PhUSE US Connect 2019

Paper DS06

Evolving CDISC to the Next Decades: The CDISC Proof of Concept

Peter Van Reusel, CDISC, Austin, USA
Sam Hume, CDISC, Austin, USA

ABSTRACT

Over the past 20 years, CDISC and the CDISC community have established a globally accepted standard for clinical studies. The CDISC standards have become the de-facto standard between clinical trial sponsors, services providers and regulatory agencies.
These standards have proven their value and use over time, but it is generally acknowledged that the implementation of data standards by various companies did not deliver the return of investment that was initially expected. The current CDISC foundational standards are normative; they describe the structure and business rules of the datasets and metadata. The actual data concepts however are not standardized in a meaningful way.
CDISC and the CDISC Community is embarking on an exciting journey to explore how the biomedical concepts can be represented as linked informative standards which also describe the meaning between concepts and the current foundational standards. This presentation will discuss the scope and the approach of the CDISC 360 project.

INTRODUCTION

The CDISC foundational standards define research data and metadata structures, but writing these standards as documents has yielded more text than metadata. Gaps in standards metadata limit automation opportunities. The inherent flexibility provided by the standards supports a broad range of implementations, but that flexibility also allows for inconsistencies that make scaling automation difficult. The lack of a conceptual foundation for the standards further contributes to these inconsistencies. The relationships that would be expressed by these concepts remain largely implicit in the current versions of the standards.

The CDISC 360 Project seeks to implement standards as linked metadata with a conceptual foundation providing the additional semantics needed to support metadata driven-automation across the end-to-end clinical research data lifecycle. New software tools will consume this new metadata to ease standards implementations while increasing data processing efficiencies.

CDISC 360 will demonstrate that the feasibility of standards-based metadata-driven automation as a start towards realizing the primary benefits expected of the CDISC standards: substantially improved efficiency, consistency, and re-usability across the clinical research data lifecycle. These benefits drive the return on investment in the CDISC standards implementations expected by CDISC stakeholders.

CDISC 360 OBJECTIVES

CDISC 360 seeks to demonstrate how standards enable metadata-driven end-to-end automation. This Project will develop proof-of-concept extensions to the CDISC standards metadata as well as the accompanying software tools that automate the end-to-end application of the standards. This approach will be demonstrated via 3 specific use cases.

CDISC 360 SCOPE

The CDISC 360 Project implements end-to-end standards-based metadata-driven processing demonstrated by implementing portions of the CDISC Type 1 Diabetes TAUG.

USING THE CDISC TYPE 1 DIABETES TAUG SELECT THE FOLLOWING:

- 1 or 2 statistical end points
  - Analysis of Glycated Hemoglobin
  - Summary of Hypoglycemic episodes
- ~3-4 ADaM datasets
  - ADSL(Subject-Level Analysis Data (ADSL))
  - Hemoglobin A1C Analysis Dataset (HbA1c Analysis Dataset)
  - Hypoglycemic Episodes Analysis Dataset (Hypoglycemic Episodes Analysis Dataset)
  - Hypoglycemic Episodes Summary Dataset (Hypoglycemic Episodes Summary Dataset)
- ~7-8 SDTM datasets
  - DM (Demographics, so support standard variables in ADSL)
  - VS (Vital Signs, for height and weight in ADSL)
  - CM (Concomitant Medications, to support stratification by background treatment, and for treatments of hypoglycemic events)
  - LB (for Hemoglbin A1C data)
  - CE and FACE (for data on hypoglycemic events)
  - EX, ML (for data about meals and study treatments relative to hypoglycemic events)
  - Trial Design datasets (for arms, visit schedule, definition of hypoglycemic events as disease milestones)
- ~15 CDASH CRFs
  - CDASH CRFs needed to support SDTM datasets above. One CRF will support collection of data about hypoglycemic events that will be mapped to multiple SDTM domains.

METADATA / DATA PROCESSING:

- Use Case 1: Create end-to-start specification
  - Produce a standards-based, machine-readable specification
- Use Case 2: Generate start-to-end metadata
  - Use standards specification to generate study metadata artifacts
  - Demonstrate the ability to generate study metadata given a specification
- Use Case 3: Transform data start-to-end
  - Use machine-readable metadata to generate study data artifacts
  - Demonstrate the ability execute data transformations given the study metadata

EXPECTED CDISC 360 OUTCOMES

Apply the 80/20 rule to ensure the Project automates 80% of the end-to-end metadata and data processing needed to generate study artifacts suitable for a regulatory submission.

- Demonstrate standards metadata-driven automation
- Model and document the new standards metadata developed for CDISC 360
- Standards gap analysis
- Technology gap analysis
- Assess the feasibility of producing a solution that scales
- Roadmap to develop and implement the next generation of CDISC standards

CDISC 360 RISKS

Insufficient dedicated project staffing resources represents the most substantial risk to the CDISC 360 Project. The Project's scope and schedule demand the focused attention of a dedicated team, in addition to the partners and volunteers providing support to the project. Most CDISC staff carry a range of responsibilities, but project success demands dedicated attention from staff team members. Excessive scope or scope creep represents another substantial risk to the project.
Complexity represents a substantial project risk, especially in the analysis parts of the lifecycle where greater gaps exist in the standards metadata needed to drive automation. To mitigate this risk, use case 2 and use case 3 will be split up in 2 distinct processing cycles:

1. Start-to-end metadata and data automation from end points to CDASH to SDTM
2. Start-to-end metadata and data automation from SDTM to ADaM to TLFs

Accomplishing cycle 1, while falling short of the full project objectives, represents a major accomplishment with substantial benefits to the stakeholders. Cycle 2 has much higher associated risk. Breaking the deliverables into 2 cycles ensures that the project produces valuable results despite an incomplete or delayed conclusion to cycle 2. The project team includes a communication resource to mitigate against the risk of insufficient community buy-in due to a lack of transparency with the broader stakeholder community.

OUT OF SCOPE

- Graphical interface for automation features
- Complete implementation of new standard extensions in SHARE
- Complete validation or testing of automation components
- Full study implementation - the Project represents a subset of a study

CONCLUSION

CDISC is hoping to engage the clinical data standards community to widely collaborate on this important initiative. The expected project outcomes will provide the industry with insight into the next generation of clinical data standards. Aside from the technology and standards gaps, this project is also expected to introduce a way to think about and work with standards that is unfamiliar compared to current practices. This will require a shift in our mindsets.

CONTACT INFORMATION

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

Peter Van Reusel
CDISC
401 West 15th Street, Suite 800
Austin, TX 78701
Work Phone: +32 476545917
Email: pvanreusel@cdisc.org
Web: ww.cdisc.org