An Approach to CDISC SDTM Implementation for Clinical Trials Data

William T. Chen, Merck Research Laboratories, Rahway, NJ
Margaret M. Coughlin, Merck Research Laboratories, Rahway, NJ

ABSTRACT

The Clinical Data Interchange Standards Consortium (CDISC) is a non-profit international organization established to define standard data formats to support the exchange of clinical data among and within organizations such as pharmaceutical and medical device companies, contract research organizations, universities, medical research institutions, and regulatory agencies. The CDISC Study Data Tabulation Model (SDTM) defines the standard format for tabulation data. Generally, there are three approaches for implementing SDTM within the pharmaceutical industry: pure SDTM, submission-only, and database-only. The pure SDTM approach means all study data will be SDTM-compliant, starting from data capture and ending with data analysis and submission. The submission-only approach leaves the current practice alone and creates the SDTM-complaint datasets for submissions. The database-only approach feeds the study data into a clinical database, which is SDTM-compliant. This paper discusses the database-only approach to implementation of SDTM version 1.1 and SDTMIG (Implementation Guide) version 3.1.1 and explores the pros and cons of the current SDTM standard.

INTRODUCTION

Clinical Data Interchange Standards Consortium (CDISC) was formed in 2000 with the mission to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare. It first published the Study Data Tabulation Model (SDTM) version 1.0 in June, 2004, followed by SDTM Implementation Guide V3.1. The following year SDTM version 1.1 was published in April, 2005, followed by SDTM Implementation Guide version 3.1.1 in August, 2005.

Meanwhile, in July, 2004, FDA issued a draft guidance, which was finalized in November 2005, that specified a new submission structure based on the electronic Common Technical Document (eCTD) format originally defined by the International Council for Harmonization (ICH). SDTM was recommended as the preferred submission standard. In December, 2006, FDA published a proposal in the Federal Register that would mandate the use of SDTM for all Data Tabulation submissions with a two-year transition period. In August, 2007, FDA published the draft guidance Pharmacogenomic Data Submissions – Companion Guidance and SDTM was again recommended as the preferred submission standard.

To meet the expected requirement, many companies have started to adopt the SDTM standard. The development of the SDTM standard is time-consuming and complex. The implementation of such a standard is equally challenging, especially when there are still gray areas within the standard. This paper will discuss three general approaches to implementation of SDTM and our hands-on experience with one of them.

COMPARISON OF IMPLEMENTATION APPROACHES

There are three alternatives to implementing the SDTM standard in data collection that this paper will discuss. They are pure SDTM, submission-only, and database-only. In the pure SDTM approach, an organization builds its entire data collection system and the underlying database based on the SDTM standard. This alternative comprehensively embraces industry standards. The submission-only approach may be the most parsimonious in the application of industry standards in that only when the studies are to be submitted to regulatory authorities is the collected data mapped to the SDTM standard. Finally, the database-only approach is a hybrid that builds the study database based on the SDTM standard while the data collection system does not necessarily need to be SDTM-compatible.

Each of these approaches has its own pros and cons. The following table summarizes some of the pros and cons of each approach.

SDTM Implementation	Pro	Con
Pure SDTM approach (most comprehensive implementation of industry standards)	• Same terminology is used within the company and outside the company • Data is compatible across studies. It is easier to combine the data for submission and other analysis activities • Facilitates outsourcing which becomes easier and more transparent • Control and upgrade of the standards are centralized to one organization	• Need to develop the data collection system and database internally. This is time-consuming and expensive. • Need to reengineer processes and organizations in order to enable the new data collection system and database. • Harder to detect data issues with data residing in different domains, especially when REL-REC is involved. • Control and upgrade of the standards are enforced by one organization • May lose the company-specific submission advantage built over the years from company-centric submission design
Submission-only approach (most parsimonious implementation of industry standards)	• Current process and systems remain intact • No need to deal with the complexity of SDTM for majority of clinical personnel • Need to map only studies which are ready for submission. • May be cost-effective and efficient	• May delay the submission due to complexity of the SDTM mapping • Need to maintain two set of datasets and validated analysis programs. One is for submission and the other one is for internal use • Potential for miscommunication between the agencies and sponsors if some sponsor's staff unfamiliar with SDTM terminology.
Database-only approach (hybrid implementation of industry standards)	• Can choose any cost-effective system and process to collect study data • Supports standardization of medical terminology to industry standards • Facilitates combining data from multiple sources • Control and upgrade of the standards are centralized to one organization	• Need to build a new database. This is time-consuming and expensive. • Need to reengineer processes and organizations in order to enable the new database • Control and upgrade of the standards are enforced by one organization • Lose the company-specific submission advantage built over the years from company-centric submission design

Figure 1 – SDTM Implementation Approaches

The next section will provide details on our experience using the database-only approach.

SDTM IMPLEMENTATION
We piloted SDTM in 2006 and implemented in 2007 using the hybrid database-only approach. This approach reengineers the whole process, reorganizes activities, sets up a group to enforce standards for sponsor-specific domains, creates derivations to implement the SDTM, and requires a training investment in staff.

The database, Clinical Data Repository (CDR), implements FDA's JANUS data warehouse model and CDISC's SDTM clinical data standard. It is the repository for collected data, derived data, and protocol level metadata. CDR supports the audit trail, business logic/database derivations, frozen files, study pooling, vocabulary leveling, medical monitoring, and metadata-controlled blinding/unblinding. Another feature of CDR is fully enabled standards-based data exchange with other internal and external systems. For analysis and reporting (A & R), data is extracted from CDR and transformed into SDTM and SDTM+ SAS$^®$ datasets in another environment on the SAS platform. This achieves the industry standard format now required for SAS datasets.

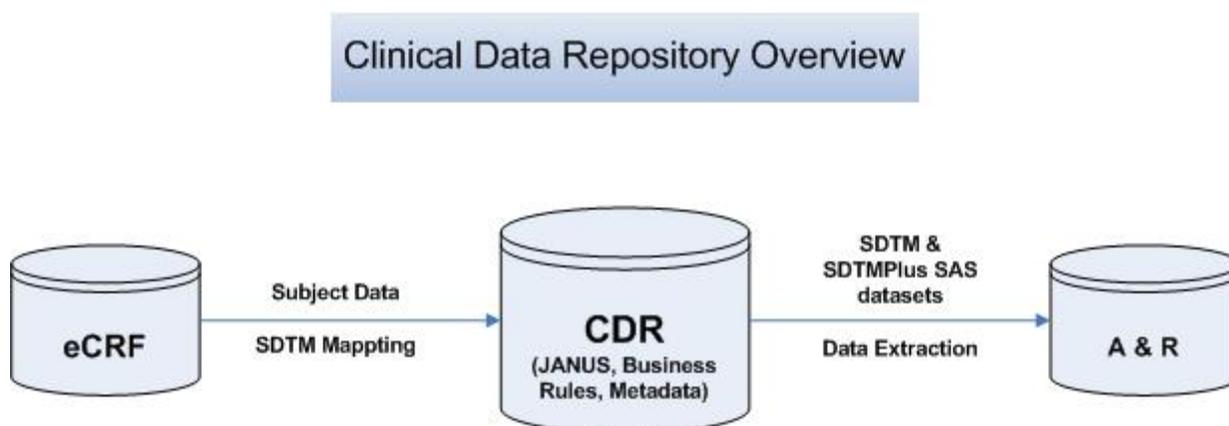 Below is a high-level CDR implementation process flowchart:

Figure 2 – Database-only SDTM Implementation

For the implementation, we start with the eCRF for electronic data capture (EDC). The patient data is then mapped to SDTM format prior to loading into CDR. Derived variables are created in the CDR based on business logic (rules). This business logic is stored as metadata in the system. Then, study metadata is applied. For example drug formulation is stored as study metadata. Upon data extraction, the data is transformed from the CDR to SAS SDTM and SDTM+ formatted datasets, each dataset representing an SDTM domain. Finally, after data extraction, analysis and reporting is performed on a SAS platform.

The next few sections will provide details on SDTM+, business logic for derived variables, and standards oversight.

SDTM EXTENSION
SDTM+ is our extension of the SDTM standard generalized to include derived variables needed for analysis and reporting. Since SDTM does not allow the addition of new variables it is necessary to represent additional data for analysis and reporting purposes in SDTM+ datasets.

BUSINESS LOGIC
The business logic is used for derivations. Each piece of logic, or rule, describes the purpose of the rule, the input variable, the derived variable names, and how to derive the values. The following is a brief description of the business rule to derive the start day relative to subject reference start date (--STDY) in our environment:

Business Rule Description	Derivation
Start relative day should be calculated on all domains that contain a date field which describes when an event or intervention started. Start relative day is an integer number of days, and should not be 0. This business rule does not make use of time.	• If Start Date is blank, then it cannot be calculated. • If Start Date is prior to the Subject Reference Start Date, then --SDTY = --STDTC – DM.RFSTDTC • Otherwise, --SDTY = --STDTC – DM.RFSTDTC + 1

Figure 3 – Example of business logic for a derived variable

STANDARDS OVERSIGHT COMMITTEE
An oversight committee was created to provide guidance on the end-to-end development of data standards. This group ensures consistency of data capture, handling and output standards. Forms design to enable data analysis is an important focus of the group. It assists in the creation of new SDTM domains when necessary. All data capture and mappings must be reviewed and approved by this committee before patients can be enrolled in studies.

USER ACCEPTANCE TEST
Correct SDTM mapping and business rules application is essential to successful implementation of SDTM. As a result, user acceptance testing (UAT) for SDTM mapping, business rule derivations, and phasing is a critical integrated part of the new process. We used a primarily SAS-based approach to user acceptance testing in order to leverage our core competencies. The SAS-based UAT process allows for easily replicated and documented testing. Once the testing phase has successfully concluded, we can be confident in a successful SDTM implementation.

During UAT, the SAS programmer performs the testing. Other technical personnel are involved with the implementation of SDTM mapping, business rules and phasing on test data. SAS is used by the tester to compare mock or test data to the SDTM mapping specifications. Once the mapping has passed testing, SAS programs then check that business rule logic and phasing are implemented correctly in CDR. Results can easily be saved as SAS log and lst files as evidence of successful testing. These files also serve to document any issues that may be found during testing and facilitate issue tracking and resolution with others involved in the UAT.

CONSIDERATIONS FOR IMPLEMENTING PURE AND DATABASE-ONLY SDTM APPROACHES
Successful implementation of SDTM requires new processes, organizations, and training. The process starts with the data collection, data mapping and business logic applications. It ends with communicating with FDA for sponsor-specified domains.   A new organization is required to ensure the consistent implementation of sponsor-specified domains and business logic. It also serves the purpose of being a subject matter expert to interpret CDISC standards. Training is essential for implementing the new process and also educating people in new terminology.

SDTM includes highly structured domains and rules for sponsor-specified domains. For sponsor-specified domains, the consideration will be how much normalization is needed so that they can meet the rule requirements and, at the same time, maintain the feel and look of CRFs. It is important to have normalization that is easy to understand and groups the data together. Otherwise, when it comes time to analyze the data, it will be difficult to pull the data together.

Another important factor for consideration is whether one should add new data to SDTM pre-specified standard domains or create a custom domain. Challenges exist with either choice and there are no clear decision criteria for custom vs. standard domain implementation. Until more guidance is provided on this topic, implementation will remain variable.

To eliminate potential implementation problems, one may want to start with a pilot team. The team should have members from different therapeutic areas and clinical operations. Time allowing, it is helpful to test out the process and database with studies from different therapeutic areas before full-scale implementation.

CONCLUSIONS

SDTM standards are evolving. Adopting the SDTM standards requires careful planning, reengineering, and corporate culture change. At the same time, these standards create the venue for sponsors and regulatory authorities to speed up submissions and provide the opportunity for agencies to more readily pool study data across companies. They also create business opportunities for software vendors and CROs due to the use of universally accepted standards.

However, current standards leave gray areas for interpretation. Sponsor companies may implement similar data items in different domains. Until further clarified, variations in SDTM implementation will exist. SDTM also does not address the therapeutic needs for specific domains. CDISC may want to create a special body to guide the sponsors on gray areas and create therapeutic-specific domains. This should make the implementation of SDTM standards easier and help speed up submission reviews by regulatory agencies.

REFERENCES:

Gillespie, Amy, Varughese, Mary. 2007. "Developing Reporting & Analysis Methodologies," PharmaSUG Proceedings.

Kenny, Susan J., Litzsinger, Michael A. 2005. "Strategies for Implementing SDTM and ADaM Standards," PharmaSUG Proceedings.

Kubick, Wayne R. Ruberg, Stephen, Helton, Edward. "Towards a Comprehensive CDISC Submission Data Standard," Drug Information Journal (May 1, 2007), pp.373-382

Shostak, Jack. 2005. "Implementation of the CDISC SDTM at the Duke Clinical Research Institute," PharmaSUG Proceedings.

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

William T. Chen   Margaret M. Coughlin, MPH
126 E. Lincoln Avenue   126 E. Lincoln Avenue
P.O. Box 2000, RY34A-320   P.O. Box 2000, RY34A-320
Rahway, NJ 07065   Rahway, NJ 07065
Phone: (732) 594-3959   Phone: (732) 594-3781
Email: william_chen@merck.com   Email: margaret_coughlin@merck.com