Discussion on current guidance on hypernyms as synonyms

Current guidance on the use of hypernyms:

Broader synonym

When a synonym is more general than the FSN, and there is no context in which it has the same meaning as the FSN, the synonym should be inactivated. The description inactivation value of 723278000 | Not semantically equivalent component (foundation metadata concept) | is used.

For example,

- FSN: Sprain (morphologic abnormality)
- SYN: Joint injury - more general meaning than the FSN

However, a more general synonym is acceptable when there is a context in which the synonym has the same meaning as the FSN.

For example:

- FSN: Entire fundus uteri (body structure)
- SYN: Fundus in the context of obstetrics - same meaning as the FSN

Feedback ⊙ ⊙

Past guidance:

Broader synonyms

When a synonym is more general than the FSN, and there is no common context in which it has the same meaning as the FSN, the concept should be retired as ambiguous.

Example:

FSN: |Sprain (morphologic abnormality)|

SYN: |Joint injury|

However, a more general synonym is considered valid when there is a context where the more general synonym has the same meaning as the FSN.

Example:

FSN: |Entire fundus uteri (body structure)|

SYN: |Fundus| in the context of obstetrics.

These valid broader synonym terms should be marked using the "degree of synonymy" field in the language reference set, with value "near synonymous (depending on context of use)".

Discussion issue:

A concern has been raised as a result of a recent authoring decision in relation to the concept 398042001|Accidental dural puncture (disorder)|.

Briefly, this concept has been inactivated on the grounds that its FSN and preferred terms were hypernyms for its real meaning, as revealed by its former taxonomic position under 33211000|Complication of anesthesia (disorder)|.. A replacement concept 781129002|Accidental puncture of dura **during anesthesia** (disorder) | has been created, and there is a MAYBE from the inactive original to the new active current…but there has been resistance to the suggestion that the shorter '*accidental dural puncture*' hypernym, familiar to anesthetists might usefully also be added back as a synonym of the new concept.

The general need for hypernyms has been discussed in an email thread by EAG members:

Email discussion:

Jim Case <jca@snomed.org> Mon, Oct 7, 12:39 PM

Jeremy,

Thanks for forwarding this. ...there is the potential [for this topic] to cause heartburn amongst those who expect implied context to be carried along with hypernyms. The most critical aspect of this is the fact that we no longer (at least at this time) have the ability to assign a "degree of synonym" in RF2, which was one of the driving forces to change the editorial guide. We should not be allowing for things that we cannot do, so to speak. If we want to revisit this in detail, then this should probably be turfed to the MAG as it is a modeling and tooling issue. But we can certainly come up with some recommendations from the editorial standpoint if the situation warrants.

Guillermo Reynoso October 7,2019

I think we should at least start the process leading to restore the degree of synonymy refset, as it would be helpful in many use cases, particularly for those that have language authorities (e.g.

the Royal Academy of the Spanish Language) or extensions that add interface-oriented descriptions or say, patient terms (although I prefer them in a separate extension/Language Refset as the Netherlands do).

ROGERS, Jeremy (NHS DIGITAL) Oct 8, 2019, 9:25 AM

As to "...we no longer (at least at this time) have the ability to assign a "degree of synonym" in RF2, which was one of the driving forces to change the editorial guide"

Was that a tooling constraint, or something more principled that put a block on that mechanism being carried forward? I could imagine that it would not necessarily always be correctly used even if that functionality was available. In which case I could also entirely sympathise with an IE Editorial Stance along the lines of:

1. Terms that advertise as true synonyms but that are in fact hypernyms are intrinsically dangerous, especially in the context of clinical coding tools that are in the habit of presenting terms that exactly match the input search phrase but without exposing the true taxonomic position of the underlying concept. See also <link to collected thoughts on the pros and cons of contextually resolved hypernyms as terms that are user selectable and/or also persisted to the interoperating, human readable EHR>
2. For this reason, SI itself will NOT be adding hypernymic synonyms to its core content.
3. SI acknowledges the potential use of a language refset variant in which the familiar binary 'acceptable' and 'preferred' values for acceptability are augmented by a New! Core! possible ternary value of 'hypernym', for which possibility a new metadata concept ID under 900000000000549004|Acceptable (foundation metadata concept)| will be added to the International Edition. But to restate, SI will NOT itself be taking advantage of that functionality.
4. Although NRCs may choose to use this, SI and Ed Comm draw attention to the pitfalls of doing so (see (1))

Guillermo Reynoso Tue, Oct 8, 12:05 PM

Just adding some historical perspective...

The original language refsets had a column for degree of synonymy, then the column was moved to a separate refset (degree of synonymy) [the values were maintained in the environment we used before the WB, and later in the WB, the data was maintained but was never published, and got lost in the transition between environments).

The reason degree of synonymy was introduced between 2007-2011 was due to **resistance to adopt a pure synonym policy in the IE** (emphasis added), while there was interest to maintain more general and more specific terms, as well as near-synonyms. The compromise was to at least flag them using the degree of synonymy refset as synonyms, near synonyms and "non-synonym" synonym. The policy in the early 2000's was to inactivate a concept if there was a term that was not a synonym (as it was considered ambiguous). That policy was discontinued in 2005 with the addition of the "Refer-to" links for inappropriate descriptions, and the eventual introduction of the degree of synonymy refset...

Part of that work was related to some older standards focused on terminological systems and cross-language equivalents, like:

ISO 25964

https://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en

ISO 12620

http://semanticweb.kaist.ac.kr/org/tc37/pdocument/standards/ISO%2012620%20_1999.pdf

ISO 1087-1:2000

https://www.iso.org/obp/ui/#iso:std:iso:1087:-1:ed-1:v1:en

Some of the definitions there influenced the criteria for translating synonyms in the translation guidelines discussions. For example, the Spanish Edition is more strict in the criteria for synonym inclusion. However, implementation experience [reference?]is that usability could be improved in applications that require better interface terminology features by adding quasi-synonyms/near synonyms and more general terms, etc. in extensions to the translation (or using new description types in addition to synonym, like entry or index term, etc.)

Using those references applied to SNOMED CT, a synonym would be a term that represents the same concept as the FSN (the main entry term), while near-synonym is a term that represents the same concept as the FSN, but for which interchangeability is limited to some contexts and inapplicable in others.

The degree of synonymy idea was derived from the degree of equivalence, as mentioned in annex A of ISO 12620:1999:

A.2   term-related information

A.2.1 term type
DESCRIPTION: An attribute assigned to a term.
NOTE: *Term types* can include:

A.2.1.1   main entry term
ADMITTED NAME: **head term**
DESCRIPTION: The concept designation that has been chosen to head a terminological record.
EXAMPLE: "radix" in annex C, figure C.1

A.2.1.2   synonym
DESCRIPTION: Any term that represents the same or a very similar concept as the main entry term in a term entry.
EXAMPLE: "dynamicizer" in annex C, figure C.2
NOTE: Synonymy is generally relative, i.e., synonyms rarely cover all aspects of the same concept in all instances. The resulting *degree of synonymy* (A.2.10) is treated using the conventions defined for *degree of equivalence* (see A.3.1)

A.2.1.3   quasi-synonym
ADMITTED NAME: **near synonym**
DESCRIPTION: A term that represents the same or a very similar concept as another term in the same language, but for which interchangeability is limited to some contexts and inapplicable in others.
EXAMPLE: The distinction between synonyms and quasi-synonyms can be subjective or strongly context-oriented. For instance, some can designate "domain" and "subject field" as synonyms, whereas others would argue that "domain" is broader, but is usable in many of the same contexts and is therefore a quasi-synonym.

A.2.1.4   international scientific term
DESCRIPTION: A term that is part of an international scientific nomenclature as adopted by an appropriate scientific body.
EXAMPLE: *Homo sapiens*

A.2.1.5   common name
DESCRIPTION: A synonym for an international scientific term that is used in general discourse in a given language.
NOTE: Common names are generally formed based on metaphor, analogy, and function without reference to the classification rules applied to scientific nomenclatures. Although common names are widely used in general language, they are used in technical and scientific writing as well. Common names vary from language to language and frequently regionally within languages.
EXAMPLE: *Kalmia latifolia* is commonly called "mountain laurel" in the northern United States, and "calico bush" or "sheep's bane" in the south.

'ROGERS, Jeremy (NHS DIGITAL)' Wed, Oct 9, 2:13 AM

"The original language refsets had a column for degree of synonymy, then the column was moved to a separate refset (degree of synonymy) [the values were maintained in the environment we used before the WB, and later in the WB, the data was maintained but was never published, and got lost in the transition between environments)"

Question: Is that data lost for all time then? Or could it be salvaged and maybe pressed into service to provide some useful examples of what could go wrong if you allow hypernyms and other flavours of non-synonymous synonyms, but have no mechanism for distinguishing them from true synonyms?

Very happy to try and help work this up into something closer to a solution than a revisiting of a very old problem.

It is obviously a very long standing topic with much prior art. Personally, I'm a fan of rigidly pure synonymy in the core reference artefact. But, I'm also quite happy for folk to add around that rigid core some kind of capability to support contextually-resolved flavours of non-synonymy – such as hypernyms, non-unique abbreviations and near-patient terms. But folk who do that need to know the pitfalls involved, and the best practice ways of attempting to avoid falling into them.

The centre's role might therefore be limited to collating guidance on those pitfalls and the techniques for avoiding them, and offering at least one central solution for encoding non-synonymy in the hope that those outside the centre who choose to go down this path might be nudged toward do so in a common way. From Guillermo's potted summary of the ancient lore in this territory, flagging non-synonyms in a novel (and relatively small) cRefset makes more sense than rolling this into a normal language refset but as an additional column.

There is I think also underneath all this a clinical debate to still be had about whether EPR recording should be moving to a world in which clinicians can usually find the correct concepts using e.g. hypernmic search phrases, but they can't persist the utterance – or subsequently render it back to themselves or to any other clinician – by means of the same term. I know at least some EPR suppliers to generalist clinicians who are already choosing to restrict re-display of already entered EHR content to always using only the Preferred Term, on exactly the kind of

safety grounds underpinning the anxieties over routine use of e.g. hypernyms and non-unique abbreviations if and when these travel outside the clinical context in which they might reasonably be expected to be "naturally" resolved to their true meaning.

In the specific case of [the] original query, for example, some lexical searches against the search expression 'accidental dural tap' would already match on the PT of the existing new concept 'accidental dural tap during anaesthesia' since it shares all three search tokens in common and even in the right order. So presumably the underlying clinical grumble isn't that they can't find a code to match the jargon term they want to search for, but rather that they must subsequently endure the longer but less ambiguous term whenever they review the EHR. In an interoperating world where stuff is going to move around and much more widely, IMHO clinicians may well have to learn to live with a new obligation to be somewhat more precise in their naming of things.

Jim Case <jca@snomed.org> Oct 9, 2019, 6:57 AM

We are in total violent agreement on this. I have often "complained" about calling the additional descriptions for concepts synonyms, when often they were not. I agree that attempting to constrain descriptions on concepts in the core to true synonyms is an admiral, but possibly unattainable goal; however the provision of technical mechanisms for users to add near synonyms in their context of use is also very valuable. Developing guidance for the "proper" way to do this and a listing of the potential and real pitfalls of doing so might be of equal or greater benefit. I see a policy/position paper coming out of this along with guidance on how to do it properly. It dovetails well with the existing policy on patient-friendly terms.

With regards to returning clinical utterances back to clinicians, companies such as IMO [are] doing just that. The focus of SNOMED CT, IMHO, is to provide a set of synonyms that facilitates the identification of the proper concept, without necessarily providing the preferred colloquialism, idiom or dialect variant.

Guillermo Reynoso Oct 9, 2019, 8:55 AM

The degree of synonymy/equivalence includes non-synonymy as a valid value, used in the editing environment to handle compromises at the International Edition English level (allowing it

to keep non-synonyms or near-synonyms but flagging them). Given the tooling adaptations that might be required, this might still be an option for the core (as the two other options described below are more radical, but that is a technical aspect, what we need is to get consensus on the appropriate way and guidance.

Our experience with the Netherlands patient terms is that they are better handled as an extension to the main NL extension, in a separate module. There are some popular terms that you cannot just add as synonyms in the main content, and they are usually ambiguous and bound to many similar concepts (e.g. think the current abdomen-abdominopelvic discussion thread vs. belly). The patient terms extension is based on the NL extension, adding descriptions, changing acceptabilities, or excluding some medical terms.

An add-on extension of use case-specific terms and preferences is more robust than a degree of synonymy refset, and avoids including controversial "synonyms" in the main content. However, maintaining more than one extension at the same time is a significant cost, and might work better when the use case is quite different from the main one (e.g. patient terms) rather than when we are trying to improve retrieval for our usual health professional users.

Something that might be worth considering would be to add another description type (entry term, index term, etc.) that would not imply synonymy at all. This is supported in RF2 (there is a description refset descriptor that lists the allowable description types, currently FSN, synonym and definition), but some implementations might have issues if they hardcoded the current usage.

My current preference would be to add a description type at the extension level, to manage just one extension (and simplifying dependencies), and leaving pure synonyms in the extensions without contamination. I think that is as far as SNOMED CT can go.

Resources like IMO (or other UI-oriented lexicons) usually collect recognizable strings that are language and care setting dependant, and map them to other terminologies like SNOMED CT with the compromise that the string ID will be preserved even if the meaning (the mapping to the reference terminology) might change as the target terminology evolves (e.g. inactivates the mapped concept and replaces it). Or a string iID (a descriptionID?) can map to a list of candidate concepts. That I think is the boundary we cannot cross. We cannot be a string-based terminology resource, but we can use RF2 mechanisms to provide clear indication of valid

synonyms and eventual use-case oriented associated descriptions (e.g. non synonyms, near synonyms, lexical or graphical variants, etc.)

My team is trying to recover at least an old version of the degree of synonymy content (not necessarily the most recent) to get some examples. We might need to reach out to the SI technical team for old environment backups that might contain it…

Jim Case <<u>jca@snomed.org</u>> Oct 9, 2019, 9:03 AM

Thanks everyone for their contribution to this. I think there are a couple of items that need clarification and list of appropriate/viable options to consider.

1. Do we want to reintroduce the notion of degree of synonymy to SNOMED CT?
2. What is the scope of "allowable" degree of synonymy?
3. What types of synonymy should be supported in the "core" (by this we mean the clinical core and international edition content)?
4. What are the mechanisms, technical architectures that can be used to manage degrees of synonymy?
5. How should degree of synonymy (including non-synonymy) be handled by extensions?

It would be helpful to try and flesh out some of the options under each of these topics in advance of the discussion in KL. If there are other heading to consider, please add them as well.