PhUSE 2016

Paper CD04

Assessing CDISC Therapeutic Area User Guides in a machine readable format

Johannes Ulander, S-Cubed, Copenhagen, Denmark
Niels Both, S-Cubed, Copenhagen, Denmark

ABSTRACT

In 2015 we did a comparison of all the Therapeutic Area User Guides available at the time, where we analysed the consistency of mapping disease diagnosis and symptoms to the SDTM Medical History domain. The process to compare was manual and time consuming and the analysis illustrated that four significantly different approaches had been used to model medical history. This presentation will take the next logical step in this project and suggest a biomedical concept to provide a consistent approach for mapping disease diagnosis across the multiple Therapeutic Areas. We will share the analysis on whether that biomedical concept can work across the multiple Therapeutic Area User Guides and assess the benefits of expressing the standards in a machine readable format. At the same time illustrate how this can improve the deployment of standards within the industry.

INTRODUCTION

In 2015 we did a comparison of all the Therapeutic Area User Guides available at the time and found that several approaches had been used for modelling information related to a subject's Medical History. All approaches were in line with the SDTM Implementation Guide, however, it becomes a challenge for a sponsor to maintain a simple mapping route from the internal data collections standards to the SDTM model, as the different implementations require different relationships between collected data points.

What we will try to do in this paper is explore and discuss the possibilities of using a Metadata Repository as the foundation for developing more precise standards. When we talk about machine readable format, we are trying to make use of the CDISC standards that are released via CDISC SHARE in XML or semantic web formats (like RDF). At the same time the tools that are used should be able to utilise these formats, in a way that the underlying complexity of the relationships is not directly visible for an ordinary user. That will allow the user to focus on the task of building standards that are fit for purpose.

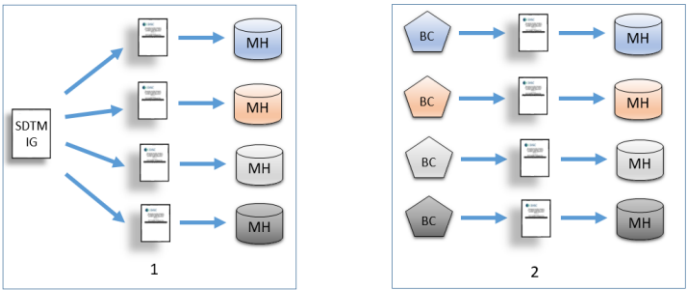
It is our experience with mapping data points, by annotating CRFs, that there can be situations, where several mappings are valid according to the rules of SDTM. Oftentimes the choice of which exact mapping to choose, is down to personal preference and experiences of the mapper. We've all been there when doing SDTM annotations, when you've done an annotation and when you or someone else revisit it later, it turns out that it could be done slightly different. It is a subjective decision dependent on the mapper and the mapper's experience. It is therefore impossible to be 100% consistent with mapping across studies and the result is that there can be different flavours of SDTM. Utilising an MDR should make mappings far more consistent,

When doing this assessment, we are doing it as subject matter experts in CDISC standards and clinical data standards, and hence our approach is from a technical view, rather than medical.

THOUGHTS ON BIOMEDICAL CONCEPTS

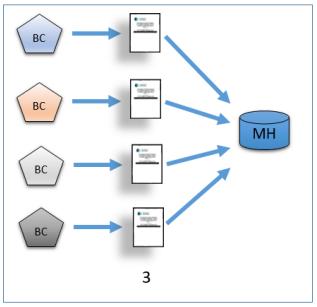
During the recent years, development of clinical data standards has been moving towards being based on the idea of Biomedical Concepts. There are many opinions on about how to use and describe Biomedical Concepts and below we try to explain how we believe that they should work in connection with the development of TAUGs.

Figure 1 shows the current setup, where each TAUG is individually derived from the SDTMIG. The result of this is that you have different flavours of SDTM MH domains. Creating a Biomedical Concept for each TAUG is not desirable if it would still produce the same variability (figure 2.)

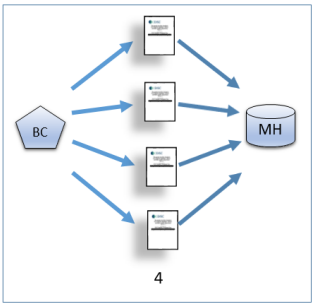


Different colours represent flavours of standards

An acceptable approach would be having a Biomedical Concept for each TAUG, if it would generate the same output in SDTM. (Figure 3) However, it should be feasible to create one Biomedical Concept that would ensure that we have no variability across the Therapeutic Areas.



3

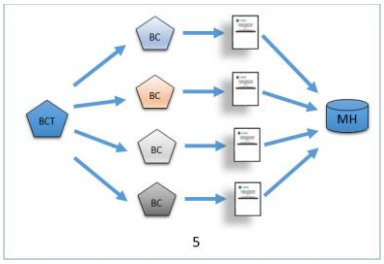


4

MOVING FROM THEORY INTO PRACTICE

The above describes the high level idea of how the Biomedical Concept assures the alignment of implementation in the TAUGs. So what we want is one Biomedical Concept that will make sure that the outputted SDTM MH domains are aligned across Therapeutic Areas. (Figure 4)

From a tool perspective, the implementation should become a merge of Figure 3 and 4, as you will need something to identify each of the implemented TAUGs in your tool. You would still have a master Biomedical Concept that works as the glue to make sure that the implementations of each Therapeutic Guide are still aligned. This means that you create a Biomedical Concept Template as the master Biomedical Concept. (Figure 5). This is very similar to how the SDTM model works together with the SDTM Implementation Guide.



5

To create this Biomedical Concept Template for Medical History, we must be prepared for what we've found in the TAUGs, which is a complex set of information, such as diagnosis, study indication, type of disease, family history, whether any conditions are acute or chronic and what are only symptoms (which we discussed in last year's paper. See reference 1.)

We started out simple, by adding the usual information from the SDTM MH domain, which should be needed for describing a disease.

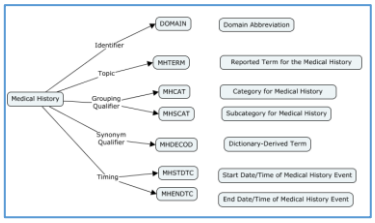


Figure 6. Visualisation of the biomedical concept template based on SDTM MH

By doing that, you start wondering what you should do with all the other information, such as diagnosis date, whether it is a symptom or related to the subjects family history. This simple visualisation in figure 6 has many assumptions made and probably only make sense to someone who is familiar with SDTM. The full biomedical concept to describe all assumptions, definitions and details, will be impossible to visualise in one single picture.

To illustrate this complexity (figure 7), we made a simple graph in Neo4j expressing how the variables in the SDTMIG MH domain are inherited from the Identifiers, Timing and General Observation Classes of the SDTM model. So the figure excludes all general and specific assumptions from the SDTMIG (e.g. that the variables MHSTAT and MHREASND are related.)

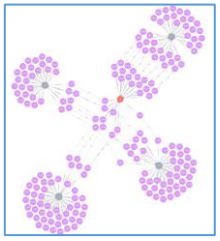


Figure 7. Visualising the SDTMIG MH Domain relationship with the SDTM Model in Neo4j

ADDING SDTM REPRESENTATION IN THE BIOMEDICAL CONCEPT

But what is important here, is that a Biomedical Concept should include the representation of a specific attribute in SDTM. So it includes an exact definition of how, e.g. date of diagnosis, should be stored in SDTM.

By including the SDTM representation in the Biomedical Concept it is possible to control the way medical history becomes represented in SDTM. And it is apparent that if one starts the mapping activity of a new TAUG by fitting it into your existing Biomedical Concept Template, it would inherit an existing representation from the Biomedical Concept Template. When a new version of SDTMIG becomes relevant, the mapping to SDTM in the Biomedical Concept Template should be controlled via versioning and all Biomedical Concepts derived from the same template can be updated at the same time. This also allows for an automated process of impact analysis.

STANDARDS MANAGEMENT AND IMPACT ANALYSIS

As an example of automated impact analysis we are using the change of test code for Blood Urea Nitrogen in the SDTM Controlled Terminology as an example, as there were no good examples for medical history. The submission value test code for Blood Urea Nitrogen was BUN in version 2015-12-18, and was changed to UREAN in the 2016-03-25 version. We got help from Dave Iberson-Hurst to use the Glandon MDR tool to assess the impact (see reference 2 and 3).

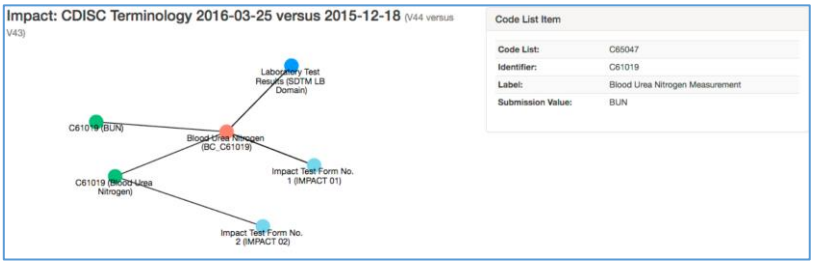


Figure 8. Impact analysis from Glandon (see reference 2)

What is illustrated by this, is that by having version controlled Biomedical Concepts in an MDR, it is possible to automate the process of impact analysis.

We will still need thorough analysis to understand whether something we are trying to add to a concept is a synonym of something we already have created, or if it is truly something new, but by having automated impact analysis it will help in controlling the standards.

We are dependent on tools that can help us in keeping track of all this information, so we don't get lost.

USING SOFTWARE TO WORK WITH BIOMEDICAL CONCEPTS

One of the biggest features of the SDTM model is its ability to streamline data, so that data that is collected in similar ways can get a common expression in SDTM, e.g.:

- using a specific variable target for well-defined information (such as AGE or --LOC)
- the use of relative timing variables (--ENRF, --STRTPT etc.)
- having open variables that can be used to create traceability between the CRF and SDTM datasets (e.g. --SPID and --CAT.)

As we will need to specify our concept to a greater detail and the current SDTM model is open for interpretation, we are not able to streamline the information into SDTM. We need to use a direct way of representing each unique part of the concept.

So what you need is something that details the minimum of needed information, just like observations work in SDTM, that can expand when necessary, which is possible in e.g. XML format, but where it is possible to specify exactly what it is that you are adding. But as SDTM is a 2 dimensional standard (rows and columns) and needs relationships explained in a separate data structure, it would be possible to use e.g. a URI to expand an observation and express a direct relationship to another observation.

CONCLUSION

During our work to try and analyse the TAUGs with regards to MH and find a way to represent them in a MDR friendly structure, we have made a number of observations, which we think can be relevant to be aware of in future TAUG development as well as for development of SDTM model.

We believe that it is necessary to develop a Biomedical Concept Template, which is used to derive Biomedical Concepts from when working with MH for the different TAUGs. Such a template should include how data points shall be represented in SDTM. The biomedical concepts need to be controlled in a metadata repository, thus allowing for a systematic version control and impact analysis.

REFERENCES

1. Phuse Conference 2015. CD03: Therapeutic Area Standards and their Impact on Current SDTM Implementations
2. Blog post, Dave Iberson-Hurst: Light, A Tunnel and another Grand Départ http://www.assero.co.uk/2016/light-a-tunnel-and-another-grand-depart/
3. Glandon http://www.assero.co.uk/glandon/

ACKNOWLEDGMENTS

The authors would like to thank Dave Iberson-Hurst, Assero, for giving us the opportunity to test our thoughts in practice in the Glandon Metadata Repository tool.

RECOMMENDED READING (NOT REQUIRED)

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Johannes Ulander/Niels Both
S-Cubed ApS
Lille Strandstraede 20C, 5
Copenhagen / 1254
Denmark
Web: http://www.s-cubed.dk