Paper CM02

A Dive Into the World of Biomedical Concepts

Berber Snoeijer, ClinLine, Leiderdorp, The Netherlands

ABSTRACT

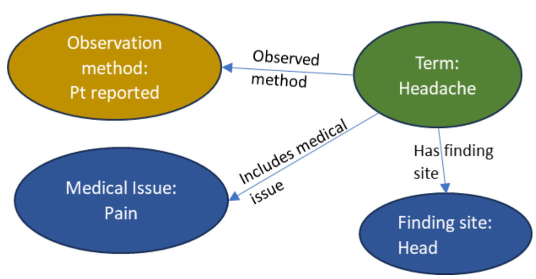
What are biomedical concepts and how can we practically use them? This paper describes the aspects and use of biomedical concepts from various dictionaries. This includes CDISC COSMoS and it's representation in CDISC USDM. One of the difficulties is the linkage (or mapping) of biomedical concepts from one dictionary to another. This can be handled by a combination of automation for exact matches and manual one-off review of non-exact matches using automated documentation and verification in these cases.

INTRODUCTION

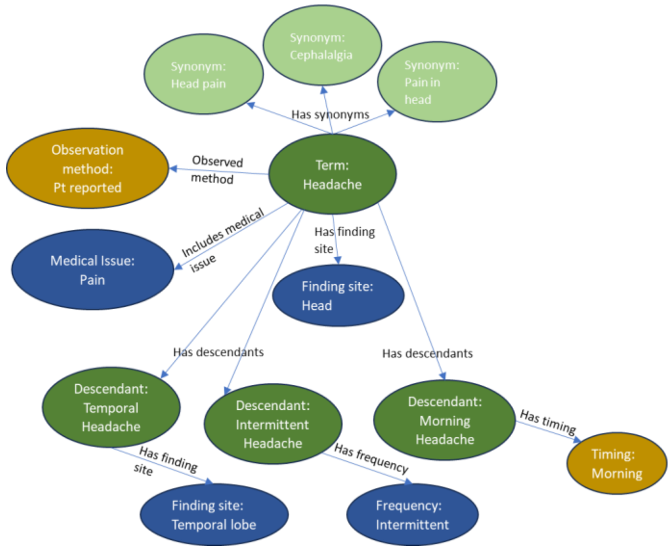
Biomedical concepts can be used in our own clinical trial practices and to get access to data from external real-world data sources. If well-defined, they include more specific details informing and enabling set-up of information collection systems like electronic data capture systems. In addition, they give detailed information on analysis level to enable automated identification of issues on different aggregation levels. However, the capture of biomedical concepts is not uniform and the principle might therefore be confusing. Moreover, mapping from one biomedical concepts dictionary to another is complex if not well understood and not well-documented.

WHAT ARE BIOMEDICAL CONCEPTS?

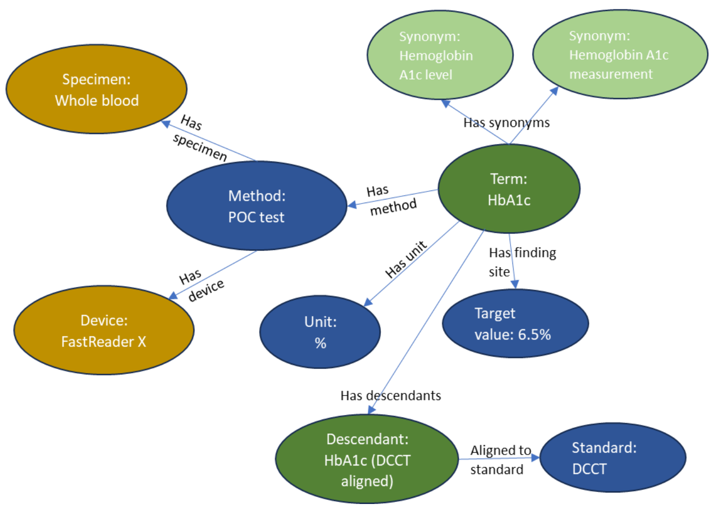
Biomedical Concepts are observations or measurements that can be obtained from a healthy volunteer or patient. These can be for example observations on the condition of a patient or a biological assessment such as a lab measurement. The observations can come both from regular clinical practice as well as from clinical trials. When we look closely to an observation or measurement, we can distinguish several distinct characteristics. For example, a simple condition like headache already includes information on the location (head) and the condition (pain). In addition, the biomedical concept could also include how the information was obtained. For example self-reported by the patient, based on a questionnaire or based on medical examination. See the example in the picture below.



A biomedical concept may have synonyms. Synonyms have exactly the same characteristics as the original term. For headache we have for example synonyms like "Head pain", "Cephalgia" and "Pain head". A biomedical concept can also have descendants (children) and ancestors (parents). Descendants inherit all the characteristics of the parent and, in addition, include some characteristics that were not specified for the parent concept. Contrary, ancestors include only a subset of the characteristics of their children and do not have any other additional characteristics.
Below you see some descendants of Headache like "Morning headache" which includes additional timing information, "Intermittent headache" which includes frequency information and "Temporal headache" which specifies more detailed finding site information. Note that in the latter case the child BC has a more specific description of the finding site information that was already specified for the parent. So in this case also the characteristic itself is a child of the main finding site characteristic.



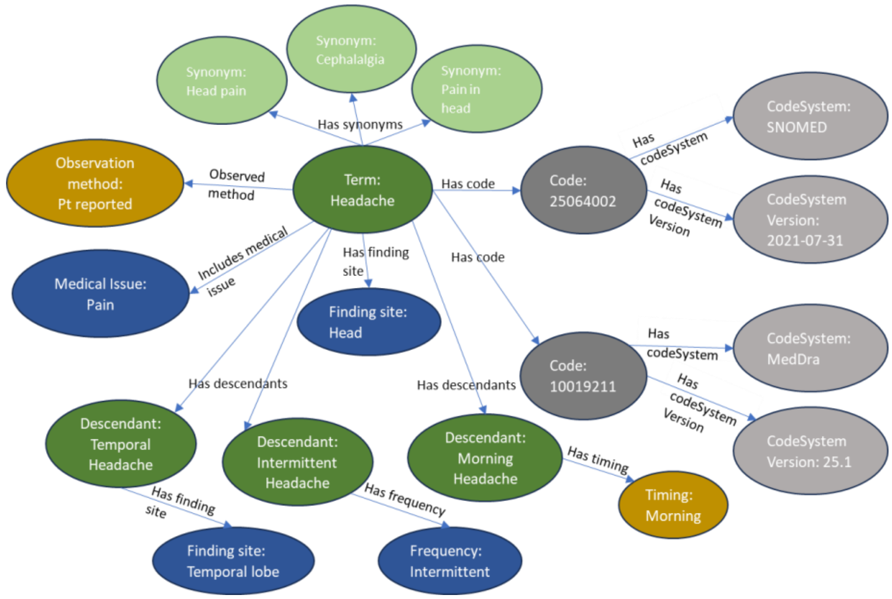
The same principle can be applied to a lab measurement as you can see in the example for HbA1c below.



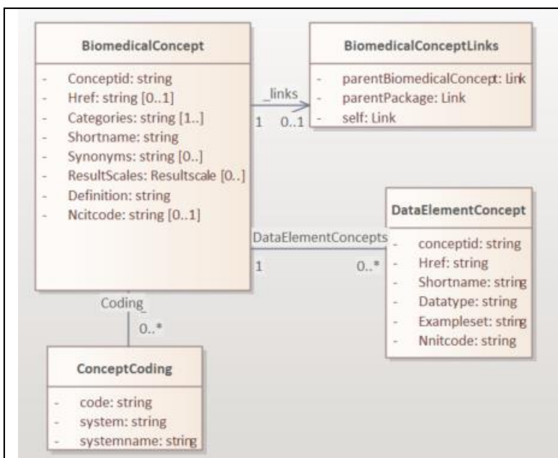
HOW ARE BIOMEDICAL CONCEPTS REPRESENTED IN DICTIONARIES?

All Biomedical concepts that are represented in a dictionary have a unique identifying code in that dictionary. So the example above for HbA1c has a unique code in SNOMED as well as in LOINC. The ancestors and the descendants have their own unique code.

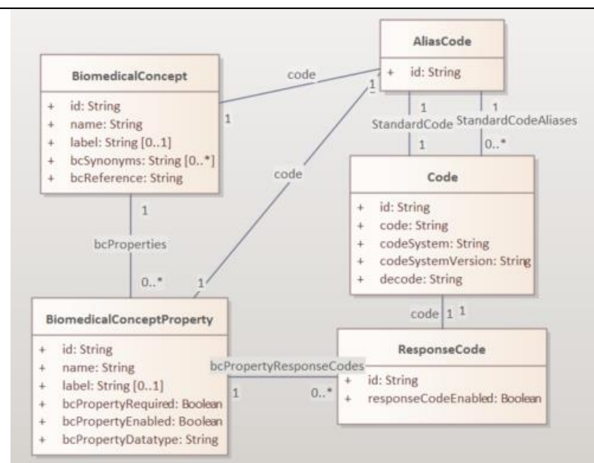
Dictionaries keep evolving. Therefore, for traceability reasons, it is important to note the version of the code system that is used. Some codes are newly introduced in a newer version while other codes are deprecated. See the example below for headache using different coding systems and noting the system versions.



There is no uniform way to store biomedical concepts and their relationships in dictionaries. From a technical perspective, these can be relational(-like) databases, document based databases or graph databases. From a structural perspective, they differ significantly.
CDISC Biomedical concepts are defined in their COSMoS project which “kicked off in 2022, taking a pragmatic, iterative approach to creating biomedical concepts and representing them in the Foundational Standards as dataset specializations with Value Level Metadata definitions” (See www.cdisc.org/cdisc-biomedical-concepts). In addition, “CDISC is collaborating with TransCelerate as a part of TransCelerate’s Digital Data Flow Project to develop a Study Definition Reference Architecture that will serve as a standard model for the development of conformant study definition technologies” (See www.cdisc.org/ddf). The resulting Unified Study Definition Model (USDM) includes the same biomedical concept principles as in COSMoS.
COSMoS is designed as the dictionary storing the CDISC controlled terminology for Biomedical Concepts, while the USDM is dedicated to capture study design information. The USDM is therefore slightly different in the set-up and naming of classes. The USDM allows to make a distinction between the main code which will often be the CDISC coding and for linkage to the corresponding code in other coding systems via the AliasCode class. See figure below.
Both representations of biomedical concepts include the option to define the potential responses or results of the corresponding measurement (See ResultScales in COSMoS and the ResponseCode class in USDM) which is needed for properly informing the data capture systems.

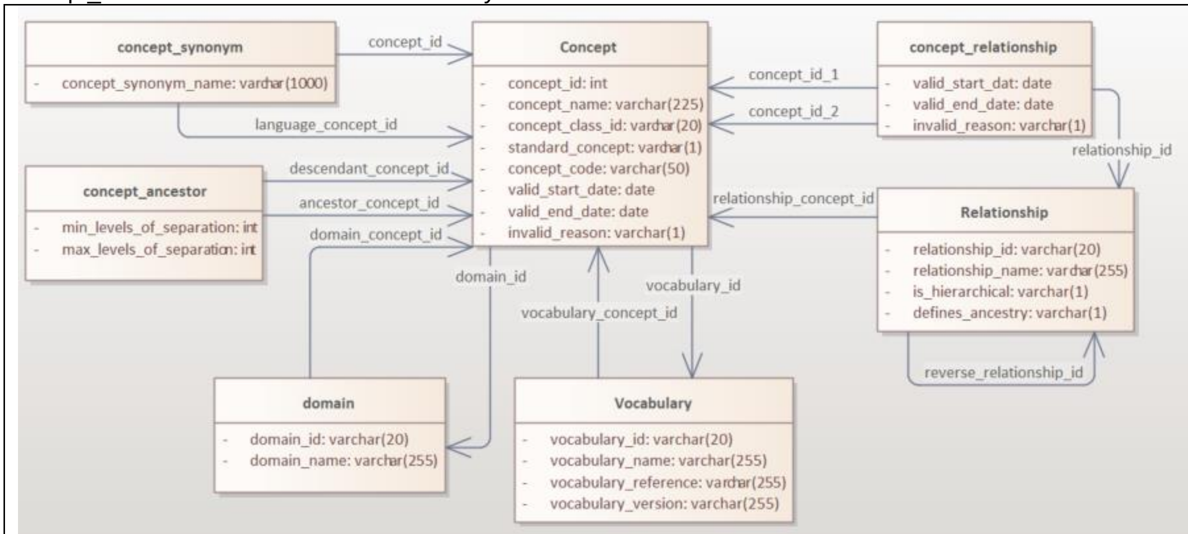


Representation based on CDISC COSMoS v2.0 API specifications.



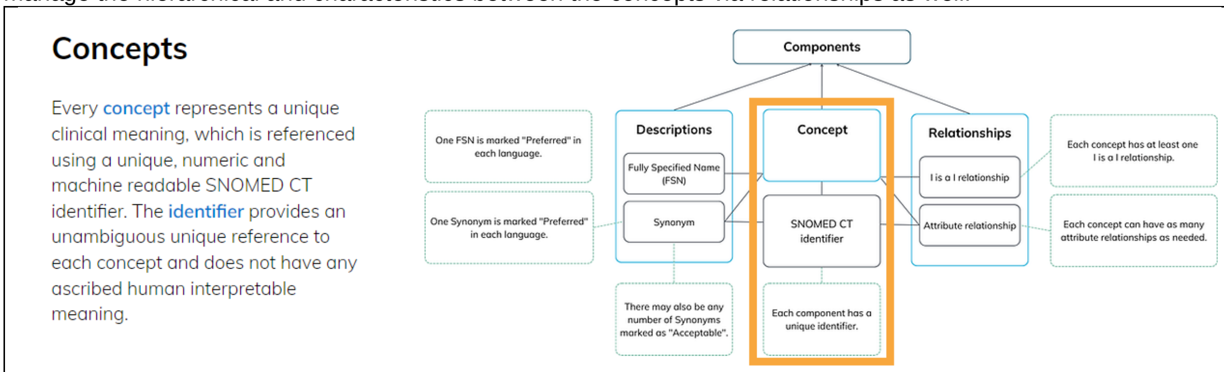
Representation based on CDISC USDM v2.0 API specifications.

The OMOP biomedical concepts are stored in their Athena dictionary (see https://athena.ohdsi.org). The picture below shows the logic of these biomedical concepts as represented in OMOP. OMOP does not have separate classes for properties or responses but defines that as concepts as well. The corresponding relationship then indicates whether it is a property, response, ancestry or mapping of the main concept. All relationships are defined in both ways to facilitate easy search in both directions. OMOP includes concepts of many dictionaries including SNOMED, LOINC and MedDRA (the latter only if you have the proper license). These are for the greater part mapped to each other via the mapping relationships. Although already defined in the concept_relationship table, synonyms and ancestors or descendants are also represented in the additional concept_synonym and concept_ancestor tables which facilitates easy searches and selections.



Athena Biomedical concept relevant vocabulary classes

The SNOMED concepts are differently structured. They do not have a separate property class but include the characteristics in the main concepts when defined (See www.SNOMED.org/five-step-brieving). However, they manage the hierarchical and characteristics between the concepts via relationships as well.



www.SNOMED.org/five-step-brieving

LOINC is more focused on measurements and has a more hierarchical structure that relates to that like analyte/component name, sample type, method type and result scale.

HOW CAN WE MAP BIOMEDICAL CONCEPTS FROM ONE DICTIONARY TO ANOTHER?

As you can see from the examples above the structure of the biomedical concept libraries varies amongst sources. So to start with, firstly the mapping of classes and attributes from one library to another should take place. To ensure traceability and interoperability, the corresponding mapping information should be well-documented and computer-readable.

Even more challenging is the actual mapping of the concepts. Although there will be an overlap in terminology between dictionaries (every dictionary will include headache as a concept), there will be also a great part that cannot be exactly matched. In those cases it is key to find the closest match. This can be an ancestor mapping, an alias mapping or a match validated by a clinical expert. Using a more specific concept requires information to be added that was not reported in the first place which is of course not feasible. Therefore, except for the alias mapping, these kind of matches will in most cases imply that a less specific concept is used and thus that information is lost in the

mapping process. Therefore, it is crucial to keep the original coding and information alongside with the mapped information both to show the provenance as well as to be able to refer to in case of analysis or safety issues. In the current OMOP version this is very well handled by including variables for source values and corresponding concept ids.
The mapping process can be automated by covering the exact matches directly including the aliases and then the automatic identification of non-exact matches. For the non-exact matches, suggestions for ancestors and other mappings can be automated which then can be handled manually in a system which records the mapping decisions and authorization of these.

CONCLUSION

The structure and focus of different dictionaries for biomedical concepts is not matching. Standard mappings like those facilitated by the OMOP Athena dictionary will help in this process. However, the remaining non-exact matches should be closely reviewed and documented. Automation will help in this process to ensure provenance, traceability and interoperability.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Berber Snoeijer
ClinLine
Van der Valk Boumanweg 15
2352 JA Leiderdorp
Work Phone: +31 6 21197621
Email: b.snoeijer@clinline.eu
Web: www.clinline.eu