Proposal for precise modeling of entities in the U.S. Core Data for Interoperability

October 2020

`Version 1.0`

Edited by:

Daniel J. Vreeman, PT, DPT, MS, FACMI

RTI International, Inc.

Table of Contents

Situation

The Office of the National Coordinator for Health Information Technology (ONC) has published the United States Core Data for Interoperability (USCDI), which is a standardized set of health data classes and constituent data elements for nationwide, interoperable health information exchange. The 21st Century Cures Act Final Rule (from ONC) and the Interoperability and Patient Access Final Rule (from CMS) both require data exchange of USCDI content. Widespread availability of this standardized data has significant potential for advancing research as well, and NIH has encouraged the use of USCDI clinical data in research studies too.

Future versions of the USCDI will be drafted and finalized based on public data element submissions to the ONDEC system. (ONC is accepting submissions for USCDI Draft V2 through October 9, 2020).

Once finalized, new versions of the USCDI will feed into the Standards Version Advancement Process (SVAP). The SVAP allows health IT developers in the ONC Health IT Certification Program to voluntarily update their products to include National Coordinator-approved newer versions of select standards without waiting for future rulemaking.

Background

The USCDI Version 1 (V1) contains 2 kinds of entities:

- `Data Class`: an aggregation of various Data Elements by a common theme or use case
- `Data Element`: the most granular level at which a piece of data is exchanged

Data of Birth is an example `Data Element` that illustrates USCDI's intended level of granularity. The USCDI would not contain `Data Elements` for the smaller components Day, Month, or Year, because **Date of Birth** is the unit of exchange.

Assessment

The USCDI V1 is a landmark specification of core clinical data that will be made available by providers and payers for nationwide exchange. Recognizing USCDI's significant promise for directing future interoperability efforts, we seek to improve its foundation for continued expansion. In particular, we believe that improving the clarity of USCDI entity definitions are necessary for the industry to interpret them consistently. We seek to add precision to their specifications to enable a principled approach for users to understand, implement, extend, and refine them with future submissions for new USCDI content.

Issues in the USCDI V1

The current USCDI entity definitions are prone to misinterpretation because they do not align with the common notion of `class` in computer programming and ontologies. Typically, `class` represents something of a "template", or blueprint, with attributes (e.g. fields) that all instances (e.g. objects) of that class contain. The present usage of `class` in the USCDI simply means `collection`.

Further, the USCDI V1 entities lack a connection to prevailing or intended data `shapes`. Here we use `shape` as a generic term referring to the commonalities in data format for a particular kind of information that is represented across specific data models for information exchange, such as FHIR (or CDA or HL7v2, etc) or analytic Common Data Models (such as OMOP, PCORnet, and Sentinel). Across the prevailing exchange and analytic models, a vital sign like "heart rate" has the same shape (i.e. an `Observation`) as a lab test result like "hematocrit". This shape is quite distinct from the shapes for an immunization, a procedure, or diagnosis.

Without identifying a data `shape`, we cannot appropriately use terminology standards. For example, neither CPT codes, UCUM, nor SNOMED CT *Finding* terms should be used to identify an Observation because their conceptual model is for other kinds of concepts.

Further, the USCDI V1 does not follow a consistent pattern in distinguishing `Data Classes` and `Data Elements`. "Heart Rate" is a `Data Element`, but Laboratory is `Data Class` with one `Data Element` called `Tests` and another called `Values/Results`. In one case the measurement "variable" is a `Data Element` unto itself, and in another case it is an allowed value for a `Data Element` (Tests). This lack of consistency is confusing.

ONC's choice of the term `Data Element` is understandable because it is commonly used. But it is also problematic because `Data Element` has copious interpretations by different people.

The ISO 11179 standard provides one notion of `Data Element` that is precise and applicable across domains. The FHIR standard defines Resources with "attributes" that are called (data) elements. FHIR's use is consistent with ISO 11179's approach. FHIR does not, for example, define "heart rate" or "hematocrit" as a data element. Rather, a more generic `shape` is defined for any kind of test or measure: the `Observation` resource. This specific shape consists of attributes like `Observation.code` for a term from a standard terminology to identify what the observation is, `Observation.value` to carry the result value, `Observation.subject` to identify who it is about, `Observation.performer` to identify who made the observation, etc. These attributes are the elements.

Desired capabilities in future USCDI versions

We believe that the USCDI will be most helpful when it can be precisely understood and has a clear structure for its entities. To fulfill its purpose in defining core core data for exchange, we believe the USCDI should be capable of specifying data content with different types of precision. Consider the implications of these different statements:

- Every system should be able to send clinical measurement results
- Every system should be able to send the code identifier, result value, units of measure, reference range, physiologically relevant time, and observed specimen for each clinical measurement result
- Every system should be able to send results for the subset of clinical measurements known as vital signs
- Every system should be able to send a "heart rate" clinical measurement result

We believe that the USCDI will be most helpful to advancing interoperability when it can clearly differentiate between these kinds of statements. The need for clarity is now, as new submissions in various shapes and sizes are coming into the USCDI review process.

Recommendations

In light of these issues and desired capabilities, we offer two primary recommendations to improve the clarity of the USCDI.

First, we propose a more precise approach to defining the USCDI entities. Our proposed model is more fully described in the accompanying appendix. A key feature of this approach is aligning the definitions of USCDI entities with prevailing data shapes of exchange specifications and common data models. Another feature is to provide a clear mechanism for defining entities with various levels of specificity. And finally, this approach clarifies how to appropriately connect content from terminology standards and other syntactic standard (e.g. the USPS address specification) to different components of the model.

Over time, as industry alignment and capabilities for standardized exchange grow, it is expected that additional `Data Elements` will be enumerated for existing `Data Classes` in the USCDI. With our proposed approach, these details can be added in a principle way thereby making the "shape" more precisely drawn. Likewise, new `Data Classes` with clear delineations from existing types may be added to expand the kinds of health information represented, thereby making the picture more complete.

Second, we recommend that the USCDI publication identify exemplar technical specifications that ONC deems to have successfully represented the `Data Class` or `Data Element`. For example, we might expect references to specific FHIR profiles from the U.S. Core Implementation Guide or templates from the Consolidated CDA Templates for Clinical Notes specification. The intent of such linkages is to give users examples of how these entities have been represented in technical specifications. In ONC's USCDI version update process, entities included in the USCDI, are required to have sufficient maturity and representation in technical standards. We propose that those specifications are documented so that users can easily find them.

We hope these proposals are useful for advancing the conversation of how the USCDI can best enable nationwide health information exchange.

If you have questions, comments, or refining ideas, please contact us.

Appendix: Proposed Model for USCDI Entities

USCDI Entities

Data Element

Consistent with the notions outlined by the ISO 11179 standard and an Element as used in FHIR, at a minimum, a `Data Element` contains:

1. **Definition**
2. **Representation** (i.e. a label or identifier)
3. **Specification of permissible values**

A USCDI `Data Element` would rarely be used on its own. Rather, `Data Elements` are typically assembled into larger structures that provide context and meaning. You might consider them as an attribute, property, or component field of a larger data structure. When represented in a Common Data Model (CDM) such as OMOP or the PCORnet CDM, a USCDI `Data Element` would most closely correspond with a field in a database table.

Sometimes the permissible values are simple data types, like an integer or date. Other times, the permissible values may have more complex data types (which in some approaches are considered data elements in their own right). Examples of a complex data type would be something like **address** (street, city, postal code, etc. in the specific format defined by the USPS) or **human name** (prefix, given, family, etc).

Yet we can't fully understand **Address** in the context of health information without knowing more. It could be part of a data structure for Facilities, or alternatively, part of the demographics for Patient. Therefore, in USCDI, a `Data Element` is labeled in connection with its parent structure. That parent structure is called a `Data Class`.

In the example above, USCDI would label the `Data Element` as **Facility.Address** (or something similar). As another example, many kinds of DateTime attributes are important in health information. For example, the performed DateTime of a procedure, or clinically relevant time of a test result. USCDI would label these `Data Elements` as **Procedure.performedDateTime** or **Observation.effectiveDateTime** (or something similar).

It is not the purpose of the USCDI to specify all possible `Data Elements` for a particular `Data Class`. Nor is it to descend into the granularity of primitive data types (unsigned integers, positive integers, etc). The purpose is to pragmatically identify the key attributes necessary to support national data exchange. In the prior example, one can see how knowing the effectiveDateTime of a test result is crucial to its meaning in the context of clinical care. The USCDI approach recognizes that agreed transaction standards and implementation specifications will further specify additional details.

Data Class

A `Data Class` is a composite data structure comprised of enumerated `Data Elements` that serve as essential attributes necessary to support national data exchange. A `Data Class` is therefore a template for how specific instances are stored and exchanged in health IT systems.

You might say that a `Data Class` outlines the "shape" that particular kinds of health data have. In this way, a USCDI `Data Class` is analogous to a FHIR Resource or a database table in a Common Data Model. Yet, the USCDI `Data Class` definitions are neither exhaustive of the attributes an operational system needs nor exclusively bound to a particular exchange specification or data storage format.

For example, the **Patient Demographics** `Data Class` outlines the essential attributes including `Data Elements` for **Patient.Name**, **Patient.Birth Sex**, **Patient.Date of Birth**, etc.

As industry alignment and capabilities for standardized exchange grow, it is expected that additional `Data Elements` will be enumerated for existing `Data Classes` in the USCDI, thereby making the "shape" more precisely drawn. Likewise, new `Data Classes` may be added to expand the kinds of health information represented, thereby making the picture more complete.

Data Profile

The USCDI also defines subtypes of data belonging to a particular `Data Class` (shape) that are called `Data Profiles`. A `Data Profile` adds precision to the `Data Class` definition to characterize a specific subset of information that is useful for a particular purpose.

A `Data Profile` inherits the structural attributes (called `Data Elements`) comprising the parent `Data Class`. A `Data Profile` is defined by specifying constraints to those `Data Elements`, or optionally specifying additional `Data Elements` with constraints on them. Taxonomic (e.g. based on a terminological attribute or relationship) or functional attributes can be used to define the scope of a `Data Profile`.

Example: laboratory test results

Many electronic health record systems (EHRs) and CDMs separate laboratory test results from other kinds of tests, measurements, and observations. Yet, they typically a lab result like "hematocrit" shares the same data shape as a clinical measurement like "forced vital capacity". We can use a `Data Profile` to clarify how the shape is similar but the data content differs.

Observation Data Profile: Laboratory Test Results

To create a `Data Profile`, first we must establish a human understandable definition of the subset, such as:

> `Tests, measures, and observations about a specimen removed from the subject.`

Next, we characterize this `Data Profile` as instances of the **Observation** `Data Class`. Therefore, it inherits the specific `Data Element` attributes of `Observation.code`, `Observation.value`, etc. We create computer-understandable meaning for the `Data Profile` by specifying that the `Observation.code` is drawn from the set of LOINC terms with the attribute `CLASSTYPE` =1 (Laboratory Class).

Example: systolic blood pressure measurement

Vital signs represent an important subset of clinical measurements, including heart rate, respiratory rate, diastolic blood pressure, and systolic blood pressure.

Observation Vital Signs Data Profile: Systolic Blood Pressure

To more precisely specify a `Data Profile` for systolic blood pressure measurement, first we establish a human understandable definition (inherited from the LOINC concept model):

> `A quantitative measure of intravascular systolic pressure within the arterial system.`

Next we add specificity by further constraining the permissible values for specific `Data Elements`. Systolic blood pressure is a measurement that belongs to the **Observation** `Data Class` (and may also belong to a **Vital Signs** `Data Profile` should one be defined). Therefore this `Data Profile` inherits the specific `Data Element` attributes of `Observation.code`, `Observation.value`, etc.

To define a Systolic Blood Pressure `Data Profile`, we constrain possibilities for the `Data Elements` to more precisely identify this measurement :

- `Observation.code` must be drawn from the set of appropriate terms for systolic blood pressure from LOINC (e.g. those belonging to LG33053-6)
- Since systolic blood pressure measurements are numeric quantities, then we also specify the applicable standard for coding of units of measure as well:
    - `Observation.valueQuantity.code` is bound to the **mm[Hg]** code from UCUM to represent "millimeters of mercury".

Data Collection

As the USCDI evolves, it may be useful to group entities that span different kinds of data structures. A `Data Collection` is an enumerated set of `Data Classes` or `Data Profiles` that are organized together for a particular purpose. Presently, no such collections are defined in the USCDI, but we anticipate the need will arise.

For example, USCDI V1 contains a `Data Class` for identifying medications (as substances). Over time, the USCDI may evolve to cover the broader structures for recording how medications are used, including ordering, dispensing, administration of medications, and recording statements of medication use. If these structures became different `Data Classes` (due to their different `Data Elements`), it may be useful to group them together under a **Medication** `Data Collection`, for example.