terminology – Display Standards for Coded Information, 2007. www.cuisecure.nhs.uk/CAPS/Clinical%20Noting%20and%20Terminology/Terminology%20-%20Display%20Standards%20for%20Coded%20Information.pdf

In text parser matching, the user enters notes as unconstrained free text and the system matches words and phrases against the SNOMED CT database. This requires the system to identify and match SNOMED CT concepts as well as build postcoordinated expressions based on sanctioned attribute relationships from within the text.

The first step in the process is to identify the context, so as to identify the appropriate form, navigation or other subset to constrain the options and simplify the task.

Having entered text, the system may present a set of choices, which can be refined or elaborated in various ways (Fig. 13.1):

- Adding unstructured text to the expression to give the expression further meaning.
- Browsing alternative matches and refining selected matches; this may include reviewing the parents, siblings, and child concepts in a navigation hierarchy or qualifying attributes.
- Matching a SNOMED CT expression from within a passage of text and leaving some of the text itself unencoded but associated with the encoded expression.

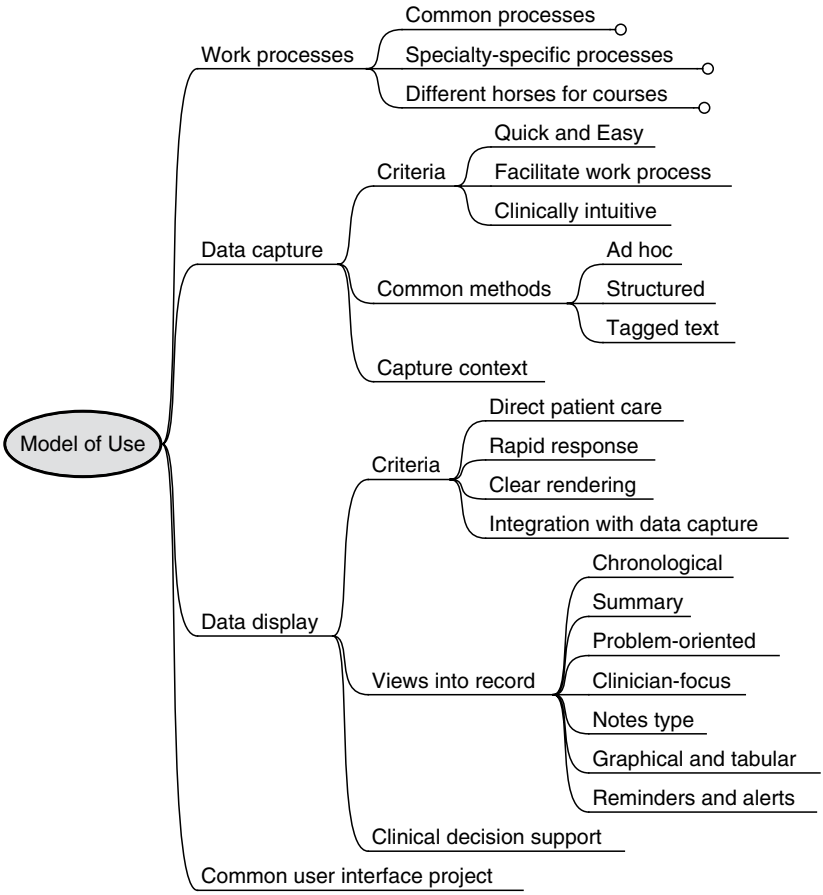

Fig. 13.1 Model of use

- Adding a qualifier to a SNOMED CT expression, using the qualifiers offered by the system, such as the severity of a condition. 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."
- Adding or selecting numerical or date and time values for a SNOMED CT expression. For example, the user should be able to add the value "38.9" and the unit "Degrees centigrade" to the concept of "Body temperature."

13.2 Model of Meaning

The model of meaning is a representation for reporting and analysis purposes, and represents our understanding of the world, so that we can reason about it in general and individual patients in particular. Computers are limited in their ability to process data and require information in a common, standardized format. The model of meaning provides such a format for data processing and reasoning. There is an enormous advantage for reporting and clinical decision support in working with a common model of meaning. This is the rationale for the development of reference models such as the HL7 RIM and the EN13606 (openEHR) reference model. Each model of use needs to be convertible into a model of meaning to make it computable.

The HL7 TermInfo report provides a set of recommendations on how to bind SNOMED CT with HL7 Version 3,[5] and expresses this requirement as follows:

> Every application has its own data entry screens, workflow, internal database design, and other nuances, and yet despite this, we talk of semantic interoperability. In order to achieve interoperability, and enable a receiver to aggregate data coming from any of a number of applications, it must be possible to compare data generated on any of these applications. In order to compare data, it helps imagine a canonical or normal form. If all data, regardless of how it was captured, can be converted into a common form, it becomes possible to compare.

This issue applies not only when we wish to exchange data between applications (semantic interoperability), but just as much within EHR systems, where a single system is used to support different models of use (semantic operability).

A number of suppliers have chosen to use a standard reference model as the basis of their applications architecture. The use of the HL7 RIM for this purpose has been termed RIMBAA (RIM-based Application Architecture). The EN 13606 (OpenEHR) reference model is also used in this way. It is also practicable to use a common proprietary architecture, which has a direct mapping to standards such as the HL7 RIM or EN 13606.

For storage and analysis, we need information to be in a form that allows it to be reused in a very wide range of different ways. The best way to do this is to store the data in a form that reflects the model of meaning. The storage system needs to keep information safe and secure, but the value of a system depends on what you can get out of it. The primary purpose is to enable questions to be answered accurately and efficiently. Accurate reports are precise and complete; efficient reports are quick and timely.

[5]HL7. Using SNOMED CT in HL7 Version 3; Implementation Guide. www.hl7.org/Special/committees/terminfo/index.cfm

Users need to be able to count and extract groups of patient records for innumerable reasons. The process of specifying a report involves:

- Selection and exclusion criteria to identify the records required. The first problem for the user is to identify what codes to use.
- The sort order in which to display the results.
- The content and format of each record displayed.
- The summary data at the top and bottom of the report (headings, totals, percentages).
- Where to send the report – screen, printer, or file.

Data retrieval for a user is particularly difficult if the underlying structure of the data is not the way that the user thinks it is, yet this is often the case. If the database is organized using a common model of meaning, then this has to be taught to everyone who wants to interrogate the data.

A second requirement is that users become familiar with the layout of their data collection screens and picking lists and may well regard this as "the way the system works." They reasonably expect to use the same lists and groups for their reports. The ability to report data in the way they are structured on data collection screens is a basic requirement. However, this can be quite awkward if the data are stored using a general-purpose model of meaning; system designers may need to go to some trouble to resolve this issue.

Another difficulty is the need for multidimensional analysis. A trauma surgeon may want to know:

- How many fractured shaft of femur did I see last year?
- How many open reductions and fixations did I perform (on any bone)?
- How many times did I use a locked intramedullary nail?

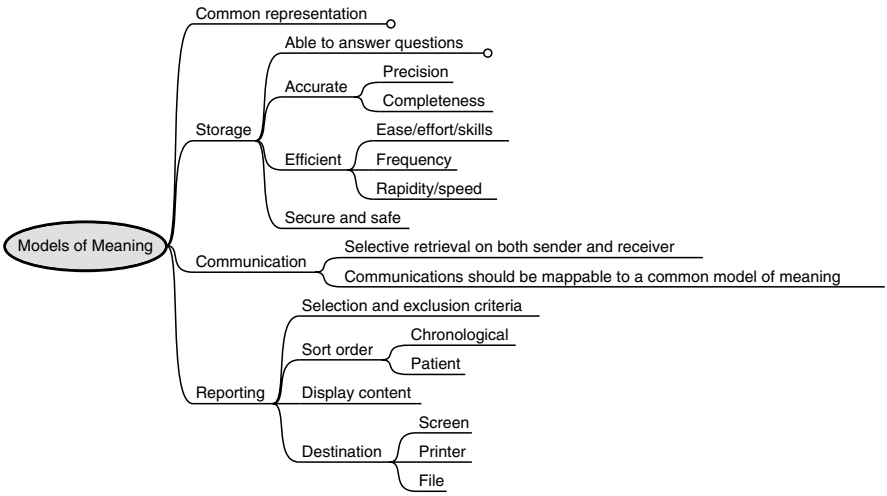

Fig. 13.2 Model of meaning

A simple code with a meaning such as "open reduction and fixation of fracture of the shaft of femur using a locked intramedullary nail" might appear at first sight to allow these questions to be answered, but in practice it is difficult to answer this sort of question in this way. This is why multiaxial coding systems such as SNOMED CT are needed, but they also have to be implemented properly in electronic patient record systems (Fig. 13.2).

13.3 Structural Models

In spite of the overlap between them, we need to bear in mind the fundamental differences between structural models such as HL7 RIM and clinical terminologies such as SNOMED CT.

The basic idea of a structural model is to provide a common framework that represents clinical information in a consistent standardized way, and relates each entry to common meta-data such as its subject, author, date/time, and location. Simple structural models can handle the basic structure of each entry quite easily, but become increasingly complex as the granularity of the data become finer and more detailed.

Structural models may be built around standardized reference models, such as the HL7 V3 RIM or the EN13606 Reference Models (including openEHR).

The basic structure of the HL7 V3 RIM is quite simple with its backbone of Act, Role and Entity, linked with ActRelationship and Participation association classes, although this top-level simplicity covers up a good deal of complexity. The HL7 Clinical Statement model provides a more refined pattern for representing clinical information, which is used in CDA Level 3.

The EN13606/OpenEHR reference model covers much the same domain as the HL7 Clinical Statement pattern, but is based on the traditional structure of medical records, with record components such as folder, composition, section, item, cluster, and element, as well as participants such as subject of care, healthcare professional, organization, and software or device. Many people find this easier to work with than having to translate these into the HL7 concepts of Act, Entity and Role. It is practicable to use EN13606 terms at the analysis level and to map these to the HL7 RIM-based artifacts as a subsidiary step.

EN 13606/OpenEHR Archetypes provide a standardized approach for representing and sharing clinical data specifications. Each archetype defines how the EHR reference model hierarchy is organized to represent the data for one clinical entry or care scenario. Because these archetype definitions are represented in a standardized form, using a formal language known as Archetype Definition Language (ADL), they can be shared and used across record-sharing communities to define how locally organized clinical data should be mapped consistently (even if the data originate from multiple systems).

EN13606/OpenEHR Templates are aggregations of constrained Archetypes, for a specific purpose. An Archetype usually contains a lot of detail that is not required in every use case. A Template contains just the functionality that is required.

The HL7 Clinical Statement pattern and the EN13606 reference models both contain a number of predefined attributes, which provide a structure to the model, but both allow the same information to be structured in more than one way. For example, any number of different coding schemes can be used and there is a good deal of flexibility in the way that data are handled. To take a simple example, both models allow the term "family history of asthma" to be handled either as a single concept, "family history of asthma," or as a composite of "family history" + "asthma."

13.4 Terminology Binding

Terminology binding is the process of establishing links between elements of a terminology such as SNOMED CT and an information model. Terminology binding is principally concerned with what can be said (concepts) – not how it is said (terms).

We often find situations where there are several possible ways to express the same meaning, due in part to the overlap between HL7 and SNOMED CT. It is impossible to draw a clean dividing line between the two.

The Terminfo Guide recommends:

The terminology should be used for specifying:

- Specific concepts and value sets
- Inclusion and exclusion constraints based on the SNOMED CT concept model
- Subsets/refsets including navigation hierarchies
- Simple semantic relationships, such as laterality
- Postcoordinated expressions at various levels of nesting

The structural information model should be used for specifying:

- Instance information and meta-data for any clinical statement such as dates and times, people and places, numbers and quantities
- Grouping and organization of the record framework including the record structure, the way that items should be grouped together, and anchors for terminology components, such as codes
- Differences due to the work process for a specific use case

There remain some grey areas, where the choices are not clear-cut, such as how best to handle issues such as context, negation, and uncertainty.

Recommendations should be widely understandable by implementers who are familiar with the use of SNOMED CT and HL7 V3; be able to be applied consistently and cover common scenarios.

Recommendations should result in instance representations that can be converted, by following a set of computational rules into a standard normal form.

We need to be able to confirm with tools that an instance conforms to the recommendations and that existing tools and applications, either in their current form or with reasonable enhancements, can produce the recommended instances.

In particular, the model should not require a combinatorial explosion of precoordinated concepts. For example, the model should not require the creation of the crossproduct of "Allergic to" and all drugs and substances.

Where more than one approach appears to be viable and broadly equal in respect of the criteria above a single approach is recommended to avoid unnecessary divergence. Where one approach has already been successfully implemented and the other has not, the implemented approach is recommended. Optionality is restricted where possible to simplify the delivery of semantic interoperability (Fig. 13.3).

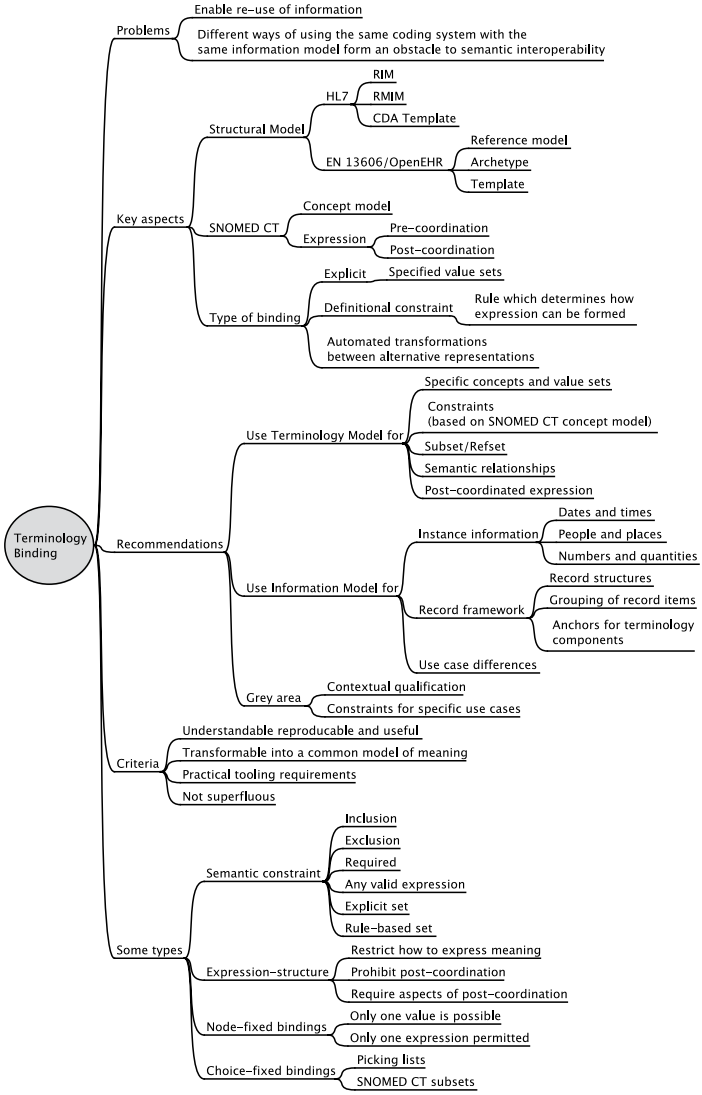

Fig. 13.3 Terminology binding