Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240665

Scalar Values in SNOMED CT: A Proposed Extension

Julien EHRSAM*[a,b,1], Christophe GAUDET-BLAVIGNAC*[a,b] and Christian LOVIS[a,b]
[a] *Division of Medical Information Sciences, Geneva University Hospitals, Geneva, Switzerland*
[b] *Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland*
*These authors contributed equally
ORCiD ID: Julien Ehrsam https://orcid.org/0000-0003-1506-558X

Abstract. Representing numeric values such as scalars holds great importance for accurately depicting clinical data. While the result value itself will always be represented using an integer, decimal, or other scalar format, it needs to be linked to its corresponding data element. In SNOMED CT, as in most other terminology systems, this is done through an attribute relationship. While some scalar values are already included in this way, they only represent a small fraction of possibilities. Our intention is to expand the scope of scalar representation by validating new attributes using a previously established method. The result is a list of five attributes validated for local representation of scalar values, improving semantic representation and interoperability.

Keywords. SNOMED CT, post-coordination, scalar values, semantic enrichment

1. Introduction

Numeric values such as scalars are an important facet of clinical data and are necessary for representing many types of data elements, from results to sizes. Many works mention the need for representing scalar values [1,2], but a solution applicable at scale is lacking. Widely considered as the most complete representation of medical concepts [1], SNOMED CT (SCT) still contains gaps which limit representation and reuse possibilities, such as representation of scalar values. Integers and decimals were previously represented with children of the |Number (qualifier value)| concept such as |2500 (qualifier value)|, but have since been inactivated, as scalar values do not need to be represented by a specific concept [3]. This has been replaced [4] by Unicode text expressions preceded by a hash (#), approved in the range of 10 attributes to refine the |Pharmaceutical / biologic product (product)| hierarchy, such as |Count of base of active ingredient (attribute)|. In fact, all 16'071 concepts which have an attribute with a scalar value are part of this hierarchy. Some numerical concepts remain in SCT, such as quantitative cut-offs used to define qualitative results or concepts containing a scalar value in the concept name but not in any attribute relationship [5]. These are generally

[1] Corresponding Author: Julien Ehrsam; E-Mail: julien.ehrsam@hug.ch.

results of measurements such as |Fetal heart baseline rate range above 200 (finding)|, are not fully defined and are gradually being removed.
In the long term these will hopefully reappear as fully defined concepts with new ways to represent scalar values through an attribute relationship. In the meantime, we propose a method to add to the currently approved attributes to expand the representation of our data. Currently, when using SCT to represent clinical patient data, a link must always be kept to the clinical data warehouse (CDWH) to represent scalar values. By having all available patient data represented in SCT directly in a graph database, the process is simplified as querying can be done directly within the semantic knowledge graph, allowing for actionable and representative information retrieval [6]. Indeed, representing clinical data in data models does not allow for semantic querying, and incomplete semantic representation does not allow full data discovery. The proposed method can be a first step to fully representing clinical data with semantic standards, bridging the current gap and allowing full semantic representation, increasing the analytical power through the use of numeric queries. This can also improve semantic interoperability.

2. Methods

The process to find and validate a new attribute for post-coordination resides in an established method [7]., applied when a data element which cannot be represented using current SCT rules is encountered. This includes frequently encountered scalar values such as lab results, dates, and measures. Discussions during focus groups are central to the process. According to this method, attributes approved by SNOMED International should be used first, but as previously mentioned this does not apply to most scalar values. If none are available, then an unapproved attribute can be used. This process is time-consuming and requires knowledge of SCT's structure and grammar rules. For example, which hierarchy to refine is not a simple question to answer. Lab test results should instinctively be findings. However, a finding is the interpretation of an observable entity or a procedure with a result value (Figure 1). Observables can be viewed as questions to which a quantitative or qualitative answer can be given, the association of the two making the finding [6,8,9]. Since the aim is not creating new precoordinated concepts but post-coordinated expressions, we must refine observations and procedures, not findings.

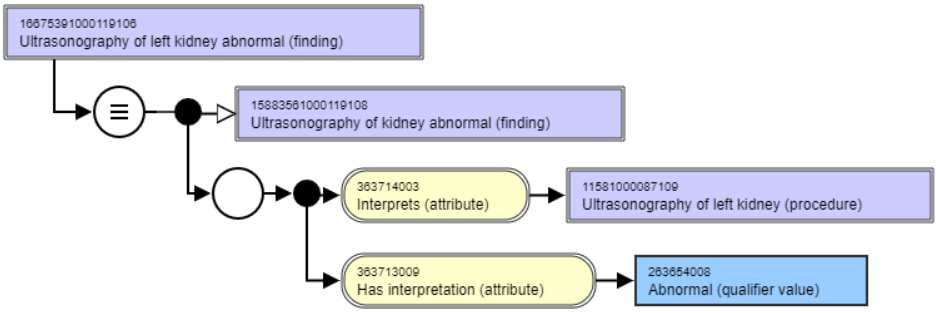

Figure 1. Clinical findings interpret observable entities or procedures.

A quantity is defined as the association of a numeric value with a unit [10], therefore, to refine an observable entity or a procedure quantitative result completely, a unit must

be attributed to it. As in other terminologies such as LOINC [11], this is done by expanding the domain of |Units (attribute)| to represent units for laboratory procedures, by adding << |Laboratory procedure (procedure)| AND <<|Evaluation procedure (procedure)|, as described in previous work [7].

Assessment scales and scores are represented heterogeneously in SCT. Some detail the final score with a finding with no attribute relations, such as |Glasgow coma scale, 9 (finding)|. Observable entities exist for some scores but not all, as do procedures describing the assessment. These often have no link to the score itself. As the most complete list of scores exists in the |Staging and scales (staging scale)| hierarchy, this was the one chosen to be extended, allowing representation of scores which are not yet detailed in SCT. This approach applies only to scores which have a numeric result.

3. Results

The results described below are examples of attributes which have been validated for representing scalar values, with corresponding examples.

3.1. Sizes

- Attribute: 103373006 |With size (attribute)|
- Type: new attribute approval
- Domain: <<260787004 |Physical object (physical object)|
- Range: << 255506008 |Dimensions (qualifier value)| OR >=#0
- Definition: Attribute used to represent the situation when an object is described using a numeric or adjectival descriptor concerning its size, such as large, 15G.
- Example: 360003007 |Guedel airway (physical object)| : 103373006 |With size (attribute)| = #0

3.2. Dates

- Attribute: 410671006 |Date (attribute)|
- Type: new attribute approval
- Domain: <<71388002 |Procedure (procedure)|
- Range: >#19000101
- Definition: Attribute used to represent the dates of procedures, such as surgical interventions, consultations, hospital admissions. The value must be an integer in the international date format of YYYYMMDD.
- Example: 281036007 |Follow-up consultation (procedure)| : 410671006 |Date (attribute)| = 20240615

3.3. Lab test results

- Attribute: 79409006 |Resulting in (attribute)|
- Type: new attribute approval
- Domain: << 108252007 |Laboratory procedure (procedure)| AND <<386053000 |Evaluation procedure (procedure)|

- Range: >=#0
- Definition: Attribute used for representing the outcome of laboratory procedures with quantitative results.
- Example: 70901006 |Creatinine measurement (procedure)| : 79409006 |Resulting in (attribute)| = 85.2

3.4. Measurement results

- Attribute: 246205007 |Quantity (attribute)|
- Type: new attribute approval
- Domain: << 363787002 |Observable entity (observable entity)|
- Range: >=#0
- Definition: Attribute used for representing quantitative results of measurements of bodily functions or sites, such as BMI or heart rate.
- Example: 251670001 |Baseline fetal heart rate (observable entity)| : 246205007 |Quantity (attribute)| = 204

3.5. Scores

- Attribute: 246262008 |Score (attribute)|
- Type: new attribute approval
- Domain: <<254291000 |Staging and scales (staging scale)|
- Range: >=#0
- Definition: Attribute used to represent the results of scales and scores with a numeric value.
- Example: 450731007 |Rockall score (assessment scale)| : 246262008 |Score (attribute)| = 2

4. Discussion

The ability to adequately refine concepts with scalar values while maintaining SCT's grammar rules is crucial. To represent data elements such as lab results, sizes, and dates, a new method must be used. Many possible use-cases exist, and this work focuses on cases encountered in practice and is therefore not exhaustive. Using unapproved attributes has drawbacks, as these may become validated for a different use or inactivated by SNOMED International. A regular check needs to be made to avoid these scenarios.

For lab test and measurement results, the post-coordinated expression created would become a finding if made into a pre-coordinated concept in the authoring platform and submitted to the national SNOMED center. This would also be semantically more accurate, but the concept definition would have to be completed to include cardinalities for example, to allow for proper classification. However, due to the large number of new concepts which would need to be created, a different approach was chosen. This can cause problems in retrieval and querying, as results are instinctively thought of as findings, and will be searched as such. As this has not been implemented, no solution has been established for this yet. The massive scale of values which will be created is another limitation of this work. The storage in a graph database has yet to be implemented.

An important limitation of this work is that it is not yet fully implemented. Certain attribute-value pairs are used to describe the entire variable name, as in section 3.1, where the equipment in the list is simply labeled as "Guedel 0". For these cases the attributes have been added to the local SCT classifier and are therefore already usable in our local query system. For data elements to which a value is associated outside the variable name however, this has not yet been implemented. This work presents the first step towards such representation, based on a series of examples encountered, but has not yet been applied at large scale to the local CDWH.

5. Conclusions

Using already existing but unapproved SCT attributes, it is possible to represent scalar values while staying compliant with the SCT grammar and post-coordination rules. This should, for example, permit queries which involve numeric operators, such as "All Glasgow Coma Scores with a result <10", improving reusability and interoperability. This expands the scope of representation possible using SCT and allows for inclusion of lab results and other clinically important elements in a graph database for easier querying and exploration.

References

[1] Benson T, Grieve G. Principles of health interoperability: SNOMED CT, HL7 and FHIR. Third edition. London: Springer; 2016. doi: 10.1007/978-3-319-30370-3

[2] Lee D, Cornet R, Lau F, De Keizer, N. A survey of SNOMED CT implementations. Journal of Biomedical Informatics. 2013 Feb;46:87–96, doi: 10.1016/j.jbi.2012.09.006

[3] Cornet R, Chute CG. Health Concept and Knowledge Management: Twenty-five Years of Evolution. Yearb Med Inform. 2016 May;25:S32–S41, doi: 10.15265/IYS-2016-s037

[4] SNOMED International Proposal for Representing Concrete Domains in RF2 [Internet]. SNOMED International; 2020 [cited 2024 Mar. 18]. Available from: https://confluence.ihtsdotools.org/mag/concrete-domains-community-of-practice-consultation/snomed-international-proposal-for-representing-concrete-domains-in-rf2.

[5] Rector AL, Brandt S. Why Do It the Hard Way? The Case for an Expressive Description Logic for SNOMED. Journal of the American Medical Informatics Association. 2008 Nov;15:744–751, doi: 10.1197/jamia.M2797

[6] International Health Terminology Standards Development Organisation. SNOMED CT® Technical Implementation Guide January 2015 International Release [Internet]. IHTSDO; 2015 [cited 2024 Mar 7]. Available from: https://www.nrces.in/download/files/pdf/doc_TechnicalImplementationGuide_Current-en-US_INT_20150131.pdf.

[7] Ehrsam J, Gaudet-Blavignac C, Mattei M, Bauman M, Lovis C. Semantics in action: a guide for representing structured EHR content with SNOMED CT. Available at SSRN: https://ssrn.com/abstract=4755395. 2024;

[8] Wade G. Implementing SNOMED CT for Quality Reporting: Avoiding Pitfalls. Appl Clin Inform. 2011 Dec;2:534–545, doi: 10.4338/ACI-2011-10-RA-0056

[9] Schulz S, Miñarro-Giménez JA, Martínez-Costa C, Karlsson D, Rosenbeck Goeg K, Marko K, Van Bruwaene B, Cornet R, Jaulent MC, Hamalainen P, Dewenter H, Fatollah Nejad R, Thun S. ASSESS CT - Assessing SNOMED CT for Large Scale eHealth Deployments in the EU. 2016 [cited 2024 Mar 8]; Available from: https://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf.

[10] NHS Digital. The SNOMED CT Encoded Unified Test List (UTL) Editorial Principles v1.3 [Internet]. 2021 [cited 2024 Mar 7]. Available from: https://nhsengland.kahootz.com/gf2.ti/f/972002/57976005.6/PDF/-/UTL_SNOMED_Editorial_Principles_v1.3.pdf.

[11] Knowledge Base [Internet]. LOINC. [cited 2024 Mar 18]. Available from: https://loinc.org/kb/.