White Paper

From retrospective mapping to prospective standardization:

A comparison of integration strategies to achieve semantic data interoperability

Catherine Staes BSN, MPH, PhD

Keith Campbell, MD

Department of Veterans Affairs, Veterans Health Administration (VHA)

Office of Informatics and Analytics (OIA)

Knowledge Based Systems (KBS)

November 12, 2017

Executive summary

Background: Preserving the meaning of information when exchanging electronic health record data (i.e., semantic interoperability) is critical for delivering safe patient care and leveraging standards-based clinical decision support. Given that individuals often receive health care from more than one health system, integration of data from multiple sources is needed to 'view' a patient's complete health record and avoid erroneous clinical decisions based on incomplete or inaccurate information, such as decisions that lead to performing unnecessary tests or giving a patient a drug to which they are known to be allergic. To date, the strategy for achieving semantic interoperability between the clinical systems of the Department of Defense (DoD) and the Veteran's Administration (VA) has been to 'map' millions of data elements used in the respective EHRs to standard terminologies (e.g., SNOMED CT, LOINC, and RxNorm). 'Round trip testing' of the mapped concepts has identified problems with the quality of the mappings for bidirectional use. New strategies are required to achieve semantic interoperability to support safe patient care, both before *and* after the two organizations start using of a single vendor for their electronic health record systems. The use of logical definitions and terminology system extensions to manage concepts used in the delivery of care can overcome key challenges with the mapping strategy.

Objectives: The objectives of this report are to: a) describe the current mapping approach (i.e., retrospective mapping) and illustrate common mapping scenarios that result in poor quality mappings, b) describe a new approach for semantic integration using logical definitions and dialect extensions to represent concepts used in the delivery of care (i.e., prospective/native standardization), and c) describe the impact of the new approach on the problems observed as a result of the mapping strategy.

Methods: To describe the current mapping approach, we reviewed reports submitted by a consulting terminology expert who evaluated the process and outcomes from the multi-year mapping efforts, summarized key features of the mapping methods that threaten quality, and identified examples to illustrate mapping challenges. To describe the new approach, we explain the strategy for representing concepts required for interoperability, internal use, or integration of historical data, and we present basic models for representing concepts and managing requests for new concepts. Finally, we applied the new approach to the problems identified from the mapping strategy and discuss strengths and limitations.

Results: A major threat to quality concerned the requirement that local source terms be mapped to a <u>single</u> standardized terminology element; no creation of logical expressions was allowed to represent target concepts. The quality of the mappings were also impacted by incomplete and different mapping rules used by the two organizations. Ongoing resources are required to assess and maintain mappings over time.

The new approach to data integration involves the use of description logic to model and manage concepts from standard terminologies to support clinical care. Instead of mapping an existing local code or term to a standard code with the goal of creating semantic equivalence, challenges created by mapping were avoided by directly representing concepts using standard

codes or logical expressions that conform to a description logic model. We described the strategy for representing ‘things’ (i.e. meanings) about patient care using existing single SNOMED, LOINC, and RxNorm concepts, computable logical expressions based on SNOMED CT, LOINC, or RxNorm that are added to extensions managed by an organization, organization-specific UUIDs to support specific local needs, and by adding ‘names’ (i.e., new dialect-specific synonyms) to already-existing ‘things’. Finally, ‘alternate identifiers’ for SNOMED, LOINC or RxNorm concepts are useful for integrating historical data so source terms (linked to alternative identifiers) are accurately represented using standard concepts defined using description logic.

Conclusion: The new approach will require additional expertise, tooling and processes to support the strategy, but delivers many advantages, most notably improved representation of ‘things’ documented during clinical care and avoidance of the challenges derived from incomplete and variable mapping rules. This strategy is aligned with the new paradigm for electronic health records for the VA and the DoD. This strategy requires central management of terminology but allows localization to meet the needs of a particular facility or domain of users (e.g. nursing) and supports integration with the large volume of historical information that will continue to be important to ensure safe care delivery.

Table of Contents

Background

Semantic interoperability of EHR data (i.e., the ability to preserve the meaning of clinical information shared between systems) is critical for delivering safe patient care and leveraging standards-based clinical decision support. Given that health care for an individual is often delivered by more than one health system, integration of data from multiple sources is needed to 'view' the complete health record. *Bi-directional information sharing* is necessary for care-coordination between health care system and to communicate information for clinical decision making while avoiding erroneous clinical decisions based on incomplete information, such as decisions that lead to performing unnecessary tests or giving a patient a drug to which they are known to be allergic. Thus, the **primary goal of semantic interoperability is preservation of the meaning of information for bidirectional use.**

To achieve interoperability, heterogeneous clinical systems must understand both the structure (*syntax*) and meaning (*semantics*) of the clinical information being exchanged. In other words, systems must have a shared understanding of a) the models used to represent clinical concepts, and b) the meaning of the concepts included in those models. Without these two features, information may be viewable by humans, but not integrated for viewing trends, clinical decision support, analysis, reporting, and other uses. For example, the ability to share patient data using a web browser, where the semantics of the data are interpreted by the human who reads the text presented, demonstrates syntactic integration. Entries in Problem Lists from two systems can be viewed together, but duplicates or conflicting information cannot be automatically resolved. Interoperability requires harmonization of data on a semantic level.

In this report, we describe results of a large-scale mapping-approach to interoperability undertaken by two large US healthcare organizations. These two organizations have been mapping data elements used in their respective EHRs to standard terminologies (e.g., SNOMED CT, LOINC, and RxNorm) as specified by the Office of the National Coordinator for Health Information Technology (ONC).[1] **The goal of mapping is to transform content between a source and a target system to meet the purpose for the integration.** More specifically, mapping is designed to transform clinical meanings (i.e. names and/or codes) used in one healthcare delivery system into coded clinical meanings that can be understood by a healthcare system using a different scheme of clinical meanings (i.e. names and codes). In other words, a mapping allows a target system to understand clinical data generated by a source system. Since these two organizations both provide clinical care for overlapping populations of patients, information loss during the transformation process can affect patient care and is considered an unacceptable outcome. Note that transformations for other mapping use cases may have a higher tolerance for information loss, particularly when the use of the mappings is uni-directional. For example, loss of information when classifying diagnoses for billing (i.e., mapping SNOMED CT to ICD codes) does not impact patient safety.

The ability of mapping efforts to preserve the meaning of clinical content depends on multiple factors, including:

- the content domain

- the standard terminologies used as the target
- the use of logical expressions to represent target meanings (i.e., by linking separate concepts to represent a single meaning)
- the ability to use more than one terminology to represent target concepts for a domain

These factors are supported by findings from the following studies:

- Using SNOMED CT, the proportion of source concepts with exact matches to a single SNOMED CT (the target) concept ranged from 19-90%, depending on the domain; and logical expressions (based on more than one SNOMED CT concept) were required to achieve an acceptable representation in over 40% of the domains.[2]
- 98.5% of the clinical terms suggested for a problem list could be expressed using SNOMED CT, but the clinical terms required the creation of expressions and use of logical models.[3]
- A combination of SNOMED CT and RxNorm satisfied most criteria for encoding common allergies and provided sufficient content coverage.[4]

Thus, to successfully represent clinical content, past studies have shown the need to create logical expressions and integrate terminologies (such as SNOMED CT, LOINC, and RxNorm). On the basis of these studies, as well as other pragmatic concerns, we propose founding an Informatics Architecture upon a description-logic based integration of **S**N**O**MED CT, **LO**INC and **R**xNORM (i.e., SOLOR), which enables the creation of logical expressions when necessary to accurately reflect the intended clinical semantics. L**ogical expressions can be used to represent the meaning of concepts with formal definitions, allowing a description logic classifier to algorithmically detect concept equivalence and identify subtypes (i.e., 'children' of the concept). The classification results and concept definitions can then enable patient and population-based inferencing required for clinical decision support.** In contrast to the "mapping" strategy for interoperability, which attempts to correlate meanings (names and/or codes) between systems as 'exact' or inexact (i.e., broader than or narrower than) matches, the use of logical expressions allows for exact representation (i.e., preservation) of the meaning of clinical content.

The SNOMED CT Concept Model provides formal editorial guidance for structuring description-logic based concept definitions (i.e., logical expressions) using SNOMED CT concepts to specify IS-A relationships and attributes. This approach is already being used to integrate domains in LOINC with SNOMED CT.[5] We advocate using this same concept definition approach for enabling semantic interoperability between clinical systems and combining these definitions with a standardized observation result model, as well as standard models of goals, requests, and actions.

Assessing readiness of mappings for bidirectional use for clinical care

Bidirectional use of mappings requires that the clinical meanings (i.e. names and/or codes) used in one healthcare delivery system can be understood by a healthcare system using a different scheme of clinical meanings (i.e. names and codes). Assessing readiness requires an understanding of the relationship between local codes (names), standard codes (concepts), and the 'things' the codes are representing that pertain to clinical care (Figure 1).

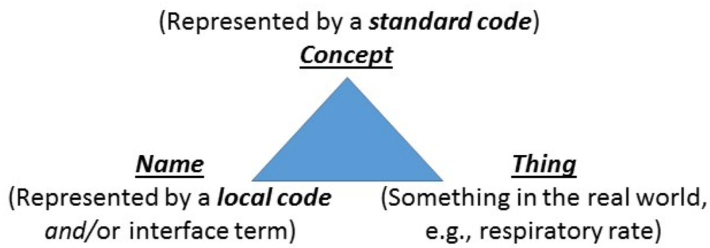
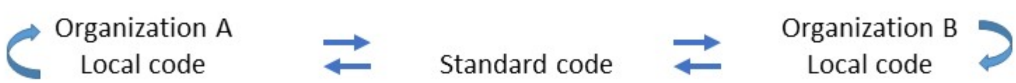


Figure 1. Key components of the semiotic triangle[6]

In the situation where two organizations have equivalent 'things' that need to be shared, round trip testing (RTT) can be used to assess the readiness of mappings for bidirectional use. RTT assesses use of standards-based mapping across organizations by domain and standard, where the standard code mediates the local codes. **To be successful, exact mappings must be shared between the two or more organizations**. Exact matches between organizations' mappings can only be obtained if: a) the two organizations are representing the same 'thing', and b) a standard code exists to represent the 'thing', and that code is used for the mappings. If these criteria are met, then the following bidirectional use can be supported:



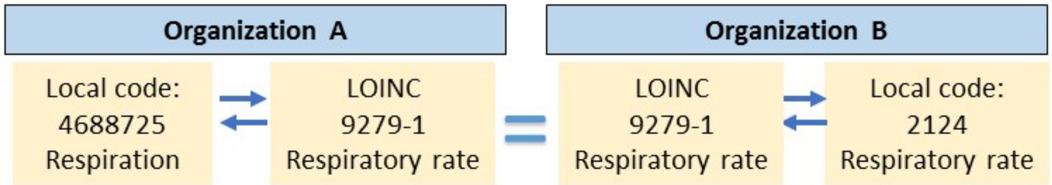
In the situation where two organizations are representing 'things' that are different in breadth and level of granularity with respect to each other, then exactly matched concepts will not be available from the organizations. For example, organizations often use different granularity and breadth for document names. In this situation, standard codes and expressions with definitional knowledge (such as, document name X 'is-a' child of document type Y) can support aggregation and understanding of the meaning of the document names shared. However, round trip testing is not the appropriate method for assessing the quality of mappings because equivalent 'things' don't exist between the two systems.

Successful Round Trip

Successful round trip testing (RTT) occurs when:

- local codes from different organizations that represent the same 'thing' (i.e., have the same meaning) are given an exact map to the identical standard code.

The following scenario illustrates successful RTT (Note: similar colors represent exact matches):



Unsuccessful Round Trip

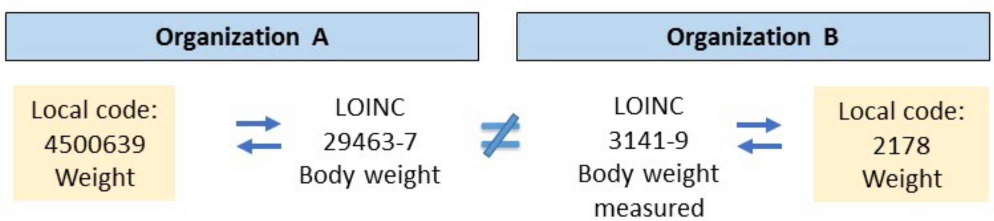
Unsuccessful round trip testing occurs when:

- Local codes from different organizations that represent the same 'thing' (i.e., have the same meaning) are NOT given an exact map to the identical standard code.

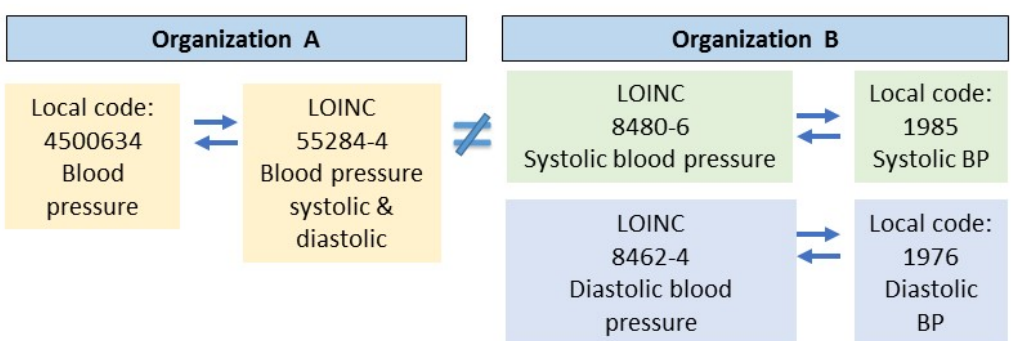
The following scenarios contribute to unsuccessful round trip testing:

- Errors exist in a mapping file (e.g., transcription errors or transposed mapping file entries).

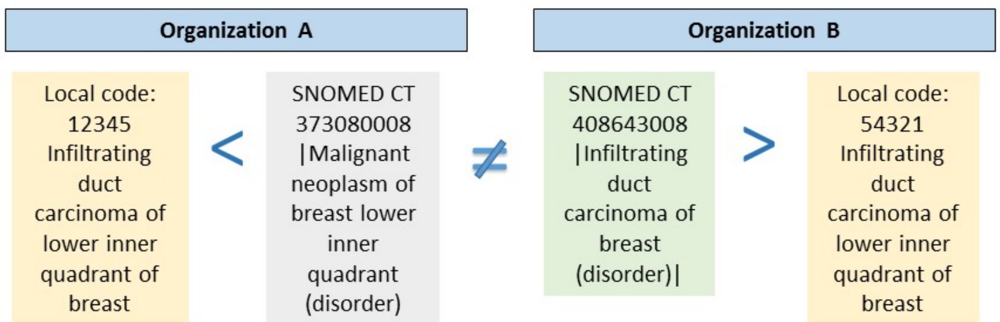
- Organizations share a common 'thing' but independently map their local codes for that 'thing' to a different standard code. In this example, local codes for equivalent 'things' are mapped to different yet arguably plausibly correct standard codes.



- Differences in information models lead to differences in the standard codes selected for the mappings.



- Differences in the breadth and/or level of granularity of local codes used by organizations compared to the available standard codes creates the opportunity for different standard codes to be selected when mapping. If a local code is more specific than that of the codes offered in a standard terminology, then the local code may get mapped to a 'broader than' standard codes. In the situation presented below, the two local codes represent the same meaning, but the exact meaning is not currently represented by a single code in SNOMED CT. There are two standard codes that are 'broader than' the local codes, either of which could be selected during a mapping effort. In this scenario, the two organizations mapped to different 'broader than' codes.



Description Logic

To meet national goals for data sharing, organizations must be harmonized to the ONC's selected data standards, mainly SNOMED CT, LOINC and RxNORM. These terminologies will be represented in SOLOR which has description logic components that allow for a new approach to harmonization and the ability to share post-coordinated expressions.

Description logic (DL) uses a formal knowledge representation language to model 'concepts' (classes) and 'roles' (properties) and their relationships, and is widely used for creating logical definitions of terminology content. The formal semantics allows humans and computer systems to exchange data without the ambiguity created by mapping. The formal semantics also makes it possible to use logical deduction to *infer* additional information from the facts stated explicitly and computably. This inference capability is an important feature that distinguishes DLs from other modelling languages such as UML.[7]

The capability of inferring additional knowledge increases the modelling power of DLs but it also requires that the modeler understands how to use DL and, above all, has good tool support for classifying content and computing conclusions.[7] The computation of inferences is called *reasoning* and an important goal of DL language design has been to ensure that reasoning algorithms of good performance are available.[7] Since DL conforms to the Web Ontology Language (OWL) EL2 profile, existing or proposed tooling can be shared to manage, classify, and query knowledge represented using DL.

SNOMED CT uses description logic to represent formal, computable concept definitions.

SNOMED CT concepts are based on the SNOMED CT concept model and have relationships ('is-a', 'attribute') and attributes with values that are both defining and qualifying. In addition, the SNOMED CT model includes hierarchies for domains of content and specifies the allowable attributes for each domain. For example, concepts in the 'Clinical finding' hierarchy may include attributes such as 'Finding site', 'Due to', 'Severity', and so forth. Concepts in the 'Evaluation procedure' hierarchy may include attributes such as 'Component', 'Measurement method', and so forth. Clincal content can be represented as either single concepts or logical expressions. Single concepts use a single SNOMED CT concept identifier; and logic expressions have a unique identifier but contain more than one SNOMED CT identifier.

Objectives

The objectives of this report are to: a) describe the current mapping approach (i.e., retrospective mapping) and illustrate common mapping scenarios that result in poor quality mappings, b) describe a new approach for semantic integration using logical definitions and dialect extensions to represent concepts used in the delivery of care (i.e., prospective/native standardization), and c) describe the impact of the new approach on the problems observed as a result of the mapping strategy.

Comparison of approaches for semantic integration

To better understand the value of semantic integration using description logic, it is useful to review the processes and outcomes of a large-scale mapping project. The mapping approach will then be contrasted with the alternative approach of coding clinical information either directly using standard codes or by creating logical expressions (i.e., description logic definitions) using standard codes. The latter approach is 'native standardization'.

Interoperability through mapping

Mapping approach and challenges of a large-scale mapping project

In this report, we describe results of a large-scale mapping-approach to interoperability undertaken by two large US healthcare organizations. These two organizations have been mapping data elements used in their respective EHRs to standard terminologies in order to enable interoperability. The two organizations followed mapping processes and guidance received from a Coordinating Office. The Coordinating Office selected the domains of content to be mapped and specified the standard codes to be used based on guidance from the Office of the National Coordinator for Health IT (ONC) (Appendix A). The Coordinating Office defined mapping rules and set priorities, and worked with the two organizations to review and 'sign-off' on the mappings. Each organization carried out its mapping project independently, but followed the same overall operating and quality control procedures and were given the same set of domains and priorities. Both organizations used teams that involved a mix of internal and contracted personnel. While there were similarities in the mapping work, there were also important differences (Table 1).[8]

Table 1. Description of selected features of the mapping effort implemented by two organizations

	Organization A	Organization B
Content to be mapped	~80,000 local codes/terms (i.e. potentially higher granularity)	~12,000 local codes/terms (i.e., potentially less granular)
Tooling	Used a mixture of tooling (LEAP, RxNav), UMLS tables, and manual processes	Used some tooling (Clue, RxNav, Workbench, RELMA), but mostly a manual process
Mapping rules	Applied rules defined by the Coordinating Office (i.e., single concept target) but created new rules for domains of content • Rules varied by domain • Rules evolved over time • Relationship types NOT added; they mapped only to the closest match attempting to make an exact match	Applied rules defined by the Coordinating Office (i.e., single concept target) but created new rules for domains of content • Rules varied by domain • Rules evolved over time • Relationship types were added, indicating whether the mapping was 'narrower than', 'broader than', 'equal to', 'unmappable', or tagged as legacy, out of scope or low frequency.

The following features of the mapping effort directly impacted mapping quality:

- Local source terms were to be mapped to a single standardized terminology element. **No creation of logical expressions was allowed** to represent target concepts. More than one source term could map to a single target; one-to-many mappings were not permitted.
- Organization A mapped to the **closest matched** target, while Organization B added a relationship type of **'exact', 'narrower than (NT)', 'broader than (BT)', or 'unmappable',** and tagged others as legacy, out of scope, or infrequent.
- While some tools were used to enable mapping, the specific tools and their use differed between the two organizations, and mapping efforts were overwhelmingly **manual processes**.
- To implement the mapping and quality assessment procedures, teams developed domain-specific business and mapping rules. Sometimes,
    - **rules varied** between the two organizations. For example, mapping of allergen reactants differed. For example, Organization A used UMLS CUIs to map medication allergens to RxNorm, and considered Drug classes and non-drug allergens to be unmappable. In contrast, Organization B mapped medication allergens to RxNorm, Drug classes to NDF-RT, and non-drug allergens (food or chemicals) to UNII codes. Similarly, one organization applied a rule for laboratory test names based on the following assumptions: If "no specimen is found in the source term and an unspecified source is not available in LOINC, map to most common specimen or serum/plasma (use best clinical judgment)" and "For Microbiology terms, when no method is indicated in the source term, the default will be to culture when clinically probable".[8]

- necessary **rules were missing.** For example, mapping rules for medications did not specify which RxNorm term types should be used.

To date, the two organizations have mapped millions of local codes and terms to standard codes using the approaches defined above. While some of the mappings produced semantically equivalent standard codes for equivalent 'Things', many of the standard codes selected were either "narrower than", "broader than" or "unmappable" in comparison to the 'things' they were representing. While a high percentage of each organizations' local terms could be mapped, one-way and round-trip testing showed limited success with semantic interoperability. Preliminary results of the RRT showed success ranging from 9% for document names to 63% for medications. In addition, the percent of mappings mediated by shared codes was low for several domains for the two organizations: 14 LOINC codes mediated only 8.6% or 40% of documents, five LOINC codes mediated only 29.4% or 45.5% of vital signs, and 5,731 RxNorm codes mediated approximately 63% or 33% of medications.

Observations from the mapping efforts

Here we present specific examples that illustrate the implications of the approach undertaken, particularly the requirement to map to a single target code, that may explain the limited success in semantic interoperability.

Information loss when mapping (problem list entry example)

Currently, Organization A uses MEDCIN[9] for its problem list entries. MEDCIN is an interface terminology that supports complex expressions that can be fully represented using post-coordination of SNOMED CT concepts. In the example below, the clinicians in Organization A can include the stage of the cancer when they document cancer on a problem list (Figure 2). The stage of cancer may be important for care management, to trigger clinical decision support, or for cohort identification. Therefore, preserving the stage of the cancer when information is shared between Organizations would be important. Currently, SNOMED CT does not have a single concept that combines prostate cancer and a given stage of prostate cancer. Currently, the stages are represented as separate 'findings' that must be associated with the prostate cancer (disorder). In other words, a single current SNOMED CT code cannot represent the full meaning of "prostate cancer stage IIB".

In contrast, at Organization B, mapping is avoided because only single SNOMED CT coded terms are currently allowable in the problem list. While this current policy may simplify the mapping task, clinicians may prefer more expressive terms when updating a patient's problem list. The widespread use of MEDCIN may be evidence of the need for expressive 'problems' or 'health issues'.

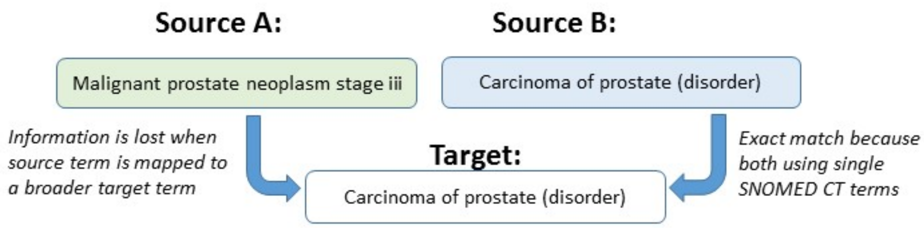


Figure 2. Single target mappings proposed for prostate cancer Problem List entries

The frequency of adverse clinical outcomes due to these kinds of mapping errors is unknown; however, some adverse events related to mapping errors are known to have occurred in the past and future preventable adverse events can be expected if no change occurs. For example, there is a potential safety risk if a veteran's premalignant condition is mapped to a benign condition in his transfer records. This kind of inaccurate mapping could result in a provider not receiving prompts for annual follow-up, potentially leading to preventable adverse outcomes.

Inappropriate model (CDS recommendation example)

To map the source term "Need for vaccination hepatitis B" following the single target strategy, the mapping team initially made the decision to use the broader term "Procedure needed (situation)". Subsequently, the Coordinating Office recommended using a narrower existing SNOMED CT term ("Requires course of hepatitis B vaccination (finding)") (Figure 3).

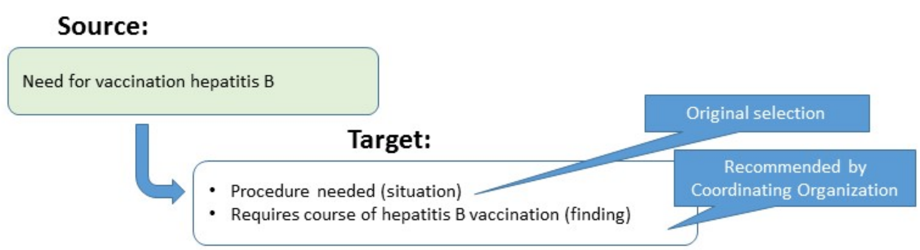


Figure 3. Single target mappings proposed for terms representing the need for hepatitis B vaccination

In fact, neither selection is appropriate for different reasons. "Procedure needed (situation)" is too broad and non-specific, and the currently-existing "Requires course of hepatitis B vaccination (finding)" is not modelled according to the currently-recommended approach. There are plans to review the current (finding) hierarchy and move concepts to the (observable entity) or (situation) hierarchies where appropriate. For example, the (situation) hierarchy is used for influenza vaccination concepts. Vaccinations are procedures that are more appropriately modeled in the situation with explicit context hierarchy which would allow for the context of the procedure (e.g., needed, refused, or done) to be used as an attribute.

Using logical expressions and the (situation) hierarchy for vaccine recommendations is a more extendable approach that limits combinatorial explosion. This approach could meet needs as new vaccines become available or new situation contexts need to be documented. In addition, this approach could improve alignment with the intent of the source concept, which may in fact be recommending a single dose of vaccine, not a 3-dose course of hepatitis B vaccine.

Variation in mapping business rules (Allergen example)

As described in Table 1, the two organizations followed different rules to map allergens. Organization A used UMLS CUIs to map medication allergens to RxNorm, and they considered drug classes and non-drug allergens to be unmappable (Figure 4). In contrast, Organization B defined mapping rules that allowed different code systems to be used as the target. For example, medication allergens were mapped to RxNorm, drug ingredients were mapped to RxNorm or SNOMED CT, drug classes were mapped to NDF-RT, and non-drug allergens including food or chemicals were mapped to UNII codes. As shown in figure 5, the differences in the business rules between the two organizations produced different results. Accurate documentation of allergens is critical for patient safety because administration of medications or exposures to products for which the patient is already known to be allergic can result in preventable adverse events, including death. Information about allergies should be shared between clinical systems where patients are seeking care. Mappable and known allergens should not remain unmapped and unavailable for processing by CDS systems.

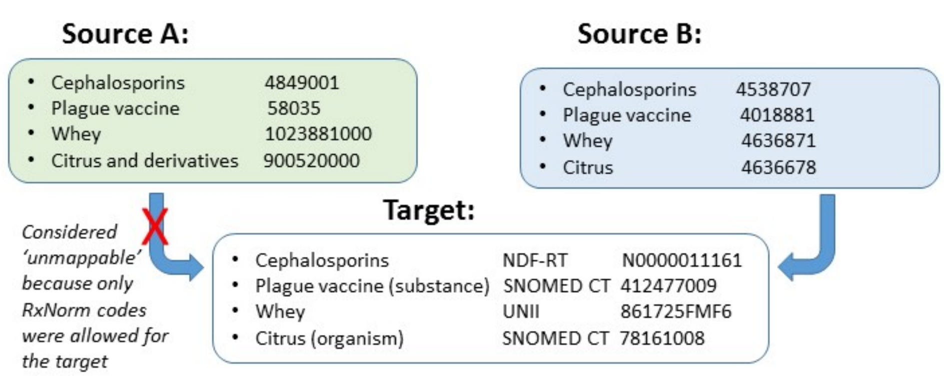


Figure 4. Sample mappings of allergens produced by different mapping business rules

Variation and incomplete mapping business rules (Medication example)

The business rules for mapping medications varied between the two organizations. At Organization A, the team was instructed to map medications to an exact or slightly broader target in RxNorm. They were instructed to use a generic target, but if the source had text for a brand name, then use the brand drug for the target.[8] In contrast, at Organization B, exact maps were done if the source term had an exact match to an RxNorm term. If there were slight deviations in the source text but an equivalent match was not found in RxNorm, a note of "not an exact match" would be included.[8]

The business rules were also incomplete. Neither organization was instructed to use specific term types for target codes, even though only four term types would be appropriate for representing prescribed drugs (Semantic Clinical Drug (SCD), Semantic Branded Drug (SBD), Generic Pack (GPCK), and Brand Name Pack (BPCK)). Table 2 shows a sample of mappings produced by one of the organizations. If Term Type rules had been in place, the target using a Semantic Clinical Drug Form (SCDF) would not have been allowed.

Table 2. Sample of medication-related mappings output from the process

Local_Code	Local_Description	Target_Code	Target_Code_Description	TTY	Comment
3000265280	Flunisolide (Nasarel Eq.) Spray 29 mcg-0.025% Nasal	603778	flunisolide 0.029 MG/ACTUAT Nasal Spray	SCD	
15985826	Flunisolide (Nasalide) Inhalation Not Specified 0.025 mg per mL Nasal	1359830	flunisolide 0.029 MG/ACTUAT Nasal Inhaler [Nasalide]	SBD	(retired)
15447192	Flunisolide (Aerobid Eq.) Spray 250 mcg Nasal	752375	flunisolide Nasal Inhaler	SCDF	
15479465	Attapulgite (Kaopectate) Tablet 750 mg Oral	209636	attapulgite 750 MG Oral Tablet [Kaopectate]	SBD	
15459554	Kaopectate or Eq. Tablet 750 mg Oral	198482	attapulgite 750 MG Oral Tablet	SCD	

Given the potential safety risks associated with exchanging information about medications, it is critical that mappings be accurate. Accurately representing the clinical drug or the branded drug is more important than normalizing the term type of SCD because the graph of RxNorm concepts can be traversed. Manually mapping drug names can pose a safety risk given the lexical challenge posed by drug names that are similar to each other and unfamiliar to the personnel doing the mappings. Manual mapping requires expertise and formal processes, measurements of reproducibility, and demonstrations of compliance with relevant standards.[10] It is difficult to train or find personnel to perform mappings if they do not yet have a clinical or pharmacy background, and persons with the needed clinical or pharmacy background may be costly to hire. For example, if the drug metroprolol in one health system were to be mapped to methotrexate by another system, a patient may receive the wrong drug when presenting to a pharmacy to pick up a renewal. These kinds of lexical challenges are known to cause human errors when prescribing or dispensing drugs, so the need for human judgement to perform mapping adds yet another opportunity for error.

Requirements for quality assessment and management of mappings over time.

The mapping process results in the creation of files linking source codes to target codes that were known to exist at the time the mapping task was completed. Terminology is not static, however. New versions of terminology code systems are published twice a year (SNOMED CT, LOINC), monthly with weekly updates from the NLM (RxNorm), or at other intervals. A new version may add new content (for example, leafs to an existing SNOMED CT hierarchy) or may correct a problem that should trigger an investigation of a previous mapping. Similarly, local terms are regularly added for use in electronic health records, and the new local terms will require mappings to support interoperability. If the newly-released content is correcting a problem relevant to the existing mappings, or addresses previously-mapped content, then mappings must be updated.

Currently, mappings are updated in response to system update cycles which may or may not correspond with updates to the underlying terminology system. In addition, quality management processes do not usually include ongoing surveillance and regression testing as updates occur. Safety of the mappings could be addressed by adopting principles from an international standard, 'IEC 62304 Medical device software – Software life cycle processes'.[10] This standard specifies life cycle requirements for the development of medical software and software within medical devices, which is applicable for the management of terminology used in electronic health records. The standard provides a "framework of life cycle processes with activities and tasks necessary for the safe design and maintenance of medical device software."[10] The standard calls out the need for risk analysis (systematic use of available information to identify hazards and to estimate the risk), risk control (process in which decisions are made and risks are reduced to, or maintained within, specified levels), and risk management (systematic application of management policies, procedures, and practices to the tasks of analyzing, evaluating, and controlling risk). In the context of mappings, QA resources should be applied based on the probability and type of safety risk (e.g., fatal vs non-fatal) associated with the concepts being mapped. For example, mappings associated with potentially high risk chemotherapy drugs should require more quality analysis than mapping associated with lower risk moisturizing skin lotions.

While processes can be implemented to manage risk and quality assessments and the tasks for updating existing mappings, the effort associated with the mapping activity could be avoided altogether if standard terminologies, such as SNOMED CT, LOINC, and RxNorm, were used natively to document care. The changes introduced in the new version of the terminology would be directly applied and available for use.

Interoperability using logical definitions and dialect extensions

Approach using logical definitions and dialect extensions

In contrast to mapping, an alternative approach to data integration involves the use of description logic (DL) to model and manage concepts from standard terminologies to support clinical care. We call this approach prospective standardization. Instead of mapping an existing local code or term to a standard code with the goal of creating semantic equivalence, several of the challenges created by mapping can be avoided by directly representing concepts using standard codes or logical expressions that conform to a description logic model. [11]

Going forward, the terminology available for documenting care can be based on the following strategies:

A. Represent 'things' (i.e. meanings) about the care of the patients that are widely used and relevant for interoperability using:
   - existing single SNOMED CT, LOINC, and RxNorm concepts
   - computable logical expressions based on SNOMED CT, LOINC, or RxNorm that are created and added to extensions managed by an organization.

B. Represent 'things' only of internal value (i.e. not relevant for interoperability) using Organization-specific concepts that are created to support a specific local need.

- An organization-specific UUID (Universally Unique Identifier) can be used for concepts that are relevant for administrative or billing decisions that do not impact direct patient care. For example, when administrative personnel need to calculate a rating for eating disorders, they need to select among a set of complex concepts to generate a rating score, such as: "Self-induced weight loss to less than 80 percent of expected minimum weight, with incapacitating episodes of at least six weeks total duration per year, and requiring hospitalization more than twice a year for parenteral nutrition or tube feeding". This information is derived from review of the clinical record, but the goal is not to store this information back into the clinical record as a fully-specified concept.

In addition to needing concepts to represent 'things', sometimes there is a need for additional 'names' (for existing 'things' in a standard terminology) to be available as an interface terminology. For instance, historically-used terms may not be present in a standard terminology but may be preferable to a given organization. To meet this need:

C. Add additional 'names' to already-existing 'things'. New dialect-specific synonyms could be added to an existing concept to meet the needs of interface terminology systems (e.g., nursing terminologies, patient friendly terminologies, and so forth). For example, if an organization has a preferred dialect of 'femur fracture' over 'fracture of femur', then a new synonym can be added to the current concept 'Fracture of femur (disorder)' (Figure 5).

The above strategies address standardization going forward, but may not address all the needs to standardize existing records to ensure historical information is available for clinical decision support, integrated views and analysis, and population-level reporting. Therefore, <u>retrospective standardization</u> may be required to handle existing source concepts, particularly when you care about semantic equivalence, but do not need to add a synonym. To support this need,

D. Add 'alternate identifiers' to SNOMED CT, LOINC, or RxNorm concepts or expressions built using SOLOR tooling. This strategy is useful for integrating historical data so source terms (linked to alternative identifiers) are accurately represented using standard concepts defined using description logic. Any existing local code can be added as an identifier, thus linking the local code's unique ID to the SNOMED CT concept ID. This approach will only work if the local identifier adheres to concept permanence desiderata principles (i.e., the meaning will not change over time).
    - For example, as shown in Figure 5, a new alternative identifier corresponding to the source system identifier could be added to the current concept 'Fracture of femur (disorder), but no new description is necessary if the existing local 'name' does not need to be displayed.

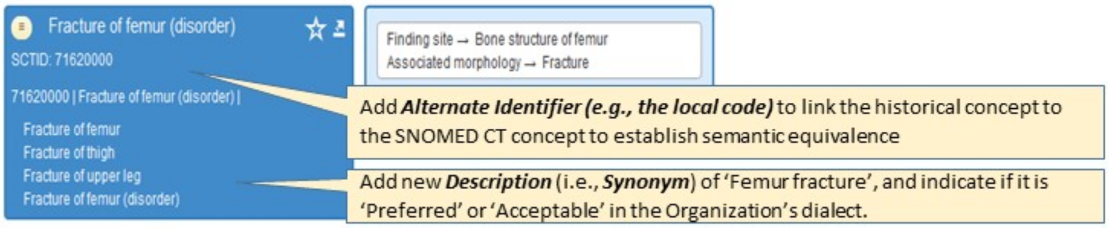


Figure 5. Snippet of a SNOMED CT Concept with a new Alternative Identifier and Synonym.

This approach has several additional strengths. First, a generic strategy is being used to handle each situation described above, supporting software reuse and avoiding accidental complexity by developing similar but different strategies for similar processes. A description of the process for handling new requests is shown in Figure 6 and the basic model for representing concepts in shown in Figure 7.Second, over time, an organization can transition to use the new content in lieu of the original non-native content. An organization can identify core information required for information exchange, decision support and reporting needs, and prioritize these needs. Third, this approach will contribute to the community and help SNOMED CT evolve. SNOMED CT does not benefit and evolve from mapping efforts.

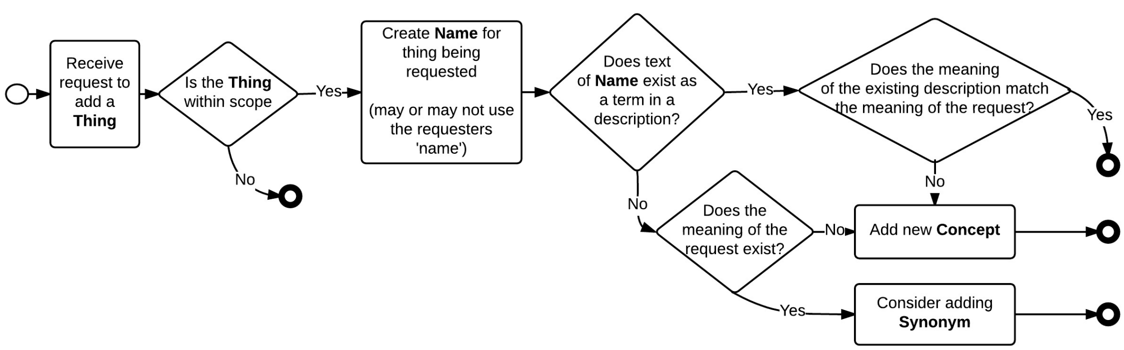


Figure 6. Process for handling the request to add a new 'thing' to a terminology

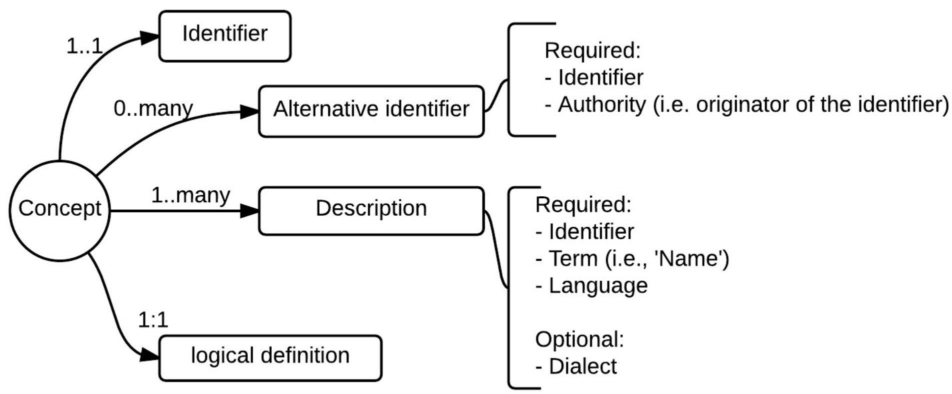


Figure 7. Basic model for representing concepts in the terminology

Proposed solutions to problems observed during mapping

The use of description logic and logical expressions allows source concepts that are represented in a structure to be more complete in content, more correct in meaning, and more consistent in representation. If both Organization A and Organization B represent their content using a logical expression model (such as provided by the SNOMED/LOINC integration effort[5]), the content can be integrated.

Here are examples of how the problems described above can be addressed:

Information loss when mapping (problem list entry example)

To overcome the data loss that occurs when selecting a single current target SNOMED CT code, either newly-defined single concepts with all the required semantics can be added to an organization’s extension, or logical expressions can be defined following business rules. Either way, the goal is to create computable logical expressions using the native standards and to provide the clinical meanings required for clinical care. Missing content can be added to an Organization’s extension, but content development business rules should be developed and followed.

Inappropriate model (CDS recommendation example)

There is an alternative to mapping “Hepatitis B vaccine needed” to a single target code resulting in the erroneous selection of a concept that is too broad (“Procedure needed (situation)”) or that is modelled incorrectly (“Requires course of hepatitis B vaccination (finding)”). The alternative would be to either create a single additional concept that expresses all of the required semantics, and add it to the organization’s extension of SNOMED CT (figure 8); or represent the concept using a logical expression (Figure 9). These two options are illustrated in the figures below as mock-ups in a demonstration environment and are not in the SNOMED CT International Release. With this strategy, the source term has complete representation as a SNOMED CT logical expression and its DL components can be represented within the formal ontological structure. There would be no data loss from Organization A's MEDCIN-based terms and the concept would now be available for Organization B as well.

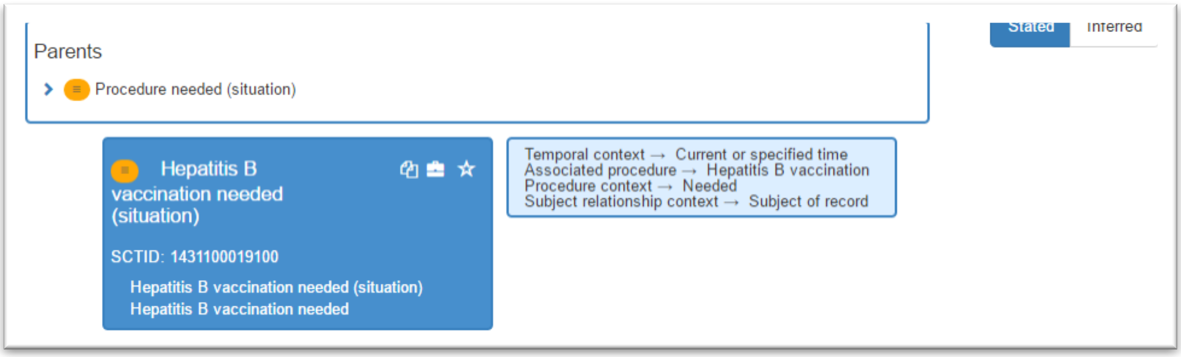


Figure 8. Mock-up of a single concept with all the required semantics needed for "Hepatitis B vaccination needed (situation)"

express '**Hepatitis B vaccination needed (situation)'**
=== 243796009 |Situation with explicit context (situation)| :
{ 363589002 |Associated procedure (attribute)| = 16584000 |Hepatitis B vaccination (procedure)|,
408730004 |Procedure context (attribute)| = 410525008 |Needed (qualifier value)|,
408731000 |Temporal context (attribute)| = 410512000 |Current or specified time (qualifier value)|,
408732007 |Subject relationship context (attribute)| = 410604004 |Subject of record (person)| }

=== 417451006 |Procedure needed (situation)| :
{ 363589002 |Associated procedure (attribute)| = 16584000 |Hepatitis B vaccination (procedure)|,
408730004 |Procedure context (attribute)| = 410525008 |Needed (qualifier value)|,
408731000 |Temporal context (attribute)| = 410512000 |Current or specified time (qualifier value)|,
408732007 |Subject relationship context (attribute)| = 410604004 |Subject of record (person)| }

Figure 9. Mock-up of logical expressions for "Hepatitis B vaccination needed (situation)"

This strategy is consistent with two current SNOMED CT concepts ("Influenza vaccine needed (situation)" and "Anti-D globulin needed (situation)"), and the model used to document that hepatitis B vaccination is done (Figure 10). Using the structures illustrated, a modeler could replace the Procedure context (attribute) of 'needed' with an attribute of 'Refused' or 'Done', or can replace the Associated Procedure (attribute) with a new vaccine. This strategy of using logical expressions is extendable and avoids the combinatorial explosion that could occur if single-concepts for every vaccine and context were to be represented by a single SNOMED CT code. In addition, this strategy supports the use of templates for defining new concepts required for an interface terminology.

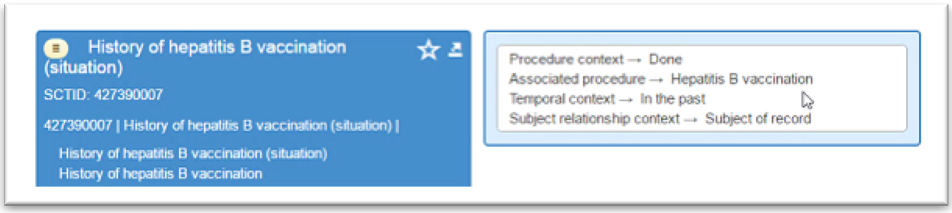


Figure 10. Semantics used to describe "History of Hepatitis B vaccination (situation)"

Variation in mapping business rules (Allergen example)

Domain-specific rules for mapping will no longer be necessary. Instead, the majority of the concepts required will be available from just three core code systems (SNOMED CT, LOINC, and RxNorm) that are integrated and managed using description logic and SOLOR tooling. Allergens for specific medications can be represented using RxNorm; all other allergens, including classes of medication, can be represented using SNOMED CT (Figure 10). Each of the local codes (e.g., Organization A code 4849001 and Organization B code 4538707 for Cephalosporin) could be added to the equivalent SNOMED CT code as alternative identifiers to enable interoperability with historical data. If clinicians require the term 'Cephalosporins', rather than the current SNOMED CT preferred term of 'Cephalosporin-class of antibiotics', then a new synonym could be added and specified as the Preferred term in for the Organization's dialect.

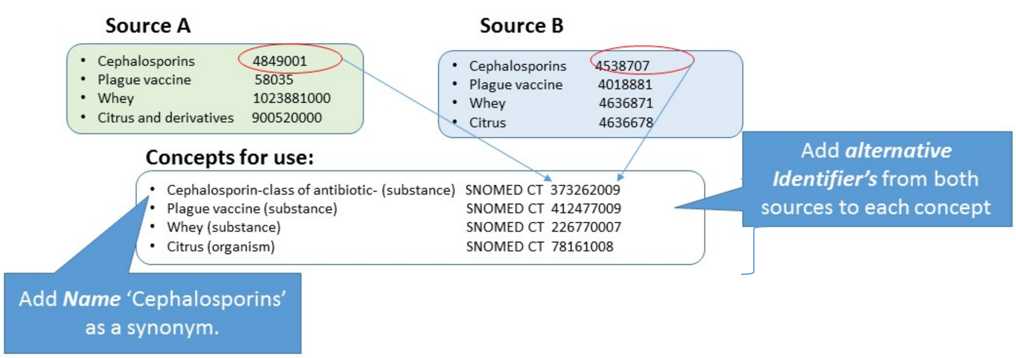


Figure 10. Example of allergens represented in SNOMED CT with additional identifiers and synonyms in the extension to support interface needs and integration of historical data.

Variation and incomplete mapping business rules (Medication example)

Business rules for mapping medication names will no longer be necessary. This is important because any mapping other than a one-to-one simple renaming of a drug concept has important patient safety implications.

Using the new strategy, an existing source concept identifier can be added in SOLOR as an alternate identifier to the same concept derived from RxNorm. Alternate drug identifiers are anticipated to become less important over time, as the RxNorm concepts become predominant in the record and historical medication information becomes less relevant for a patient's care. In addition, RxNorm content can be managed using description logic, which will allow for querying and classification for decision support, such as allergy checking and drug-drug interaction checking.

Since RxNorm does not include drug classes, drug classes will be represented using SNOMED CT, with synonyms and alternate identifiers assigned to SNOMED CT concepts where appropriate.

Requirements for quality assessment and management of mappings over time.

Quality assessment and management of mapping files will no longer be necessary as a separate activity from terminology development itself. By eliminating this mapping step, we reduce the complexity of the overall system, making it easier to scale and to validate content. Limited resources can be used more efficiently. The tasks associated with maintaining the quality of the terminology available will be addressed by the procedures already required for maintaining the SOLOR environment and the extensions. These processes will need to be conformant with IEC 6230:2006 Medical device software – Software life cycle processes.[9] When native standards are used, the synchronization challenges with new releases of terminology associated with the use of mapping files can be avoided.

Anticipated Challenges

Implementing the new strategy will have challenges that involve people, tooling, and processes. Some high level challenges include, but are not limited to, the following:

- People: To model concepts and define the associated description logic, personnel must: a) have practical clinical knowledge, b) have an appreciation for logic and formal reasoning that are required to create and maintain the logical expressions that form the heart of this approach, and c) have current knowledge about continually evolving SNOMED CT policies that may impact decisions. The high learning curve for modelling in SNOMED CT can make it challenging to hire, train, and retain qualified staff, particularly without advanced tooling to support this work. Tooling is critical for supporting the people involved.
- Tooling: A suite of tools is required to develop, manage, and monitor the quality of content over time. For example, tools are needed to create, edit, name, evaluate, and search for logical expressions to support this approach. When creating new logical expressions, modelers would benefit from templates to create expressions for pre-defined contexts, such as 'Procedure needed …', 'History of…', and so forth. Templates can be an assistive tool that enforces modeling rules and simplifies the editing task. Tooling can support the auto-generation of human readable names for a new expression, which is particularly needed if the expression is used in an interface. A classifier is required to search for equivalent expressions in a data store. Finally, tools are needed to audit and visualize content to identify errors or patterns

that can guide quality monitoring and training policies. For example, dashboards can assist with early error detection or throughput and outcome quality assessment. Any tooling needs to be agile to enable updates as policies change.

- Processes: Processes are required to enforce reproducibility, minimize variation, and increase efficiency with modeling. Standard procedures, templates, and logical expression archetypes will need to be developed, documented, and used to ensure accuracy and consistency across modelers. Defining the boundary between concepts that can or cannot be developed using predefined templates may be a challenge, but this problem can be addressed through practical experience using this approach. Enforcing reproducibility is a primary goal that can be achieved using an iterative dual independent review process[12]; but alternative strategies may be designed to achieve the same outcome at a lower cost, particularly depending on the tooling available. For example, audit information (by modeler and outcome) can guide policies about the frequency and intensity of the review activities required among the modeling team members. Dual independent review may be necessary for content that is complex; whereas, highly-templated content developed by an experienced modeler may only require a second reviewer to check a sample of the models developed. After developing content, processes are also required to manage content and issues that arise over time, such as, possible changes in the meaning of alternative identifiers used as synonyms.

Going forward, these and other challenges must be acknowledged and addressed.

Summary

Interoperability by mapping

Achieving semantic data integration by way of mapping local codes/terms from different organizations to a concept represented by a single standard code from a nationally-accepted standard has had limited success. Potential harmonization has been shown for a limited set of terms for some domains (e.g., medications) but a large proportion of the resulting mappings are not semantically compatible between the two organizations. The requirement that a single target concept must be used when mapping will never fully satisfy semantic equivalence for all domains. The limitations of the single target and manually-processed approach adversely affects data quality, limits data sharing capabilities, and may pose a preventable safety risk.

<u>Mapping is a retrospective effort at interoperability that will always be out of date and introduces new opportunities for errors.</u> The process results in the creation of mapping files of source and target terms that are to be used to interpret codes exchanged between departments. In Table 3, we present examples of challenges when mapping that impact correctness, completeness, consistency, and timeliness, and describe potential solutions.

Table 3. Potential problems observed or identified during mapping efforts and potential remediation strategies

Mapping challenge	Potential remediation strategy
Human subjectivity introduced when selecting 'broader than', 'narrower than', or 'closest match' (impacts correctness)	Allow computable logical expressions, rather than a single target concept, to define a source term
Information is lost when a single target is required, and no equivalent target code exists (impacts correctness)	Allow computable logical expressions, rather than a single target concept, to define a source term
Many 'unmappable' source concepts (impacts completeness)	• Allow computable logical expressions, rather than a single target concept, to define a source term • Create organization-specific concepts in SOLOR using DL and standard models and manage in the extension.
Difficulty ascertaining the definition of local codes and terms (impacts correctness)	Define each concept with a fully-specified name using Description Logic to clarify the definition of the concept .
As new versions of terminology is released, mappings may become out of date because maintenance of the mappings and the terminology are separate tasks.	When we use the native standardization option, terminology will be maintained as a single system. Thus, as new versions of terminology are released, maintenance will be handled once, not twice.
Problems with quality: consistency	Use standard terminologies, particularly the core of SNOMED CT, LOINC, RxNorm
Mapping process undertaken independently by the two organizations	Native standardization avoids the need for mapping
Mapping rules varied and different tooling used	Native standardization avoids the need for mapping

Interoperability by logical expressions and shared models

The use of logical expressions and shared models support interoperability. Instead of mapping an existing local code or term to a standard code with the goal of creating semantic equivalence, several of the challenges created by mapping can be avoided by directly representing concepts using standard codes or standards-based expressions that conform to a description logic model. [11]

Going forward (i.e., for <u>prospective standardization</u>), the terminology available for documenting care can be based on 'things' represented by existing single SNOMED CT, LOINC or RxNorm concepts, or computable logical expressions based on concepts from these code systems. Organization-specific concepts that are created to support a specific need, but not necessary for interoperability, can be managed using Organization-specific UUIDs. Additional 'names' can be added to existing concepts using the dialect functionality in SNOMED CT, to meet the needs of interface terminology systems (e.g., nursing terminologies, patient friendly terminologies, and so forth). Finally, 'Alternate Identifiers' to SNOMED CT, LOINC, or RxNorm concepts or expressions built using SOLOR tooling can be used to integrate historical data so source terms are accurately represented using standard concepts defined using Description Logic.

This approach has several additional strengths. First, a generic strategy is being used to handle each situation described above, supporting software reuse and avoiding accidental complexity by developing similar but different strategies for similar processes. Second, over time, an organization can transition to use the new content in lieu of the original non-native content. An organization can identify core information required for information exchange, decision support and reporting needs, and prioritize these needs. Third, this approach will contribute to the community and help SNOMED CT evolve. SNOMED CT does not benefit and evolve from mapping efforts. The challenges with this approach can be addressed.

Conclusion

In conclusion, achieving semantic interoperability to support the safe delivery of clinical care requires a change in strategies from mapping to the use of logical definitions and dialects managed in an extension.

References

1. HealthIT.gov | the official site for Health IT information. https://www.healthit.gov/
2. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. *Journal of the American Medical Informatics Association.* 2014-02-01 00:00:00 2014;21(e1):e11-e19.
3. Campbell JR, Xu J, Fung KW. Can SNOMED CT fulfill the vision of a compositional terminology? Analyzing the use case for Problem List. *AMIA Annual Symposium Proceedings.* 10/22 2011;2011:181-188.
4. Goss FR, Zhou L, Plasek JM, et al. Evaluating standard terminologies for encoding allergy information. *Journal of the American Medical Informatics Association : JAMIA.* Sep-Oct 2013;20(5):969-979.
5. LOINC. Regenstrief and the IHTSDO are working together to link LOINC and SNOMED CT. 2016; https://loinc.org/collaboration/ihtsdo. Accessed September 28, 2016.
6. The Meaning of Meaning. (Available from : https://en.wikipedia.org/wiki/The_Meaning_of_Meaning).
7. Krötzsch M, Simancík F, Horrocks I. A Description Logic Primer 2013; http://arxiv.org/pdf/1201.4089v3.pdf. Accessed September 28, 2016.
8. Wade G. 1004-B: Initial Assessment of the Current Mapping Process; Task 0004: Life-Critical System Analysis Support. Report published by Business Inforamtion Technology solutions, Inc, for Department of Veterans Affairs, Veterans Health Administration (VHA), Office of Informatics and Analytics (OIA), Knowledge Based Systems (KBS). Informatics Support Services HI 14-48 VA701-14-F-0176. 2015.
9. 2015AB UMLS MEDCIN Source Information 2016; https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MEDCIN/, 2016.
10. ISO. IEC 6230:2006 Medical device software – Software life cycle processes 2006; https://www.iso.org/obp/ui/#iso:std:iec:62304:ed-1:v1:en.
11. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology.* 2007;25(11):1251.
12. Levy DH, Dolin RH, Mattison JE, Spackman KA, Campbell KE. Computer-facilitated collaboration: experiences building SNOMED-RT. *Proceedings. AMIA Symposium.* 1998:870-874.

Appendix A. Data domains and targeted standard terminologies used when mapping between the two organizations

Data Domain	ICIB Standard
Allergies (Medication Allergies)	Drug - UMLS CUIs (or RxNorm)
Medications	RxNorm
Immunizations	CVX
Problem Lists	SNOMED CT
Vital Signs	LOINC Vitals Subset
Documents	LOINC Document Types
Pre- and Post-Deployment Assessments	None
Other Past Medical History (e.g., Travel)	Uncertain - Possibly SNOMED CT
Meaningful Use	Various (LOINC, SNOMED CT etc.)
User Experience/eHMP/JLV	Various (LOINC, SNOMED CT, ICD9/10 etc.)
VA VLER Partner Documents (CCDs, unstructured docs, CCDA)	LOINC Document Types
Results - Lab Chemistry & Hematology	LOINC
Results - Lab Anatomic Pathology	LOINC
Results - Lab Microbiology	LOINC
Results - Radiology Reports	LOINC
Encounter Data - Appointments	Encounter DXs - ICD10 or SNOMED CT
Encounter Data - Admissions	Encounter DXs - ICD10 or SNOMED CT
Procedures	CPT4/HCPCS or SNOMED CT (either)
Mobile Applications	Various
Demographics	Ethnicity & Race - CDC Race codes
	Preferred Language - ISO 639-2 alpha-3 codes
Social History	SNOMED CT
Family History	SNOMED CT
Scanned & Imported Paper Records & Non-Radiology Images	LOINC (for type); PDF-A, Text, JPEG, etc. for file type
Plan of Care - Pending Orders (Multiple Types)	Meds Orders (RxNorm)
	Consult Orders (None)
	Lab Orders (LOINC)
	Rad Orders (LOINC)
Radiology Images	DICOM
Payers	Insurance Types - LOINC
Questionnaires (General & Standard Instruments)	None
Functional Status	SNOMED CT
Providers	Provider Types - NUCC Taxonomy
Advance Directives (Metadata Only)	SNOMED CT Advanced Directives Type
Medical Equipment	UMDNS-SNOMED CT
Veterans Benefits Administration	Various