Questions and Observations from the Organizers Regarding CDISC

This document follows a brief exploration of the CDISC website and an exploration of the standards which seemed relevant to data description. CDISC produces a large number of specifications, and we wanted to be clear that we were looking at the right ones for the purposes of this workshop.

Naturally, this has raised some questions. First, this document attempts to summarize our understanding of how CDISC is organized, and then we will list out the questions which occurred to us.

# Summary

Several of the CDISC specifications seemed relevant to a discussion of how CDISC data is organized. The following picture was found to be quite useful.



All of the standards are based on/implementations of the CDISC ODM (Operational Data Model) XML. Of these, we have a “Dataset XML” used to encode data, which is structured according to the “Define XML”. Of these, there seemed to be several different types of data which would be of interest, all of which could be encoded using Dataset XML. The SDTM (Study Data Tabulation Model) is clearly of relevance. The SEND messages (non-clinical data) is also likely to be of interest, as these seem to be used, based on the following quote:

“The SENDIG is intended to guide the organization, structure, and format of standard non-clinical tabulation datasets for interchange between organizations such as sponsors and CROs and for submission to regulatory authorities. The SENDIG is based on and should be used in close concert with CDISC Study Data Tabulation Model (SDTM) v1.5, which is included in the document package.”

 Also, the ADaM (Analysis Data Model) specification also seems to be of interest. Another quote:

“ADaM defines dataset and metadata standards that support:

* efficient generation, replication, and review of clinical trial statistical analyses, and
* traceability between analysis results, analysis data, and data represented in the Study Data Tabulation Model (SDTM).”

Another CDISC work product which looks also to be of relevance is SHARE, a registry of biomedical concepts equipped with tools for mapping concepts across domains.

Some of the semantic specifications also appear to potentially be of interest: the glossary and controlled terminology.

# Questions

1. What are we missing? Are there significant places where CDISC describes data which are not covered in the specifications mentioned?
2. We did not find any explicit process model – at least not a generic one. How does CDISC describe business processes? We did see standards addressing the design of experimental protocols, but are there also messages describing the historical conduct of these? What are we missing here, if anything?
3. Which of the data sets identified are most commonly used? Which would be relevant for those wishing to reuse data encoded and modelled using CDISC?