Clinical Modeling Primer & FAQ

Whenever a group undertakes the task of defining specifications for the data associated with a complex clinical topic there arises a need to seek that knowledge from the front-line clinical experts that are most familiar and experienced with that topic. But when Physicians, Nurses and clinical researchers first get introduced to clinical modeling projects they are confronted with a host of unfamiliar terms and concepts. This document intends to provide a ‘primer’ and introduction to the world of clinical model from the perspective of a newcomer from the clinical community.

What is a Subject Matter Expert?

One of the first terms to become familiar is Subject Matter Expert (SME), sometimes pronounced ‘ShMee’. Clinical SME’s are the Physicians, Nurses and researchers who provide the clinical expertise to the project. Modeling SME’s provide an understanding of the underlying computer languages, architectures and patterns that will be used to translate the clinical knowledge provided by the clinicians and translate that into machine understandable formats. Terminology SME’s, often referred to a Terminologists, provide the skills necessary to author any new clinical concept definitions that are required by the project. The bottom line is that clinicians will only have to be vaguely familiar with these other technical tasks, because each team will have members that provide that skill to the group.

What is a Clinical Model?

The term ‘Clinical Models’ refers to a layered collection of definitions that together define a specification for the appropriate set of data-elements for specific clinical events, actions and phenomena. Clinical models contain rules regarding how to combine the clinical concepts in order to produce meaningful clinical descriptions, e.g. a chest x-ray shows a lung mass in the base of the left lower lobe of the lung. In the database world, this would be referred to as the ‘Schema’. Models define the structures, fields, data-types and legal values required for specific clinical situations. Models act as a contract between the capture of information at the point of care, and all other future downstream uses of that information.

There are four 'layers' make up the full stack of clinical knowledge, the bottom three are what are referred to as models and the top-most is composed of the logical rules and definitions of computable clinical metrics which are enabled by standardized data-definitions.

Why do Clinical Models matter?

It has often been said that clinical modeling isn’t the ultimate goal, but a means to a collection of important and worthwhile goals which Clinical Models make possible. Models enable consistent and reliable exchange and process across diverse systems.
Clinical Models make the following possible:

- Reliable Exchange of Interoperable Medical Records
- Consistent execution of Clinical Decision Support Rules
- Predictable and accurate Clinical Quality Metrics
- Robust integration of Evidence Based Medicine

A good analogy would be that nations don’t build roads for the love of asphalt, they build roads because of the need to transport people and goods from one place to another and roads enable transport. Similarly clinical models are important for their ability to enable other capabilities that can then be ‘Portable across systems and durable over time.

**What are the parts that needed to define ‘Clinical Knowledge’?**

Clinical Knowledge is a somewhat vague ‘catch all’ term that is composed of many elements, of which Clinical Models are a very an important part. Models provide a specification of what information will be captured, exchanged and processes. Classical computer software relies on a known set of fields, organized into structures, so that logic and formulaic calculations can be expressed in a machine executable language.

It’s easiest to understand the full ‘stack’ of clinical knowledge as being composed of four ‘layers’. The bottom three layers represent the definition of the Clinical Models and the top-most being the logical intelligence.

**What are Concepts and Clinical Terminologies?**

Consistent processing of clinical information depends on an understanding of explicit meaning that can shared across systems. This is addressed through the use of standardized clinical terminologies, sometimes referred to as Lexicons or Ontologies, which provide formalized definitions of clinical Concepts with assigned identifiers.

The term ‘Concept’ is used to define and identify a single specific term used in medicine within a clinical terminology. A concept is a unique, definable idea that can be simple or complex. Concepts are defined with a preferred name, unique computer friendly identifiers, known synonyms, a definition, and a set of known relationships to other Concepts. Typically multiple
layers of hierarchical categorizations and known synonyms are also defined which are useful for search and processing.

Examples of a Concepts would be:

- 10200004 | Liver structure (body structure) |
- 71620000 | Fracture of femur (disorder) |
- 303673002 | Computed tomography of pelvic organs (procedure) |

What are the challenges related to Clinical Terminologies?

If we lived in a ‘perfect world’, there would be a single, universally adopted, clinical terminology that provided all of the relevant concept definitions needed by medicine - all from a single source. This perfect terminology would define sufficient machine understandable meanings and context for each concept and would be based on a formal architecture for performing reliable and consistent computation for logic and metrics. Unfortunately, we haven’t yet reached the ‘perfect world’ when it comes to clinical terminologies.

SNOMED Clinical Terms (CT) is generally acknowledged as being the terminology source that comes the closest to meeting the full needs of the community, and therefore has attracted the greatest scope of investment toward continued evolution and improvement, but it still must be understood that it is far from being fully sufficient. For the foreseeable future each new modeling project must accept that new clinical concepts will need to be authored and published as part of that effort.

Many projects may also leverage narrower and more focused terminologies, such as the one from the US National Cancer Institute (NCI), the NCI thesaurus, which provides a far more extensive and refined concept definitions related specifically to cancer. There is however a trade-off. Using two incompatible terminologies within the modeling stack imposes severe challenges to the computability of the clinical information.

The most practical ‘ultimate’ solution to a many-terminology world is to author machine understandable definitions for non SNOMED CT concepts using an expression language that is compatible with SNOMED CT. When SNOMED CT definitions are provided for non SNOMED CT concepts their effective computability becomes identical to that of native SNOMED CT concepts. So while this solution isn’t exactly perfect, it gets us closer.
How do new SNOMED CT Concepts get created?

When it is discovered that a needed Concept doesn’t exist, they are authored using terminology specific tools, such as TermSpace or SnoOWL, and the published through a SNOMED CT National release center of an organization who has been assigned a SNOMED CT Extension Namespace. Concept authoring is a highly exotic task and is only undertaken by dedicated ‘Terminologists’. Clinical SME’s need never worry that they would be asked to directly understand this process.

Over time these new concepts will be included in future public releases of SNOMED CT. Public release of new concepts greatly increases that scope of use and value of the models that are defined by each project and should likely be the ultimate goal of the majority of our efforts.

Concept authoring efforts that target public releases should consider collaborating with an organization such as the Healthcare Services Platform Consortium (HSPC) who has already borne the burden of registering a SNOMED CT Extension and establishing the necessary relationships, policies, and procedures necessary for curating new SNOMED CT concepts into public releases. The burden of recreating those channels for each individual project would likely be prohibitive.

**Clinician Summary:**

*Each project will include new SNOMED CT Concepts. These concepts will be authored by professional terminologists based on the needs that are uncovered during the clinical modeling effort. Clinical subject matter experts (SMEs) will experience little in the way of burden for this task, beyond supplying good definitions to facilitate authoring.*

What are Value-sets and Groupers?

Individual concepts are often only useful when they are grouped together into sets for defining specific options for a clinical ‘facts’. For example, when defining the ‘Shape’ of a Breast-Abnormality, there are a finite number of options (Irregular, round, Lobulated, etc.). This definition for the four legal options for Breast-Abnormality-Shape is known as a Value set definition and these are distributed as what is often referred to as a Ref-Set, using the same file format (RF2) as are all forms of new SNOMED CT content.

There are two common purposes for creating a set of associated concepts. The most obvious is for use in the information capture process to present a list of options to a clinician. The second is less externally visible and it comes into play when systems are processing information to determine if a concept is one of many concepts that signify a specific meaning in a certain context. When concept-sets define meaning at run-time the set is often referred to as a ‘Grouper’. It should be noted that the fundamental process for defining and publishing sets-of-concepts is identical whether their intended use is as a Valueset or Grouper.

**Clinician Summary:**
Individual Concepts are most commonly collected into sets that represent the possible `answers` for a given clinical `question`. If the question is `Shape of Legion?`, then there will be a finite number of options from which to choose. The agreed list of answers to each question is called a Value-set.

**Defining Questions, Answers, Qualifications and Context:**

Douglas Adams famously pointed out that the ‘Answer to the Ultimate Question’ (i.e. 42), had no actually meaning without the ‘Question to the Ultimate-Answer’, and this quandary is also prevalent within clinical modeling.

Consider any date stored in a data-field. Absent of knowing that the date represents a birth-day, time-of-death or scheduled appointment, that data-value has no practical use. In the SQL database world the meaning of a given data-element is provided by the combination of a field-name and human readable documentation to guide other engineers in the appropriate use of that information.

Modern clinical modeling dedicates substantial effort into specifying machine understandable definitions for the meaning, use and context of the individual data-elements themselves. Using blood-pressure as an example, the two numeric values that make up the answer to the question ‘Blood-Pressure’ would further be qualified with the units of measure, method of administration and any important qualifications such as ‘at-rest’, ‘prone position’ or ‘while exercising’.

Other important examples of Context and Qualification are negation, certainty and subject of reference. Negation, more appropriately referred to as Presence / Absence is of obvious importance because the worst possible processing error would be to wrongly interpret Diabetes Mellitus in remission as Diabetes Mellitus. Nearly as detrimental would be confuse family history of Diabetes Mellitus to direct presence of a condition in the patent themselves.

Taken together there are a number of carefully considered fundamental pattern that have been designed into the architecture and languages used to express clinical models that are intended to address these cross-cutting concerns that all modeling projects must consider. But in most cases, the clinicians and other subject matter experts can simply rely on the modeling nerds to be responsible for appropriately applying established patterns to provide this aspect of the modeling specification.

**What is Negation?**

When a Physician or Nurse takes a history, most of the answers to the most common questions will be negative. Most people aren’t diabetic or Hepatitis C positive. Negation is the art of expressing a computable statement of documenting the belief of a ‘qualified’ Absence of a certain condition.
The biggest challenge with documenting the believed Absence of a condition is the rarity of certainty. If a patient states they have never had Hepatitis C, then it’s very likely they have never been diagnosed and treated for it, but that doesn’t mean they don’t have the virus. It also doesn’t preclude that they might have chosen to withhold this information.

If you contrast the question / answer scenario against the situation where a Hepatitis C test is performed which returns a negative, one can see that a much higher degree of certainty can be expressed. No test is 100% perfect, but a lab result is vastly more reliable than a verbal history.

Now consider the challenge of creating a successful CDS rules set to determine when a patient’s signs and symptoms indicate the possibility of ten possible causes, one of which is Hepatitis C. The clinical logic will attempt to prioritize each of the possible conditions using information from patient history, previous lab results and any sexual / social indications that are in the available medical record.

In the context of ruling out Hepatitis C the indication by the patient that they do not have the condition should be computed as scoring a very low weighting factor. Similarly, a negative Hepatitis C test four years ago carries less weight than a test three months ago, but a very recent test that was given after the presentation of symptoms would be scored at near certainty.

The challenge and ‘art’ of defining computable negation qualifiers is in the complexity of the problem coupled with the importance of success. Negation is primarily a responsibility borne by the modeling experts on the team. By applying a number of established ‘patterns’, that are re-used throughout the clinical model hierarchy, as a matter of course, most of the common negation scenarios can be accomplished. Clinical SME’s need only concern themselves directly when there is a need to identify additional qualifiers that are unique to a given medical scenario.

The human brain processes these types of realities as a matter of course, but computer systems are far more literal, and computable medical records must provide the necessary specific context and qualifying values to provide the system the necessary guidance to most appropriately evaluate complex, and sometimes conflicting, clinical information.

What are CIMI Clinical Models?

A clinical model is a machine understandable definition of the structure, and the details of each clinical data-point within that structure. The modeling layer is the top-most layer that makes up a fully complete clinical schema.

A CIMI clinical model is one that is defined using the semantic specifications that are being produced by the Clinical Information Modeling Initiative (CIMI) within HL7. CIMI is not the first modeling syntax in the healthcare space, others such as OpenEHR and CEML have come before it. CIMI’s goal is to provide a solution that is informed by previous work, but which is also aligned with all the various concerns that are shared with other emerging efforts, such as CQL and FHIR.
Clinical models have many common synonyms; Archetypes, Classes and Structures, to name just a few. The term model also generally encompasses all the various ‘layers’ of information on which it rests, including the clinical concepts and Value-sets which are associated with to individual fields of information within the model.

Much of the complexity of modeling is the desire to define as much of the information in a way that it can be shared with other models. It would be wasteful and ultimately inconsistent for a Skin-Wound model and a Skin-Incision model to each define their own set of Skin-Location values. The Skin-Location values would be defined once, and then shared by both the Wound and Incision models.

In the case of CIMI, models are expressed through the combination of two file formats BMM/ADL but will typically be authored using tooling that distances the author from the gory details of the underlying technical language.

**Clinician Summary:**
*The details of clinical modeling syntax should never be the concern of a Clinical SME. It will be the responsibility of modeling experts to express accurate and fully valid models that capture the clinical truth that is provided by the Clinical SME’s.*

**FHIR Resources, Profiles and Extensions**
In the case of FHIR Profiles, there are a number of ‘imperfect’ strategies for authoring, but tooling is being created to produce them automatically from CIMI Models.

One particular technical reality for FHIR Profiles, that differs from CIMI Models is the FHIR mandates that any server publish the list of FHIR Profiles it supports and will return the details of those definitions on-demand. So, there is no method for keeping the content of FHIR Profiles from being distributed between deployed systems other than to never release them to anyone.

FHIR Profiles do however define a machine and human readable method for declaring licensing policies within the header. So, which distribution of profiles cannot be kept ‘private’ legal ownership can be preserved and licensing policies enforced through traditional means.

**Clinical Quality Language (CQL) for Logic, Decision Support and Quality Metrics**
This layer sits on-top of the Schema and allows for meaning to be derived from information using a scripting language - CQL. There is where IF / THEN and formulaic calculations take place to evaluate the data found in the available fields to infer truth across multiple different data-elements. The three layers that make up the Schema don’t provide any mechanisms to evaluate data, they simply define what data must be captured and exchanged.

Another aspect where the Logic-Layer differs from the Schema layers is in the ability for logic to be more narrowly useful to a single step in the care process. The Schema needs to be shared
by ALL participants in the larger workflow to provide much value, whereas Decision-Support and Metrics are almost always specific to a specific clinical event or stage. IE, the decision support logic for the diagnostic processes and treatment processes are leveraged at very different phases of care, and by different clinical communities.

What is a KNART?

CIMI Clinical-Models, FHIR Profiles, and CQL rules and metrics, will likely be published and distributed as embedded content within a KNART package file. The KNART specification allows for a rich and robust wrapper of metadata to be provided that expresses in machine and human readable form a great many important information elements related to the use and dependencies of the package, and these include licensing and ownership policies.