



44TH
Annual Meeting



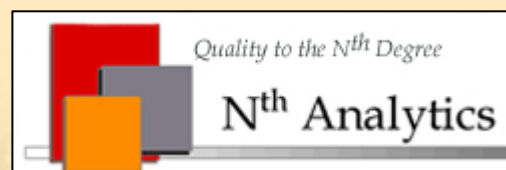
Boston 2008

CDISC SDTM Implementation Process

Michael Todd

President

Nth Analytics



Overview

- CDISC has several initiatives
- Only SDTM will have the force of law
 - The FDA announced plans to require implementation of SDTM as a federal regulation
- Therefore, near-term, only one will have significant impact
- We will focus on SDTM in this presentation



Pace of Change

- The pharmaceutical industry changes very slowly
 - Regulated industry
 - Two powerful groups: Government / Industry stasis
- Slow implementation of SDTM is entirely expected
- Probability of other CDISC initiatives having near-term impact is low



Convergence of Forces

- Post Vioxx era is unique time when significant forces converge to produce change
 - FDA: Focus on safety
 - Industry:
 - Generic competition, cost pressures, resistance to price increases, resistance to me-too drugs, costly, safety-related failures
 - Increased efficiency becomes important
 - Perceived benefits from investment in new methodologies
 - Cost-effective only with widespread implementation



FDA: Focus on Safety

- FDA has been hammered by Congress and media (including blogosphere) for high-profile safety failures
- We see an increased emphasis on cross-compound class effects:
 - Epilepsy compounds and suicidality
 - Tumor necrosis factor (TNF) blockers and cancer
- Analyses of the sponsors' entire clinical database required to address these concerns



Regulatory Basis for use of SDTM



FDA to Mandate Use of SDTM

- Federal Register

- Vol. 71, No. 237 /Monday, December 11, 2006 / The Regulatory Plan / Page 72784
- “The Food and Drug Administration is proposing to amend the regulations governing the format in which clinical study data and bioequivalence data are required to be submitted”
- “...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.”
- 2 year implementation period

- A federal regulation has the force of law



eCTD Guidance

- The eCTD guidance is the current regulatory basis for the use of SDTM in FDA submissions
- FDA Guidance: “Providing Regulatory Submissions in Electronic Format – Human and Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications”
 - <http://www.fda.gov/cder/guidance/7087rev.pdf>
 - Module 5 Datasets: ‘See associated document “Study Data Specifications” for details on providing datasets and related files (e.g., data definition files, program files).’
- Study Data Specifications
 - <http://www.fda.gov/cder/regulatory/ersr/Studydata.pdf>



eCTD: Study Data Specifications SDTM

- Regulatory basis for use of SDTM:
 - “Specifications for the Data Tabulation datasets of human drug product clinical studies, are located in the Study Data Tabulation Model (SDTM) developed by the Submission Data Standard working group of the Clinical Data Interchange Standard Consortium (CDISC).”
 - “This folder is reserved for the datasets conforming to the SDTM standard.”
 - “The latest release of the SDTM and implementation guides for using the model in clinical trials is available from the CDISC web site.”



eCTD: Study Data Specifications DEFINE.XML

- “The data definition file describes the format and content of the submitted datasets.”
- “The specification for the data definitions for datasets provided using the CDISC SDTM is included in the Case Report Tabulation Data Definition (define.xml).”
- “The latest release of the Case Report Tabulation Data Definition Specification is available from the CDISC web site (<http://www.cdisc.org/models/def/v1.0/index.html>).”
- “Include a reference to the style sheet as defined in the specification and place the corresponding style sheet in the same folder as the define.xml file.”



eCTD: Study Data Specifications Annotated CRF

- “This is a blank case report form annotations that document the location of the data with the corresponding names of the datasets and the names of those variables included in the submitted datasets.”
- “The annotated CRF is a blank CRF that includes treatment assignment forms and maps each item on the CRF to the corresponding variables in the database.”
- “The annotated CRF should provide the variable names and coding for each CRF item included in the data tabulation datasets.”
- “All of the pages and each item in the CRF should be included.”
- “The sponsor should write ‘not entered in database’ in all items where this applies.”
- “The annotated CRF should be provided as a PDF file. Name the file blankcrf.pdf.”



SDTM Implementation



SDTM Implementation: Costs and Benefits

- Staggering cost of compliance
 - Analogous to Y2K
 - Will require a federal mandate to achieve critical mass
- However, with widespread implementation, industry should realize substantial cost savings:
 - ETL tools will largely automate creation of reporting and analysis datasets
 - Industry-wide standard databases will enable development of commercial reporting and analysis tools
 - Will shave weeks off timelines
 - SDTM will greatly simplify data transfer between clients and CROs



SDTM Implementation: Phases

- Phase I: understand the rules of SDTM
 - that has been done.
- Phase II: tool implementation
 - ongoing
- Tools cannot do it all
- Implementation requires individuals with data management and SDTM experience to map with authority



The Role of the Data Manager



Data Manager: Key to Mapping Success

- Reviews the SAP, Protocol and CRFs
- Applies his/her experience with study and CRF design
- Uses CDISC SDTM Guidance rules to annotate CRFs with SDTM domains and variables
- Decides which SDTM domains are required based on CRF data
- Assigns CRF data to SDTM variables
- Creates new protocol specific domains when needed
- Creates supplemental domains when needed
- Participates in validation of SDTM domains



Sample Mapping Spreadsheet

Microsoft Excel - Book1

File Edit View Insert Format Tools Data S-PLUS Window Help Acrobat

C5 = Subject Identifier for the Study

	A	B	C	D	E	F	G	H
1	Domain	VariableName	VariableLabel	Type	Origin	Role	Comments	Core
2	DM	STUDYID	Study Identifier	Char	CRF	Identifier	[default]	Req
3	DM	DOMAIN	Domain Abbreviation	Char	Derived	Identifier	[default]	Req
4	DM	USUBJID	Unique Subject Identifier	Char	Sponsor Defined	Identifier	[default]	Req
5	DM	SUBJID	Subject Identifier for the Study	Char	CRF	Topic	%USUBJID(VARNAME=SUBJID)	Req
6	DM	RFSTDTC	Subject Reference Start Date/Time	Char	Sponsor Defined	Timing	%RFSTDTC	Exp
7	DM	RFENDTC	Subject Reference End Date/Time	Char	Sponsor Defined	Timing	%RFENDTC	Exp
8	DM	SITEID	Study Site Identifier	Char	Derived	Record Qualifier	SUBSTR(DEMOG.INVSITE,5,3)	Req
9	DM	INVID	Investigator Identifier	Char	Derived	Record Qualifier	DEMOG.INV	Perm
10	DM	INVNAM	Investigator Name	Char	Derived	Synonym Qualifier	%INVNAM	Perm
11	DM	BIRTHDTC	Date/Time of Birth	Char	CRF	Result Qualifier	%ISO_DATETIME(DATE=DEMOG.DMDOB DT, TIME=0)	Perm
12	DM	AGE	Age in AGEU at RFSTDTC	Num	Derived	Result Qualifier	%AGE	Exp

Sheet1 / Sheet2 / Sheet3

Ready



Validation: SDTM Structure

- SDTM compliance checks
 - Conformance with Implementation Guide rules*
 - Metadata structure and domain models
 - Keys, topic variables
 - Variable names, labels, type etc. are correct
 - All required variables have values, etc.
 - All expected variables are present
 - All collected relevant and derived data

* Study Data Tabulation Model Implementation Guide v3.1.1 – Section 3.2.3 Conformance



Validation: Source Data

- Verify metadata
 - Data manager is the expert for this task
 - Includes a manual review process
- Techniques
 - Independent programming
 - For each source dataset, verify data converted correctly to SDTM for a random sample of subjects



Validation: All Source Data Mapped?

- Leverage metadata
- Use metadata to query source data structure and determine differences
 - List of source data variables not mapped
 - List of source data values not mapped
 - Source data variables differing from SDTM variables



SDTM: New Job Roles



Mapping Specialist

- New role, but really an old role in a new time
- The mapping specialist decides how to convert the raw data into SDTM domains and variables
- Senior position
- Historically done by SAS programmers or statisticians



Data Manager/Mapping Specialist: Job Descriptions

Clinical Data Manager	Mapping Specialist
Ensure the completeness, accuracy and consistency of the data	Authoring of mapping specifications from source (raw) data to target (CDISC SDTM)
Building, testing and validation of clinical database using standardized software packages	Development of CDISC SDTM domains utilizing a data conversion tool
Development of data specifications and CRF design	Quality Control of CDISC SDTM domains
Protocol review regarding what, when, and how data are collected	Creation of data definition files and Case Report Forms annotated to CDISC SDTM



Mapping Specialist: Qualifications

- In-depth knowledge of CDISC SDTM V3.1.1 and V3.1.2 Implementation Guides
- Clinical data management experience
- Knowledge of ETL tool
- Basic SAS knowledge



Data Integration Specialist: Job Function

- Develop conversion jobs in ETL tool
- Write any SAS macros required for complex derivations that can't be handled directly in the tool
- Execute the job and review log for anomalies



Data Integration Specialist: Qualifications

- Knowledge of clinical data
 - Not as much as mapping specialist
- ETL experience
- Intermediate SAS skills, especially macro language and PROC SQL
 - Not as much as traditional SAS programmer
 - ETL tool is doing most of the work
 - Mostly writing code fragments
- SDTM knowledge
 - Not as much as mapping specialist



Summary

- Convergence of forces will drive SDTM implementation
- As implementation becomes widespread, organizations using ETL tools will have a decided edge
- Data management expertise is critical to successful SDTM automation

