Paper RS10

The CDISC Study Data Tabulation Model (SDTM): History, Perspective, and Basics

Fred Wood
Principal Consultant, Octagon Research Solutions

ABSTRACT

The CDISC (Clinical Data Interchange Standards Consortium) SDTM is a standard for submitting data tabulations to the FDA in support of marketing applications. In July of 2004, this standard became a Study Data Specification (1) referenced in the eCTD Guidance (2). The FDA has announced on several occasions their intention to make the SDTM required by regulation (3-6). This paper/presentation will focus on the SDTM and the SDTMIG (SDTM Implementation Guide: Human Clinical Trials), but will also describe the regulatory climate around the SDTM, and provide an update on recent activities with SEND (Standard for Exchange of Nonclinical Data), an implementation of the SDTM for animal toxicology and pharmacology studies.

INTRODUCTION

NOTE TO READERS

In an attempt to make wading through the alphabet soup of acronyms, some of the more frequently used ones are listed at the end of this paper.

CDISC BACKGROUND

From its inception in 1997, CDISC has recognized the need for the establishment of standard data models to improve the process of electronic acquisition and exchange of clinical trials information for the benefit of all medical and pharmaceutical stakeholders. This is reflected in the mission statement:

> "CDISC is an open, multidisciplinary, non-profit organization committed to the development of worldwide industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trials data and metadata for medical and biopharmaceutical product development. The mission of CDISC is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare."

CDISC STANDARDS

While the focus of this paper is the SDTM (developed by the CDISC Submission Data Standards [SDS] Team), it is important to understand that the SDTM is just one of many standards being developed by CDISC. Others include the following:

- The Lab Model (LAB) was created for the transfer of lab data from vendors to sponsors. It can be implemented in ASCII, SAS®, or XML. The Lab model was the first CDISC model to become an HL7 (Health Level 7) standard.
- The Operational Data Model (ODM) is a vendor-neutral, platform-independent, XML-based format for interchange and archive of data from clinical trials, including study data, metadata, and administrative data. A sub-team has also formed with representatives from the SDS and ODM Teams to map the SDTM into ODM XML.
- The Analysis Data Model (ADaM) is a set of guidelines and examples for analysis datasets used to generate the statistical results for submission to a regulatory authority such as FDA. It specifically addresses needs of statistical reviewers. The CDISC ADaM Team has developed a number of analysis-level standards, using the SDTM as the foundation. A sub-team formed with representatives from the ADaM and SDS Teams conducted a pilot project in 2006, and a summary of this has been posted on the CDISC website (cdisc.org). In addition, an ADaM Implementation Guide was made available for public review in the first part of 2008. During the upcoming year, the ADaM and SDS Teams will be working to further harmonize the two models.
- The Protocol Representation Group (PRG) is identifying standard elements (including a subset of those in SDTM) of a clinical trial protocol. It is an HL7 initiative with a CDISC team to support the standards-development activities. A machine-readable model is an ultimate goal for this team.
- Standard for Exchange of Nonclinical Data (SEND). SEND is an implementation of the SDTM specific to animal toxicology and pharmacology studies. More details on SEND are provided in a separate section below.
- The CDISC Terminology (CT) Team is working with representatives of government (FDA, NCI), academia, and pharmaceutical companies to identify and/or define standard lists of values for use in the clinical-data lifecycle. This work supports the SDTM, SEND, and other CDISC standards.

- Clinical Data Acquisition Standards Harmonization (CDASH) consists of a set of content standards (element name, definition, and related metadata) for a basic set of global data collection fields (based upon the SDTM) that will support clinical research studies.
- The CDISC define.xml standard is now a Study Data Specification (1) referenced in the eCTD Guidance (2). It is a replacement for the traditional define.pdf, allowing much greater flexibility in the metadata describing the submitted data in a machine-readable format. The SDS-Team-based Metadata Subteam conducted a pilot project in 2006 and will be publishing an implementation guide for the submission of define.xml, which will also include recommendations for the annotation of CRFs.

DATA SUBMISSION BACKGROUND

Case Report Tabulations (CRTs) originated in FDA regulation of the late 1980's, originally intended to allow sponsors to submit data tabulations to the FDA instead of copies of every subject CRF. Recently, as documented in the FDA Study Data Specification (1), which is referenced in the eCTD Guidance (2), the definition has expanded to include data listings, patient profiles, data tabulations, and analysis datasets. Since the first three are essentially different views of the same data, FDA presentations have stated that if SDTM is the format for the data tabulations, patient profiles and data listings are not required (7, 8). This is because review tools can generate the other two views.

SDTM BACKGROUND

The SDTM originated as the Submission Data Model (SDM), developed by the CDISC Submission Data Standards (SDS) Team, which began meeting in 1999. It has been comprised of approximately thirty members who come from a cross-section of the pharmaceutical industry, including most major pharmaceutical companies, numerous small to mid-level pharmaceutical companies, CROs, and service providers. The SDS Team collaborates regularly through biweekly teleconferences, quarterly face-to-face meetings, and literally thousands of regular email communications. SDS Team meetings have been regularly attended by three key FDA observers representing the Office of Business Process Support, biostatistics, and medical review.

The concept that developed into the Study Data Tabulation Model (SDTM) v1.0 / Submission Data Standards (SDS) v3.1 was initially presented to the SDS Team by FDA liaisons in October 2002. Prior to that time, the SDS Team had developed the Submission Data Domain Models Version 1.0 (v1, 2001), and Version 2.0 (v2, early 2002), and was just about to publish v2.1. All of these versions focused on safety domains. While the v1 and v2 data concepts had been well received by industry, it was recognized there were a couple of major shortcomings. The revolutionary concept that the FDA proposed, termed Version 3.0 by the Team, addressed those major shortcomings by 1) providing a standard for all clinical trial data, not just safety data, and 2) providing a standard based on data modeling principles, rather than data management / operational database principles. This concept provided a consistent approach for modeling all clinical trial data using three primary data classes (Events, Interventions, and Findings).

The preliminary draft version of the SDTM concept was published in June of 2003 as the Submission Data Domain Models Version 3.0, or better known as SDS v3.0. The first version intended for implementation was published as two documents in June of 2004: the SDTM v1.0 (the model), and the SDTMIG v3.1 (the implementation guide). During the period from June 2003 to June 2004, there were a number of enhancements leading to the final approved version, many the result of a joint FDA/industry pilot to test SDS v3.0, and two public review/comment periods. Version 1.1 of the SDTM was published in April of 2005, followed by Version 3.1.1 of the SDTMIG four months later.

SDTM v.1.2 and SDTMIG v3.1.2 were posted for public comment in July 2007. Version 3.1.2 of the SDTMIG contained many additions and clarifications to v3.1.1, with more than thirty noteworthy changes for which comments were solicited. In addition to these, the SDS Team received close to a thousand comments during the 60-day comment period, many of which communicated minor fixes. The Team spent the remainder of 2007 and early 2008 addressing these.

REGULATORY EVENTS RELATED TO THE SDTM

Shortly after the publication of the first production versions of the SDTM and the SDTMIG in 2004, the FDA recognized the SDTM as an approved method for submitting the data-tabulation component of Case Report Tabulations (CRTs) in the then-draft eCTD guidance. Health and Human Services (HHS) announced in the December 2004 regulatory agenda (3) that the agency was moving towards requiring the submission of clinical trial data in a standard format.

In its 2004 Critical Path Report, the FDA presented its diagnosis of one of the scientific challenges underlying the medical product 'pipeline problem'. The report then laid out a path forward, beginning with extensive outreach and consultation with public and private stakeholders. Stakeholders confirmed this diagnosis and provided examples of scientific investments that could revolutionize medical product development. In January of 2005, Dr. Janet Woodcock, acting Director of CDER, met with the CDISC Board of Directors (BOD) to review the FDA's Critical Path Initiative. She clearly stated that industry adoption of the SDTM for submission of data for all clinical trials in marketing applications was a significant component of the FDA's Critical Path Initiative, and asked the CDISC BOD for recommendations on how CDISC and the FDA could enhance their collaboration to promote quicker industry adoption of the SDTM.

On February 1, 2005, the FDA conducted a public meeting to review the status of industry adoption of the SDTM. Multiple FDA liaisons reviewed the importance of the SDTM and its value to both industry and FDA. They also confirmed that sponsors submitting data in the SDTM would not be required to submit data listings and patient profiles. FDA review tools will automatically produce data listings and patient profiles from properly formatted SDTM datasets. Additionally, FDA agreed to provide advance training to reviewers on the SDTM, and the review tools, when sponsors notify FDA in advance that they are submitting in the SDTM format. The FDA has recognized that ensuring this training occurs is a necessary component of SDTM implementation in industry.

In March 2006, the agency the FDA released The Critical Path Opportunities List (9) based on feedback from stakeholders and the special insights of FDA's product reviewers. Item 44 on that list is entitled Development of Data Standards, and it mentions the SDTM in the following text:

> "CDISC is paving the way by developing its Study Data Tabulation Model for describing observations in drug trials. That model could someday encompass observations needed for other types of trials. Health Level 7 and CDISC are working to create standards that can be used for the exchange, management, and integration of electronic healthcare information to increase the effectiveness and efficiency of healthcare delivery."

Since late 2006, a number of Federal Register announcements have further indicated the FDA's commitment to the SDTM. On September 29, 2006, the FDA announced the withdrawal of the three Electronic Submission Guidances for eNDA, eANDA, and eAnnual Reports (10). This notice designates eCTD as "preferred format for electronic submissions" and notes that beginning January 1, 2008, any electronic submission going to CDER must be eCTD.

In December of 2006, the FDA announced that there would be Notice of Proposed Rulemaking (NPRM) regarding the SDTM in March of 2007 (3a). From this notice:

> "The proposal would revise our regulations to require that data submitted for NDAs, BLAs, and ANDAs, and their supplements and amendments be provided in an electronic format that FDA can process, review, and archive. The proposal would also require the use of standardized data structure, terminology, and code sets contained in current FDA guidance (the Study Data Tabulation Model (SDTM) developed by the Clinical Data Interchange Standards Consortium) to allow for more efficient and comprehensive data review. "

The timing given in this notice was March of 2007. It has since has been changed two additional times: in April 2007, a date of November 2007 was given (5), and in December 2007, a date of September 2008 was given (6).

SDTM BASICS

The current CDISC Submission Data Standard consists of two documents: the Study Data Tabulation Model Version 1.1, published in April 2005, and the Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.1.1, published in August 2005. The first describes the model, while the second provides guidance on model implementation, including 1) domain models and examples with real data for commonly submitted datasets, 2) a set of assumptions to aid in interpretation and application of the intended implementation, 3) more detailed descriptions of the Trial Design Model (TDM) tables, and 4) a more detailed discussion on representing relationships within and across submission datasets. The SDTM is built around several key concepts. These are described in the following paragraphs.

Domains

Domains are groups of related observations. These observations are grouped by topic in datasets. Datasets and domains are usually the same, but some domains contain more than one dataset.

Observations

Observations can be described by a series of named variables. Each variable, which normally corresponds to a column in a dataset, can be classified according to its role. Most variables in a domain begin with a prescribed prefix.

Example: In Study ABC001, Subject 1234-0001 had a heart rate of 100 bpm on Study Day 6. This would be represented in a dataset as follows:

STUDYID	USUBJID	VSDY	VSTESTCD	VSORRES	VSORRESU
ABC001	1234-0001	6	HR	100	BEATS/MIN

Observation Classes

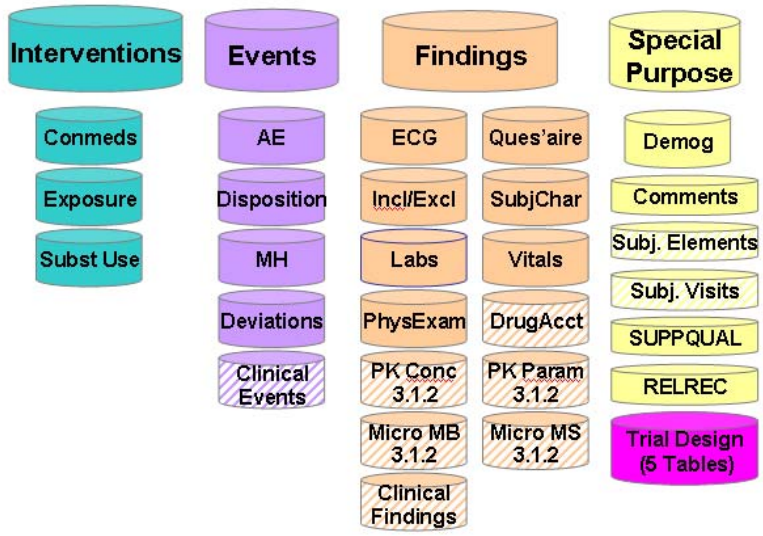
An observation can be classified as one of three major types: Interventions, Events, or Findings. They can be described as follows:

- **Interventions:** investigational treatments, therapeutic treatments, and surgical procedures administered to the subject or animal. One record per constant dosing/treatment interval.
- **Events:** occurrences or incidents independent of planned study evaluations occurring during the trial (e.g., adverse events) or prior to the trial (e.g., medical history). One record per event.

- **Findings:** observations resulting from planned evaluations (e.g. lab tests, ECGs, microscopic findings). One record per finding result or measurement.

Each observation class has a defined set of standard variables. These standard variables, along with Identifiers and Timing Variables, both of which can be used in all observation classes, are the building blocks for constructing SDTM domains. Variables other than the standard ones for each class must be represented in a SUPPQUAL dataset (to be discussed later). Figure 1 shows the observation classes for the domains modeled in the SDTMIG.

Figure 1. Fitting V3.1.2 Domains into Observation Classes



Lighter colored domains were new in the draft version of SDTMIG v3.1.2

Variable Roles

Every variable has been assigned a Role that describes the type of information conveyed by each variable within an observation. These Roles are defined in the SDTM:

- **Topic Variable** - Identifies the focus of the observation. There is only one per dataset.
- **Identifier Variables** - Identify the study, the subject, the domain, and sequence number of the observation.
- **Timing Variables** - Describe the start and end of the observation, and/or when it was collected.
- **Qualifier Variables** - Describe the attributes and results of the observation. These can be further subdivided into Grouping, Result, Synonym, Record, and Variable Qualifiers.

Variable Metadata

Each dataset or table is accompanied by metadata definitions (define.pdf or define.xml) that provide information about the variables used in the dataset. Included are the variable name, variable label, the {data) type, controlled terms or format, the origin (e.g., CRF, derived), the role (as described above), and any comments about a variable or its data.

Additional Datasets and Tables that Need to Be Submitted

In addition to data submitted in accordance with the three observation classes, there are a number of special-purpose datasets that are also part of the SDTM. Included are the Demographics dataset, the Comments dataset, the Supplemental Qualifiers datasets, the RELREC (Related Records) dataset, and seven TDM datasets. The preferred method for submitting Supplemental Qualifiers is as one SUPP-- dataset per domain, with the hyphens representing the two-letter domain code.

The Demographics dataset includes a set of standard variables that describe each subject in a clinical study. The Comments domain is a fixed domain that provides a solution for submitting free-text comments related to data in one or more domains or collected on a separate CRF page dedicated to comments. The Comments dataset is similar to the SUPPQUAL dataset but it allows for one comment to span multiple variables (COVAL-COVALn) in order to accommodate comments longer than 200 characters. It is flexible in that it permits comments to be related to a subject, to a CRF page or SDTM domain, or to specific parent records in a domain.

Relationship Tables

In order to understand the primary relationship datasets, Supplemental Qualifiers and RELREC, it is necessary to understand that every record in the observation-class datasets has a unique set of keys consisting of STUDYID (Study ID), USUBJID (Unique Subject ID), DOMAIN (SDTM Domain), and --SEQ. The latter is a sponsor-defined numeric identifier unique within

study, subject, and domain; the two hyphens indicate the two-letter domain code. Knowing the values for these four columns allows precise identification of a single record within a submission. Both RELREC and SUPPQUAL use this same concept to identify a parent record to which either a non-standard variable (Supplemental Qualifiers) or another independent (parent) record (RELREC) is related. These tables use STUDYID, USUBJID, RDOMAIN (Related Domain), and two other variables that point to the related record. These two variables are IDVAR and IDVARVAL, which describe the parent record's unique identifier variable and its value. Other variables aside from --SEQ, such as Grouping Qualifiers and Identifiers, can be used in IDVAR to express relationships other than one to one. An example would be using --GRPID (Grouping Identifier) to relate multiple ECG measurement records to a single SUPPQUAL record.

The Supplemental Qualifiers special-purpose dataset is used to submit values for variables not presently included in the general-observation-classes and the Demographics domain. In addition to the keys described in the previous paragraph, each Supplemental Qualifiers record also includes the name of the Qualifier variable being added (QNAM), the label for the variable (QLABEL), the actual data value for each instance or record (QVAL), the origin (QORIG) of the value (e.g., whether it was collected via CRF, or derived), and the Evaluator (QEVAL, to specify the role of the individual who assigned the value, such as an Adjudication Committee or the sponsor).

Another common reason for using a SUPPQUAL dataset is to capture attributions. An attribution is typically an interpretation or subjective classification of one or more observations by a specific evaluator, such as a population flag that classifies a subject or their data according to their evaluability for efficacy analysis. Since it is possible that different attributions may be necessary in some cases, SUPPQUAL provides a mechanism for incorporating as many attributions as are necessary. For example, if two individuals provide a determination on whether an adverse event is treatment emergent (e.g., the investigator and an independent adjudicator) then separate QNAM values should be used for each set of information, perhaps AETRTEM and AETRTEM2. This is necessary to ensure that reviewers can join/merge/transpose the information back with the records in the original domain without risk of losing information.

The Related Records (RELREC) dataset is used to describe collected relationships between records in two (or more) datasets, such as an Event record and an Intervention record, or a Finding record and an Event record. One example would be the collection of an Adverse Event number on a Hospitalization or Lab page. RELREC should be not be used to determine associations after the fact (e.g., as part of the analysis process). Relationships are described by creating RELREC records for each of the related observation records, and then by assigning a unique character identifier value for the relationship. Each RELREC record contains same keys as SUPPQUAL to identify a record (using --SEQ in IDVAR) or group of records (using --GRPID in IDVAR). RELREC uses an additional, unique variable, RELID, the relationship identifier, which is the same for all related records. The value of RELID can be any value chosen by the sponsor.

<u>The Trial Design Model (TDM)</u>
The TDM allows description of key aspects of the planned conduct of a clinical trial in a standardized way. These standardized descriptions will allow reviewers to:

- clearly and quickly grasp the design of a clinical trial
- compare the designs of different trials
- search a data warehouse for clinical trials with certain features
- compare planned and actual treatments and visits for subjects in a clinical trial

Modeling a clinical trial in this standardized way requires the explicit statement of certain decision rules that may not be addressed, or may remain vague or ambiguous, in the usual prose protocol document. Prospective modeling of the design of a clinical trial should lead to a clearer, better protocol. Retrospective modeling of the design of a clinical trial should ensure a clear description of how the trial was interpreted by the sponsor.

The TDM is built upon the concepts of Elements, Arms, Epochs, and Visits. They are shown in Figure 2, and can be described as follows:

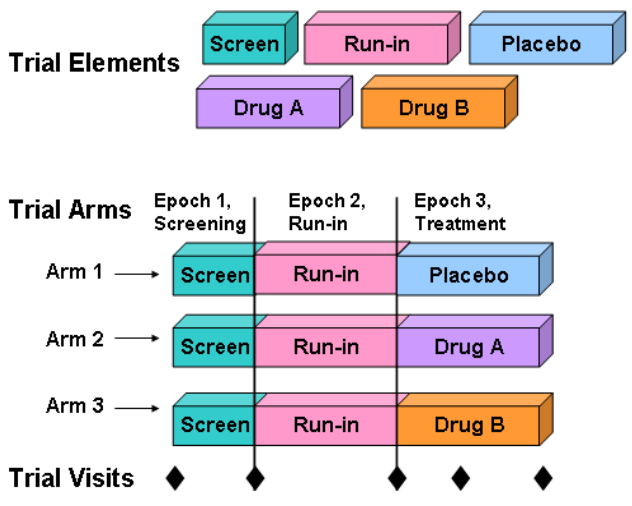


Figure 2. Key Concepts of Trial Design

- An Element is the basic building block for a discrete treatment period (or constant-dosing interval) within a clinical trial. Planned Elements are described in the Trial Elements (TE) table, which includes a description of the treatment(s) are planned for subjects in that Element, rule for starting the Element, a rule for ending the Element, and a planned duration (if it's fixed and not conditional).
- An Arm is a planned sequence of Elements, and is usually equivalent to a treatment group. The Trial Arms (TA) table describes the sequence of elements within each Arm, as well as rules for

branching and transitions. The Demographics domain contains the name of the Arm to which the subject was randomized.

- A Visit is defined as a clinical encounter that encompasses planned and unplanned trial interventions, procedures, and assessments that may be performed on a subject. Planned visits are described in the Trial Visits (TV) table. A Visit has a start and an end, each described with a rule. Different Arms can have different planned-visit schedules. A Visit need not be nested within a single Element. In other words, it may start in one Element and end in another.
- The term EPOCH is used to describe a period of time that cuts across Arms, and applies to the trial as a whole Epochs may contain the same number or a different number of Elements for each Arm. Epochs are defined in the TA table.

The TDM also includes the Trial Inclusion/Exclusion (TI) and Trial Summary (TS) datasets. The TI table is used to describe the protocol-specified inclusion/exclusion criteria used for all subjects. This is contrast to the IE domain, which contains the subject-specific exceptions to those criteria (i.e., criteria not met) for subjects included in the submission data. TS is used to submit trial-level summary information, examples of which include the blinding schema, drug indication, and trial objectives.

There are two subject-level tables that describe the actual element sequence (Subject Elements) and visit schedule (Subject Visits). These tables include the actual dates for the start and end of each element or visit, and are flexible enough to handle both planned (and therefore in the TE and TV tables) or unplanned elements and/or visits. These tables have been removed from Trial Design (Section 7) in SDTMIG v.3.1.1 to Special-Purpose Domains (Section 5) in SDTMIG v3.1.2, since they represent data about subjects, and are not related to the planned design of the trial.

THE STANDARD FOR EXCHANGE OF NONCLINICAL DATA (SEND)

As noted in the Introduction, SEND is an implementation of the SDTM for nonclinical studies. The work on this standard began in 2002, and an FDA pilot project was announced in January of 2003. The pilot tested Version 1.0 for data from acute, subchronic, and carcinogenicity studies. Input from the pilot, as well as efforts to more closely align this implementation with that for human clinical trials, resulted in the Version 2.x standards. The latest of these is v2.3, posted in November 2005. As expected, some SEND domains are identical to those described in the SDTMIG for human clinical trials (e.g., ECG, Vital Signs); however, there are a number of domains specific to nonclinical research such as Organ Weights, Microscopic Findings, and Food and Water Consumption.

There was a resurgence of SEND activity in May of 2007. A group of representatives from pharmaceutical companies, contract research organizations, technology vendors, and the FDA resumed regular meetings to improve the SEND model. There was interest in developing SEND for use a data transfer standard (vendor to sponsor) in addition to its original purpose as a data submission standard. New subteams, along with one existing team, now include the following:

- The (Existing) Safety Pharmacology Subteam has remained active since its inception in 2005. They have been defining new domains for Safety Pharmacology studies, including Respiratory Measurement, Hemodynamic Measurements, and Central Nervous System Tests.
- The (New) Re-Baselining Subteam has been working on operational aspects of the model, as well as on enhancements and updates to the current implementation guide to improve consistency with the SDTMIG. One example is the need to address more complicated trial designs (e.g., Latin square) for safety-pharmacology studies, and the addition of a Demographics domain to replace or add to the existing Subject Characteristics domain.
- The (New) Terminology Subteam has been defining comprehensive sets of standard terminology, leveraging the work from the CDISC Controlled Terminology and Glossary Teams, as well as with the National Cancer Institute's Enterprise Vocabulary Services (NCI EVS).
- The (New) Reproductive Toxicology Subteam has renewed prior efforts to develop domains for reproductive toxicity studies. These studies can be complex due to 1) the staggered timing of the phases of gestation and weaning within treatment groups, and 2) the relationships that need to be maintained between mating partners and between parents and offspring, possibly through multiple generations.
- The (New) FDA-Pilot Group will be responsible for developing the specifications for the regulatory pilot, which was announced in October 2007 (11). This group consists of approximately seven sponsor companies who will be submitting nonclinical data to an IND or NDA in both the current PDF format as well as the electronic SEND format.

THE FUTURE

REGULATORY CLIMATE

The FDA has been developing the Janus data warehouse as the repository to store all submitted clinical-trials data. Janus will provide a stable foundation for FDA's growing list of standard review tools, developed through Cooperative Research and Development Agreements (CRADAs) with various vendors. The tools have been developed to utilize the SDTM, rather than vice versa. Most of these tools create tabular and/or graphical views of the SDTM data via canned reports, although custom reports can be run as well. The tools routinely bring demographics and treatment data into all views of subject data, utilizing the fixed relational database structure of Janus (upon which much of the SDTM is based). Many tools also provide built-in

hyperlinks that allow drilling down from group summaries to individual-subject data, and to navigate from graphs to tables. Some allow for various "what-if" scenarios, such as allowing the elimination of outlying values from the calculation of means.

The National Center for Toxicological Research (NCTR) will be housing a Janus database for nonclinical data, and representatives from this FDA center will be involved in the SEND Pilot Project (mentioned above). This center has also been involved in the FDA's Voluntary Genomics Data Submission (VGDS) Project.

Without the submission of data in a standard format, the above capabilities would not be possible. The consistent message from the FDA at many meetings over the past two years, including the DIA Annual Meetings and the CDISC Interchanges has been to submit data in SDTM format.

DATA SUBMISSION STANDARDS

We can expect to see the development of standards for more types of data and a greater level of harmonization of standards, both within CDISC and between CDISC standards and those of other standards-development organizations. Some of the major efforts are described below.

SDTMIG Domains
Within the CDISC SDS Team, domains for oncology studies and pharmacogenomics data are currently under development.

Device Teams
The SDTM is the basis for domains for devices, and these are being developed by two teams. The team for diagnostic devices has been active since early 2007, and it is expected that several domains will be posted for comment and finalized this year. A Devices Team for implantable devices had its first teleconference in November 2007, and models are expected to be developed through 2008. Both teams have active involvement from representatives of the Center for Radiological Devices and Health (CDRH).

HL7 Harmonization
Efforts are well underway to harmonize the SDTM with other CDISC standards, and to harmonize CDISC standards with those of Health Level 7 (HL7) standards. Harmonization is being accomplished through work on the BRIDG (Biomedical Research Integrated Domain Group) model. BRIDG was developed to provide an overarching model that could readily be understood by clinical-data domain experts. It would provide the basis for semantic interoperability between these various standards. In this way, there could be interchange of data between clinical research, biomedical/clinical research, and patient healthcare. In 2007, a team began work on mapping the SDTM to the BRIDG model.

Trial Design Model
The TDM in the SDTM is being harmonized with the HL7 Protocol Representation Group (PRG). The Trial Design domains from the SDTM have been mapped to the BRIDG using UML (Universal Modeling Language) diagrams as part of the BRIDG Model. The Trial Design Subteam of the SDS Team and PR Group released Part 2 of the TDM for comment in late 2005. TDM2 added the domains to describe the scheduling of Trial Activities such as assessments, interventions, and administrative actions. Since the comment period, the TDM has undergone some significant changes, including mapping to the BRIDG, and a new release is planned for 2008.

CONCLUSION

The regulatory status of the SDTM not withstanding, it is to a sponsor's advantage to move toward implementation of the SDTM for submission datasets. Once FDA reviewers become accustomed to being able to easily navigate through submitted data using the dedicated tools, being unable to do so for subsequent submissions will lead to inefficiencies in the review process.

Since becoming the recommended standard for the submission of clinical and preclinical trial data to the FDA in marketing applications in 2004, the SDTM has begun to be used by sponsors for their upstream processing to support Clinical Study Reports and Integrated Summaries. This paper covers the basics and background of the SDTM. Material presented in the live session will expand on information presented here and provide additional implementation examples.

REFERENCES

1. Study Data Specifications. Current version: 1.4. August 1, 2007; Available via http://www.fda.gov/cder/regulatory/ersr/Studydata.pdf
2. Guidance for Industry: Providing Regulatory Submissions in Electronic Format -- Human Pharmaceutical Product Applications and Related Submissions using the eCTD Specifications. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Revision 1, April 2006.

3. HHS Regulatory Agenda. Announcement in Federal Register of Proposed Rule Stage for "Submission of Standardized Electronic Study Data From Clinical Studies Evaluating Human Drugs and Biologics". Federal Register Vol. 69, No. 238, Page 73134. December 13, 2004.
4. Federal Register Notice (2006) Electronic Submission of Data from Studies Evaluating Human Drugs and Biologics. Vol. 71, No. 237, Monday, December 11, 2006.
5. Federal Register Notice (2007) Electronic Submission of Data from Studies Evaluating Human Drugs and Biologics. Vol. 72, No. 82, Monday, April 30, 2007.
6. Federal Register Notice (2007) Electronic Submission of Data from Studies Evaluating Human Drugs and Biologics. Vol. 72, No. 236, Monday, December 10, 2007.
7. Gensinger, G. (2006). "FDA Experience with Submission in SDTM Format". DIA Annual Meeting, June 2006.
8. Gensinger, G. (2006). "FDA Experience with Submission in SDTM Format". CDISC Interchange, September 2006.
9. Critical Path Opportunities List, U.S. Dept Human Services, FDA, March 2006.
10. Guidances on Providing Regulatory Submissions in Electronic Format; Withdrawal of Guidances. Federal Register Vol. 71, No. 189, Page 57548. September 29, 2006.
11. Federal Register Notice (2007). Electronic Nonclinical Study Data Submission; Notice of Pilot Project Vol. 72, No. 191, Wednesday, October 3, 2007.

ACKNOWLEDGEMENTS

The author would like to thank the following:

- CDISC, their staff, and members for their commitment to creating industry data standards *that enable information system interoperability to improve medical research and related areas of healthcare.*
- Members of the CDISC SDS Team for their countless hours invested in bringing the SDTM and its Implementation Guide to realization.
- Members of the CDISC SEND Team for their energy and enthusiasm in developing industry-wide standards for nonclinical data for both operational and submission use.
- FDA Liaisons to the CDISC SDS and SEND Teams for their continued support.
- The leadership and members of the Clinical Data Management Department at Procter & Gamble Pharmaceuticals, whose vision for data standards allowed my involvement with the SDS Team and CDISC, beginning in 1999.

SOME OF THE ACRONYMS USED IN THIS PAPER

ADaM	Analysis Data Model
BRIDG	Biomedical Research Integrated Domain Group
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
CRT	Case Report Tabulations
CT	Controlled Terminology
eCTD	electronic Common Technical Document
HL7	Health Level 7
NPRM	Notice of Proposed Rulemaking
ODM	Operational Data Model
PRG	Protocol Representation Group
SDS	Submission Data Standards
SDTM	Study Data Tabulation Model
SDTMIG	SDTM Implementation Guide: Human Clinical Trials
SEND	Standard for Exchange of Nonclinical Data
TDM	Trial Design Model
XML	eXtensible Markup Language

CONTACT INFORMATION

Fred Wood
Principal Consultant
Clinical Data Strategies
Octagon Research Solutions, Inc.
585 East Swedesford Road, Suite 200
Wayne, PA 19087
610-535-6500 x5418
fwood@octagonresearch.com