PhUSE 2016

Paper CD17

Understanding the potential of Share: a visual business case and technical use case

Author: Peter Van Reusel
Business & Decision Life Sciences, Brussels, Belgium

ABSTRACT

Over the years, our industry has embraced CDISC data standards. There is a continuous growing consciousness that these data standards hold a promise of efficiency and automation, but we are not seeing many successful examples just yet.

The rate at which the data standards are being updated, released and implemented has significantly increased to a point where we are moving away from the use of spreadsheets for this metadata management towards the use of Metadata Repositories (MDRs). The current MDRs are often seen as clunky, complex and only accessible to a small number of data standards experts. The learning curve is too steep for end users that do not have a deep understanding of the metadata, systems or requirements.

However, by configuring the advanced metadata models in Share, there is an enormous potential to make the existing data standards much more valuable. On the one hand for the scientific end-users and on the other hand for the data manager and statistical programmers which will more focus on value added tasks rather than repetitive work.

The presentation will demonstrate how visualizing this metadata can help clinicians and statistician understand the value of metadata. It is these scientists that will determine what clinical assessment data we will collect and how this data is used for analysis. Their input will be used by next generation MDRs to create true metadata driven processes.

The second part of the presentation will simulate how potential tools can leverage the selection of the clinician and statistician by using interoperable metadata to automate substantial parts of the clinical data business process.

Finally, we will discuss what additional metadata is needed to enable such an automation by showing a gap analysis between current available metadata and metadata which has not been standardized yet.

INTRODUCTION

Sponsor Clinicians and Statisticians are responsible for defining the design and purpose of clinical studies. They provide input to the operational strategy and feasibility of clinical research studies, participate in defining the key components of the clinical protocols, assesses the performance of techniques used for endpoint measures just to name a few key activities.

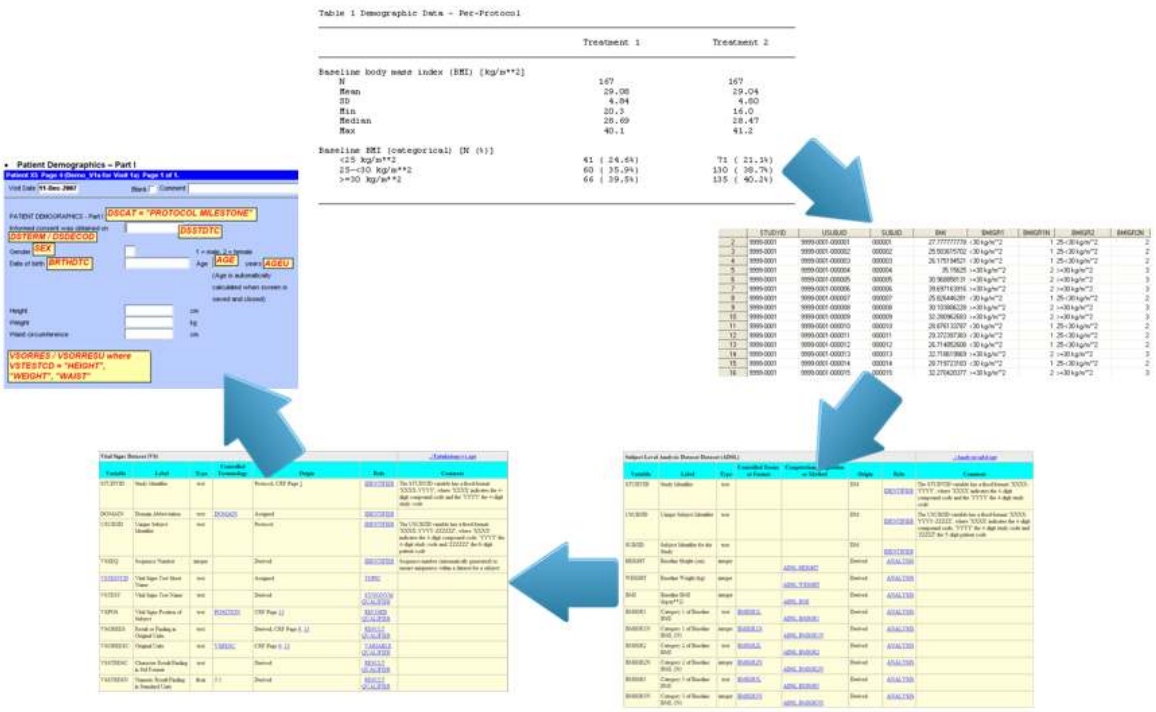
Even-though they have a responsibility to initiate and plan clinical studies, we should not assume they are data standards experts. In reality, it are the data standards experts, data managers and statistical programmers which will use data standards to ensure regulatory requirements and an efficient cost-effective process.

CDISC data standards are evolving over time. More and more data model versions and Therapeutic Area User Guides (TAUGs) are being released at a very fast pace. Companies need to organize themselves to keep up to date with evolving data standards and specific regulatory requirements, specifically the US FDA and Japanese PMDA.

CURRENT STATE OF DATA STANDARDS USAGE

The industry has come a long way over the last decade in the understanding and implementation of industry wide CDISC Data Standards. About 85% of current FDA submissions are based on one or more versions of SDTM and the adoption of ADaM is one the rise as well. Because FDA and PMDA are endorsing the CDISC SDTM and ADaM standards as mandatory requirements, it is expecting that 100% of FDA CDER and PMDA submissions will be CDISC compliant.

With the extensive use of standard metadata, the submitted study data is becoming self-explanatory. Regulatory reviewers and other clinical data stakeholders can trace the study results back to the origin of the data by using the analysis result metadata, analysis metadata, tabulation metadata and annotated CRFs.



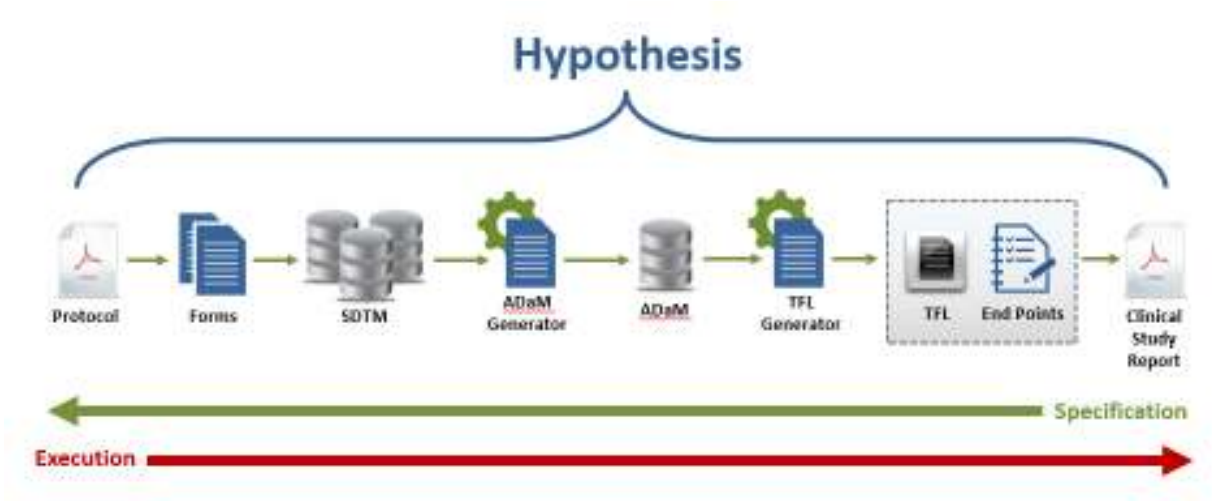
This transparency would not be possible a decade ago, certainly not on such an industry wide base.

However, there is also a consensus that the implementation and use of these data standards have proven to be very costly. Companies are spending massive resources to convert legacy data for submissions and to update their systems, processes and people to use the standard data. There was an expectation that the industry is not only using data standards for regulatory compliance, but also to gain efficiency in re-use of data, increasing quality and improved cross-functional communication. The current data standards did not deliver on their promise of a more cost effective process. Sponsors are using expensive and clunk metadata repositories (MDRs) to store and handle their data standards.

The currently available commercial MDRs require a high level of data standards expertise and persistence of end-users. MDR users are bogged down in their daily work by administering details such as overlapping and inconsistent metadata, version control and more than often they are focusing on small snippets of individual metadata rather than metadata on a conceptual meaningful level. The current CDISC standards are organized in tables of metadata which are not conceptually linked with each other. For example, domain metadata is not related with variable metadata is not related with codelist metadata.

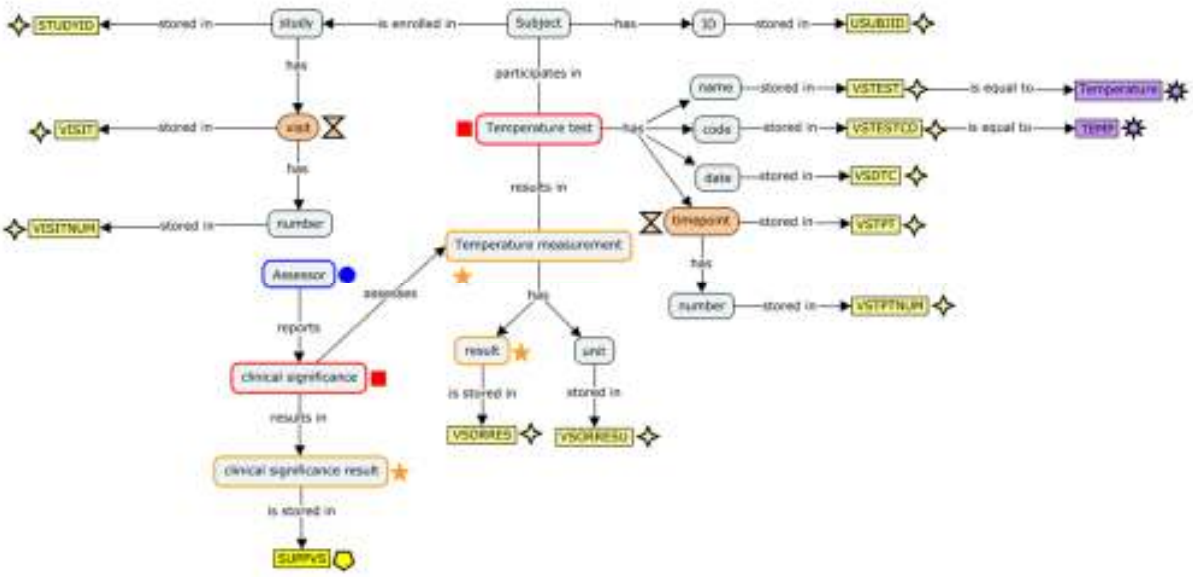
SPECIFYING BY BROWSING

In a more ideal world, there would exist user friendly systems that would allow non data standards experts to generate study specification, without needing to understand the underlying metadata models. A statistical end-user would design and specify the study statistical endpoints without being bothered by the underlying analysis results metadata and ADaM dataset structures. A clinical end-user would choose the data points to collect from an extensive set of existing data collection modules or forms. The underlying system should be able to construct the underlying metadata models simply by understanding how these endpoints or data collection modules are semantically connected to their individual tables, variables, values and code lists.



In turn, this concept specification should automate the generation of all or large parts of the clinical protocol, the EDC system, the SDTM database, the ADaM datasets and the Analysis Results Metadata.

This goal will be impossible to achieve without providing conceptual meaning to the metadata. CDISC and other standards organization have created concept maps that semantically link clinical data not through tabular metadata but through modelling diagrams. These modelling diagrams create relationships between different elements. An example of such a concept map is seen below.



Such a concept map can be stored, represented and used in a database and can be used by programs to generate metadata definitions. In the example above, a program can generate the table, variable and code list metadata for a parameter like temperature test by reading the databased UML model. These relationship are flexible, defined and dynamic. In the above example, it is perfectly possible to create a relation between the parameters and its electronic mapping rule to the SDTM or ADaM model.

The CDISC Share repository is capable of storing any metadata in this way. Currently CDISC has uploaded large part of the existing data models in the Share system and the existing CDISC metadata is already available electronically. However, the conceptual metadata to link the data models to a meaning is not available because these links have not been created by the industry today. There is a very substantial effort needed on an industry level to actually create these definitions.

CONCLUSION

Before we can reach the ultimate goal of automatically generating substantial parts of the clinical data process, the industry data standards needs to evolve and be enriched by semantical information and links. Only with this sematic metadata will future systems be able to generate code and configure systems that collect data, clean data, represent data and finally analyse data.

An effort is needed on an industrial level and the industry will need to create a precompetitive cooperation framework to achieve these goals.

CONTACT INFORMATION

Contact the author at:
Peter Van Reusel
Business & Decision Life Sciences
St Lambertusstraat 141 Rue Saint-Lambert
1200 Brussel – Bruxelles
Mobile: +32 476 545917
Email: peter.van.reusel@innovion.be
Web: www.businessdecision-lifesciences.com