CDISC 360: Continuing the journey towards end to end automation

The opportunity

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 POC 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

Critical success factors

- Prove the feasibility of standards-based end-to-end automation
- Document the new standards metadata required to support end-to-end automation

High-level timeline

The following high-level timelines frame the key phases of the POC. The POC itself will use an agile methodology and timebox certain phases within the project. Timeboxing ensures the on-time completion of a specified project phase, but does not guarantee scope. Therefore, the scope of the project will be adjusted as each phase progresses to ensure the completion of the highest priority tasks. Should deliverables critical to the success of the project slip, a review of the overall timelines will be conducted by the project steering committee.

#	Stage	Start	End
1	Initiation, scoping, and staffing	Nov 2018	Mar 2019

2	Planning, recruiting external participants, funding	Mar 2019	Jun 2019
3	Create end-to-start specification	Jun 2019	Oct 2019
4	Generate start-to-end metadata	Oct 2019	Feb 2020
5	Transform data start-to-end	Feb 2020	Jun 2020
6	Analyze results and publish POC report	Jun 2020	Sep 2020

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 POC represents a subset of a study

Expected CDISC 360 outcomes

Apply the 80/20 rule to ensure the POC 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 the POC
- Standards gap analysis
- Technology gap analysis
- Assess the feasibility of producing a solution that scales
- Roadmap to CDISC 2.0 implementation