



Setting the  
Global Standard  
for Clinical Data

# “Tracking the Progress made by CDISC in Implementing New Standards”

**CLINICAL DATA INTERCHANGE  
STANDARDS CONSORTIUM**

**Mike Todd**

**CEO**

**Nth Analytics**

**Email: [info@nthanalytics.com](mailto:info@nthanalytics.com)**

**Website: [www.nthanalytics.com](http://www.nthanalytics.com)**

# “Tracking the progress made by CDISC in Implementing New Standards”

- ❖ Outlining the New Developments in CDISC Standards
- ❖ Discovering the Ways Standards will Change Organizational Management of Data
- ❖ The Competitive Advantages that Stem from Adherence to the Standards

# Outlining the New Developments in CDISC Standards

# Clinical Data Interchange Standards Consortium

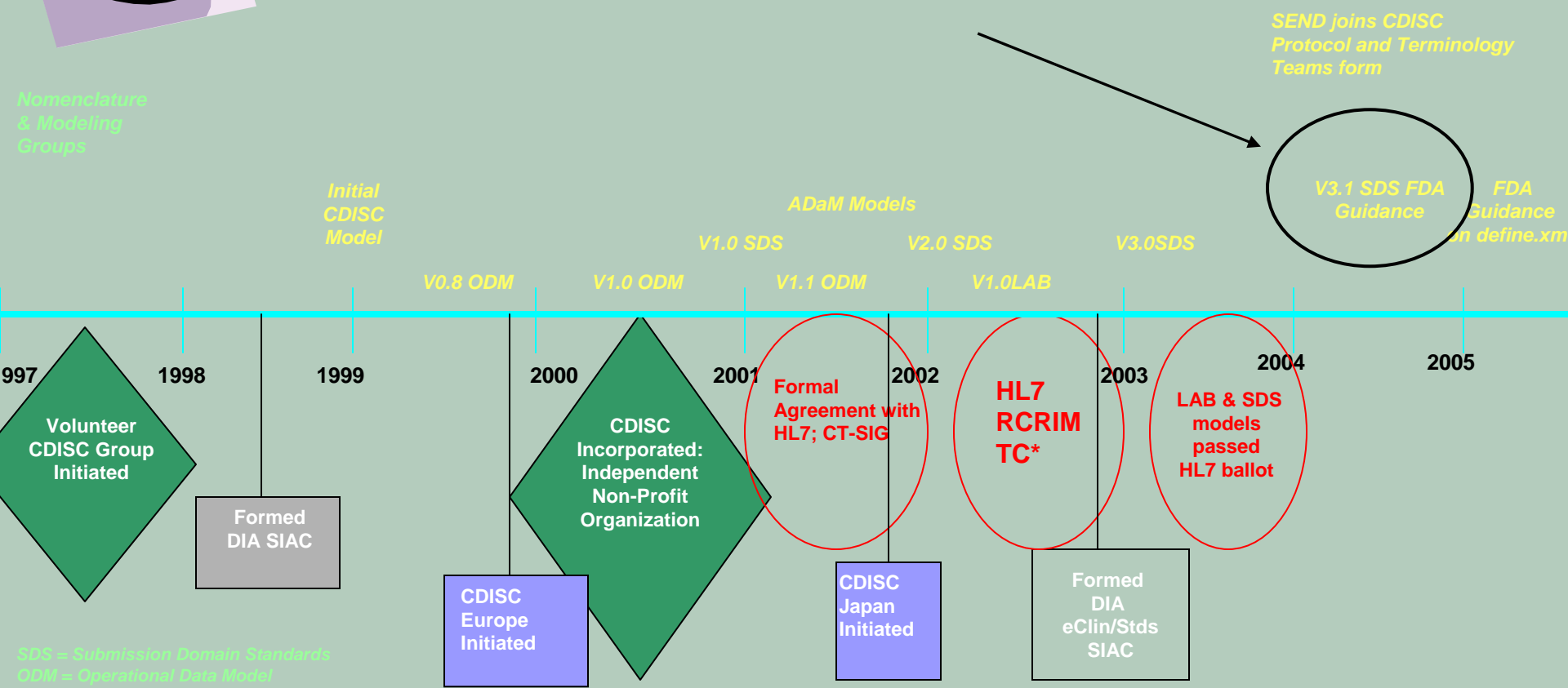
## ***MISSION:***

***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 Timeline

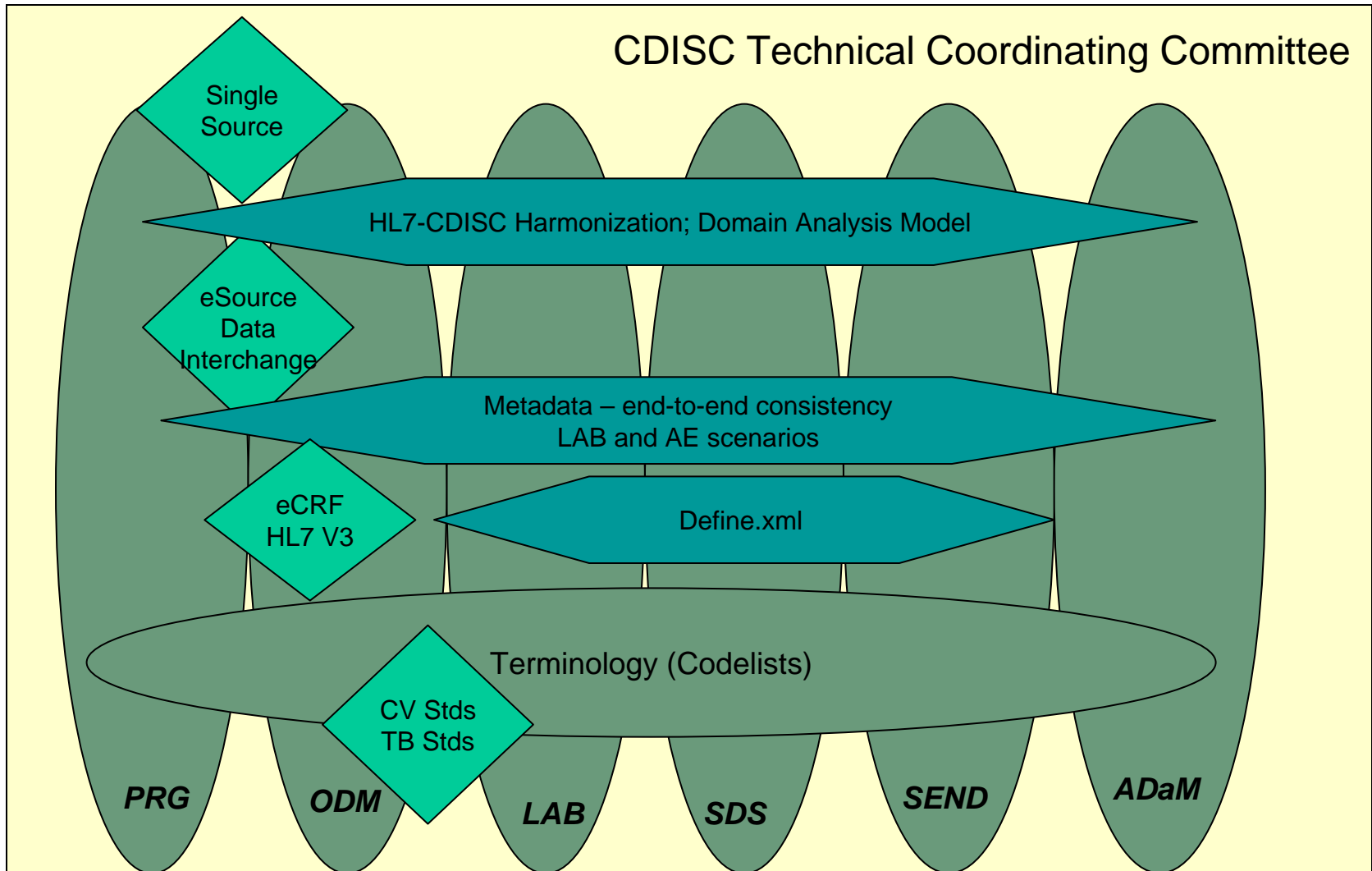
Nomenclature & Modeling Groups



SDS = Submission Domain Standards  
 ODM = Operational Data Model  
 LAB = Laboratory Data Model  
 ADaM = Analysis Dataset Models  
 SEND = Standards for the Exchange of Non-clinical Data

**\*RCRIM TC: Regulated Clinical Research and Information Management Technical Committee; Co-chaired by HL7, CDISC and FDA**

# CDISC Teams and Projects - 2005

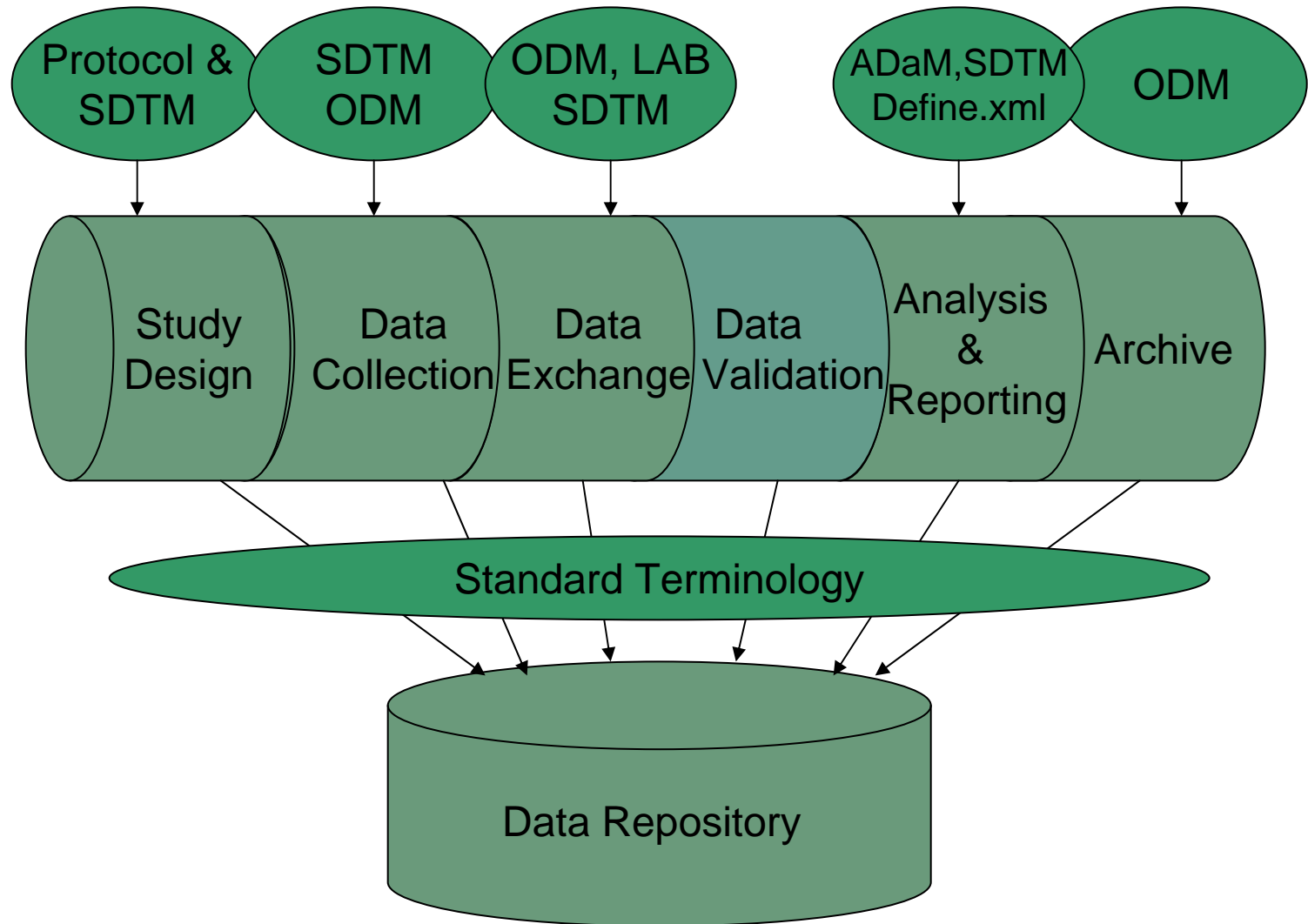


Maintenance, Member Relations, Education and Implementation Groups, Glossary

# CDISC and HL7

- HL7 is a standards development organization dealing with data standards for all health care operations
  - hasn't dealt much with the nuances of clinical trials
- CDISC is oriented to biopharmaceutical drug development only
  - hasn't dealt with health care applications important to HL7 such as reimbursements and order processing

# Opportunities to Apply Standards





# Protocol Representation Group

- Initiated in mid-2003 as HL7 Project to work with domain experts to develop a standard for clinical trial protocol representation
  - Joint activity of CDISC and HL7 RCRIM
  - Comprised of > 20 representatives from major pharmaceutical companies, CROs, NIH/NCI, FDA, technology providers
- Goals:
  - Revise and post new version of protocol elements for review, including SDTM Trial Design Model elements
  - Support development of glossary elements relevant to protocol
  - Develop storyboards for 3 use cases: (SDTM, Trial Registry, Structured document authoring)
  - Goal to develop a structured XML representation of protocol derived from HL7 RCRIM DSAM
- Status:
  - Identified core elements
  - Supported DSAM modeling efforts
  - Working on creation of HL7 messages and/or CDA representation
  - Working with SDS (Trial Design), ADaM (Analysis Plan), and NCI
  - Developing implementation examples and storyboards.

# CDISC Operational Data Model (ODM)

- **Production Version 1.2.1 (XML Schema)**
  - Operational data/metadata/audit trail definition, exchange & archive
  - Supports use of ODM for eSubmissions metadata (define.xml)
  - New, improved extension mechanism
  - Harmonized with CDISC LAB model
- **Goals**
  - Revise ODM Mission to reflect need for stability, support and interoperability
  - Extend core model to support eCRFs, eSDI, submission data, compatibility with protocol model
  - Support ODM/HL7 RIM Harmonization
  - Contribute to CDISC cross-team projects
  - Cost-effective means of archiving, exchange and data retention of electronic data
- **Status**
  - Mission revised
  - Scope of ODM 1.3 defined; development in process (year's end)
  - Mapping to RIM and HL7 datatypes completed
  - ODM requirements represented in the BRIDG
  - Define.xml added to FDA study data specification as new XML standard for metadata
  - Under consideration for data submissions, and EDC Archive
  - Supporting RCRIM/NCI eDCI project

# ODM

- Models to support data acquisition, interchange and archiving of operational data
  - Includes audit trail
  - XML schema
  - Machine-readable metadata is as important as consistency and documentation.
  - Each ODM instance includes all the information, as machine-readable metadata, needed to import the clinical data into a clinical database
- SDTM supports the metadata flow from the operational database to regulatory submission

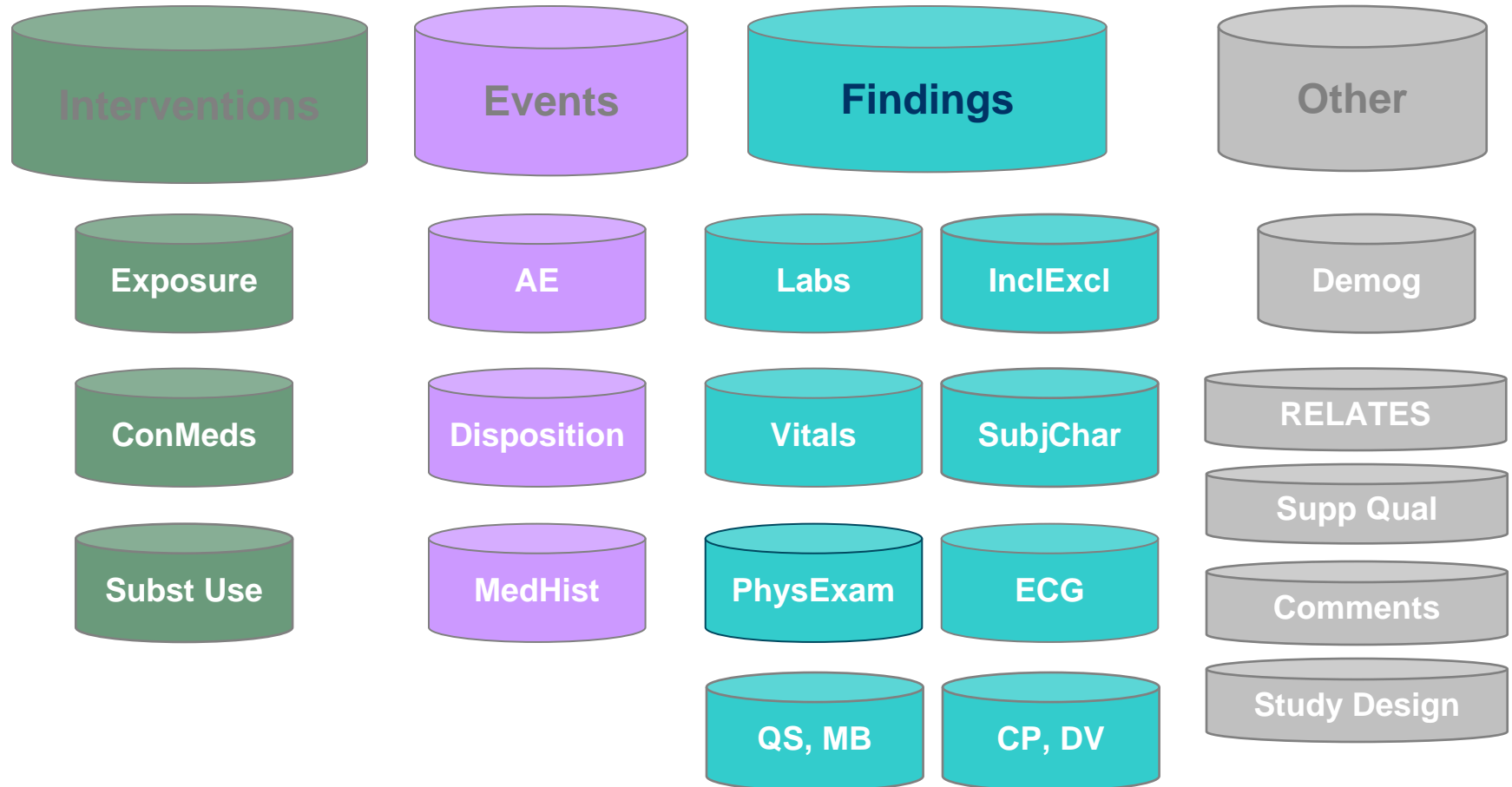
# CRF as a Metaphor for ODM

- ODM instance contains metadata and data.
- Conceptually, ODM metadata is an annotated case report form (CRF) formatted as XML.
  - Like an annotated CRF, ODM metadata describes the study events, such as visits, and types of data collected, such as adverse events, physical exams, or efficacy.
  - ODM metadata also describes the fields on the CRF with their attributes such as numeric or character, length, codelists, and so on.
  - The data part of an ODM instance contains the actual clinical data, formatted as XML, and laid out by subject, study event, type of data, and field
    - precisely the way clinical data are arrayed in a paper CRF.
- CRF is designed to transport data, from site to sponsor and to archive data in a file cabinet
- The XML version of a CRF, that is the ODM, is designed to transport data and to archive data. The analogy between CRF and THE ODM extends even further. T
  - But, the hierarchical, subject-study event structure of a CRF makes it a clumsy database for queries.
  - An ODM instance is clumsy database for the same reason, even though it is a good data transport and archiving format.

# Study Data Tabulation Model (SDTM)

- Production Version 1.1/3.1.1
  - SDTM 1.1 final released May 05
  - SDTM IG 3.1.1 final released Sept 05
  - Referenced as FDA eCTD Guidance specification since 21 July 04
- Goals
  - Release one maintenance update (as needed) per year
  - Complete trial design model (planned assessments and interventions)
  - Support definition of controlled terminology for SDTM
  - Define new domains (DV, MB, PK, DA) and applications (Devices, etc.)
  - Work with ADaM to clarify relationship with analysis datasets
  - Provide guidance on use of SDTM in operational databases
  - Support cross-team projects
- Status:
  - SDTMIG 3.1.1 recently released
  - Ongoing work on pre-specified events & interventions (e.g., Signs & Symptoms)
  - Recommended submission format for e-Submissions by CDER (CDER pilot in process)
  - First submission in Sept 2004 – several more planned for 2006
  - Supporting FDA during early stages of transition process.

# SDS V3.1 Fundamentals: CDISC Standard Domain Models and General Observation Classes



# The FDA without standards

- Submissions standards vary from sponsor to sponsor and within sponsor
- Interpretations of the guidance vary
- FDA reviewers must use multiple tools
- Lost time spent understanding the format
- Decodes and terminology are not standard
- Some sponsors are still sending hard copies

## **Review #1**

**1,100 pages**

**14.5 lbs**

**3 trees**

## **Review #2**

**100 pages**

**12.5 ounces**

**2 branches**





# How Standards Help the FDA

- In July 2004, the CDISC Study Data Tabulation Model (SDTM) was announced as the **FDA standard format for clinical trial submission data**
- **WebSDM™**, the Web Submission Data Manager, was developed for the FDA to offer SDTM-compliant data import, validation, review, and collaboration
- WebSDM™ helps both pharmaceutical companies and regulatory agencies take advantage of the efficiencies of standardized data and standardized processes to reduce cost and to increase the speed and quality of data capture and review.



## Select Application

User: BATCHAPP Administrator [admin]

### Available Applications:

ID	Name	Sponsor	Description	Created	FDA Review	Application Number	Application Type	Drug Name	Submission Type	Serial Number
<input type="radio"/> 0	Wonderdrug	Company 1		03/08/2005 20:03:35 EST	OND/OPaSS		NDA	Wonderdrug	Original Application	
<input checked="" type="radio"/> 1	Wonderdrug ISS	Company 1	Integrated safety summary data for Wonderdrug	03/24/2005 13:37:11 EST	OND/OPaSS		NDA	Wonderdrug	Original Application	
<input type="radio"/> 2	MaxiWonder	Company 2	MaxiWonder ISS	03/31/2005 11:54:29 EST	OND/OPaSS		NDA	MaxiWonder	Original Application	

Select Application

**NDA data loaded into a software program developed under a Cooperative Research and Development Agreement**

**Principal Investigators: Randy Levin, M.D., CDER and Wayne Kubick Vice-President, Lincoln Technologies, Inc**



# Study Data Domains

[Preferences](#) [Settings](#) [Feedback](#) [Exit](#) [Help](#)

[Home](#) [Select](#) [Domains](#) [Screening](#) [Subject Lists](#) [Reports](#) [Advanced](#) [Load&Check](#) [Run History](#)

User: BATCHAPP Administrator [admin], Application/Study: Wonderdrug ISS/ISS forWonderdrug

Application: Wonderdrug ISS      Study: ISS forWonderdrug      Sponsor: Company 1      Last Data Load: 03/25/2005 14:13:49 EST  
Last Run Name: Re-Run of Initial load of Wonderdrug IS

[View Complete Error Log](#)

Domain	Subjects	Description	Report	Download Rows	Variables	Structure Errors	Consistency Errors
<a href="#">AE</a>	<a href="#">2072</a>	<a href="#">Adverse Events</a>		<a href="#">10733 rows</a>	<a href="#">65</a>	0	<a href="#">29</a>
<a href="#">AE_SUPP</a>	<a href="#">2072</a>	<a href="#">Supplemental Qualifier Vars</a>		<a href="#">66593 rows</a>	<a href="#">10</a>	0	0
<a href="#">CM</a>	<a href="#">2801</a>	<a href="#">Concomitant Meds</a>		<a href="#">51027 rows</a>	<a href="#">61</a>	0	<a href="#">12</a>
<a href="#">CM_SUPP</a>	<a href="#">2783</a>	<a href="#">Supplemental Qualifier Vars</a>		<a href="#">454710 rows</a>	<a href="#">10</a>	0	0
<a href="#">DM</a>	<a href="#">2826</a>	<a href="#">Demographics</a>		<a href="#">2826 rows</a>	<a href="#">32</a>	0	<a href="#">4</a>
<a href="#">DS</a>	<a href="#">2826</a>	<a href="#">Disposition</a>		<a href="#">12303 rows</a>	<a href="#">34</a>	0	<a href="#">4</a>
<a href="#">EG</a>	<a href="#">2768</a>	<a href="#">ECG</a>		<a href="#">82727 rows</a>	<a href="#">52</a>	0	<a href="#">4</a>
<a href="#">LB</a>	<a href="#">2819</a>	<a href="#">Lab</a>		<a href="#">452750 rows</a>	<a href="#">60</a>	0	<a href="#">4</a>
<a href="#">MH</a>	<a href="#">2826</a>	<a href="#">Medical History</a>		<a href="#">12331 rows</a>	<a href="#">56</a>	0	<a href="#">4</a>
<a href="#">MH_SUPP</a>	<a href="#">2816</a>	<a href="#">Supplemental Qualifier Vars</a>		<a href="#">23577 rows</a>	<a href="#">10</a>	0	0
<a href="#">OT</a>	<a href="#">2506</a>			<a href="#">16037 rows</a>	<a href="#">57</a>	0	<a href="#">3</a>
<a href="#">SC</a>	<a href="#">2818</a>	<a href="#">Subject Characteristics</a>		<a href="#">226047 rows</a>	<a href="#">32</a>	0	0
<a href="#">VS</a>	<a href="#">2797</a>	<a href="#">Vital Signs</a>		<a href="#">110680 rows</a>	<a href="#">48</a>	0	0

Maximum Error Severity Levels:



Done

Local intranet



# Variable Characteristics (AE)

Preferences Settings Feedback Exit Help

Home Select **Domains** Screening Subject Lists Reports Advanced Load&Check Run History

User: BATCHAPP Administrator [admin], Application/Study: Wonderdrug ISS/ISS for Wonderdrug

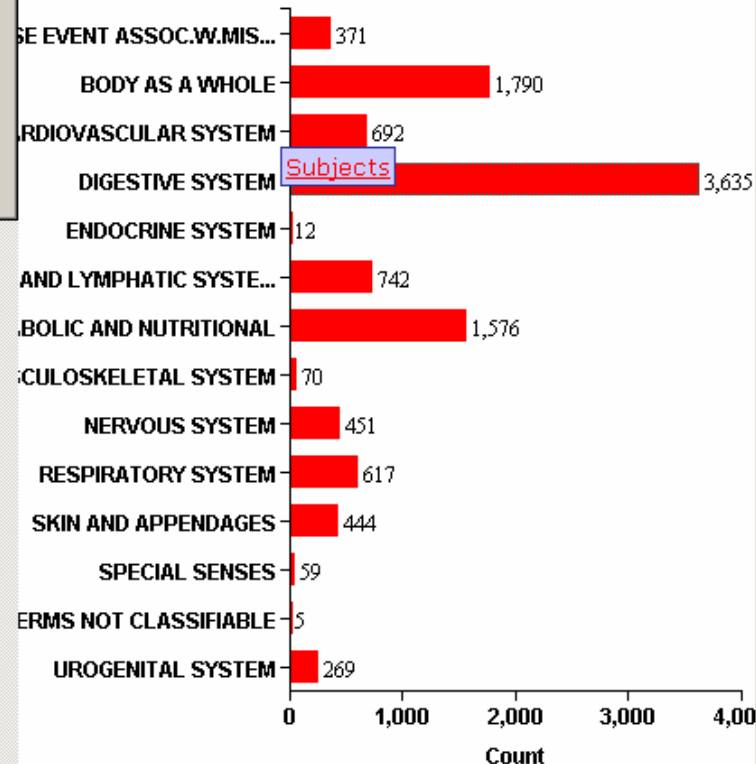
[Back](#)

## Variable Summary

Variable	N	Nulls	Mean	SD	Min	Max	Type
<a href="#">AEACN</a>	5492	5241					Character
<a href="#">AEACNOTH</a>	0	10733					Character
<a href="#">AEBODSYS</a>	10733	0					Character
<a href="#">AECAT</a>	0	10733					Character
<a href="#">AECONTRT</a>	10713	20					Character
<a href="#">AEDECOD</a>	10733	0					Character
<a href="#">AEDUR</a>	0	10733					Character
<a href="#">AEDURH_</a>	0	10733					Date
<a href="#">AEDURL_</a>	0	10733					Date
<a href="#">AEDURP_</a>	0	10733					Number
<a href="#">AEENDTC</a>	9453	1280					Character
<a href="#">AEENDTH_</a>	9453	1280					Date
<a href="#">AEENDTL_</a>	9453	1280					Date
<a href="#">AEENDTP_</a>	9453	1280	67019.613	74899.388	60	2678400	Number
<a href="#">AEENDY</a>	9451	1282	9.546	13.907	-6	378	Number
<a href="#">AEENRF</a>	0	10733					Character
<a href="#">AEGRPID</a>	0	10733					Character
<a href="#">AELOC</a>	0	10733					Character

[Help](#)

## AEBODSYS

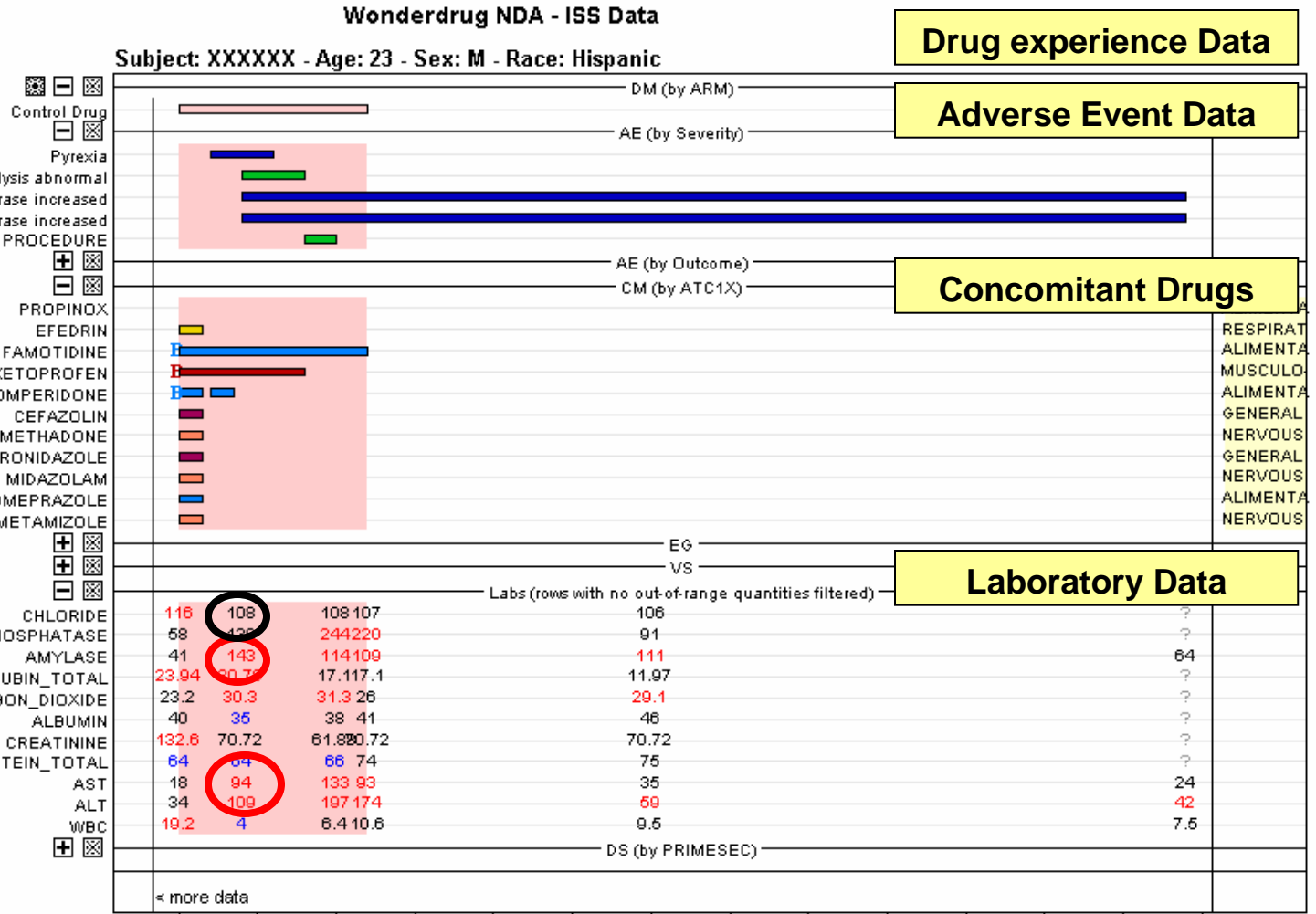


javascript:smcases("csFirstLevelDrilldown.jsp?classname=com.lincolntechnologies.webchecker.DomainSubjectListGatherer&configid=3

Unknown Zone (Mixed)

# Assessing Potential Liver Injury by Analyzing Increases in Serum Alanine Aminotransferase (ALT) and Total Serum Bilirubin (TBILI) IN ONE STEP

**Individual Patient Profile:**  
Linkage of several data tables using the same timeline



**X-axis: Days into Study**

# CDISC Analysis Dataset Modeling Team (ADaM)

- Standard models for regulatory submission of analysis datasets to facilitate statistical reviews
  - Providing input to FDA on forthcoming guidance on analysis dataset submissions
- Goals
  - Revise models to conform with SDTM concepts
  - Develop additional analysis submission models
  - Define elements for analysis-level metadata
  - Work on standards for submitting analysis programs
  - Support cross-team efforts (SDS, Define.xml, protocols)
- Status:
  - General Considerations Document posted in January; update in progress
    - Being revamped into a new format with principles and examples
  - Updated/New Models posted for Subject-level Analysis, Categorical, Change from Baseline
  - Adverse Events and linear models developed and soon to be posted' linear models in development
  - Working with SDS on guidelines for representing analysis datasets in SDTM format

# SDTM versus ADaM

- SDTM
  - Source Data
  - Vertical
  - No redundancy
  - Character variables
  - Each domain is specific to itself
  - Dates are ISO8601 character strings
- ADaM
  - Derived Data
  - Structure depends on analysis
  - Redundancy is needed for easy analysis
  - Numeric variables
  - Combines variables across multiple domains
  - Dates are formatted as SAS dates to allow manipulation



# Controlled Terminology Team

- Initiated in May 2004 to support terminology needs for all CDISC models/teams
  - Includes representation from CDISC (pharma), US Govt, Academic
  - Meet CDISC model terminology needs by surveying and evaluating available vocabularies and recommending adoption of specific terminology
- Goals
  - Support terminology needs of SDTM, HL7 and other CDISC models
  - Develop code list recommendations for other Clinical Research needs
  - Increase awareness and adoption
  - Support cross-team initiatives
- Status
  - Defined process for defining and approving CT
  - Utilizing NCI-caDSR toolset
  - Package 1 of SDTM terms assessed against EVS and ready for release; Package 2 in progress
  - SDTM 8-Char Lab TESTCD list under final internal review

# Terminology: Examples

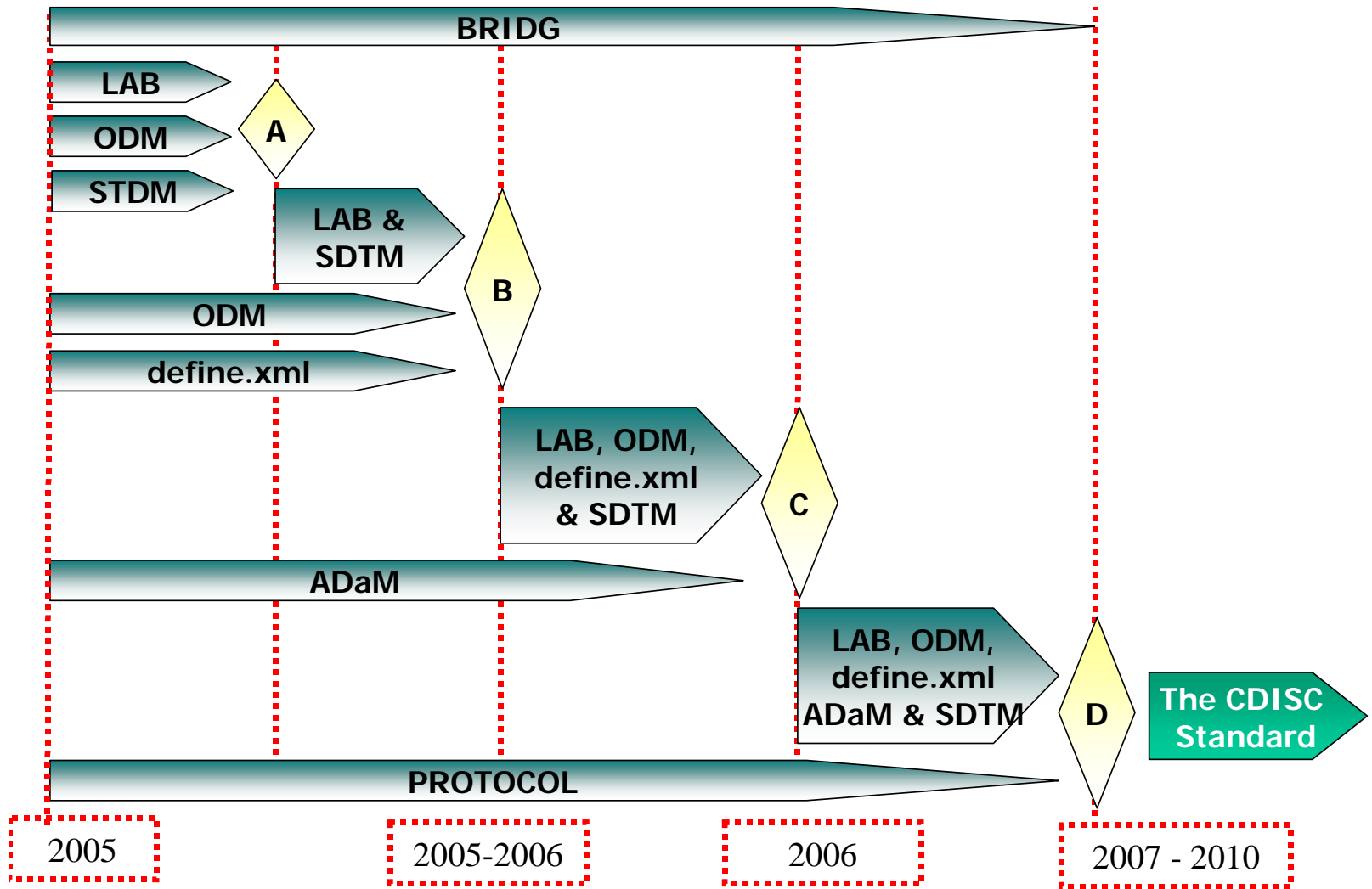
## Package-1 Content (N=32)

<ul style="list-style-type: none"><li>• Action Taken with Study Treatment</li><li>• Age Units</li><li>• Country</li><li>• Domain Abbreviation</li><li>• Ethnicity</li><li>• Dose Form</li><li>• Identification variable</li><li>• Category for Inclusion/Exclusion</li><li>• Not Done or Null Answer</li><li>• No / Yes/ Unknown Answers</li></ul>	<ul style="list-style-type: none"><li>• Outcome of Event</li><li>• Race</li><li>• Reason for Non-Completion</li><li>• Causality</li><li>• Route of Administration</li><li>• Severity / Intensity</li><li>• Sex</li><li>• Size Code</li><li>• System Organ Class MedDRA</li><li>• Relation to Reference Period</li><li>• Age Group</li></ul>	<ul style="list-style-type: none"><li>• Type of Control</li><li>• Description of Trial Design</li><li>• Diagnosis Group</li><li>• Trial Indication Type</li><li>• Trial Blinding Schema</li><li>• Standard Toxicity Grade</li><li>• Trial Phase</li><li>• Trial Summary Parameter</li><li>• Type of Trial</li><li>• Vital Signs Test Name</li><li>• Units for Vital Signs</li></ul>
--	---	---

# The CDISC Roadmap

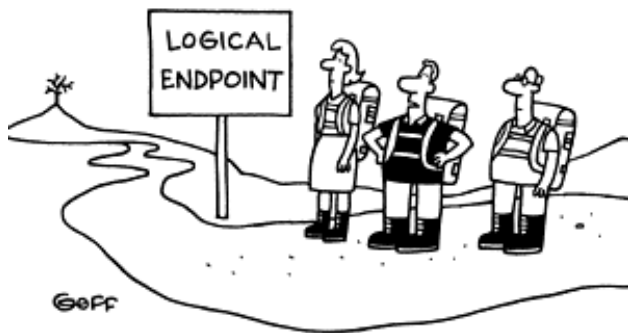
- Purpose:
  - To provide a concise, common specification of all technical products to be developed by CDISC.
- Endpoint:
  - By 2008, there will be a **single CDISC standard** for the full life-cycle of a clinical trial or study from protocol representation through the capture of source data to submission and archive, comprising a set of fully integrated and consistent models which will form logically and organically from our current set.
- Success Criteria
  - **All submissions to the FDA** are being made using the CDISC standard;
  - The set of CDISC models in use across the **full life-cycle of clinical trials**; and
  - The CDISC standard being **globally adopted**.

# Roadmap Timetable



# CDISC Conclusions

©2000 Ted Goff tedgoff@tedgoff.com http://www.tedgoff.com



"I say we take down the sign and carry it with us."

- ODM, LAB, SDTM/SDS are READY!!
  - Keep stable so industry can catch up
- Work proceeds on others
- We're moving to the CDISC Standard via the "CDISC Roadmap"
- We're working with others: FDA, NCI, others
- Data standards improved effectiveness & efficiency
- CDISC is leading the standards movement for clinical trial data.
- Learn the standards, use them and have your partners/vendors use them.
- Join CDISC and participate; support the cause....

*Get Involved Now!*

*ARE WE DONE YET??????????????????*

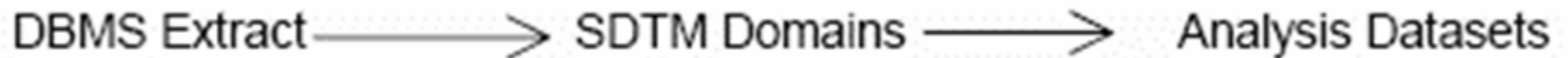
# Discovering the Ways Standards will Change Organizational Management of Data

# CDISC standards vs. organizational management of data

- **Organizational management, how does it exist in most companies?**
  - Multiple clinical trial systems and databases
  - Submission data from different departments
  - Legacy data
- **Where do you want to implement the CDISC standards?**
  - Implement from the beginning
  - Implement at the back end

# Data Flow: ADaM and SDTM

## Linear Method



## Parallel Method



## Retrospective Method



## Hybrid Method





# Implement CDISC at the beginning

- Start as early as CRF design
- Integrate CDISC formats into your database design
- Create CDISC files directly from the source database
- One set of files and variables names

# Implementing at the beginning - Pros

- Raw database = SDTM data
- 100% adherence to CDISC standards
- SDTM is immediately available
- Fewer steps to achieve a standard format
- Reporting and analysis programs streamlined
- One set of metadata

# Implementing at the beginning - Cons

- Expensive to undertake when starting at CRFs design
- Technically may not be feasible if integrating into existing system
- Why throw away perfectly good CRF designs?
- Why compromise front end systems?
- Does not answer the need for current and legacy studies
- Cross tabulation is still a problem

# Implement CDISC at the end

- Doesn't disturb the design or process of your current studies
- Simply convert all the analysis files to the standard CDISC format
- A good solution for legacy data

# Implement at the end - Pros

- No programming added to existing or legacy systems
- No knowledge needed of the existing systems
- Applied to files from any system
- Can be easily outsourced

# Implement at the end - Cons

- Does not upgrade systems to make them compliant with CDISC standards
- Process must be replicated for each
- A constant conversion activity is not efficient
- When the standards change, maintenance is required in several places.
- Additional cost for each study converted to CDISC
- There is a delay before SDTM files are available
- Data points can be lost in translation.
- Two versions of the metadata exist

# Finding a successful approach.

- Despite the stated drawbacks, a backend approach can be used successfully for a variety of studies and formats
- Implementation is flexible enough to adapt to different study designs
- Minimal programming support required for maintenance
- Reasonable cost

# Our approach

- Starting point is the **analysis files** produced from any database system
  - They contain all the required analysis data
  - They are the basis for applying the CDISC standards
- Compare the analysis files to the CDISC domains
  - Ensure that required CDISC variables are available
  - Understand all derivations



# How it works

- Create an ETL Process
  - Define how analysis data fits into the CDISC domains and standard variables
  - Match data to required, and permitted and expected CDISC data when possible
  - Provide an automated mechanism for specifying the data sources and algorithms
  - Basis for the FDA-mandated “DEFINE.PDF” documentation
  - Provide the metadata for the CDISC files

# How it works (cont.)

- Implementing an ETL Process
  - Programs read table driven metadata to translate the analysis data into CDISC formats
  - The metadata tells the code which analysis variables populate the CDISC variables
  - The metadata indicates when specialized code is required
  - All code is developed to be generic using the metadata to indicate when variations were required
  - New studies only require changes to the metadata

# Why was this approach successful?

- Does not disrupt existing clinical trial systems
- It works for legacy data
- Cost each time is minimized since only the metadata needs to change for each new study
- Delay producing subsequent SDTM files is minimized since only new study specific situations cause additional coding
- Changes to CDISC standards are made to the metadata avoiding costly programming revisions

# Why was this approach successful? (cont.)

- Code is reusable since macros are created to produce the CDISC standard formats..
- The metadata serves as a reference to the origin of a data point
- Table driven metadata provides automatic documentation

# Our challenges

- Missing raw data
  - We had to add coding to return to the raw data to capture missing data
  - Example: Inclusion/Exclusion data
- Ambiguous raw data
  - CM file, variables CMSTRF and CMENRF have controlled terminology of Before, During and After
  - Since much of the sponsor's concomitant therapy data had missing or incomplete dates, we had to provide more sophisticated routines to determine these controlled terms

# Our challenges (cont.)

- Trial Design and Subject domains
  - Trial Design and Subject domains had to be created manually since this is not captured in analysis files
- Normalized data
  - CDISC formats result in a vertical view of the data
  - Related variables reside in separate rows
  - Data must be transposed into a horizontal format to make it useful for standard TLGs to process correctly.

# Lessons Learned and Positive Outcomes

- ETL is a tool that can be used as a means to automation as well as documentation
- ETL process works most successfully when analysis files are carefully studied in relation to CDISC domains.
- Library of customized routines to convert non-standard data into CDISC version 3.1 variables
- Creating generic code that is manipulated by the metadata, decreases the amount of programming effort.

# The Competitive Advantages that Stem from Adherence to the Standards



# The competitive advantages that stem from adherence to the CDISC standards

- Complies with FDA guidance and goals towards a submission standard
- Integrated Summary Reports are easily facilitated using the CDISC standard analysis files for all studies
- One set of Metadata
- Standard macros are developed once and reused
- Reports can be standardized due to standard input files
- ADaM files can be generated from SDTM files

# Documentation & References

- SDTM & SDS V 3.1.1 documentation
  - <http://www.cdisc.org/standards/index.html>
- eCTD - Latest backbone + description, see Section 4, p. 16 of 5640dft.pdf (11 Mar 2004)
  - <http://www.fda.gov/cder/regulatory/ersr/ectd.htm>
- HL7 Regulated Clinical Research Information Management (RCRIM) - Charter, minutes, leadership, interest areas, etc.
  - <http://www.hl7.org/Special/committees/rcrim/rcrim.htm>
- For latest announcements and guidance docs:
  - <http://www.fda.gov/cder/regulatory/ersr>