# Healthcare Situational Data Semantics

# Overview

Clinical Practice Guidelines (CPGs) provide best practice definitions for a particular healthcare situation. They are developed and created by leading experts in the industry based on years of knowledge, experience, and evidence. They are invariably published in the community as documents, documents that are effective within the community of professionals with a strong shared context and deep understanding of the terms and concepts implicit in a CPG.

Bringing a CPG to life with automation requires that the context and meaning implicit in a CPG be made explicit and bound to the context and meaning implicit in the health informatics systems that enable it. The CPG is part of a federation of information, processes, guidelines, and systems that ultimately serve a particular clinician at a particular time for a particular patient.

Making context and meaning explicit is the focus of this paper. We do this by binding the situation-specific terms and context of a CPG to well-defined “building block” concepts: the meaning that underlies the terms and “situational data elements” used in any guideline, model, or data schema. These building blocks are *semantic concepts*.

By making the implicit, explicit, with semantic concepts we:

* Make the authoring of guidelines and systems easier and more rigorous by defining the contextual vocabulary in terms of shared semantic concepts relevant to the healthcare professional.
* Provide automation to help bind the data elements needed by a CPG or application to the underlying health information systems by “pivoting through” the semantic concepts.
* Provide confidence that data elements in a CPG or application mean what the healthcare professional thinks they mean; that each data element has a precise contextualized semantic definition.

For example; What is “current blood pressure” as it may appear in a CPG? What kind of blood pressure? how current? under what conditions? What values would be relevant for a particular patient being evaluated for a particular pathway? Binding “Current blood pressure” to a Hypertension CPG would make all of these assumptions specific. It would also allow HIT (Health Information Technology) to be accurately queried for the relevant values.

***Data has value with meaning and context.***

# 

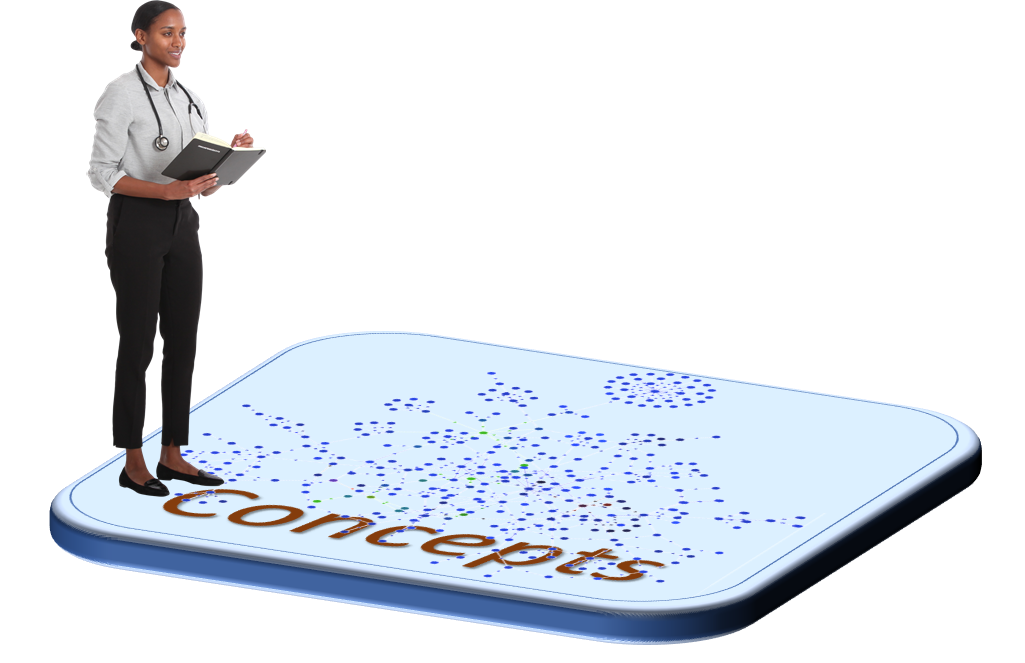
# Context and meaning

## The shared context that binds us

For humans to communicate, to make records, to share experiences, there must be some shared context - a mutual understanding of the world and experiences we share. If there were nothing shared, there could be no communications. Consider communicating with an alien: Ideas about communicating with alien life start with mathematics, surely any advanced civilization would understand Pi and other universal values - mathematics provides a shared context we think is universal. This shared context is comprised of a set of interrelated concepts - numbers, circles, addition - we have a shared understanding of what these concepts mean, they are the semantics of mathematics.

As humans, we have the shared context of our planet, life as we know it, and traditions of culture dating back centuries. Science has offed us shared concepts of methods, measurements, and physical systems. As much as we seem to differ and disagree, there is a huge body of concepts that binds us - that allows us to communicate, record, and reason.

The medical community also has an enormous body of shared context, of common context. Understanding of the body, of disease, of treatment that also has a history and culture. While there is much unknown or disputed or thought of differently, the shared context within the community is an enormous asset. Without this shared context a Clinical Practice Guideline (CPG) couldn’t be written or understood. The CPG uses our shared context and shared understanding of meaning to communicate a complex reasoning process. We all stand on a body of shared concepts.



The clinician is standing on a shared context that defines an interrelated set of shared concepts without which we would not be able to communicate.

Defining these shared concepts is called “semantics” - semantics means meaning.

## The “situational” context that blinds us

As much as we share; communication can be difficult and error-prone. Our natural languages are imprecise and depend on understanding a context that may not be fully stated. We attach words to concepts loosely and depend on that context to resolve any ambiguities. We use “community slang” that carries that context and meaning but could be confusing to an outsider, or to automation.

Attempts have been made at universal languages, of having just “one way” to say anything. Not only is it nearly impossible to get agreement on such universal ways of communicating - it would be impractical. “The patient’s blood pressure” assumes a lot of situational context - what patient, what kind of blood pressure, taken when? It just would not work to qualify everything with every assumed context - saying anything would take forever.

The way we deal with situational context is to understand terms and data elements in context, how they may “expand” into a more general shared context need not be stated when communicating a CPG to another professional with the right training. Even with this shared context, the meaning of terms has to be very carefully specified.

The problem we run into is that there are hundreds of thousands of situational context, each CPG defines its own vocabulary. Each HIT system defines its own schema, each clinical visit presents its own unique situation. Even for professionals, it can be difficult, information shared across disciplines or communities makes it worse. Once we want to use information from one situational context in another we have the difficulty and risk of translation. So while situational context makes it easier for those “in the tribe” to communicate - it can be blinding for anyone else, or any other system. Each of these situational contexts defines its own way of thinking and communicating, its own “viewpoint” of the world.

## The treachery of images (and data)

”[This is not a pipe](https://en.wikipedia.org/wiki/The_Treachery_of_Images)”, a 1929 painting by Belgian [surrealist](https://en.wikipedia.org/wiki/Surrealist) painter [René Magritte](https://en.wikipedia.org/wiki/Ren%C3%A9_Magritte), illustrates the confusion of pictures and of data - that picture (or data) about a thing, like a person or a pipe, is not the person or the pipe. When we “model data” we use terms and concepts that mirror our understanding of the world - but an SQL database about you is not the same as you. As obvious as this sounds, confusion abounds in the way we talk about and manage information. E.g. “When a temperature was taken” is not the same as “when a temperature was recorded”.

Each “kind of data” (A.K.A. “schema”) defines its own vocabulary for the situations it is intended for. It is defined in terms of a set of “situational data elements”, variables for the data schema. What is that situation? Is it for billing or diagnosis? Is it historical or current? “Metadata” about a record gets intertwined with facts about the subject (the patient).

The result is that each such schema defines its own “situational context” and vocabulary. More often than not the connection between that “data vocabulary” and real-world concepts is loose or not defined at all. The context assumed in the data is unstated. These loose connections and unstated context make sharing, integrating, and federating data complex and risky.

The combination of different communities with different perspectives, different vocabularies, using systems with different schema results in the “data problem” we know today - stovepipes and difficulty sharing, communicating, recording, and analyzing across communities and systems.

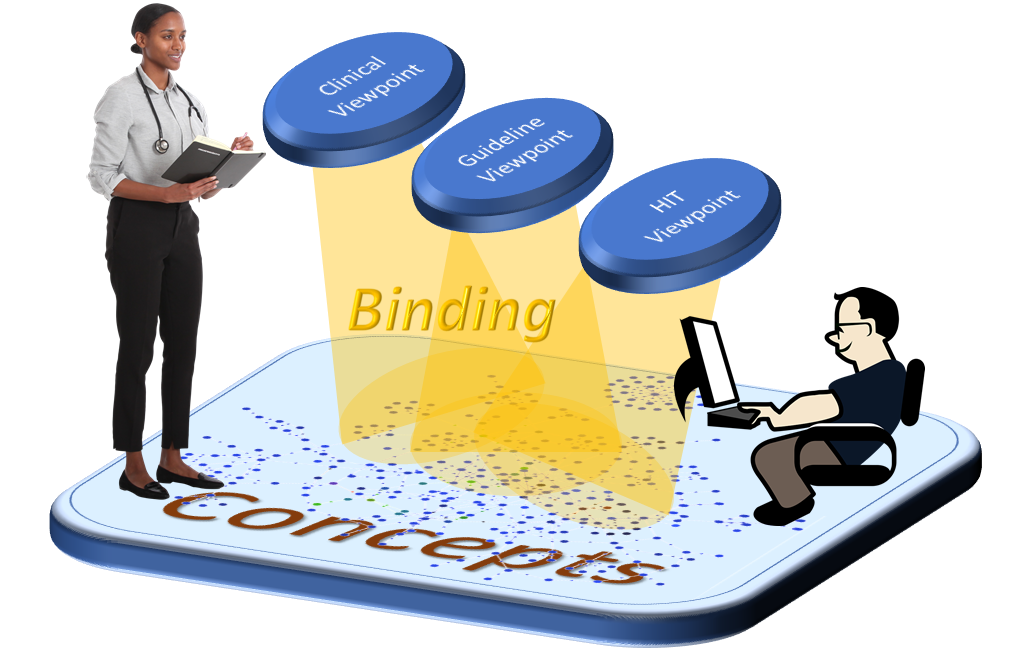
## The challenge of situational context.



The clinician sees each visit, guidelines, and HIT based on its own context and has to “connect the dots” between them. Automation doesn’t share the clinician’s experience or context and has trouble connecting the dots, so automation can’t help the clinician without a developer connecting each piece together individually, this is not scalable or safe.

## 

## Binding the viewpoints to concepts



Different communities, stakeholders, and applications may have different terms for the same concepts. To separate the concerns of the viewpoints from the way the semantics are expressed generically, data elements in the viewpoints are **bound** to one or more semantic concepts. This binding provides consistent meaning for the data elements within the selected situation or context. The binding also provides for mapping data between viewpoints, pivoting through the concepts.

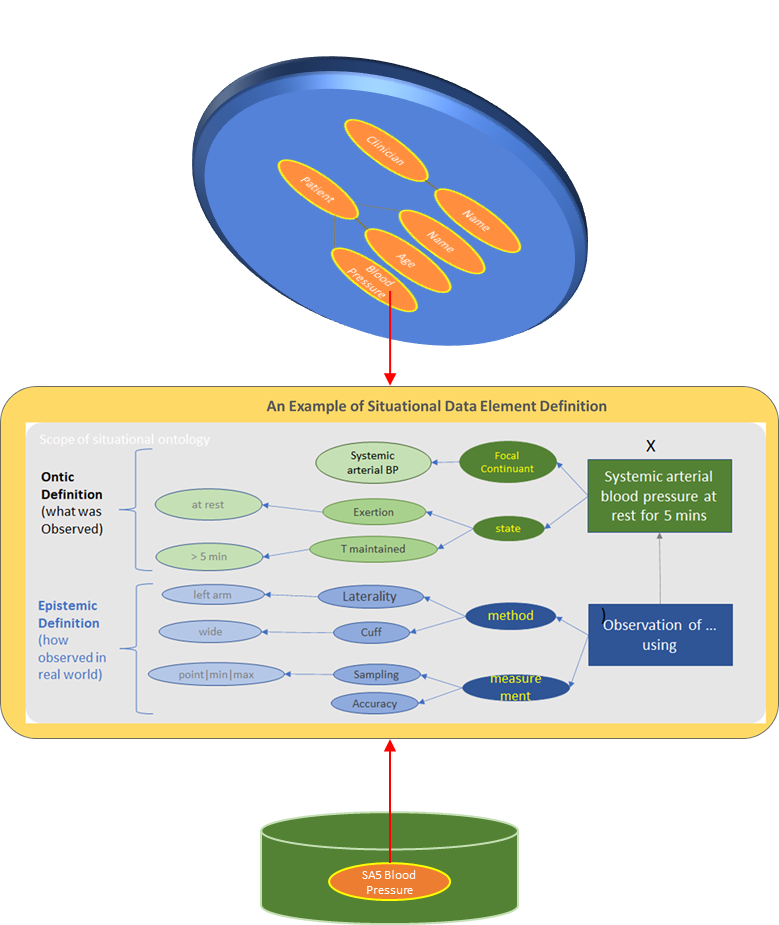
Binding does not require the same terms or “data representation” across the various viewpoints.

Binding is expressed as a set of mapping relationships defining how data represents concepts in context, binding may or may not be simple. The specifics of those relationships are detailed below.

## 

## Data Elements supporting viewpoints

Under each viewpoint there is a set of supporting “situational data elements” that may be described by a data schema. Each of these data elements can be bound to the set of concepts that describes it precisely. This provides a simplified contextual view of information while retaining semantic rigor. This connection is composed of the ***binding relationships*** which we will detail below.



The disk on top is the “supporting data” for one of the viewpoints defined by bindings to concepts. Data sources are also bound to the same concepts so that we can “pivot through” the concepts to connect viewpoints.

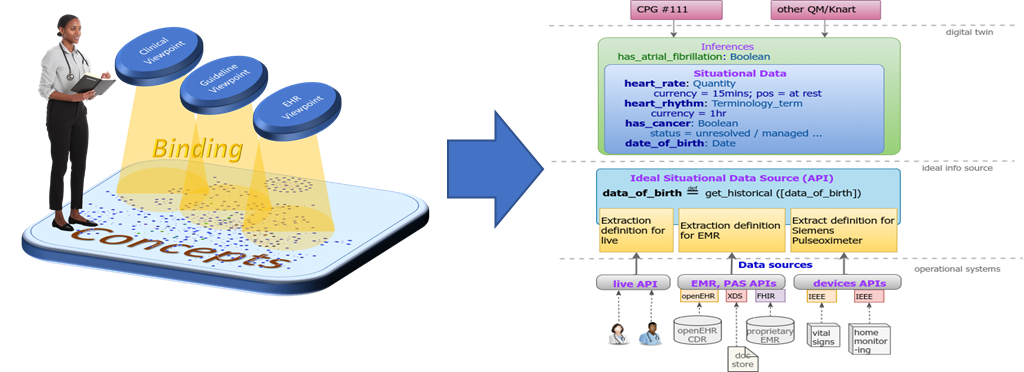
Consider that a backend system, an HIT, has this information and is also bound to the semantic concepts. All the pieces are in place to form the queries into that data source for any CPG (or another viewpoint) that also bound. The conceptual model provides an implementable “pivot point” between data, systems, and multiple viewpoints.

The connection between the viewpoint and the data can also be rendered in natural language, friendly to both clinicians and technologists. Clinicians can validate the binding and technologists can be confident it is implemented accurately. There can be any number of implementation strategies and technologies for “projecting” information from multiple data sources to applications supporting clinician-relevant viewpoints.

The situational data elements support decisions and processes that may also have some automation as part of a viewpoint.

## Automating the connections

With the above layers in pace the manual and subjective process of connecting information from multiple sources for different uses can be replaced with objective and testable automation. The data structures, APIs, queries, and transformation derived from the binding between data and concepts.



What we see in deployed systems is always a viewpoint-specific data layer. Concepts and bindings join across viewpoints, data structures, and technologies.

{{More here linking to Thomas}}

# Requirements for context and meaning

The pattern we see above; that viewpoint specific data is bound to concepts, is the basis for situational data semantics. The foundation of this is: “Data represents concepts”. Every atom of data is saying something about something “real”, or real in our context. The statement “John’s height is 1.8 meters” is a statement about a real characteristic (height) expressed in an understood unit (meters) about a real individual (John) at a specific time period (John wasn’t born 1.8 meters). Any “data representation” of that fact is just another way of saying the same thing – our challenge is understanding the many different ways of saying the same thing and saying them precisely, with all of the unsaid context and assumptions made explicit.

Making this task more challenging is that the same terms are used to mean different things is different context. A generic concept of blood pressure may be fine in one setting but not during an operation.

Another common complexity is time – while we understand that things are always changing many of our information systems and even ways of speaking are about a particular snapshot in time; “now” or “at that time when the patient was pregnant”. Untangling time, history, and what is “current” is essential.

All of the above suggest certain requirements on the “layers” we have defined above. The following enumerates essential requirements.

## Requirements for the data layer

The data layer represents the “schema”, “structures”, “signatures”, or “data elements” that are defined for specific situations, processes, or applications – the data in support of the viewpoints. The data layer is specific to the user’s (or technologies) viewpoint. This layer simplifies, flattens, and tailors general concepts to the needs of the stakeholder or application in a particular situation.

This is the layer most often defined by HIT systems, their interfaces, query languages, databases and even most standards. There are many ways to define such data structures. As there are many ways and a vast legacy of data structures; this is the one we want to put the least requirements on. Essentially, we want “structured data” with the following minimal features:

* A defined “name space”: some way to identify a particular data structure – schema, process, etc.
* Terms unique within a namespace, these terms identify, by some label or ID, some intended kind of data about some specific or category of things.
* Some concept of identity of things, e.g. a “record” about something like a person.
* Some way to identify “containers” of information – or information resources.
* Kinds of data that can “populate” data, based on the terms, about identifiable things.
* Optionally, a way to reference data that is in some other container.
* Optionally: a way to have structure (such as hierarchical containment) within a container.
* At minimum, a textual definition for terms in a namespace
* A way to attach these terms and containers to processes, decisions, and applications.

It should be noted that the above set of requirements can be met by many things, SQL, JSON-Schema, Method signatures, etc. Terms can identify “types”, “Classes”, “Tables”, “Attributes”, “Fields”, “Relationships”, etc. What we will be doing in the other layers is “binding” these terms to concepts.

Because each term in a namespace is unique we have a “hook” to bind it to what it means in terms of semantic concepts.

### Non-goals

We are specific to “structured data”, to be shared in a reliable and automated way unstructured data must be structured. Structuring unstructured data is aided by the capabilities expressed here but is another concern.

We are also limiting our concern to the data that supports people, processes, and applications – we are not defining how processes are managed, decisions are made, or diagnosis recommended. These may also be aided by semantic concepts but is not our focus – we are focusing on the information provided to and derived from these processes and applications.

## Requirements for the semantic concept layer

Everything that can influence the meaning or interpretation of the data layer, including context, has to be grounded in concepts. This puts requirements on how we organize and categorize these concepts. While we build on a long tradition of modeling and logic, there are specific needs to support the “data represents concept” pattern:

* Kinds of “things” in our world – e.g. people, instruments, blood, or a rash
* Identity; how we distinguish one individual from another
* Kinds of social constructs and obligations such as appointments, beliefs
* kinds of situations
* how people, organizations and things play roles in various situations
* observations such as a temperature taken or rash noticed as independent of the thing being observed
* inferred information such as a diagnosis
* rules and practices such as a CPG or insurance requirement
* qualifiers for other concepts such as “current”, “projected”, or “normal”
* temporality; how things change over time and have different characteristics, phases or relationships over time
* units such as weight & time
* uncertainty and likelihood
* records and “speech acts” of all of the above as distinct from the thing being recorded or spoken about

The following detail some of the more important aspects of the semantic concept layer:

### Situations

Situations reflect how we view and participate in the world. Situations may be very general like an office visit or quite specific such as a breast cancer Chemotherapy session. One or a related set of situations are part of forming context. Based on an office visit there is a patient, a particular time, a particular set of clinicians, specific actions taken. Situations provide a core “starting point” from which to branch out to the people, places, things, and times that we understand and care about.

Situations can be long-lived, such as a pregnancy. They can be population wide like the Covid-19 Pandemic. Situations can also be very small or atomic, such as the taking of a particular patient’s temperature at a particular time as part of some wider situations, such as an examination.

As all situations exist for a timeframe, situations are foundational for formalizing temporality. Concepts like “during” or “while” are bound to a situation or set of situations.

#### Vital signs and relationships as atomic situations

Semantically, every relationship or characteristic value that exists over a time period can be considered an “atomic situation”. The set of “body temperatures” a person has are each situations over a person’s life. These form a time-series of temperatures for that person over their lifetime. Considering body temperature as a time series starts to address the temporal requirements – we can understand both the temperature over time (history)and at a particular point in time (a “snapshot”).

Like characteristics, relationships are atomic situations – a person’s marriage or healthcare provider are relationships that each have their own timeframe and are situations of people or organizations being connected in specific ways.

### Individuals and roles

Consider a particular person; that person is of a certain age, has a weight at a particular tine – has a specific genome. These are characteristics inherent in that person, that individual. Differentiate these inherent characteristics of the person from the roles they may play: that person acting as a patient, a physician, or a parent in various situations. Understanding of the person as well as the many roles they may play at various times in various situations enables understanding the “whole person” while not bundling all possible roles of that person together. We want to understand a person as well as the roles they play. Roles are where individuals and situations connect.

The same individual/role separation of concern exists for things; the X-ray machine as a medical device and as an asset on the books of a hospital – very different roles (and information) about the “same thing”. Another example is a unique treatment as an medical intervention for a person and a part of a study of a medication.

### Qualifiers

As discussed below under “binding”, most concepts we deal with are “qualified”, usually in multiple ways. Temporal qualifiers include “Current”, “Latest”, Historical”, “Projected”. Qualifiers can be somewhat subjective like “best”. Qualifiers typically come in some kind of package, the type of the qualifier. We need to understand types of qualifiers and their values.

Some qualifiers are “parameterized”, such as the qualifier for “current” that may have a time limit. We need to be able to define parameterized qualifiers.

### Ontological and Epistemic concepts

Because we have been keeping written and computerized records for so long, we can get confused between some fact about a patient and the record of that fact – between the “ontological” (what exists in the world) and the “Epistemic” (how we record or communicate some fact). The ontological reflects the “reality” of the patient, it is what we are trying to understand and consider. But the source of information; where it came from, when it was recorded and who told us is also important – sometimes called “metadata”. We need both ontological and epistemic concepts as well as the important relationships between them. The same “facts” may be recorded in many different ways and different places, but it is the same fact.

### Logics and ontologies

Much of the above has been addressed in various models, including ontologies, for decades (some for centuries). However the demands placed on semantic concepts for our purpose differs somewhat from general “first order logic” (FOL) inference. FOL has as a priority deriving new facts from existing knowledge whereas our focus is understanding the same facts in various context and representations. Considering multiple contextual dimensions (including time) is inherently not first order, it is a “higher order” logic. Both the concept and binding layers must have a formal logic, but that logic is not “first order”. That said we build on “FOL” and assume all of the “FOL” capabilities of expression.

Many ontologies focus on a “current state” of things rather then the multiple characteristics, relationships, and phases something can go through over its lifetime. For healthcare in particular, time and change are essential. Few things are 100% certain, yet many systems and ontologies only deal with binary true/false.

For these reasons we do not commit to the “first order” limitations – this means that facts and rules and have facts and rules about them and that anything can change over time and may be interpreted differently by various stakeholders. Of course, a particular viewpoint or record will probably be more specific, perhaps a “snapshot” in time or a specific history – understanding that temporal context is crucial to understanding what it represents.

## Requirements for the binding layer

The binding layer is responsible for “grounding” every “data element” in a viewpoint or schema to the set of concepts that define it. A simplistic assumption is that you can just “point” to a concept from a data element; “this is what it means”. The problem is, there are billions of permutations of meaning and a simple pointer is not practical.

### The contextual view of something

During a visit a patient may engage at the front desk, with their physician, with an x-ray technician, with the payment desk. Each of these engagements demands a different set of information about the same person related to very different processes. The binding layer must be able to present this “contextualized view” of someone or something. Defining these contextualized views is the “hook” between the viewpoint and a set of concepts. A viewpoint may utilize many different contextualized views of many kinds of things. One view may also “branch out” to others. This branching out tends to look like a hierarchy of information tuned to the target viewpoint.

### Binding to situations

Most viewpoints or schema are intended for a certain class of situations. The situation of an office visit “locks down” a timeframe – a visit for gout narrows it down even more. As part of that situation there is the role of a particular person as the patient and the roles of various clinicians. Focusing on an office visit enables us to talk about “the patient” because we have the context of that situation. It allows us to talk about the current weight that was taken in that office visit. One or a small number of situations can be the focus of a viewpoint but can branch-out to many related situations.

Within a situation the roles that people or things play in that situation each have a contextualized view.

### Qualifiers

Consider the English syntax of a “qualifier”. Most data elements are qualified within a specific qualified data frame. Consider the example of “blood pressure”, there are good definitions of this concept:

**Blood pressure**: The force of circulating blood on the walls of the arteries. Blood pressure is taken using two measurements: systolic (measured when the heart beats, when blood pressure is at its highest) and diastolic (measured between heart beats, when blood pressure is at its lowest).

The above could be considered a definition for the “base concept” of blood pressure but this base concept needs to be qualified to fully define a data element. Examples of qualification include:

* There are different ways to take a blood pressure
* There are different positions and conditions of the patient
* There are real blood pressure readings; “current” and historical.
* There are average, target, and recommended blood pressures

Diagram

Description automatically generated

Some qualifiers may “select” from a fixed list, others may require parameters, such as the timeframe for “T maintained” or what is considered “current”.

Based on the above we can assert a requirement for the binding layer:

The binding of a data element must be able to qualify a base concept across any number of qualifying dimensions. Qualifiers must be able to be parameterized. Each base concept and qualifier must be formally defined.

### Restrictions

Within a contextualized view we are frequently assuming specific characteristics or relationships. E.g. a breast cancer chemotherapy CPG assumes a patient with breast cancer. It may assume specific usage profiles for specific medications.

As we are defining a contextualized view, we want to state these expectations and requirements as restrictions on the people and things we find in applicable situations. Restrictions can be kinds of things, specific relationships, specific values – anything that constrains what can be in any specific “slot” in out view. Restrictions may be simple or complex expressions focused on a particular concept in a particular situational context.

### Local terminology

The fully qualified, restricted definition of a data element would tend to be rather long and cumbersome. In addition, particular communities tend to have specific terminology. A contextualized view can assign a local, contextual, term to a concept. To make sure this local term is understood the full definition can be “unwound” using structured English, from the binding logic, in medical terms so that the clinician can be confident of what it means.

### Flattening

To be applicable across many different kinds of context and situations the concept layer is necessarily normalized – each concept individual, unique, and specific. If directly implemented this would create a complex “graph” of information that, while very flexible, would be cumbersome to use or implement. As part of a contextualized view we can “flatten” a chain of relationships, restrictions, and characteristics to simpler and more context specific data elements with local terminology.

—------------------stopped here—------- Cory Edits

## Definitions

A model of situational data is termed a *situational data model* (SDM) with individual elements termed *situational data elements* (SDEs). The meta-model of situational data models is termed the *Situational Data Meta-model* (SDMM).

In any CPG, there are typically multiple situational variables that could be formalized as SDEs. Formal representation is proposed to be in the form of an abstract data-type or structure, whose semantics are defined according to the SDMM.

## Business Benefits of Situational Data

The approach to Defining Situational Date Elements offers the following benefits:

1. Models of situational data (along with the other knowledge domain artifacts) can be distributed/shared across members of the healthcare community who want to investigate and possibly implement the CPG.

2. At a logical level, capturing data according to a published/shared situational data model is actionable and provides a common basis for understanding. CPG authors, informaticists, solution architects, system architects and enterprise resource managers can understand what information is required, and why it is required. They can then make appropriate plans and decisions based on such models and their local operating environment.

3. When the decision to implement a particular CPG has been made, this common model of understanding becomes a communication tool for the implementation team of developers, testers, trainers, operations, and the other actors necessary to deploy a successful application. As the SDM is actionable, automation techniques can be applied to create computable, executable implementations

4. As new evidence is captured for a CPG, having a structured approach for describing and capturing situational data provides a foundation for evidence-based analysis as well as a platform for simulating new approaches.

## What is an SDE?

An SDE may be defined for any *logical input data element* defined in a CPG, CDS rule-set or other application in order to perform the inferences of the CPG, including rule logic, determining the path through its decision points. AnG ‘input data element’ (or ‘input variable’) is a data item relating to the real-world operating context - i.e. the ‘situation’. Such variables may come directly from the situation in real-time (such as vital signs), or maybe historical knowledge about the subject pertinent to interpretation of the current situation.

Typical information items represented by an SDE may be:

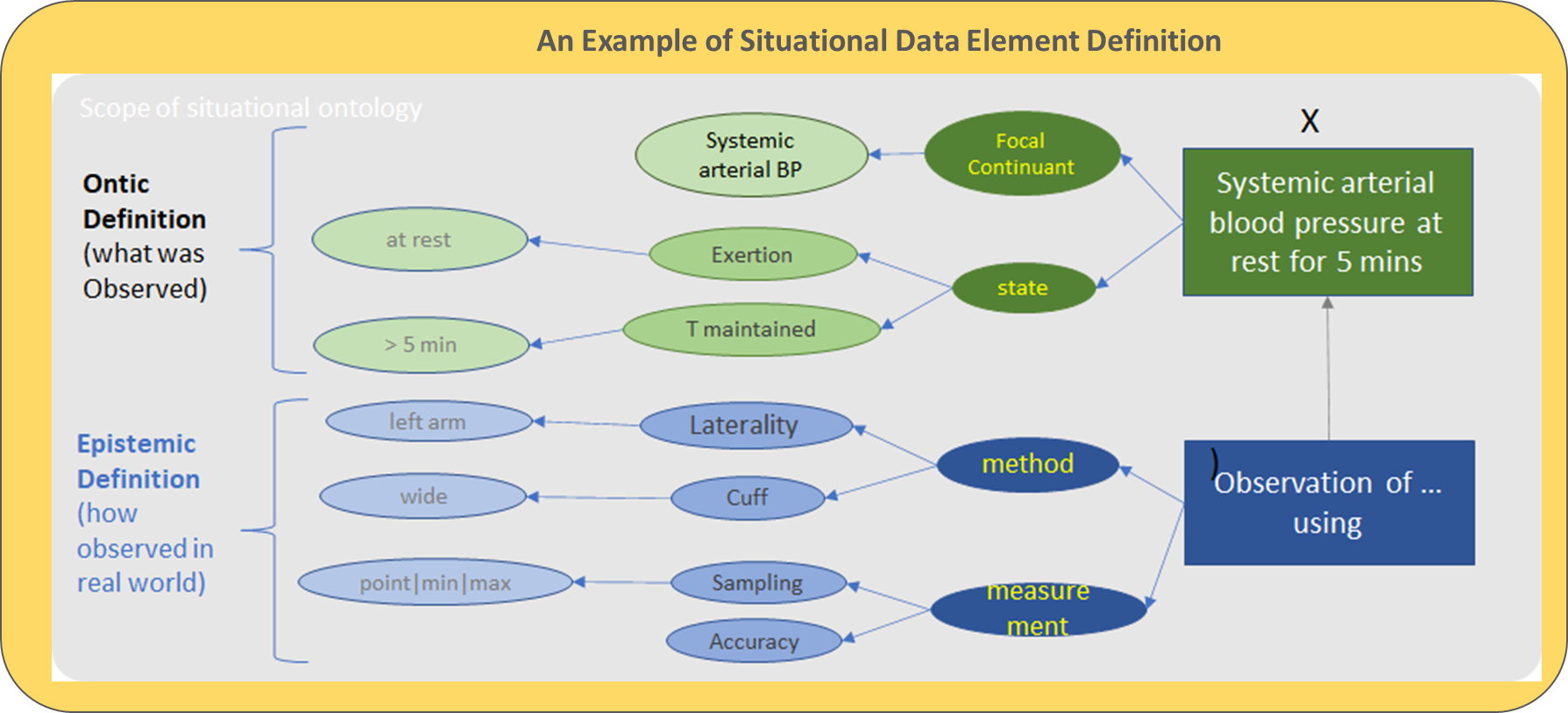
* the *state* of some continuant of the subject (e.g. blood pressure; lab result) at a point or interval of time;
* an *event* that occurred at some moment, i.e. occurrent (e.g. loss of consciousness, past procedure);
* a statement *about* the subject, including classification (e.g. previous diagnosis);
* an *aggregate fact* about the subject (e.g. number of previous pregnancies);
* an *order* for some intervention / medication etc for the subject;
* a subject *goal*;
* a stated *request* or *preference* of the subject;
* potentially other information about the subject, e.g. to do with health plan cover; see https://wolandscat.net/health-informatics/epistemic-status-classification/

For any of the above, there is potentially a complex structure of data elements making up the SDE information item. In the case of vital signs for an intubated patient, it may be a data structure consisting of samples of various vital sign data points (systolic BP, diastolic BP, heart rate, SpO2 etc) over time. A previous diagnosis may similarly be a complex statement including various dates and other items.

SDE information structures may include data points relating to how the central data were obtained or created. For observations, this is the method, e.g. using a certain kind of instrument, with a certain type of sampling, e.g. average over 5 minutes, instantaneous value etc.

For diagnoses, it might include a method or protocol for arriving at the diagnosis, such as use of a certain test result, or concurrence of particular symptoms considered to be diagnostic. For the other categories, it generally relates to how the information was obtained, e.g. via a patient questionnaire and may or may not be of interest for the SDE consumer.

The following diagram is a graphical representation of a Situation Data Element Definition for blood pressure observation with its sub-parts.



## Model of an SDE

A model of an SDE has three distinct components:

* a precise **formal model** of the data item as determined by the CPG authors;
* **meta-data** that document and assist in human communications about the SDE;
* epistemic **definitions needed by implementers**.

### SDE Formal Model

An SDE may be a single date item, typically with pre-coordinated semantics such as ‘current systolic blood pressure (sitting, at rest)’, or it may be a larger structure as alluded to above, e.g. a set of patient goals or preferences. In the latter case, we consider the SDE to be a hierarchical structure whose leaf elements may be referenced with a typical ‘dot’ notation or similar.

The general case is therefore that an SDE is formally a hierarchical data structure, with a single item being a common case.

MORE HERE

Value of formal model

- accurate and precise definition for the data used by the authors in developing the CPG.

- ability for the current implementers to customize and localize the CPG. For example, a specific device may have been defined is expensive and not available to the local setting. However, they may have another device that could be substituted. The local implementers may also have other changes they need to make to the CPG to implement the defined SDE in their environment. As the SDE is logical, they can make these decisions at the logical level and still leverage the knowledge in and value of the CPG.

### SDE Meta-data

The second definitional component of an SDE is its descriptive meta-data. Having defined the semantics of a SDE as important is that the team working together on the SDE have common names and descriptions for what they are referring to. There are three primary meta-data items: short name, long name, and description. In the above diagram, the “current systemic arterial blood pressure at rest” could be the long name, “current blood pressure” could be the short name, and the description could be “....”

Communications are simplified by the short name while still ensuring the terms used for a SDE amongst the team is accurate. The long name and description provide more detail if required. These terms are linked to the SDE Formal Model if and when the need for the definition is required.

### Epistemic Definition for Implementation

Expression of SDE in terms of lower-level primitives relating to the ideal information medium.

MORE HERE

## Examples and use cases

1. Current Diastolic Blood Pressure
2. Patient has cancer
3. The VA Covid 19 at Home (Ken)

### Patient Preferences for determination of Qualified Service Provider

One of the more difficult issues is finding qualified and timely behavioral or social service. For example, the patient may need counseling services for sexual assaul. The patient has preferences for a specific race, gender, and/or ethnicity of the counselor. This preference is for this situation only and is not applicable for other healthcare or social services.

The preferences could be:

· Race preference is Native American

· Gender preference is Female

· Ethnicity preference is Native American

Another example could be:

· Race Preference Black

· Gender preference is Male

· Occupational history preference includes military service

# Conceptual Architecture -

The semantic definition of SDEs above describes an SDE within the modelling environment of CPGs, CDS and other SDE-consuming artefacts. At execution time, however, SDEs have to be populated with values for the real world entities they represent.

We can consider the instantiated form of SDEs to constitute a digital data parallel of a (fully real world scenario) of:

**Patient state + Clinical History → Physician’s conceptualization of pertinent clinical aspects (as expressed in CPGs etc)**

In order to effect the arrow in the above formulation, i.e. to instantiate the ‘digital twin’, real world sources of data must be relied upon, including:

* EMR / patient record systems;
* departmental and other HIT systems;
* realtime devices;
* workers ‘in the room’;
* the patient.

Since these sources will be different in their details in every deployment environment (hospital, clinic etc), defining SDEs for a standard CPG or other shareable knowledge artefact (e.g. CDS rule-set) directly in terms of specific back-end systems and persons is not achievable other than on a per-site basis, and therefore not generalisable.

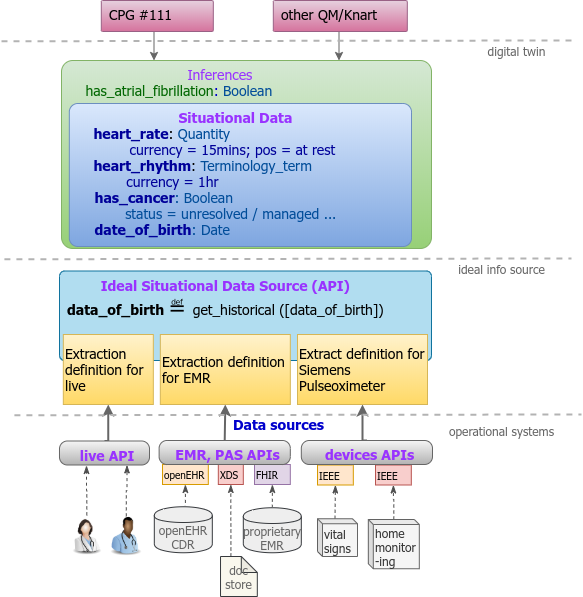
One attempt to solve this well-known problem in the past was known as the ‘vMR’, or virtual Medical Record, which is an idealised EMR against which CPG data items could be defined. There are various issues with this approach:

* The EMR is only one place SDE data might be obtained; often it is obtained directly from devices or people;
* The EMR conceptually imposes a historical model of observations, decisions, orders and actions relating to a patient, which over-specifies situational data items, which in the end simply represent states, events and historical facts about the subject.

As an alternative approach, we propose a 3 layer conceptual architecture as follows:

* SDE definition level - as used by CPGs, CDS, etc;
* Ideal patient information medium - a simple but formal model of everything that can be known about a subject according to a scientific realist ontology (which we propose); this medium may be represented by an API;
* Binding of the ideal medium to any particular deployment environment; this may be achieve in terms of API Implementation and scripts, calls to lower level APIs and so on.

The architecture is visualised below.



### The Ideal Information Source

An ideal information source for the subject is one from which any current state or event data and any historical information about the subject may be obtained. It is indexed by time, and conceptually covers real world data sources including storage systems e.g. ideal HIT + real-time devices + real-time 'in the room' observations.

The ideal information source is assumed to be structured according to a Situational Data Ontology, that captures the relationship of ontic entities and their epistemic counterparts (recordings).

Each SDE is expressed at this level in an abstract epistemic formulation that relates it to this ontology. Examples:

* has\_cancer: Boolean -> ∃ d: Diagnosis that has-index-condition c: Condition | c << V AND d.status /= |resolved| ... or similar
  + where V represents a value set of all cancers; might just be Snomed::1234|malignant neoplasm|

### Concrete Data Sources

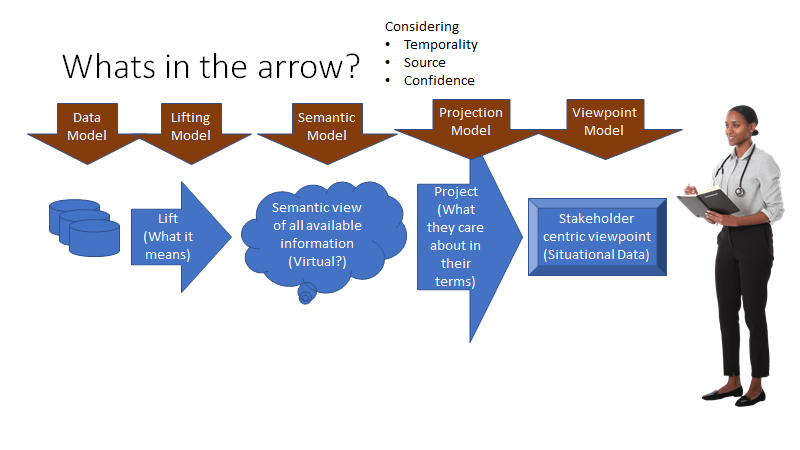
Ideal information source formulations can only be populated via concrete implementations, i.e. scripts expressing calls to APIs and / or logic that implements the sense of the epistemic specification, for a given environment / system / API standard. This is where HL7v2, CDA, FHIR, openHIT, X12, IEEE, UI form display etc come in.

<TB: DS stmt could be expanded?> DS: human readable guidelines are always based off an O perspective; CPG implementations aspire to look like that, but they really operate on the E level

# Scaling SDEs: Libraries, Reuse

Large scale use of SDE definitions and Models with repositories of CPGs, CDS rule-sets and so on immediately poses the question of how multiple Situational Data Models (SDMs), many containing common items (e.g. vital signs SDEs) are rationalised such that common SDEs are re-used.

=========== TB: revised to here 21 Jan 2022 ==============

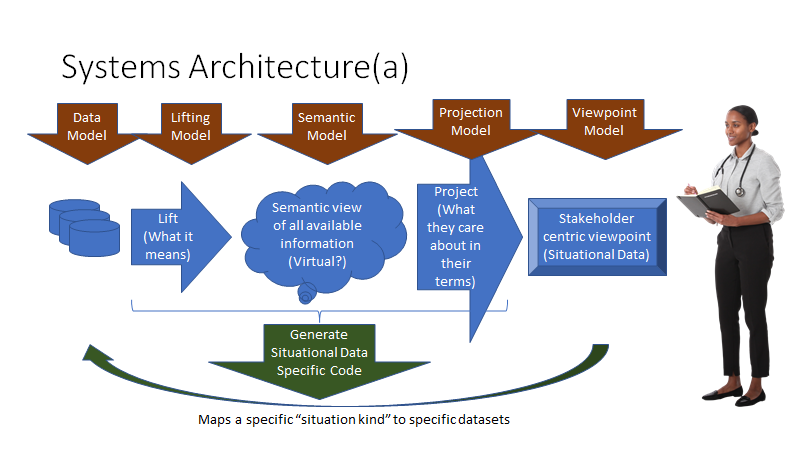
This picture is just for us to think about...

Software / System Architecture  
 Systems, interfaces, roles and functions   
 (function played by FHIR API, MDMI, etc)

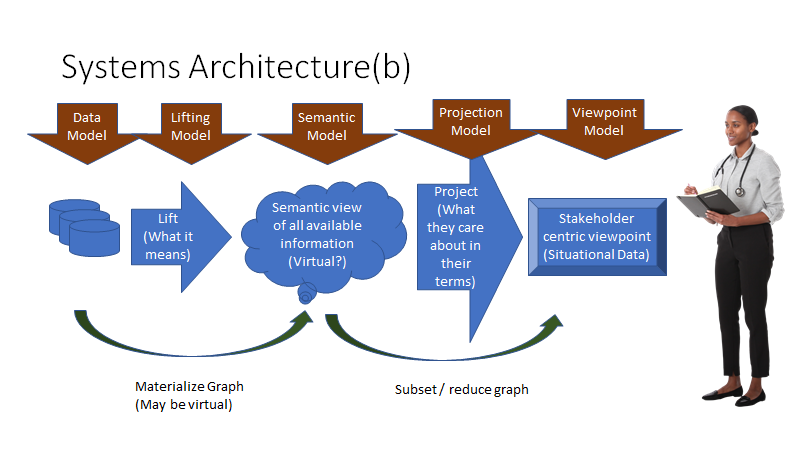
Data / Semantics Architecture

moving “clinical concepts” from source systems into a CPG space. “Systems” can be both systems of record (e.g. HIT systems) and well as applications to capture data (e.g., answer the question of “Is the parent able to stand?”)

Seman



Stakeholder centric viewpoint should be presented such that it is understood by all stakeholders. All stakeholders should include authors, semantic knowledge engineers, implementers, quality assurance / testers, and decision makers who must make decisions based on limited resources. The presentation to these stakeholders needs to provide the viewpoint in such a way so that it is consumable by all stakeholders in terms they can understand . See also CPG IG Introduction Section (<https://build.fhir.org/ig/HL7/cqf-recommendations/>) for list of actors in a CPG.



# 

# Glossary (ken to draft)

Various kinds of ‘models’.

Not sure where to put this:

|  |  |
| --- | --- |
| Proposed Properties of Situational Data Element | Definition of Property |
| Properties needed by humans and computable properties | |
| Description | Human Readable text describing the meaning unambiguously |
| Identifier | Unique, computable Identifier |
| Name | Identify used by humans for an SDE |
| DataType | Need Definition. E.g., coded, Quantity (incl units), Real, TimeSeries<Real> …  QUESTION FOR CORY: I believe timeseries can be a type of datatype; e.g. start, periodicity, end, and values in “cells” needs to reference another SDE. |
| Properties are needed to semantically define the SDE completely | |
| ReferenceOntologicalProperty | Reference to a computable expression that describes the semantic meaning of the SDE |
| ReferenceOntologyModel | SNOMED, HL7, OpenHIT, an upper ontology such as MDMI HealthcareConceptModel |
| SDEType | Whether the SDE is of type information or assertional.  - SDE of type information represents an observable or measured fact.  - SDE of type assertional represents an assertion based on other information or assertional SDEs. For example, answers to questions such as Does Patient have Diabetes or What is the Current Blood Pressure are SDEs of type assertion. |
| FocalBinding | A computable expression to further specify the validity of the range of values for the SDE, e.g., numerical range, valueset binding. I believe this lines up with TB’s B.  Best Practice: Supports scalar and Parent / Child structures |
| ValidityBindings (Preconditions) | Other SDEs where the instance of the SDE must be true. If there is more than 1 SDE contained in the validity binding, all instances must be true for the parent SDE to be valid.  Best Practice: Use when deterministic expression of all conditions must be true or if a simple arithmetic expression. If any other more complex rules are required use such as conditional rules / scoring logic / or other more complex inferencing algorithms are needed, do not use. |
| TemporalBinding | Temporal context that describes the validity of the SDE value for the situation.  Best Practice: Used primarily to insure currency of SDE. Primary example is ensuring the semantics of current. |
| EpistemicBinding | Expression of the validity for an SDE derived from a SDE representing a collection of epistemic data.  Best Practice. For example, an expression could be latest, earliest, most common, mean, max, min, etc.. Also needs to encompass TB’s Di and Ds descriptions. As an example, capability should support populating CQL expressions. |

## Fallacies, Pitfalls, and false Expectations

Localization Service - A Service that interfaces the platform independent logic of the CPG with the platform specific implementation of the execution environment.

Digital Twin - The real-time digital representation of a physical object. The Situational Variable is a digital twin which defers from the epistemic, or recorded, Variable.

Epistemic Variable - A recording of an element of data, generally in healthcare it is the historical data in an HIT.

Situational Data Server - A Service holds the Situational Variables and maintains their validity for use by other components that use (consume and populate) the variables including but not limited to Process Logic, Assertional Logic, UI Services, and Localization services.

Clinical Practice Guideline (CPG) - Clinical practice [guidelines](https://build.fhir.org/ig/HL7/cqf-recommendations/documentation-approach-04-guideline-development.html) are systematically developed statements to assist clinical practitioner and patient decisions about appropriate care for specific clinical conditions, procedures, and/or similarly scoped activities. Guidelines consist of recommendations for patient care, which are based on scientific research and data (evidence), vetted through rigorous processes of a review and synthesis by recognized domain and methodological experts and other key stakeholders (e.g. patient and caregiver advocates) to guide healthcare decisions and activities for defined scope. A guideline may consist of one or more recommendations, contextualizing information, the possible means or strategies for bringing together or orchestrating recommendations, and other relevant considerations. <https://build.fhir.org/ig/HL7/cqf-recommendations/glossary.html>

Clinical Pathway Guideline - [Pathways](https://build.fhir.org/ig/HL7/cqf-recommendations/documentation-approach-12-03-cpg-plan.html#pathways) convey the full scope of guideline recommendations in a manner that can be implemented for the purpose of guiding patient care within a singular CPG. In part, pathways serve a similar yet higher order function as strategies, and serve the function of addressing interdependencies such as decision and orchestration logic across multiple concerns. <https://build.fhir.org/ig/HL7/cqf-recommendations/glossary.html>

Ontic Variables -

Epistemic Variables -

Decision Tables / Assertional Logic / Inference (CPG)

Clinical Decision System (CDS) (OMG) – Clinical Decision Systems encompasses a variety of tools to improve outcomes, generally at the point of care.Decision Requirements Diagram (DRD) (OMG) – A DRD is the DMN diagram for displaying Decisions, Decision Services, Data Inputs and other DMN elements.

Extras

Healthcare Conceptual Model (Cory’s work)

# Saving Cory’s work

between the communities, and the definition for how the many kinds of data and the many CPGs can be federated - not by doing everything the same way, but by having a reference for contextualized meaning. That reference is “semantics” (semantics means meaning) in a structured model with well-defined ways to connect the technologies, the CPGs, and their terms to meaning, to the semantic models. As we will see below, this meaning reference provides value throughout the healthcare lifecycle.

As we depend more on the aid of automation to deliver, interpret, and sometimes to suggest diagnosis; understanding, communicating, and applying the context of the situation is crucial. Yet, the way CPGs and the related clinical information is described and contextualized has been ill-suited for automation.The implicit context recognized by experts must be made explicit. Each term and piece of data must be unequivocally tied to what it means. Critical nuance must not be lost.

CPGs and other healthcare information has grown from a tradition of meaning established by text, intended for communication between professionals, with a strong shared context, to other professionals. Each of these CPGs describes its own “situational data”, with its own terms, rules, and processes. How the term or data point in each CPG relates to what it means is implicit, how it may or may not be related to any other CPG or / and information system schema becomes a manual exercise in interpretation. Manual exercises that are time-consuming, expensive, and error-prone.

The issues with the current state are clear; in any one healthcare context, we have a hodgepodge of CPG guidance, structured data sources, and various automation “assistants” trying to assist the clinician in a context that is unstated based on terminology that is inconsistent without any structured reference to what any of it means. Another way to say this is that the information is not “computable”; appropriate for machine interpretation.

One approach to this problem has been a more general schema, vetted in standards bodies to have exactly one way to “say the same thing” for an information system. In the small, this can work, but in the large, it suffers from the diversity of communities, disciplines, and technologies. Having only “one way” to produce or vend a solution has proven inflexible and hard to establish universally.

The approach outlined in this paper separates the solution from the meaning; we define a reference for meaning and context that can be the glue between the CPGs and schema, the common understanding between the communities, and the definition for how the many kinds of data and the many CPGs can be federated - not by doing everything the same way, but by having a reference for contextualized meaning. That reference is “semantics” (semantics means meaning) in a structured model with well-defined ways to connect the technologies, the CPGs, and their terms to meaning, to the semantic models. As we will see below, this meaning reference provides value throughout the healthcare lifecycle.

...

# Archive

## Cory’s original intro

Context is essential in informing any healthcare situation. Context determines how we interpret data and how we rank its relevance. Clinical Practice Guidelines (CPGs) provide best practice definitions for a particular healthcare situation based on the context of the authors which must be communicated to and understood by practitioners. Each CPG must then be related to the particular situation at hand, with a particular patient's unique circumstances.

As we depend more on the aid of automation to deliver, interpret, and sometimes to suggest diagnosis; understanding, communicating, and applying the context of the situation is crucial. Yet, the way CPGs and the related clinical information is described and contextualized is ill-suited to such automation.The implicit context recognized by experts must be made explicit. Each term and piece of data must be unequivocally tied to what it means. Critical nuance must not be lost.

Information systems have grown from a tradition of “structural definition” of data, it could be in SQL, XML, API or anything else - these “schema” describe the structure of data and data elements managed by or communicated by systems. Meaning is established by text, tradition, and teams that “understand” what it means. It is commonly understood that it is quite difficult to “connect the dots” between independently developed schema, even if they contain the same kinds of facts.

CPGs and other healthcare information has grown from a tradition of meaning established by text, intended for communication between professionals, with a strong shared context, to other professionals. Each of these CPGs describes its own “situational data”, with its own terms, rules, and processes. How the term or data point in each CPG relates to what it means is implicit, how it may or may not be related to any other CPG or and information system schema becomes a manual exercise in interpretation. Manual exercises that are time-consuming, expensive, and error-prone.

The issues with the current state are clear; in any one healthcare context, we have a hodgepodge of CPG guidance, structured data sources, and various automation “assistants” trying to assist the clinician in a context that is unstated based on terminology that is inconsistent without any structured reference to what any of it means. Another way to say this is that the information is not “computable”; appropriate for machine interpretation.

One approach to this problem has been a more general schema, vetted in standards bodies to have exactly one way to “say the same thing” for an information system. In the small, this can work, but in the large, it suffers from the diversity of communities, disciplines, and technologies. Having only “one way” to produce or vend a solution has proven inflexible and hard to establish universally.

The approach outlined in this paper separates the solution from the meaning; we define a reference for meaning and context that can be the glue between the CPGs and schema, the common understanding between the communities, and the definition for how the many kinds of data and the many CPGs can be federated - not by doing everything the same way, but by having a reference for contextualized meaning. That reference is “semantics” (semantics means meaning) in a structured model with well-defined ways to connect the technologies, the CPGs, and their terms to meaning, to the semantic models. As we will see below, this meaning reference provides value throughout the healthcare lifecycle.

This White Paper proposes **a standard description of reference semantics for use by industry**.

## Ken’s Edit

Context is essential in informing any healthcare situation. Clinical Practice Guidelines (CPGs) provide best practice definitions for a particular healthcare situation based on knowledge for a specific context of the authors. The author’s knowledge captured in each CPG can be segmented into informational knowledge, process knowledge, and assertional knowledge. This captured knowledge in a CPG can be described using logic constructs, independent of any specific implementation language. For a myriad of reasons, all of the CPG knowledge should share the same semantics for terms. This is especially true of the captured information knowledge as it becomes foundational for describing and defining process knowledge and assertional knowledge.

CPG implementations use the captured knowledge of the CPG. Healthcare solutions traditionally have been built over the decades on a tradition of “structural definition” of data, it could be in SQL, XML, API or anything else - these “schema” describe the structure of data and data elements managed by or communicated by systems. Meaning is by structure and the teams that “understand” what the structure means.

The approach outlined in this paper is not attempting to create anything new. Our goal is to fill a gap between defining contextual information and separates the implementation from the semantics; we define a reference for meaning and context that the fills the gap between the CPGs and implementations using structural definitions. Our approach provides an actionable, common understanding of the semantics of the information defined in the CPG that is actionable into structural definitions, assertions, and processes. For us, actionability includes a large scope of techniques with the emphasis on automation processes into computable implementation artifacts.

**This White Paper provides an approach called Defining Situational Data Elements using Semantics.**The **approach is for filling the gap between the information knowledge captured in a CPG and the structural definition of data in the CPG implementations by providing an actionable common model of understanding using reference semantics. This common model connects the contextual or logic definitions in the CPG with the implementation schemas. The authors believe this approach is a generalized approach for a situation. It can be used across all CPGs. We also believe it is a framework that can be extended to include providing a common model of understanding using reference semantics into process and assertional knowledge.**