thoughtsphere

DATA TRANSFORMATION SDTM

Background

Clinical Data Interchange Standards Consortium (CDISC) is a non-profit organization that develops and maintains standard models to support the exchange of clinical research data. CDISC's Standard Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) are the preferred models for data submissions and are required by the FDA and other regulatory authorities across the globe. **Despite their wide-spread use within the industry, SDTM and ADaM datasets are difficult to produce due to the dynamic nature of study designs and therapeutic area distinctions.**

While many organizations leverage libraries and repositories to support the creation of these datasets, extensive SAS programming on a study-by-study basis is routinely required. Due to the large amount of work (and cost) involved, organizations often wait until late in a trial before starting conversion and analysis programming to avoid potential rework caused by changes to the protocol, Statistical Analysis Plan (SAP), and data sources.

Not having SDTM datasets available early in the study creates many challenges for project teams. It often leads to siloed data review processes and the creation of custom dashboards and analytics to oversee the trial. Moreover, if project teams create review outputs that do not follow SDTM and/or align with the SAP, they may misinterpret the data and miss key indicators and data anomalies.

Sponsors who outsource data management activities to CROs are also impacted by the lack of data standardization early in a trial. In the absence of on-demand standardized data transfers, sponsors often develop in-house data aggregation and transformations rules to standardize & substantiate source datasets received from CROs.

Complexities of SDTM data transformation & submission:

- The re-use of data transformation code is limited and often requires extensive customization on a study-by-study basis due to differences in metadata definitions.
- For variables that do not confirm to the SDTM model, SDTM extensions are manually created and managed to provide supplemental qualifiers for SDTM and ADaM datasets
- SAS development of SDTM and ADaM datasets require static exports of clinical data
- SAS Programmer(s) must manually revise code to adjust for protocol amendments and document all revisions accordingly
- Manual documentation workflows are required to track SAS programming revisions and approvals throughout the configuration process

> The workbench includes an advanced Smart Map engine that leverages Natural Language Processing (NLP) and AI/ML algorithms to automate the mapping process with little to no manual intervention required.

- In addition to the SDTM and ADaM data sets, additional documentation such as protocol versions and annotated CRFs must be manually compiled and included as part of submission package to regulatory agencies.

ThoughtSphere Solution

The days of waiting months if not years before having SDTM datasets are over. **Using ClinHUB, our patented data ingestion engine, data mapping can begin as soon as the database build is finalized, and data source domains and plans are defined.** That's right, creating SDTM datasets can start well before FPI.

Embedded in the ClinHUB data fabric is our Clinical Data Mapping Workbench where aggregated source data is mapped to the target data model selected (SDTM, E2B, or custom). The workbench includes an advanced Smart Map engine that leverages Natural Language Processing (NLP) and AI/ML algorithms to automate the mapping process with little to no manual intervention required.

The mapping workbench provides an **intuitive user interface** to review the auto-mapped assignments and override the results as needed with a simple drag and drop operation. The Smart Map feature applies **Human-In-the-Loop** AI to continually learn from user action(s) and apply the learnings in future mapping iterations. To create custom derivations and SDTM extensions users can take advantage of pre-written code and functions available in the embedded Software Development Kit (SDK) or create their own code from scratch.

During the life of the study, ClinHUB's datalake automatically accommodates changes to source data and flags changes to end users so that the study configuration can be reviewed, updated, and promoted quickly. **Embedded user workflows streamline the approval and promotion of variable mappings and conversion rules to provide a comprehensive audit trail throughout the process.**

In addition to being able to export SDTM datasets on-demand, developers can create ADaM datasets in ThoughtSphere's MAP module. SDTM data is also used to generate a vast library of out-of-the-box analytics and dashboards to support Risk Based Quality Management (RBQM), Data Management, and medical reviews. This not only creates efficiencies for data reviewers during study start-up, but also improves data quality and reviewer consistency within and across studies.

When preparing for data deliveries, end-users can expediate the process using **the platform's SDTM Data Export Configuration feature which allows users to catalogue and collate required documentation for data deliverables.** Key highlights of the feature include:

- Data package is generated as an archive file that incorporates supporting requisite components
- Supplemental qualifiers are automatically generated for the extension columns setup in Mapping Type
- Automatic ISO Date format conversion of source data
- Auto generated Define.XML file providing metadata context
- Transpose data from wide/column structure to tall/record structure per SDTM requirements
- Annotate transformation code expressions with a plain-English text description
- CDISC SDTM controlled terminology is pre-configured in the platform
- Ability to schedule the data package generation so it is available for download