Data Standards & Clinical Data Interchange Standards Consortium (CDISC)

Carol Baker, Jeff Abolafia, Rob Woolson
May 20, 2012

Significant Contributor:
Frank DiIorio, CodeCrafters, Inc., Philadelphia PA

Part 1

Overview and Background

Presentation Overview

- Why Data Standards and CDISC?
- Background
- CDISC Overview/Basics
- A Closer Look at the Study Data Tabulation Model (SDTM)
- SDTM Interactive Exercises
- A Closer Look at the Analysis Dataset Model (ADaM)
- ADaM Interactive Exercises
- How CDISC Affects Your Work
  - **Transitioning to CDISC**
  - **Workflow**
  - **Deliverables**
  - **Implementation Challenges**
  - **CDISC Benefits**
- Resources / Questions

CDISC Basics

What is CDISC?

CDISC stands for **C**linical **D**ata **I**nterchange **S**tandards **C**onsortium.

- Standards organization formed in 1997 as a volunteer group
- CDISC is an open, multidisciplinary, non-profit organization
- Now supported by >150 member companies including pharmaceutical companies, biotech companies, CROs/service providers, and technology providers
- CDISC has established worldwide industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trials data and metadata for medical and biopharmaceutical product development

Extracted from the Clinical Data Interchange Standards Consortium (CDISC) website [http://www.cdisc.org/about/index.htm].

The CDISC Vision

"The exchange of all clinical trial data between any two parties will be achieved by the application of the appropriate CDISC data models and standards."

The CDISC Mission

"To develop and support global, platform independent standards that enable information system interoperability to improve medical research and related areas of healthcare."

Why CDISC and Data Standards

- FDA Perspective
  - So much data so little time
  - Data management or drug evaluation
  - Putting it all together

- Sponsor Perspective
  - Corporate standard
  - Standard inputs
  - Standard outputs
  - Lab data – ugh!!!
  - Outsourcing made easier

Common Terminology

CDER Center for Drug Evaluation and Research (FDA)

CBER Center for Biologics Evaluation and Research (FDA)

eCDT Electronic common technical document

PDUFA Prescription Drug User Fee Act

Why CDISC and Data Standards

NDA Submissions

	FY2007	FY2008	FY2009	FY2010	FY2011*
NDA Total	23,310	22,308	22,148	22,443	5,763
NDA Electronic	8,771	11,272	13,297	15,497	4,283
NDA Electronic %	37.63%	50.53%	60.04%	69.05%	74.32%
NDA eCTD	2,085	7,410	11,146	14,007	3,857
NDA eCTD % of Total	8.94%	33.22%	50.33%	62.41%	66.93%
NDA eCTD % of Electronic	23.77%	65.74%	83.82%	90.39%	90.05%

*Through 12/31/10

Drug Information Association www.diahome.org 9

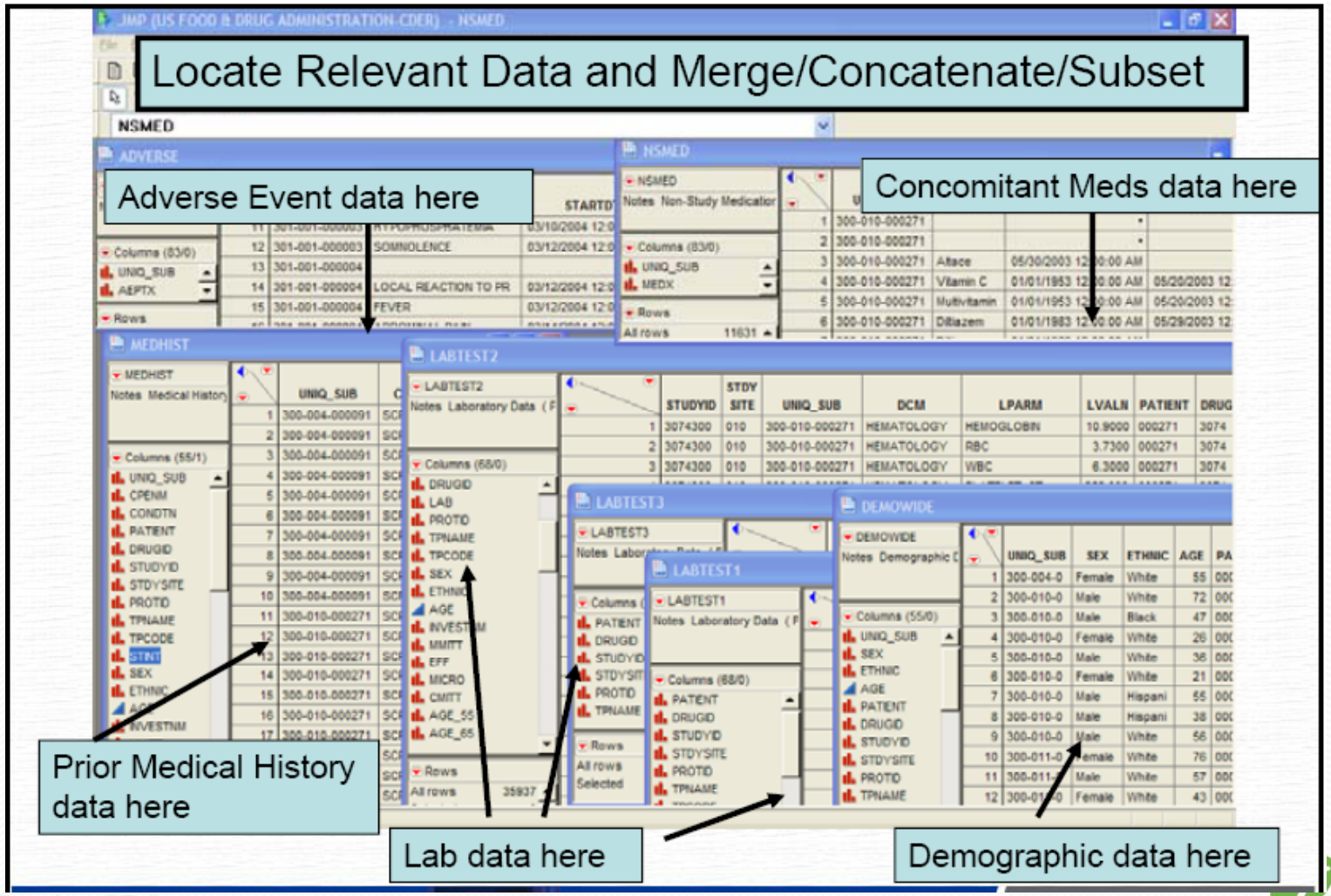
Why CDISC and Data Standards

- Need data standards for electronic submissions
- To support an NDA clinical (raw) and analysis databases submitted to FDA
- Clinical database
  - Typically 30-40 datasets per study
  - Each represents a different domain (Labs, AE)
- No easy way to re-assemble data
- Reviewers perform lots of data manipulations
- Extremely inefficient and error prone

Why CDISC and Data Standards

- Example:
  - **Simple Review Question : Who are the patients with Liver Function tests (ALT) over 3x Upper Limit of Normal?**
  - Were there any confounders?
  - Any serious hepatic adverse events?
  - Extremely demanding data manipulations required to answer this simple question

Why CDISC and Data Standards



Rho
Giving flight to research

Why CDISC and Data Standards

Study #1 – demog.xpt

SUBJID	SEX
0001	M
0002	F
0003	F
0004	M
0005	F

Study #2 – dmg.xpt

ID	GENDER
A1	Male
A2	Male
A3	Female
A4	Female
A5	Male

Study #3 – axd222.xpt

USUBID	SEX
00011	0
00012	1
00013	1
00014	0
00015	1

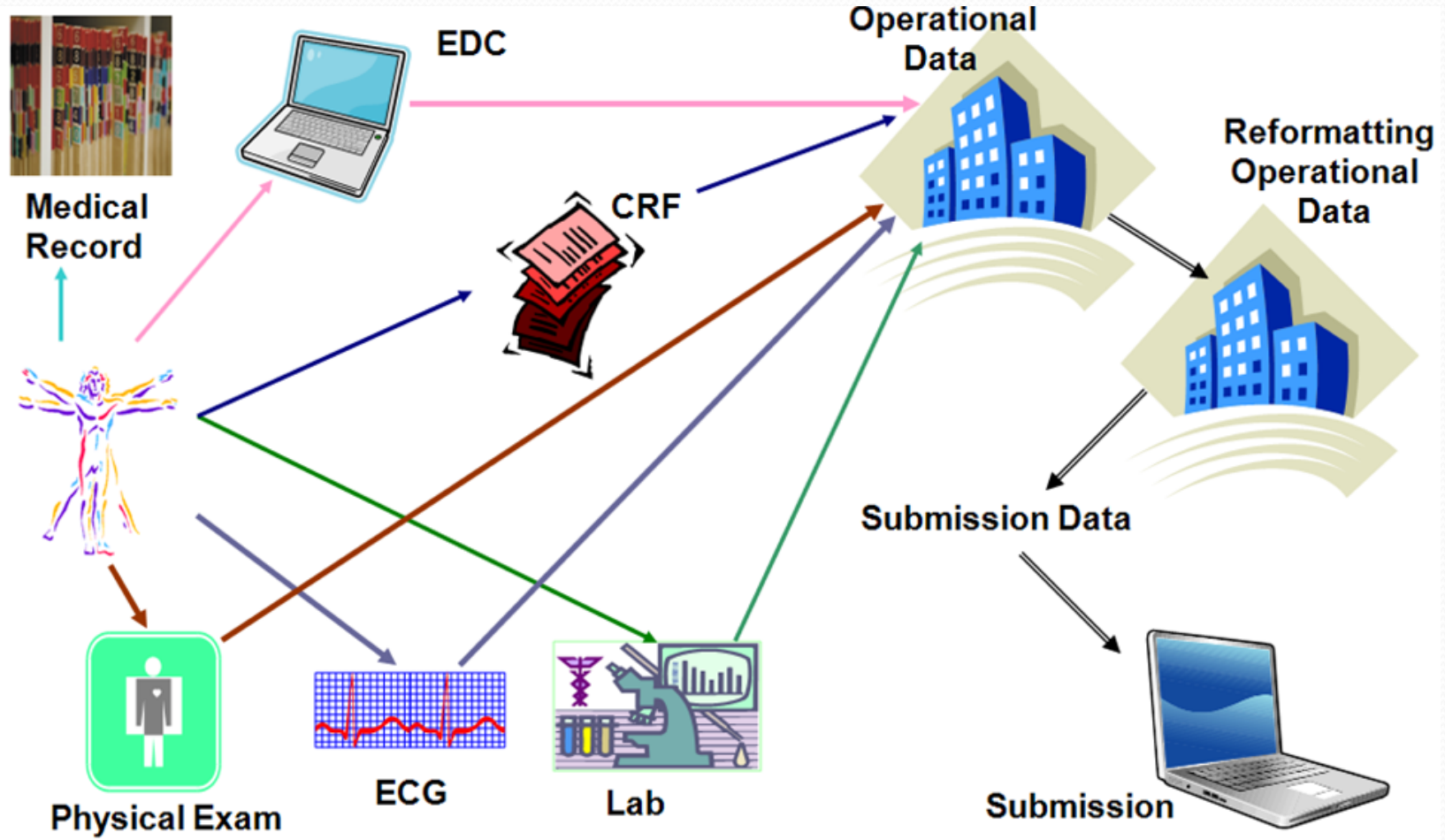
Study #4 – dmgph.xpt

PTID	GENDER
0001	1
0002	1
0003	2
0004	2
0005	1

Why CDISC and Data Standards

- Difficult to reproduce results or show what was done
- Error prone process
- Manual dataset creation (↓ productivity) → tedious, frustrating
- Great deal of time spent trying to answer one question
- Large number of issues/questions unexplored
- Inability to perform efficient and effective drug evaluation cross life cycle
- Many best practice safety questions not routinely asked during NDA review
- Reviews completed under time constraints…little time to think

Current State of Clinical Data Transfer



Rho
Giving flight to research

CDISC and Data Standards: The Vision

- **CDER Computational Science Center**
  - Formed in 2009
  - Provide CDER reviewers a more aligned and automated method for completing reviews
  - Develop modern computing tools
  - **Establish a comprehensive data standards program**
  - **Adoption and enhancement of CDISC standards**
  - Must have standards to use tools
  - **Data Standards Plan**
- CBER – has developed an infrastructure to receive CDISC submissions

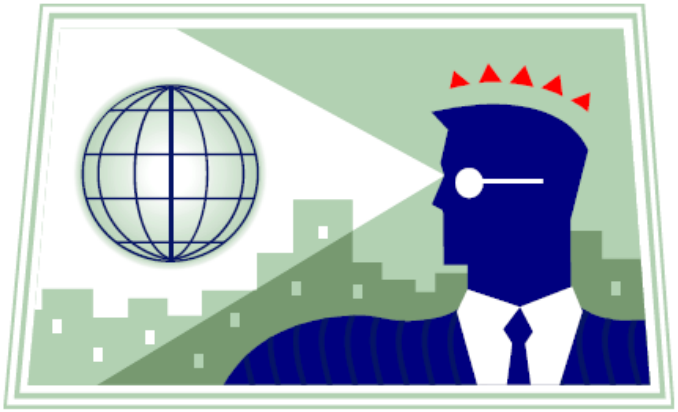
Data Standards: New Guidance

- FDA released new draft guidance in February 2012
- Promotes use of data standards in submissions to FDA
- Announces FDA's intention to propose a new Federal regulation that would require the submission of standardized electronic study data
- IND or IDE should include sponsor's plan to submit standardized data to FDA
- Submission should describe in the cover letter how study data standardization plan was implemented

CDISC and Data Standards: The Vision

Electronic submissions+ better data + better tools…

FDA staff will have the ability to access and associate all relevant data and to conduct consistent and rigorous analyses necessary to answer key regulatory questions, within a committed timeframe.



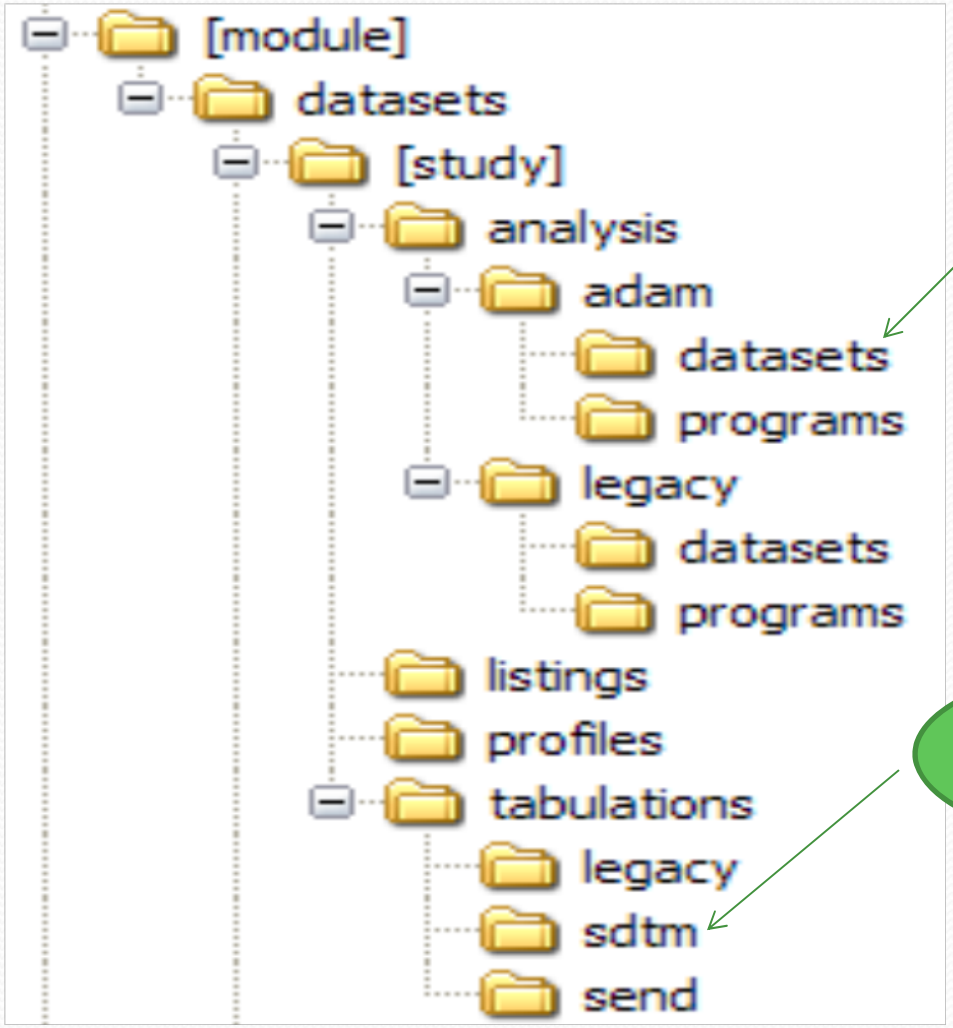
What does CDISC mean for FDA: Summary

- More standardized datasets
- Standardized documentation
- Ability to use standard review tools
- Less time to understand the submission database
- Faster and higher quality review
- **More efficacious and safer drugs**

FDA: Trends & Preferences

- Trend toward E-Submissions
- “E-Submission data needs standards”
- Number of CDISC submissions is increasing
- Preferred and requested by many reviewers
- FDA developing comprehensive data standards plan
- FDA has invested in CDISC – time/training/tools
- CDISC standards part of PDUFA plan

Submitting Clinical Data



ADaM Datasets, Define.XML, Define.PDF

SDTM Datasets, Define.XML, aCRF

How CDISC Affects Your Work

Clinical Trial Activities

Study Start-up	Study Conduct	Analysis/Reporting	Submission
Study design	Patient recruitment	Data analysis	ISS/ISE prep
Protocol development	Data acquisition	Safety analysis	Clinical-Stat report
CRF development	Site monitoring/audits	Analysis table prep	TLFs
DB structure/validation	Lab/ECG data transfer	Clinical assessments	eCTD file structure
Edit checks/validation	Site audits	Report generation	
Lab/ECG specs	Database QC/DBL		
Site/PI identification	Analysis Programming		
Patient recruitment plan	Site evaluation/initiation		
Critical documents	Initial statistical tables		
IRB approval	Study closeout		
Training of teams/sites			
Randomization plan			
Test article prep			
Statistical analysis plan (SAP)			
Analysis table shells			

Impact of CDISC Standards on Clinical Trial Activities

Study Start-up	Study Conduct	Analysis/Reporting	Submission
Study design	Patient recruitment	*Data analysis*	*ISS/ISE prep*
Protocol development	*Data acquisition*	*Safety analysis*	*Clinical-Stat report*
CRF development	*Site monitoring/audits*	*Analysis table prep*	*TLFs*
DB structure/validation	*Lab/ECG data transfer*	*Clinical assessments*	*eCTD file structure*
Edit checks/validation	*Site audits*	*Report generation*	
Lab/ECG specs	*Database QC/DBL*		
Site/PI identification	*Analysis Programming*		
Patient recruitment plan	*Site evaluation/initiation*		
Critical documents	*Initial statistical tables*		
IRB approval	*Study closeout*		
Training of teams/sites			
Randomization plan			
Test article prep			
Statistical analysis plan (SAP)			
Analysis table shells			

Clinical study activities affected by (and streamlined by) CDISC standards

What does CDISC mean for Sponsors?

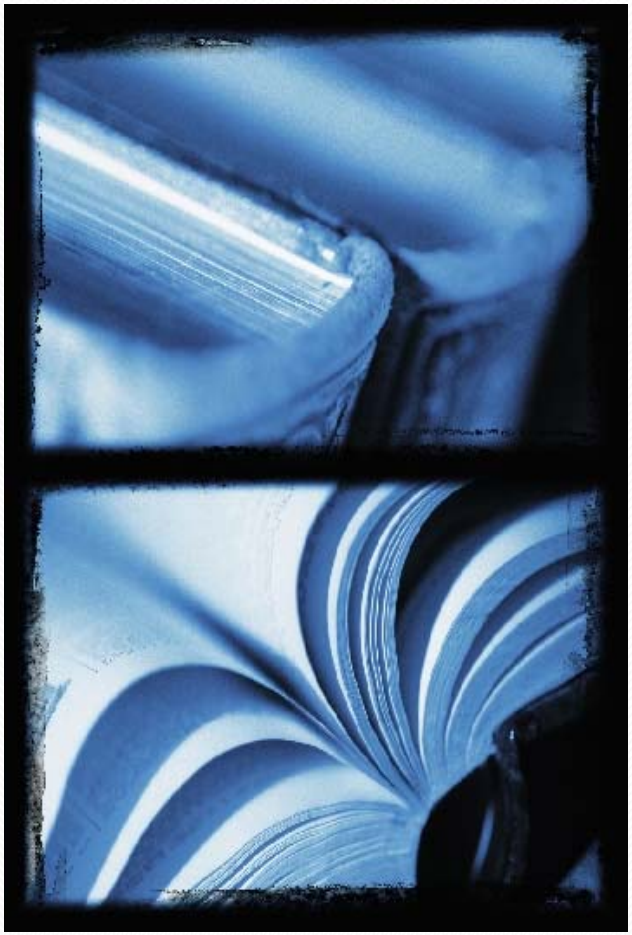
Short Term:

- More work
- Need to assimilate standards into processes
- More project parts to consider in the timeline
- Different workflow process
- Need to build new tools
- Lots of training
- New set of deliverables = More profit/business opportunities
- More coordination needed for more deliverables
- Legacy Conversions

Long Term:

- Corporate standard
- Standardized datasets across studies and therapeutic areas
- Facilitates data exchange with multiple partners
- More efficient in the long term
- Ability to use standard tools/processes
- Facilitates data integration
- Faster and higher quality review

Main CDISC Standards & Models



- Common CDISC Terminology
- CDASH
- SDTM
- ADaM
- Differences between SDTM and ADaM
- A quick look at ODM
- Other CDISC standard models

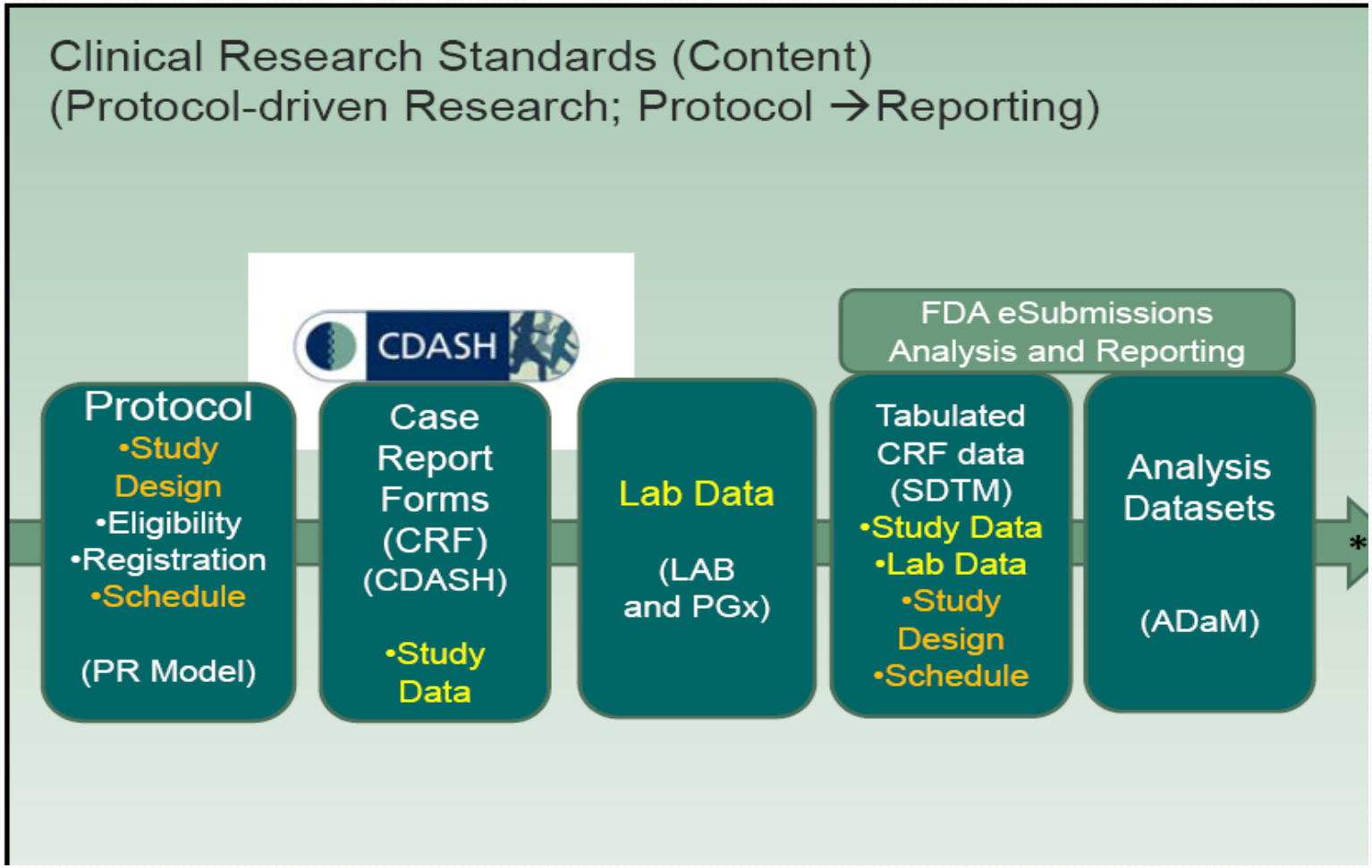
CDISC Models

- **Study Data Tabulation Model (SDTM)**
  - Version 3.1.1 and 3.1.2 accepted by FDA
  - Referenced as specification in FDA Guidance on eSubmissions for Implementation of ICH eCommon Technical Document
- **Analysis Dataset Models (ADaM)**
  - ADaM Version 2.1 and Implementation Guide Version 1.0 released 12/09
  - Accepted by FDA
- **Clinical Data Acquisition Standards Harmonization (CDASH)**
  - recommended basic standards for the collection of clinical trial data
  - Version 1.1 released January 2011
- **Operational Data Model (ODM)**
  - Production Version 1.3
  - XML schema
  - Part of eCTD data specifications
- **Protocol Representation Model**
  - Version 1.0 released May 2009
  - Spreadsheet of protocol elements with definitions; documentation; initial HL7 model
- **Laboratory Data Model (LAB)**
  - Production Version 1.0.1
  - Implementations through SAS, ASCII, XML/ODM and HL7 V3 RIM message
- **Standards for the Exchange of Non-clinical Data (SEND)**
  - Version 2.3 released November 2005. Based upon CDISC SDS V3.1
  - Included in SDTM model now referenced in FDA Guidance

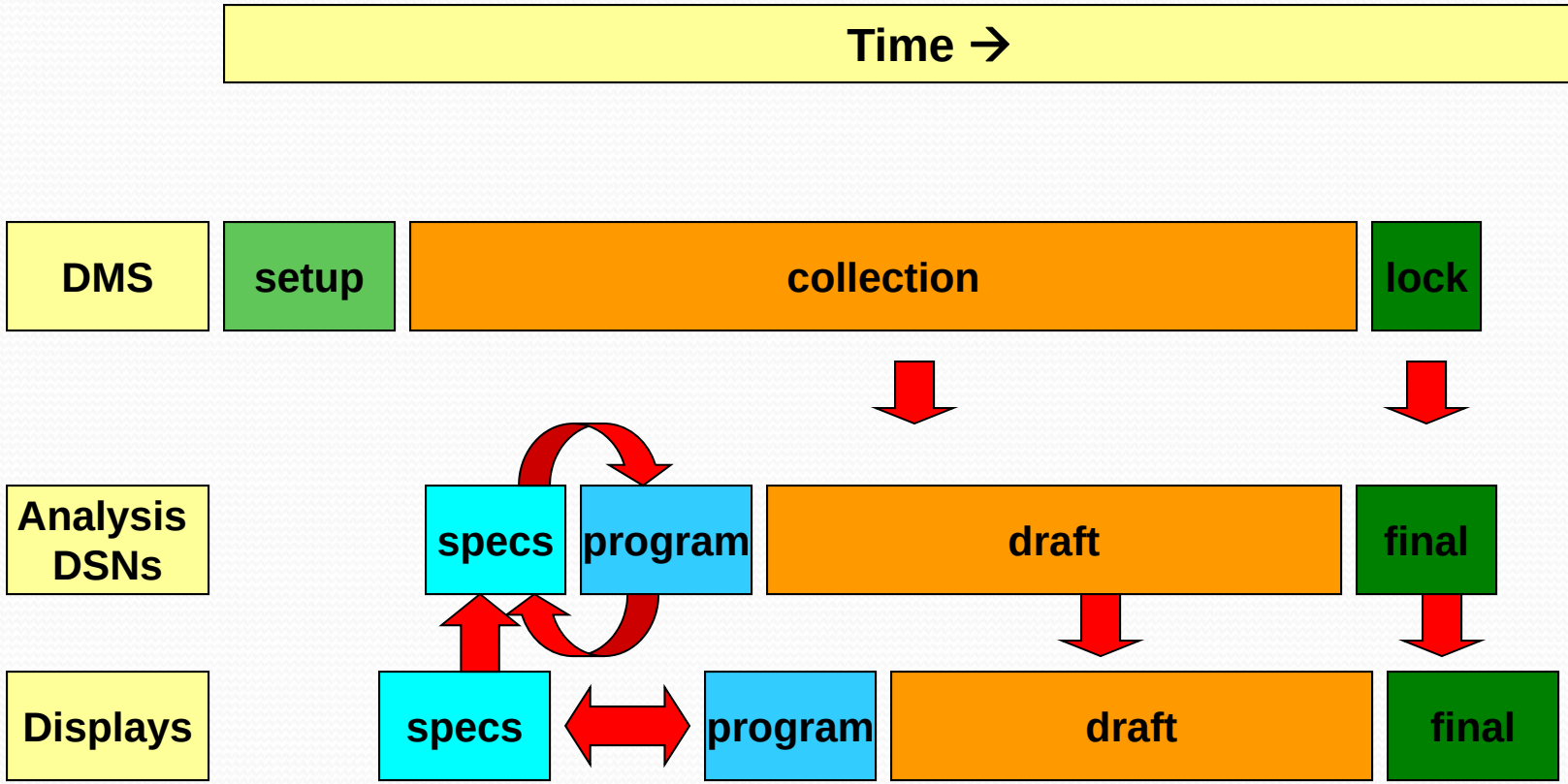
CDISC Standards & Models

Need to Know:	Nice to Know:
Clinical Data Acquisitions Standards Harmonization (CDASH) - Focus on standardizing definitions of CRF and data collection standards	**Operational Data Model (ODM)** - Production Version 1.2. - XML schema
Study Data Tabulation Model (SDTM) -Version 3.1.1/3.1.2 accepted by FDA - Referenced as specification in FDA Guidance on eSubmissions for Implementation of ICH eCommon Technical Document	**Laboratory Data Model (LAB)** - Production Version 1.0.1. - Implementations through SAS, ASCII, XML/ODM, and HL7 V3 RIM message.
Analysis Dataset Model (ADaM) - Version 2.1 -Provides guidelines and examples for analysis datasets - Accepted by FDA	**Standards for the Exchange of Non-clinical Data (SEND)** - Based upon CDISC SDS V 3.1. - Included in SDTM model
	Protocol Representation Model - HL-7 CDISC Collaboration. - Spreadsheet of protocol elements with definitions & documentation.

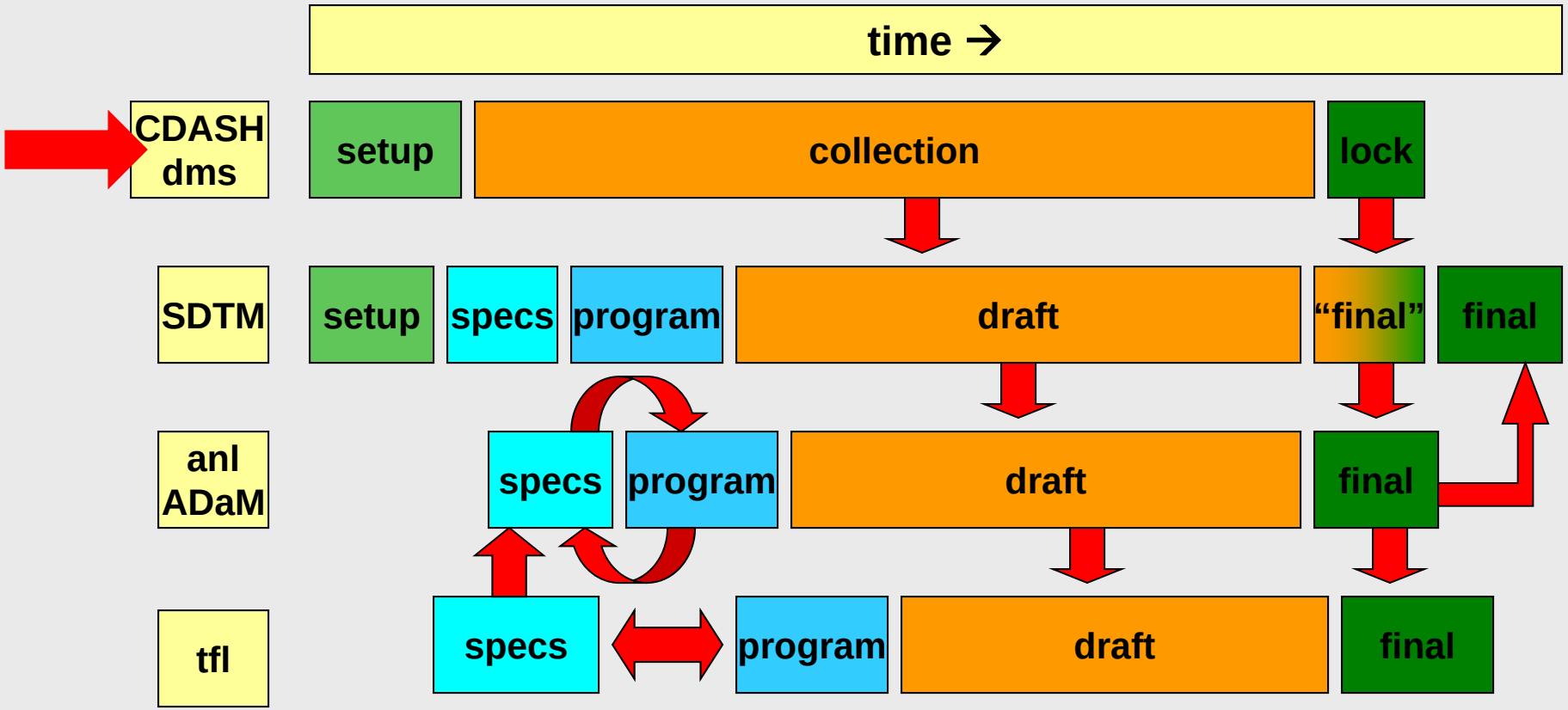
CDISC Models



Pre-CDISC Clinical Trial Workflow



Workflow for CDISC Project



Clinical Data Acquisitions Standards Harmonization (CDASH)

- CDISC released **Clinical Data Acquisition Standards Harmonization (CDASH)**
- One of many CDISC models
- Key component of end-to-end standards
- Moves standards upstream to **clinical data collection**
- Standard for data collection
- "Content" Standard
- Goal is to develop a set of global 'content standards' (element name, definition, metadata) for a core set of global data collection fields

Clinical Data Acquisitions Standards Harmonization (CDASH)

- Identifies a basic set of data collection fields present on most CRFs.
- Sponsors can add additional data collection fields to capture specific data points specified in the protocol or to satisfy certain regulatory requirements.

- Version 1 has content-driven direction for data collection across 16 domains including demography, adverse events, and safety domains common across therapeutic areas.
- Facilitates mapping data to SDTM

CDASH Domains

- Common identifying and timing variables
- Inclusion and Exclusion Criteria (IE)
- Adverse Events (AE)
- Laboratory Test Results (LB)
- Comments (CO)
- Medical History (MH)
- Prior and Concomitant Medications (CM)
- Physical Examination (PE)
- Demographics (DM)
- Protocol Deviations (DV)
- Disposition (DS)
- Subject Characteristics (SC)
- Drug Accountability (DA)
- Substance Use (SU)
- ECG Test Results (EG)
- Vital Signs (VS)
- Exposure (EX)

A Look at the CDASH Standard

	Question Text	Prompt	SDTM or CDASH Variable Name	BRIDG	Definition	CRF Completion Instructions	Information for Sponsors	Core
5	What is the date of collection?	Collection Date	DMDAT	PerformedActivity .dateRange	Date of collection.	Record the date the demographics data were collected in this format (DD-MON-YYYY).	The date of collection may be derived from the date of visit and if so, a separate date field is not needed. **For the SDTM-based dataset**, the SDTM IG variable DMDTC is derived via the CDASH Date of collection (DMDAT) and converting to the ISO 8601 format. (See AGE Additional Information for Sponsors.) This field does not map directly to an SDTM variable.	R/C
6	What is the sex of the subject?	Sex	SEX	BiologicEntity.administrativeGenderCode*	The assemblage of physical properties or qualities by which male is distinguished from female; the physical difference between male and female; the distinguishing peculiarity of male or female (NCI – CDISC Definition). {SEX} (See Section 2.2.)	Record the appropriate sex (e.g., *F* (female), *M* (male).	Collect the subject's sex or gender, as reported by subject or caretaker. This is the self-reported sex of the individual and/or is the clinician's assignment based on a physical examination. This is a phenotypic assessment and a genotypic assessment (see Section 5.6.2 Collecting Sex, Ethnicity and Race). *See the BRIDG model for complete path.	HR

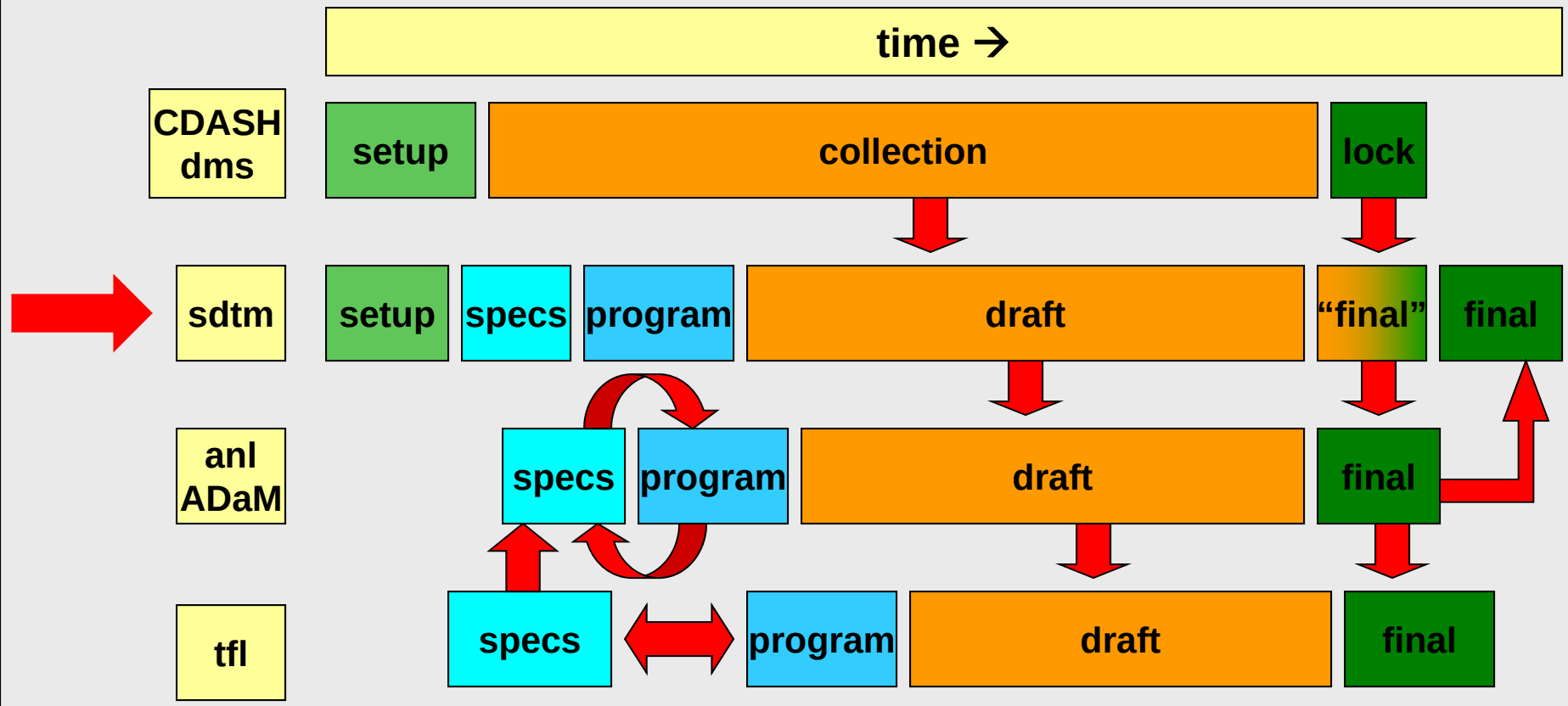
Controlled Terminology

Rho
Giving flight to research

Goals of CDASH Implementation

- Standardize study-set up
- Standardize data collection in the industry
- Offer our customers a less expensive and more efficient data management solution that is CDASH compliant
- Perform less metadata mapping from source data to SDTM. Develop SDTM-compliant metadata cheaper and faster

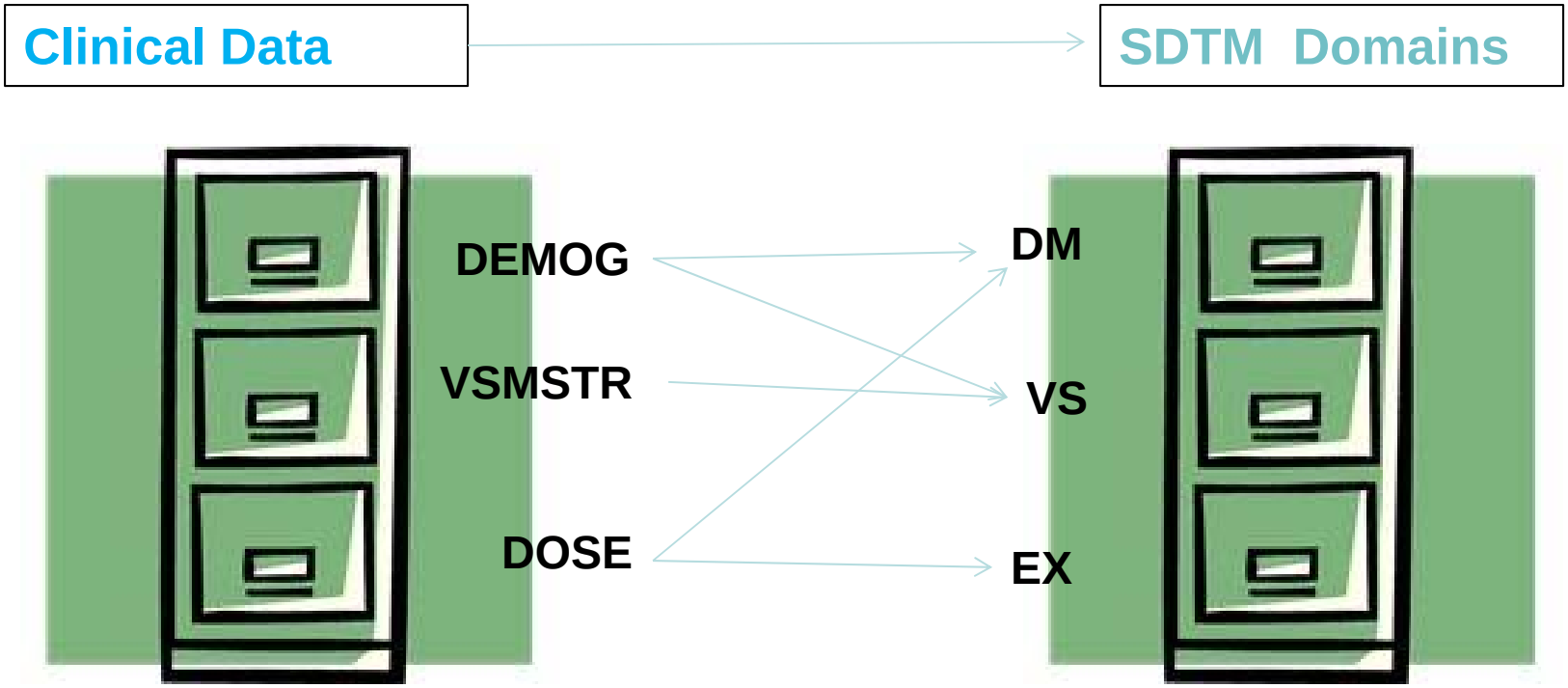
Workflow for CDISC Project



What is SDTM?

- Study Data Tabulation Model
- A standard way to represent clinical (raw) data
- Standard to submit clinical data to FDA
- Standard for exchange, submission, warehousing
- Data or observations are grouped into a series of standardized domains (e.g., AE, EG, DM, etc)
- Standardized domains
  - Standard structures
  - Standard variable names
  - Standard variable attributes
  - Controlled (common) terminology

What is SDTM?



Data is mapped from clinical database to SDTM domains

SDTM Versions

- Version 3.1.2
  - Newest version
  - Now used for production
  - Now accepted by FDA
- New versions now released twice a year

Before SDTM

- **Standard domain names, standard variables, and standard variable names were not established**
- **Overall inefficient and error-prone process:**
  - Steep learning curve for each study
  - Difficulty joining datasets
  - Allotted review time was spent “cleaning data”
  - Reviewers were required to familiarize themselves with unique domain names, variables, and variable names for each study and each dataset within a study
  - FDA Reviewers typically lack programming support and need customized tools

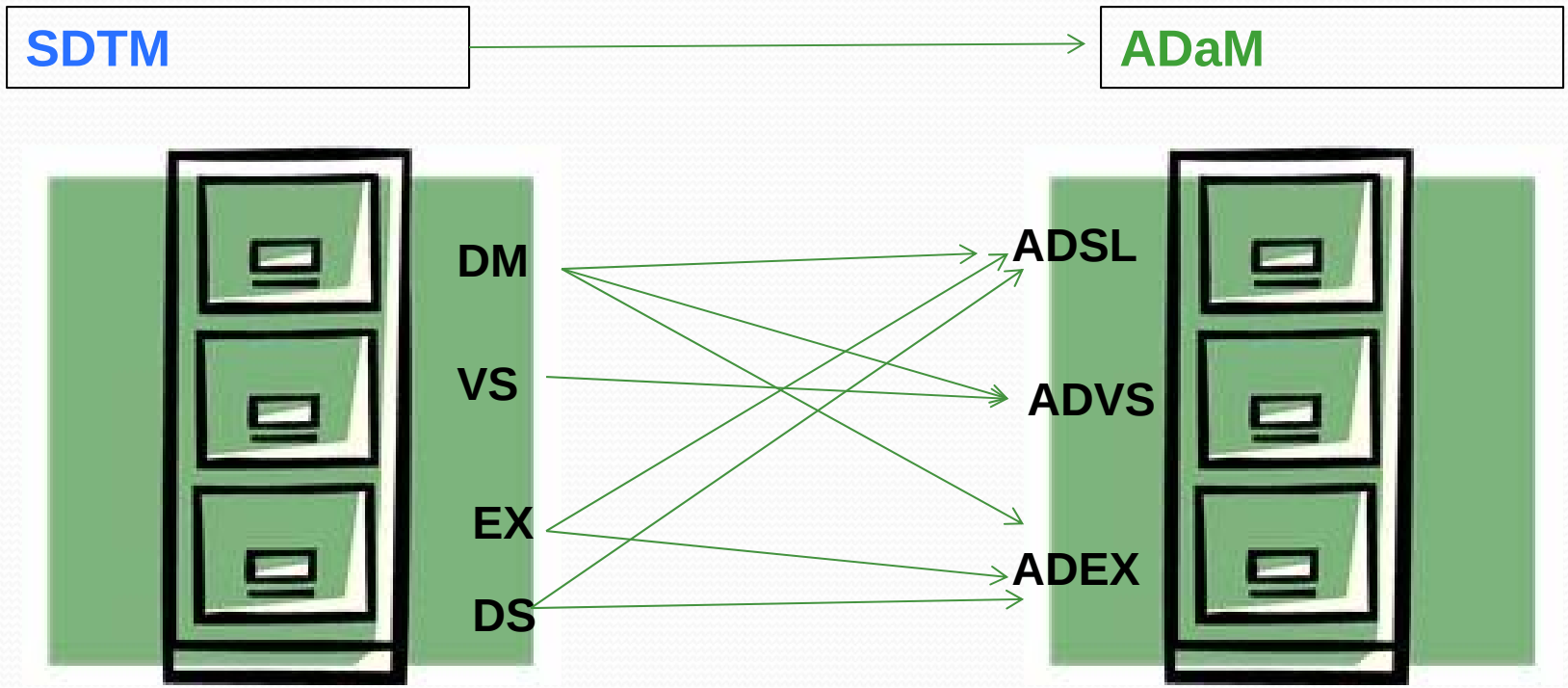
After SDTM

- **Standardization** of domains, variable names and structures
- SDTM provides standardized and structured tools to store, display, review, and analyze data
- Easy to locate data with standard domain names
- Immediate familiarity with data through standard data structures, variables, and variable names
- Increase in reviewer efficiency
- Overall more time-efficient process, increased consistency and minimal learning curve from study to study.

SDTM  ADaM

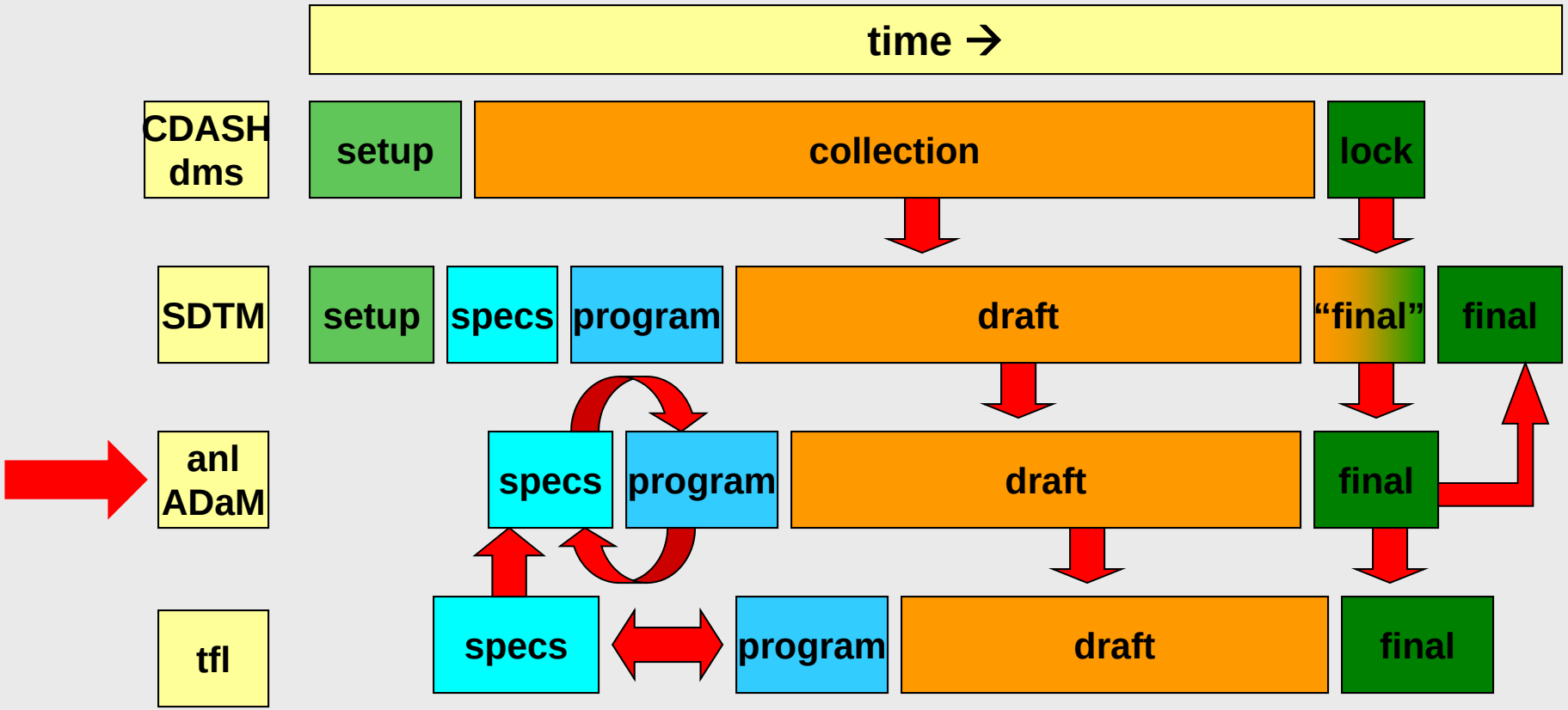
- SDTM is a standard for clinical or raw data
- Analysis datasets are also needed to support statistical analysis and data displays
- Purpose of analysis datasets is to facilitate statistical analysis and display production
- SDTM data is the source for the creation of analysis datasets
- The SDTM data is mapped to a collection of analysis datasets which are structured in a way that is conducive to performing analyses
- **ADaM** is a standard for analysis datasets

SDTM → ADaM



Data is mapped from SDTM database to ADaM datasets

Workflow for CDISC Project



What is ADaM all about?

- ADaM provides general guidelines for analysis datasets
- Describes the general structure, metadata, and content for analysis datasets.
- Design of analysis datasets driven by the study's scientific and medical objective
- Less structured than SDTM

The Purpose of ADaM

"The overall principle in designing statistical analysis datasets and related metadata is that there must be **clear and unambiguous communication** of the content, source and quality of the datasets submitted in support of the statistical analysis performed by the sponsor." [1]

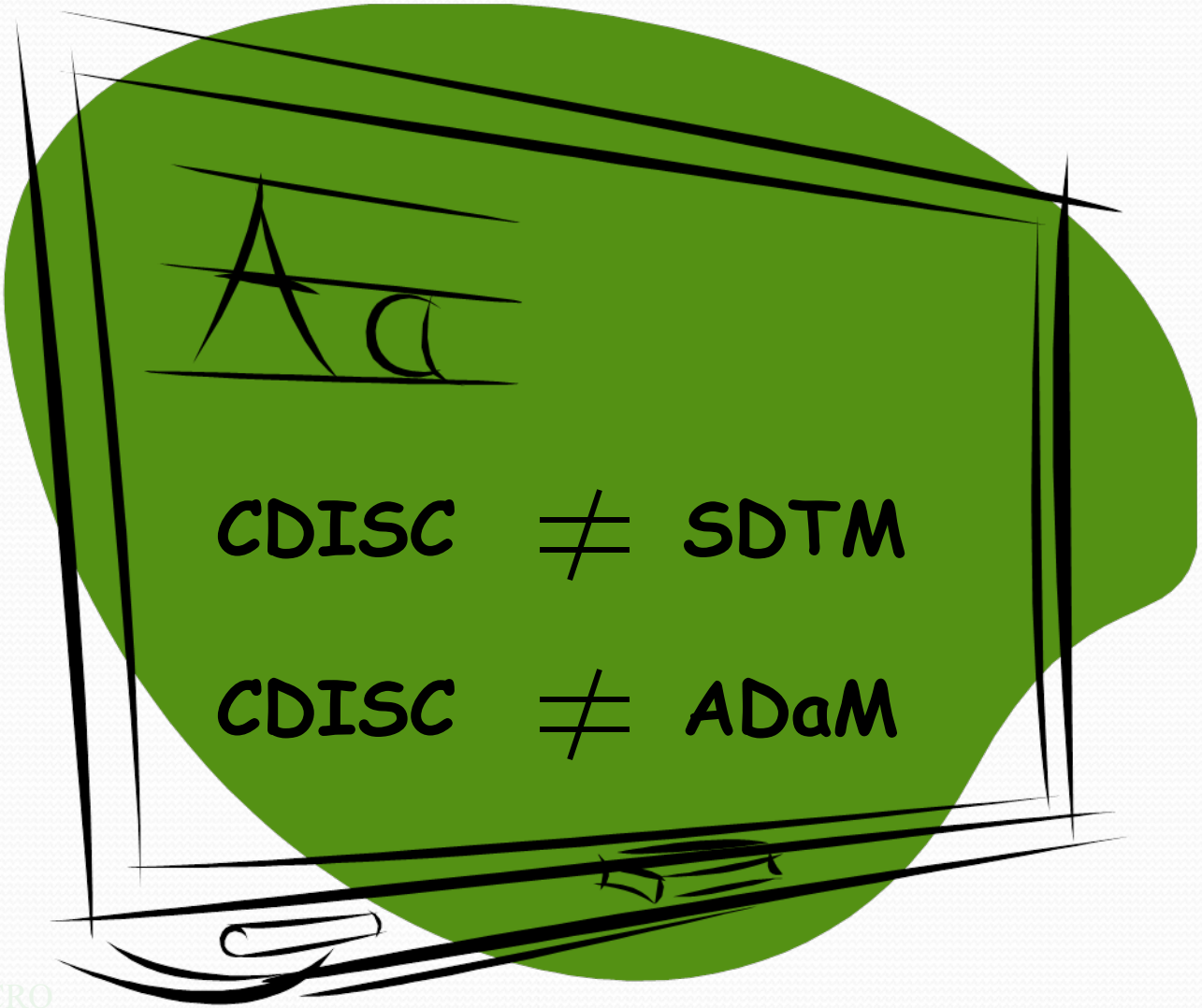
"ADaM will allow **FDA reviewers** to replicate most analyses, tables, graphs and listings with minimal or no transformations ('one proc away'), and enable them to easily view and subset the data used to generate any analysis, table, graph or listing without complex programming." [2]

[1] http://www.fda.gov/oc/datacouncil/meetings/wilson.pdf
[2] http://www.lexjansen.com/pharmasug/2003/fdacompliance/fda014.pdf

Differences between SDTM and ADaM

DIFFERENCE	SDTM	ADaM
Data origin	Source data	Derived data
Data structure	Vertical	Dependent on analysis
Redundancy	No redundancy	Redundancy needed for easy analysis
Variable type	Character	Numeric
Domain	Each domain is specific to itself	Combines SDTM variables across multiple ADaM domains
Date Format	ISO 8601 character strings	Formatted as numeric (SAS) dates to allow manipulation

Clarification



Operational Data Modeling (ODM)

Basics:

- XML schema
- Production Version 1.3 (released December 2006)
- Covers wide range of standards (SDTM, ADaM, SEND)

Details:

- Vendor neutral, platform-independent format specification of standard XML schema for the interchange and archive of data collected from various sources in a clinical trial.
- Designed to:

  (a) represent a wide range of study information to be compatible with many data management systems

  (b) be compliant with guidance published by FDA fro computer systems used in clinical trials

CDISC Operational Data Model

- Define a data model that can be used to facilitate interchange of clinical data
- Define a standard to enable exchange of metadata and data or updates to both between heterogeneous systems.
- Utilize XML as the language to build the syntax of ODM

ODM XML example

```
<MetaDataVersion OID = "v1.2.0" Name = "Version 1.2.0">
  <Protocol>
    <StudyEventRef StudyEventOID = "SE.VISIT0" OrderNumber = "1" Mandatory = "Yes"/>
    <StudyEventRef StudyEventOID = "SE.VISIT1" OrderNumber = "2" Mandatory = "Yes"/>
  </Protocol>
  <StudyEventDef OID = "SE.VISIT0" Name = "Pre-treatment" Repeating = "No" Type = "Scheduled"
     Category ="PreTreatment">
    <FormRef FormOID = "FORM.DEMOG" OrderNumber = "1" Mandatory = "No"/>
    <FormRef FormOID = "FORM.DRUGPHRM" OrderNumber = "2" Mandatory = "No"/>
  </StudyEventDef>
    ...
  <FormDef OID = "FORM.AE" Name = "Adverse Events" Repeating = "No">
    <ItemGroupRef ItemGroupOID = "IG.AE" OrderNumber = "1" Mandatory = "No"/>
  </FormDef>
    ...
  <ItemGroupDef OID = "IG.AE" Name = "Adverse Events" Repeating = "Yes" IsReferenceData = "No"
     SASDatasetName = "AE" Domain = "AE Domain" Origin = "AE Origin" Role = "AE Role"
     Comment = "AE Comment">
    <ItemRef ItemOID = "IT.TAREA" OrderNumber = "1" Mandatory = "No"/>
  </ItemGroupDef>
</MetaDataVersion>
```

Define.xml

Datasets for Study 1234

Dataset	Description	Structure	Purpose	Keys	Location
DM	Demographics	Special Purpose - One record per event per subject	Tabulation	STUDYID, USUBJID	crt/datasets/1234/dm.xpt
TE	Trial Elements	Trial Design - One Record Per Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/te.xpt
TA	Trial Arms	Trial Design - One Record per Element for each Arm	Tabulation	STUDYID, ARM	crt/datasets/1234/ta.xpt
TV	Trial Visits	Trial Design - One Record per Visit per Arm	Tabulation	STUDYID, VISIT	crt/datasets/1234/tv.xpt
SE	Subject Elements	Study Design - One Record Per Subject Element	Tabulation	STUDYID, ELEMENT	crt/datasets/1234/se.xpt
SV	Subject Visits	Study Design - One Record Per Subject Visit	Tabulation	STUDYID, VISIT	crt/datasets/1234/sv.xpt
PE	Physical Examination	Findings - One record per event per subject	Tabulation	USUBJID, PETEST, PESEQ	crt/datasets/1234/pe.xpt
SC	Subject Characteristics	Findings - One record per subject characteristic	Tabulation	USUBJID, SCTESTCD	crt/datasets/1234/sc.xpt
VS	Vital Signs	Findings - One record per subject per vital sign	Tabulation	USUBJID, VSTESTCD	crt/datasets/1234/vs.xpt
CO	Comments	Special Purpose - One record per comment per subject	Tabulation	STUDYID, COSEQ	crt/datasets/1234/co.xpt

Vital Signs Dataset (VS)

Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
STUDYID	STUDY IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
DOMAIN	DOMAIN ABBREVIATION	text		CRF Page	Identifier	DOMAIN ABBREVIATION
USUBJID	UNIQUE SUBJECT IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
VSSEQ	SEQUENCE NUMBER	integer		CRF Page	Identifier	SEQUENCE NUMBER
VSTESTCD	VITAL SIGNS TEST SHORT NAME	text		CRF Page	Topic	Vital Signs CRF Page 4, CRF Page 7
VSTEST	VITAL SIGNS TEST NAME	text		CRF Page	Synonym Qualifier	VITALS SIGNS TEST NAME
VSPOS	VITAL SIGNS POSITION OF SUBJECT	text		CRF Page	Record Qualifier	POSITION OF THE SUBJECT DURING A MEASUREMENT OR EXAMINATION.
VSORRES	RESULT OR FINDING IN ORIGINAL UNITS	text		CRF Page	Result Qualifier	Vital Signs CRF Page 4, CRF Page 7
VSORRESU	ORIGINAL UNITS	text		CRF Page	Variable Qualifier	ORIGINAL UNITS
VSSTRESC	CHARACTER RESULT/FINDING IN STD FORMAT	text		CRF Page	Result Qualifier	CHARACTER RESULT/FINDING IN STD FORMAT
VSSTRESN	NUMERIC RESULT/FINDING IN STANDARD UNITS	float		CRF Page	Result Qualifier	CHARACTER RESULT/FINDING IN STD FORMAT
VSSTRESU	STANDARD UNITS	text		CRF Page	Variable Qualifier	Default units.
VSSTAT	VITALS STATUS	text		CRF Page	Result Qualifier	VITALS STATUS
VSLOC	LOCATION OF VITAL SIGNS MEASUREMENT	text		CRF Page	Record Qualifier	LOCATION OF VITAL SIGNS MEASUREMENT
VISITNUM	VISIT NUMBER	integer		DERIVED	Timing	Clinical encounter number.
VISIT	VISIT NAME	text		CRF Page	Timing	VISIT NAME
VSDTC	DATE/TIME OF MEASUREMENTS	date		CRF Page	Timing	Vital Signs CRF Page 4, CRF Page 7
VSDY	STUDY DAY OF VITAL SIGNS	integer		CRF Page	Timing	STUDY DAY OF VITAL SIGNS

Value Level Metadata

Variable	Value	Label	Type	Controlled Terms or Format	Origin	Role	Comment
VSTESTCD	FRAME	Frame	float	FRAME	CRF Page		Vital Signs CRF Page 4
VSTESTCD	HTRAW	Height raw	text		CRF Page		Vital Signs CRF Page 4
VSTESTCD	WTRAW	Weight raw	text		CRF Page		Vital Signs CRF Page 4
VSTESTCD	MEANBP	Mean Blood Pressure	float		VSSTRESN		See Computational Method: COMPMETHOD.MEANBP

Define.xml

Demographics Dataset (DM)							dm.xpt
Variable	**Label**	**Type**	**Controlled Terms or Format**	**Computational Algorithm or Method**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text			CRF Page 7	Identifier	
USUBJID	Unique Subject Identifier	text			Sponsor Defined	Identifier	
DOMAIN	Domain Abbreviation	text			Derived	Identifier	
SUBJID	Subject Identifier for the Study	text			CRF Page 7	Topic	
RFSTDTC	Subject Reference Start Date/Time	text			Sponsor Defined	Timing	
RFENDTC	Subject Reference End Date/Time	text			Sponsor Defined	Timing	
SITEID	Study Site Identifier	text			Derived	Record Qualifier	
AGE	Age in AGEU at RFSTDTC	float			Derived	Result Qualifier	
AGEU	Age Units	text	AGEU		Derived	Variable Qualifier	
SEX	Sex	text	SEX		CRF Page 7	Result Qualifier	
RACE	Race	text	ADRACE		CRF Page 7	Result Qualifier	
ARMCD	Planned Arm Code	text	ARMCD		Derived	Result Qualifier	
ARM	Description of Planned Arm	text			Derived	Synonym Qualifier	
COUNTRY	Country	text			Derived	Result Qualifier	
DMDTC	Date/Time of Collection	text			CRF Page 7	Timing	
DMDY	Study Day of Collection	float		COMP_STUDY_DAY	Derived	Timing	

Laboratory Data Model (LAB)

- Production Version 1.0.1
- Standard model for the acquisition and interchange of clinical trial laboratory data
- Harmonized with ODM

Protocol Representation Model

- Production Version 1.0 (released January 2010)
- Identifies standard elements of a clinical trial protocol that can be further defined, elucidated or codified to facilitate study design, regulatory compliance, project management, trial conduct, and data interchange among consumers and systems.

Standards for the Exchange of Non-clinical Data (SEND)

- Production Version 3.0 (released May 2011)
- Guides the organization, structure, and format of non-clinical data (mostly from animal toxicity studies) submitted to the FDA
- Based upon SDTM
- Intended to facilitate non-clinical data from sponsor to FDA

CDISC Deliverables

Non-CDISC Project Deliverables:	CDISC Project Deliverables:
▪ Data Management data ▪ CRT data ▪ Analysis datasets ▪ TLFs ▪ Dataset specifications ▪ Annotated CRF	▪ Data Management data ▪ SDTM datasets - Domain datasets - SUPPQUAL datasets - Trial Design datasets - Controlled Terms for variables ▪ ADaM datasets ▪ TLFs ▪ Results-level metadata ▪ 2 Annotated CRFs (DM, SDTM) ▪ Define.xml for SDTM ▪ Define.pdf for ADaM

Overview of CDISC Benefits

Virtually all benefits associated with structured standards development and implementation are reaped by both sponsors and the FDA.

These benefits include:

- Increased efficiency
- Standardized automated processes
- Process development and improvement
- Reduced elapsed time for processing of regulatory submission reviews
- Time and monetary savings due to increasingly efficient data transfers among business partners
- More efficient archive and recovery procedures
- More accessible information = better communication among team members

Presentation Summary

- **Clinical Data Acquisition Standards Harmonization (CDASH)** standardizes data collection fields expected to be present on most CRFs
- **Study Data Tabulation Model (SDTM)** is a standardized way to represent raw, clinical data from a variety of sources (CRF, labs, ECG, physical exam, etc).
- SDTM data is what feeds **Analysis Dataset Model (ADaM)** datasets. The SDTM data is mapped to a collection of analysis datasets, which are structured and organized in a way that is conducive to performing analyses.
- The FDA is asking for **SDTM** and **ADaM**. We need to know the guidance and be proficient in applying the guidance to project tasks.
- **Operational Data Model (ODM)** is a standard XML schema for exchanging and archiving data and metadata

It is important to remember that in the big picture, **Clinical Data Interchange Standards Consortium (CDISC)** allows for increased efficiency, reduced processing time of regulatory submission reviews, time and monetary savings, and better communication among all team members.

Want to Get Involved
CDISC Needs You!!!

Part 2
A Closer Look at SDTM

SDTM Basic Concepts

- **What is SDTM?**
  - The Study Data Tabulation Model is a standard way to represent clinical trial data developed by CDISC.

- **What types of problems were encountered before SDTM was created?**
  - The system, overall, was inefficient and error-prone.
  - Steep learning curve for each application/study.
  - Pooling and joining datasets together was a complicated process.
  - Reviewers were required to familiarize themselves with unique domain names, variables, and variable names for each submission (causing a delay in processing).
  - Much of the allotted review time was allocated to reorganizing submitted data to varying formats before beginning the actual evaluation of the data.

SDTM Basic Concepts

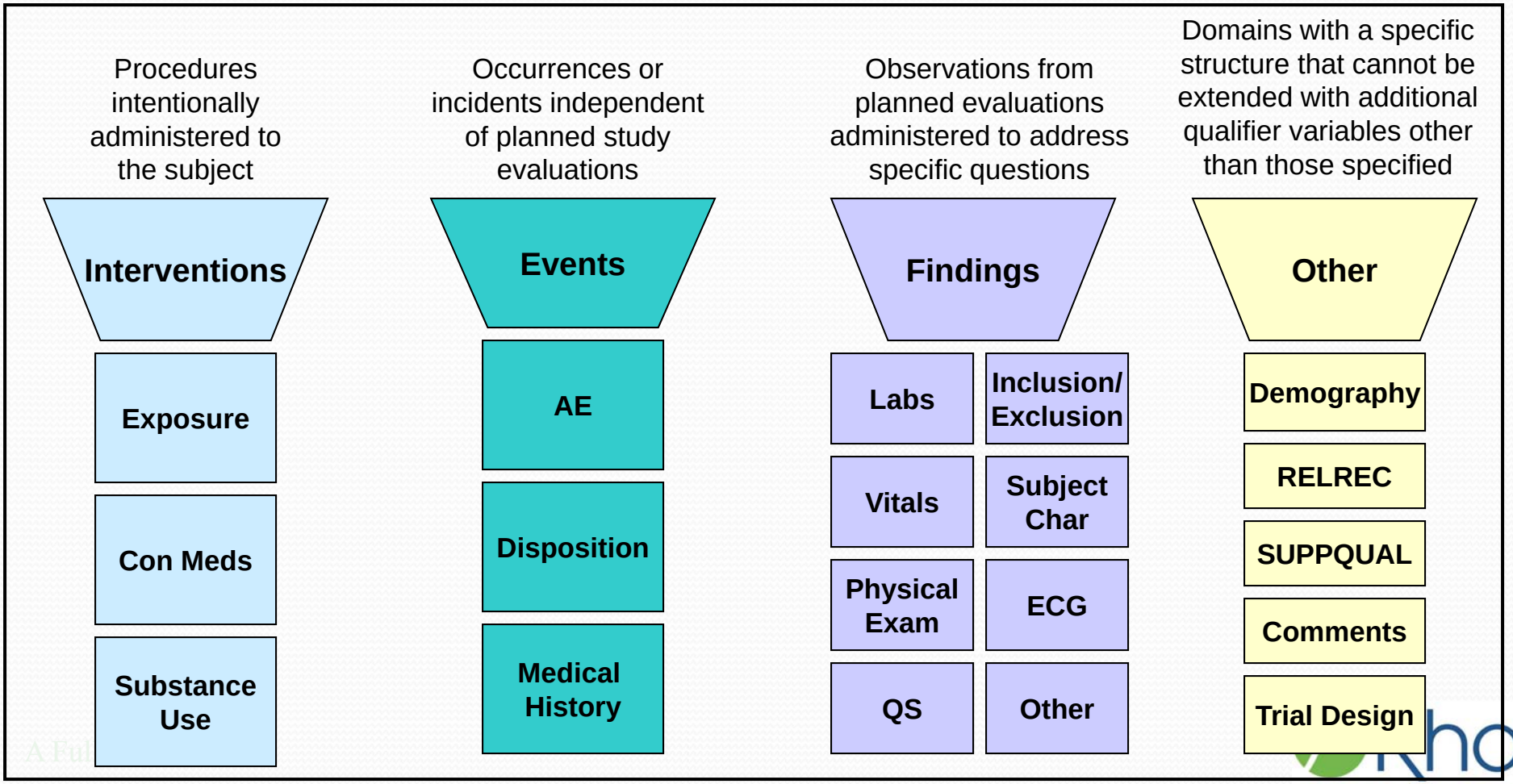
- **What exactly does SDTM do?**
  - SDTM captures all clinical data tabulations as a series of observations while enforcing standardization of domain and variable names and structures.

- **DOMAIN STANDARDIZATION:**
  - SDTM contains pre-defined domains. The two-character domain code is used throughout a submission as the dataset name, and as a prefix for most variable names in the dataset. The two-character prefix is also the value for the Domain variable found in all SDTM datasets.

- **VARIABLE STANDARDIZATION:**
  - Within each domain, SDTM provides users with pre-specified standard variables that have pre-specified attributes. The variable name, label and type are standardized and cannot be changed. The variable value is sometimes standardized through the use of controlled terminology.

General Domain Classes

Most observations collected during the study will be divided among three domain classes: Interventions, Events and Findings. Other information is captured in the Other/Special Purpose Domain Class.



General Domain Classes

- Three general classes plus other
- Each class shares:
  - Identifier variables
  - Timing Variables
  - Class specific variables
  - Additional variables **may not** be added
- Most subject level data will fit into one of these classes
- Standard domains are defined within each class
- Only submit domains that were collected
- If necessary, new domains can be created that are based on the rules for a given class

General Domain Classes

Special-Purpose Domains (defined in Section 5):

- Demographics — DM
- Subject Elements — SE
- Comments — CO
- Subject Visits — SV

Interventions General Observation Class (defined in Section 6.1):

- Concomitant Medications — CM
- Substance Use — SU
- Exposure — EX

Events General Observation Class (defined in Section 6.2):

- Adverse Events — AE
- Medical History — MH
- Clinical Events — CE
- Disposition — DS
- Protocol Deviations — DV

Findings General Observation Class (defined in Section 6.3):

- ECG Test Results — EG
- Laboratory Test Results — LB
- Questionnaires — QS
- Vital Signs — VS
- Microbiology Specimen — MB
- PK Concentrations — PC
- Inclusion/Exclusion Criterion Not Met — IE
- Physical Examination — PE
- Subject Characteristics — SC
- Drug Accountability — DA
- Microbiology Susceptibility Test — MS
- PK Parameters —PP

Findings About (defined in Section 6.4)

- Findings About — FA

Trial Design Domains (defined in Section 7):

SDTM Implementation Guide: Vital Signs

vs.xpt, Vital Signs — Findings, Version 3.1.2. One record per vital sign measurement per time point per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core	Reference
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	VS	Identifier	Two-character abbreviation for the dom	Req	SDTM 2.2.4, SDTMIG 4.1.2.2, SDTMIG Appendix C2
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req	SDTM 2.2.4, SDTMIG 4.1.2.3
VSSEQ	Sequence Number	Num		Identifier	Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.	Req	SDTM 2.2.4
VSGRPID	Group ID	Char		Identifier	Used to tie together a block of related records in a single domain for a subject.	Perm	SDTM 2.2.4, SDTMIG 4.1.2.6
VSSPID	Sponsor-Defined Identifier	Char		Identifier	Sponsor-defined reference number. Perhaps pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database.	Perm	SDTM 2.2.4, SDTMIG 4.1.2.6
VSTESTCD	Vital Signs Test Short Name	Char	(VSTESTCD)	Topic	Short name of the measurement, test, or examination described in VSTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in VSTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g."1TEST"). VSTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: SYSBP, DIABP, BMI.	Req	SDTM 2.2.3, SDTMIG 4.1.1.8, SDTMIG 4.1.2.1, SDTMIG Appendix C1
VSTEST	Vital Signs Test Name	Char	(VSTEST)	Synonym Qualifier	Verbatim name of the test or examination used to obtain the measurement or finding. The value in VSTEST cannot be longer than 40 characters. Examples: Systolic Blood Pressure, Diastolic Blood Pressure, Body Mass Index.	Req	SDTM 2.2.3, SDTMIG 4.1.2.1, SDTMIG 4.1.2.4, SDTMIG 4.1.5.3.1, SDTMIG Appendix C1
VSCAT	Category for Vital Signs	Char	*	Grouping Qualifier	Used to define a category of related records.	Perm	SDTM 2.2.3, SDTMIG 4.1.2.6
VSSCAT	Subcategory for Vital Signs	Char	*	Grouping Qualifier	A further categorization of a measurement or examination.	Perm	SDTM 2.2.3, SDTMIG 4.1.2.6
VSPOS	Vital Signs Position of Subject	Char	(POSITION)	Record Qualifier	Position of the subject during a measurement or examination. Examples: SUPINE, STANDING, SITTING.	Perm	SDTM 2.2.3
VSORRES	Result or Finding in Original Units	Char		Result Qualifier	Result of the vital signs measurement as originally received or collected.	Exp	SDTM 2.2.3, SDTMIG 4.1.5.1

Sample Domain

SDTM Implementation Guide: Vital Signs

vs.sas7bdat

	STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VISITNUM	VISIT	VSDTC
1	KRM-306	VS	KRM306-00	1	HEIGHT	Height		65	IN	1	Visit 1 (Screen	2008-04-03
2	KRM-306	VS	KRM306-00	2	WEIGHT	Weight		134	LB	1	Visit 1 (Screen	2008-04-03
3	KRM-306	VS	KRM306-00	3	WEIGHT	Weight				2	Visit 2 (Baseli	2008-04-17
4	KRM-306	VS	KRM306-00	4	WEIGHT	Weight				3	Visit 3 (Week 2	2008-05-01
5	KRM-306	VS	KRM306-00	5	WEIGHT	Weight				4	Visit 4 (Week 6	2008-05-29
6	KRM-306	VS	KRM306-00	6	WEIGHT	Weight				5	Visit 5 (Week 1	2008-07-10
7	KRM-306	VS	KRM306-00	7	DIABP	Diastolic	SITTING	64	MMHG	1	Visit 1 (Screen	2008-04-03
8	KRM-306	VS	KRM306-00	8	PULSE	Pulse Rat		60	BEATS/MIN	1	Visit 1 (Screen	2008-04-03
9	KRM-306	VS	KRM306-00	9	SYSBP	Systolic	SITTING	136	MMHG	1	Visit 1 (Screen	2008-04-03
10	KRM-306	VS	KRM306-00	10	TEMP	Temperatu		97.2	F	1	Visit 1 (Screen	2008-04-03
11	KRM-306	VS	KRM306-00	11	DIABP	Diastolic	SITTING	78	MMHG	2	Visit 2 (Baseli	2008-04-17
12	KRM-306	VS	KRM306-00	12	PULSE	Pulse Rat		63	BEATS/MIN	2	Visit 2 (Baseli	2008-04-17
13	KRM-306	VS	KRM306-00	13	SYSBP	Systolic	SITTING	160	MMHG	2	Visit 2 (Baseli	2008-04-17
14	KRM-306	VS	KRM306-00	14	TEMP	Temperatu		97.2	F	2	Visit 2 (Baseli	2008-04-17
15	KRM-306	VS	KRM306-00	15	DIABP	Diastolic	SITTING	67	MMHG	3	Visit 3 (Week 2	2008-05-01
16	KRM-306	VS	KRM306-00	16	PULSE	Pulse Rat		62	BEATS/MIN	3	Visit 3 (Week 2	2008-05-01
17	KRM-306	VS	KRM306-00	17	SYSBP	Systolic	SITTING	158	MMHG	3	Visit 3 (Week 2	2008-05-01
18	KRM-306	VS	KRM306-00	18	TEMP	Temperatu		97.6	F	3	Visit 3 (Week 2	2008-05-01
19	KRM-306	VS	KRM306-00	19	DIABP	Diastolic	SITTING	63	MMHG	4	Visit 4 (Week 6	2008-05-29
20	KRM-306	VS	KRM306-00	20	PULSE	Pulse Rat		58	BEATS/MIN	4	Visit 4 (Week 6	2008-05-29
21	KRM-306	VS	KRM306-00	21	SYSBP	Systolic	SITTING	150	MMHG	4	Visit 4 (Week 6	2008-05-29

Basic Concepts: Core Variables

- All variables are either required, expected or permissible
- **Required**
  - Must be included
  - Cannot be null for any record
- **Expected**
  - Must be included
  - Can be null
- **Permissible**
  - Can be optionally included
  - Should be used when collected or appropriate
  - Can be null

	STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AESTDTC	AEENDTC	AEMODIFY	AEDECOD
Row 1	ABC123	AE	123101	1	POUNDING HEADACHE	2003-10-12	2003-10-12	HEADACHE	HEADACHE
Row 2	ABC123	AE	123101	2	BACK PAIN FOR 6 HOURS	2003-10-13T13:05	2003-10-13T19:00	BACK PAIN	BACK PAIN
Row 3	ABC123	AE	123101	3	PULMONARY EMBOLISM	2003-10-21			PULMONARY EMBOLISM

	AEBODSYS	AESEV	AESER	AEACN	AEREL
Row 1 (cont)	NERVOUS SYSTEM DISORDERS	SEVERE	N	NOT APPLICABLE	DEFINITELY NOT RELATED
Row 2 (cont)	MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS	MODERATE	N	DOSE REDUCED	PROBABLY RELATED
Row 3 (cont)	VASCULAR DISORDERS	MODERATE	Y	DOSE REDUCED	PROBABLY NOT RELATED

	AEOUT	AESCONG	AESDISAB	AESDTH	AESHOSP	AESLIFE	AESMIE	AESTDY	AEENDY	AEENRF
Row 1 (cont)	RECOVERED/RESOLVED							-1	-1	
Row 2 (cont)	RECOVERED/RESOLVED							1	1	
Row 3 (cont)	RECOVERING/RESOLVING				Y	Y		9	.	AFTER

REQUIRED variables are variables that must be included to make the data usable. EXPECTED variables are those necessary to make a record useful in the context of a specific domain. Expected variables are assumed to be present in each submitted dataset, even if some values are null. PERMISSIBLE variables are optional variables that vary from study to study. These are variables that may be included if data was collected, but they are not essential to the overall domain or dataset.

Rho
Giving flight to research

More Basic Concepts

More Basic Concepts of the SDTM

- Non-redundancy – Demographics Selection variables like Age, Sex, Race, Treatment Group submitted in DM only
  - FDA tools apply DM Selection variables to all domains and can derive standard variables like "Days since last dose"
- 2-character domain prefix on all variables
  - Useful when performing SAS Merges
- Ability to represent relationships
  - Related datasets, records, record groups, comments
- ISO8601 date/time variables (e.g., 20040126T14:00:00)
  - Replaces SAS date and time variables and precisions
  - Allows representation of durations and truncations
- "Extra" variables can be submitted separately in SUPPQUAL
  - Merged back into parent domains by FDA review tools.

5

More Basic Concepts

Standard Identifiers Used for All Domains

Name	Label	Type
STUDYID	Study Identifier	Char
DOMAIN	Domain Abbreviation	Char
USUBJID	Unique Subject ID	Char
--SEQ	Sequence Number	Num
--GRPID	Group ID	Char
--REFID	Reference ID	Char
--SPID	Sponsor ID	Char

More Basic Concepts

Standard Timing Variables

Visits	VISIT	Visit Name
	VISITNUM	Visit Number
	VISITDAY	Visit (Study) Day
Dates	--DTC	Date/Time of Collection
	--STDTC	Start Date/Time
	--ENDTC	End Date/Time
Days	--DY	Study Day of collection
	--STDY	Start Study Day
	--ENDY	End Study Day
Timepts	--TPT	Time Point Name
	--TPTNUM	Time Point Number
	--TPTREF	Time Point Reference
Durations	--DUR,	Collected duration
	--ELTM	Elapsed Time from Time Points
References	--STRF	BEFORE, DURING, AFTER DM Ref Period
	--ENRF	BEFORE, DURING, AFTER DM Ref Period

ISO 8601 Format

- **What is it?**
  - A standardized format for representing data, date-time, time, and duration.
- **What does it look like?**
  - General form is: **yyyy-mm-ddThh:mm:ss**
  - This is the "extended format" (using the -, T, and : separators) required by CDISC.
  - Build left to right (year -> second). Missing components are represented by a hyphen or blank, depending on whether additional information to the right is available (see examples).

`2007-12-09T15:30:00`	Complete specification
`2007-12-09T15:30`	Same as above
`2007-12-09`	No time, so omit all time components
`2007---10`	Represent missing month with a hyphen

What do we do with variables that don't fit this model?

- There is always the possibility that there will be "extra variables" that will not fit the typical SDTM model. When this is the case we create a supplemental dataset. SUPPQUAL acts as a supplement to SDTM variables.

- Example of when to use SUPPQUAL: For the AE domain, a client wants to include all MedDRA terms and codes in their submission. SDTM only allows for the Preferred Term captured by AEDECOD and the Primary SOC captured by AEBODSYS. Where will we put the remaining AE codes?
    - We will capture this data in Supplemental or SUPPQUAL datasets. For Adverse Events, the supplemental dataset will be called SUPPAE. Dataset SUPPAE is a separate dataset from the main AE dataset.

SUPPQUAL Dataset

Main AE Dataset

STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AEDECOD	AEBODSYS	AESEV	AESER	AEACN	AEREL
FAKE101	AE	FAKE101-101	1	HEADACHES	Headache	Nervous system disorders	MODERATE	N	NONE	POSSIBLE
FAKE101	AE	FAKE101-101	2	NAUSEA	Nausea	Gastrointestinal disorders	MILD	N	NONE	POSSIBLE

SUPPAE Dataset

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG
FAKE101	AE	101	AESEQ	1	LOWLEVC	Low Level Code	10019211	DER
FAKE101	AE	101	AESEQ	1	LOWLEVN	Low Level Name	Headache	DER
FAKE101	AE	101	AESEQ	1	PRECODE	Preferred Code	10019211	DER
FAKE101	AE	101	AESEQ	1	PRENAME	Preferred Name	Headache	DER
FAKE101	AE	101	AESEQ	1	PRISOC	Primary SOC Code	10029205	DER
FAKE101	AE	101	AESEQ	1	SYSORGCL	System Organ Class Code	10029205	DER
FAKE101	AE	101	AESEQ	2	HIGHGRPC	High Group Code	10029305	DER
FAKE101	AE	101	AESEQ	2	HIGHGRPN	High Group Name	Gastrointestinal signs and symptoms	DER
FAKE101	AE	101	AESEQ	2	HIGHLEVC	High Level Code	10008028	DER
FAKE101	AE	101	AESEQ	2	HIGHLEVN	High Level Name	Nausea and vomiting symptoms	DER
FAKE101	AE	101	AESEQ	2	LOWLEVC	Low Level Code	10004071	DER
FAKE101	AE	101	AESEQ	2	LOWLEVN	Low Level Name	Nausea	DER
FAKE101	AE	101	AESEQ	2	PRECODE	Preferred Code	10049848	DER

Transposing Clinical Data

As data is collected during the trial it is typically stored as 1 record per subject per visit. Most domains in SDTM will change the structure of how data is stored to 1 record per subject per time point per test. When this is the case we must transpose the data.

Traditional

STUDYID	SUBJID	TEMP	BMI	SYSBP	DISBP	HR
1	11-111	37	25.4	96	100	21

SDTM

STUDYID	SUBJID	VSTESTCD	VSORRES
1	11-111	TEMP	37
1	11-111	BMI	25.4
1	11-111	SYSBP	96
1	11-111	DISBP	100
1	11-111	HR	21

Outcome of Transposing Clinical Data

Raw data as captured at site

VISIT	VISIT		SUBJECT	HT(cm)	WT(kg)	Temp©	RespRate	SYSBP	DYSBP	PULSE	BSA
Screening Visit	-1	6-Feb-06	1	177.8	87.9	35.6	16	143	90	94	.
Run-in	1	7-Feb-06	1	.	88.4	36.6	18	146	109	112	2.07

SDTM Data

STUDYID	DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VISIT	VISITNUM	VSDTC
FAKE101	VS	FAKE101-001	1	BSA	Body Surface Area	Sitting		m2	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	2	DIABP	Diastolic Blood Pressure	Sitting	90	mmHg	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	3	HEIGHT	Height	Sitting	177.8	cm	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	4	HR	Heart Rate	Sitting	94	beats/min	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	5	RESP	Respiratory Rate	Sitting	16	resp/min	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	6	SYSBP	Systolic Blood Pressure	Sitting	143	mmHg	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	7	TEMP	Temperature	Sitting	35.6	C	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	8	WEIGHT	Weight	Sitting	87.9	kg	Screening Visit	-1	2/6/2006
FAKE101	VS	FAKE101-001	9	BSA	Body Surface Area	Sitting	2.07	m2	Run-In	1	2/7/2006
FAKE101	VS	FAKE101-001	10	DIABP	Diastolic Blood Pressure	Sitting	109	mmHg	Run-In	1	2/7/2006
FAKE101	VS	FAKE101-001	11	HEIGHT	Height	Sitting			Run-In	1	2/7/2006
FAKE101	VS	FAKE101-001	12	HR	Heart Rate	Sitting	112	beats/min	Run-In	1	2/7/2006
FAKE101	VS	FAKE101-001	13	RESP	Respiratory Rate	Sitting	18	resp/min	Run-In	1	2/7/2006
FAKE101	VS	FAKE101-001	14	SYSBP	Systolic Blood Pressure	Sitting	146	mmHg	Run-In	1	2/7/2006
FAKE101	VS	FAKE101-001	15	TEMP	Temperature	Sitting	36.6	C	Run-In	1	2/7/2006
FAKE101	VS	FAKE101-001	16	WEIGHT	Weight	Sitting	88.4	kg	Run-In	1	2/7/2006

Trial Design Datasets

Trial Design Datasets

What are Trial Design Datasets?

Trial Design Datasets are datasets that explain the design of the clinical trial, Including:

- The specifics of the design of the clinical trial (phase, number of subjects)
- What will be done to subjects
- What data will be collected about these subjects

Why is there a need for Trial Design Datasets?

Trial Design Datasets create a standardized structure to help the FDA reviewer:

- Clearly and quickly grasping the design of the clinical trial
- Compare the designs of different trials
- Search a data warehouse for clinical trials with certain features
- Compare actual and planned treatments and visits for subjects in the trial

Trial Design Datasets

TA (Trial Arms) Dataset

An Arm is a planned path through the trial. Essentially, it is what treatment group a subject is assigned to.

- TA collects the treatment group the subject was assigned to and must also capture if transitions in treatment group occurred.

DOMAIN	ARMCD	ARM	TAETORD	ETCD	ELEMENT	TABRANCH
TA	0	Placebo	1	SCREEN	Screening	
TA	0	Placebo	2	RUNIN	Run-in	Randomized to Placebo
TA	0	Placebo	3	PLACEBO	Placebo	
TA	0	Placebo	4	FOLLOWUP	Follow-up	
TA	1	Treatment A	1	SCREEN	Screening	
TA	1	Treatment A	2	RUNIN	Run-in	Randomized to Treatment A
TA	1	Treatment A	3	TRT_A	Treatment A	
TA	1	Treatment A	4	FOLLOWUP	Follow-up	
TA	2	Treatment B	1	SCREEN	Screening	
TA	2	Treatment B	2	RUNIN	Run-in	Randomized to Treatment B
TA	2	Treatment B	3	TRT_B	Treatment B	
TA	2	Treatment B	4	FOLLOWUP	Follow-up	

Trial Design Datasets

TE (Trial Elements)

An Element is a basic building block of the trial design. Involves administering a planned intervention (treatment or no treatment) during a period of time.

- TE represents an interval of time that serves a purpose in a trial associated with certain activities affecting a subject (e.g. Screening event, Placebo given, Drug A (randomized to treatment).

DOMAIN	ETCD	ELEMENT	TESTRL	TEENRL	TEDUR
TE	SCREEN	Screening	Informed consent	Screening assessments are complete up to 30 days after start of element	
TE	RUNIN	Run-in	Eligibility confirmed	7 days after start of element	7D
TE	PLACEBO	Placebo	Randomization	14 days after start of element	14D
TE	TRT_A	Treatment A	Randomization	14 days after start of element	14D
TE	TRT_B	Treatment B	Randomization	14 days after start of element	14D
TE	FOLLOWUP	Follow-up	7 days after last dose	7 days after end of element	7D

Trial Design Datasets

TI (Trial Inclusion/Exclusion Criteria)

Contains ALL Inclusion/Exclusion Criteria for a Trial. Differs from the IE domain because IE contains only Inclusion/Exclusion Criteria which prevented the subject from being enrolled in the study. Here all criteria are listed.

STUDYID	DOMAIN	IETESTCD	IETEST	IECAT
FAKE101	TI	BLEED	Has Active Bleeding	EXCLUSION
FAKE101	TI	EXCL01	Require More Treatment for Asthma	EXCLUSION
FAKE101	TI	EXCL02	Had Acute or Chronic Nasal Congestion	EXCLUSION
FAKE101	TI	EXCL03	History of Bacterial Infection	EXCLUSION
FAKE101	TI	EXCL04	History of Cardiovascular Abnormality	EXCLUSION
FAKE101	TI	EXCL05	Have Marked Prolongation of QT Interval	EXCLUSION
FAKE101	TI	EXCL06	History of Risk Factors for TdP	EXCLUSION
FAKE101	TI	EXCL07	Using Conmeds that Prolong QT Interval	EXCLUSION
FAKE101	TI	EXCL08	Have Physical Obstruction of the Nose	EXCLUSION
FAKE101	TI	EXCL09	History of Nasal Ulceratoins or Perferations	EXCLUSION
FAKE101	TI	EXCL10	Had Prior Surgery of Nose or Sinuses	EXCLUSION
FAKE101	TI	INCL01	Provided Informed Consent	INCLUSION
FAKE101	TI	INCL02	At Least 12 Years of Age	INCLUSION
FAKE101	TI	INCL03	Able and Willing To Be Compliant	INCLUSION
FAKE101	TI	INCL04	Not Pregnant and on Birth Control	INCLUSION
FAKE101	TI	INCL05	History of SAR to Mountain Cedar	INCLUSION
FAKE101	TI	INCL06	Have a Reflective TNSS Score at Visit 2	INCLUSION
FAKE101	TI	INCL07	Demonstrated Positive Skin Test	INCLUSION

Trial Design Datasets

TS (Trial Summary)

Trial Summary collects the summary of the trial in structured format.

- Each record contains the value of a parameter or a characteristic of the trial (e.g. Protocol Title and Trial Objectives).
- Only collects planned events, not actual parameters related to the subject. For example, TS collects the number of planned enrolled subjects, not the number of subjects actually enrolled.

DOMAIN	TSSEQ	TSPARMCD	TSPARM	TSVAL
TS	1	AGEMAX	Planned Maximum Age Of Subjects	No Maximum Age
TS	1	AGEMIN	Planned Minimum Age Of Subjects	12
TS	1	AGEU	Age Unit	Years
TS	1	DESIGN	Description of Trial Design	PARALLEL
TS	1	INDIC	Trial Indications	Condition being treated; for example, (Asthma)
TS	1	OBJPRIM	Trial Primary Objective	To evaluate the safety and efficacy of two treatments compared to placebo
TS	1	PHASE	Trial Phase	2
TS	1	PLANSUB	Planned Number Of Subjects	579
TS	1	RANDOM	Trial is Randomized	Y
TS	1	SEXPOP	Sex Of Participants	BOTH
TS	1	TBLIND	Trial Blinding Schema	DOUBLE BLIND
TS	1	TCNTRL	Type of Control	Placebo
TS	1	TITLE	Trial Title	A Randomized, Double-Blind, Placebo-Controlled,Dose-Ranging Safety and Efficacy of Two Treatments
TS	1	TRT	Reported Name of Test Product	Actual compound name
TS	1	TYPE	Type of Trial	SAFETY
TS	2	TYPE	Type of Trial	EFFICACY
TS	3	TYPE	Type of Trial	TOLERABILITY

Trial Design Datasets

TV (Trial Visits)

Trial Visits describes the planned Visits or "clinical encounters" in a trial.

- Structured as one record per planned Visit per Arm.
- Used to explain when each Visit is supposed to start and end. For example, Screening for most studies will begin at the start of the study; however, the ending time for the Screening Visit will differ across studies, some may end one hour after the start of the screening visit began and others may continue for weeks until the next planned visit.

STUDYID	DOMAIN	VISITNUM	VISIT	TVSTRL
FAKE101	TV	1	Screening	Start of first Element
FAKE101	TV	2	Run-in	Up to 30 days after the first Element
FAKE101	TV	3	Randomization	Seven days after the start of the second Element
FAKE101	TV	4	One week post-randomization	Seven days after the start of the third Element
FAKE101	TV	5	End of treatment	Two weeks after the start of the third Element

Note re: SV (Subject Visits) and SE (Subject Elements):
SE and SV are subject-level trial design datasets used in SDTM. These are covered in depth in a separate presentation.

SDTM Exercises

STUDYID

OPERATIONAL CRF PAGE

Protocol: ABC123	{visit.label}	SITEID Site: {SITE} Screening #: {SCRNUM} Participant #: {ID}
DOMAIN=DM	**Demographics**	

SUBJID

Date of Birth: DEMO:BRTHDTD (dd) / DEMO:BRTHDTM (mmm) / DEMO:BRTHDTY (yyyy) — BRTHDTC

SEX

Gender (mark one): 1 ☐ (DEMO:SEX) Male 2 ☐ (DEMO:SEX) Female

RACE

Primary Race (mark one):

1 ☐ (DEMO:PRIRACE) White
2 ☐ (DEMO:PRIRACE) Black or African American
3 ☐ (DEMO:PRIRACE) Asian
4 ☐ (DEMO:PRIRACE) American Indian or Alaska Native
5 ☐ (DEMO:PRIRACE) Native Hawaiian or Other Pacific Islander
6 ☐ (DEMO:PRIRACE) Other, specify: DEMO:PRIOTHR

SUPPDM.RACESP

SUPPDM.RACE2

Secondary Race (If applicable, mark one):

1 ☐ (DEMO: SECRACE) White
2 ☐ (DEMO: SECRACE) Black or African American
3 ☐ (DEMO: SECRACE) Asian
4 ☐ (DEMO: SECRACE) American Indian or Alaska Native
5 ☐ (DEMO: SECRACE) Native Hawaiian or Other Pacific Islander
6 ☐ (DEMO: SECRACE) Other, specify: DEMO:SECOTHR

SUPPDM.RACE2SP

OPERATIONAL CRF PAGE

Ethnicity (mark one):

ETHNIC

1 ☐ (DEMO:ETHNIC) Not Hispanic or Latino

2 ☐ (DEMO:ETHNIC) Hispanic or Latino

96 ☐ (DEMO:ETHNIC) Unknown

DOMAIN=CO

Comments associated with this page:

DEMO:COMM

Submit Query | Cancel | Form Completion Help | Print | Rho

5.1.1 DEMOGRAPHICS — DM

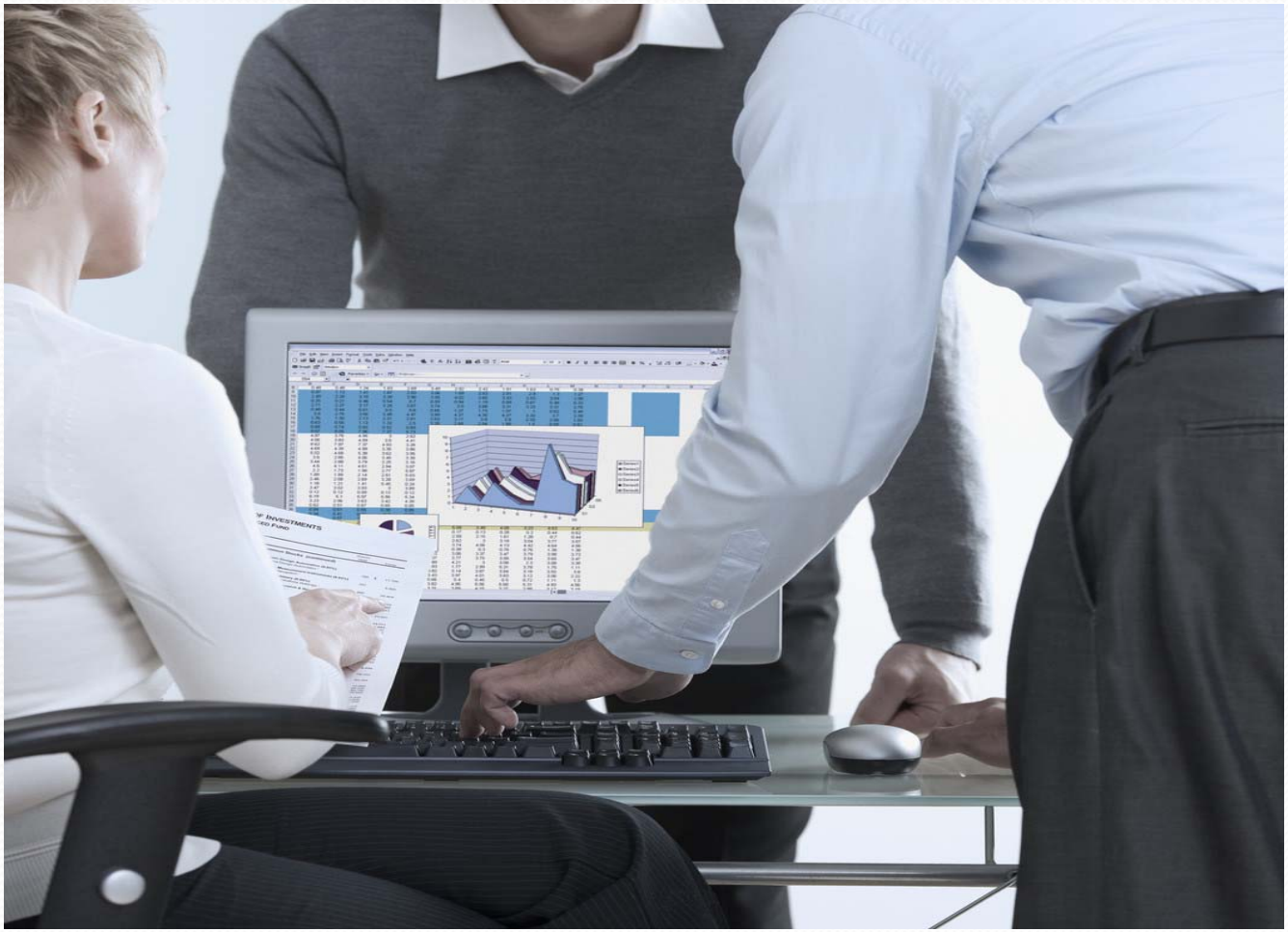
dm.xpt, Demographics — Version 3.1.2. One record per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req	SDTM 2.2.4
DOMAIN	Domain Abbreviation	Char	DM	Identifier	Two-character abbreviation for the domain.	Req	SDTM 2.2.4, SDTMIG 4.1.2.2, SDTMIG Appendix C2
USUBJID	Unique Subject Identifier	Char		Identifier	RAW.DEMO.(PROJECT|| '-'||ID) RAW.DEMO.ID is 6 character's long with the first three characters representing the site number and the last three characters representing the subject number. After the third character in ID, separate with '-'. Example USUBJID = 'ABC123-001-001'	Req	SDTM 2.2.4, SDTMIG 4.1.2.3
SUBJID	Subject Identifier for the Study	Char		Topic	Subject identifier, which must be unique within the study. Often the ID of the subject as recorded on a CRF.	Req	
RFSTDTC	Subject Reference Start Date/Time	Char	ISO 8601	Record Qualifier	Reference Start Date/time for the subject in ISO 8601 character format. Earliest date of SDTM.EX.EXSTDTC subjects who did not meet the milestone the date requires, such as screen failures or unassigned subjects.	Exp	SDTM 2.2.5, SDTMIG 4.1.4.1
RFENDTC	Subject Reference End Date/Time	Char	ISO 8601	Record Qualifier	Reference End Date/time for the subject in ISO 8601 character format. Latest date of SDTM.EX.EXSTDTC treatment. Required for all randomized subjects; null for screen failures or unassigned subjects.	Exp	SDTM 2.2.5, SDTMIG 4.1.4.1
SITEID	Study Site Identifier	Char		Record Qualifier	Unique identifier for a site within a study.	Req	
INVID	Investigator Identifier	Char		Record Qualifier	Not Collected; Not submitted	Perm	
INVNAM	Investigator Name	Char		Synonym Qualifier	Not Collected; Not submitted	Perm	
BRTHDTC	Date/Time of Birth	Char	ISO 8601	Record Qualifier	Date/time of birth of the subject.	Perm	SDTM 2.2.5, SDTMIG 4.1.4.1

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core	References
AGE	Age	Num		Record Qualifier	Calculate using SDTM.BRTHDTC and SDTM.RFSTDTC	Exp	
AGEU	Age Units	Char	(AGEU)	Variable Qualifier	='YEARS'	Exp	
SEX	Sex	Char	(SEX)	Record Qualifier	Sex of the subject.	Req	
RACE	Race	Char	(RACE)	Record Qualifier	Race of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, September 2005) for guidance regarding the collection of race (http://www.fda.gov/cder/guidance/5656fnl.htm) See Assumption below regarding RACE.	Exp	
ETHNIC	Ethnicity	Char	(ETHNIC)	Record Qualifier	The ethnicity of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, September 2005) for guidance regarding the collection of ethnicity (http://www.fda.gov/cder/guidance/5656fnl.htm).	Perm	
ARMCD	Planned Arm Code	Char	*	Record Qualifier	If SDTM.DM.ARM = 'DRUG A', then SDTM.DM.ARMCD = 'A'. If SDTM.DM.ARM = 'DRUG B', then SDTM.DM.ARMCD = 'B' would be 20.	Req	SDTMIG 4.1.2.1
ARM	Description of Planned Arm	Char	*	Synonym Qualifier	=RAW.ENRL.RANDGRP	Req	SDTMIG 4.1.2.1, SDTMIG 4.1.2.4
COUNTRY	Country	Char	(COUNTRY) ISO 3166	Record Qualifier	='USA'	Req	
DMDTC	Date/Time of Collection	Char	ISO 8601	Timing	**=RAW.ENRL.VISIT.(dmdat_yyyy,dmdat_mm,dmdat_dd) in ISO format**	Perm	SDTM 2.2.5, SDTMIG 4.1.4.1
DMDY	Study Day of Collection	Num		Timing	Not Collected; Not submitted	Perm	SDTMIG 4.1.4.1

** Indicates variable may be subject to controlled terminology, (Parenthesis indicates CDISC/NCI codelist code value)*

Part 3
A Closer Look at ADaM



Overview

- Why ADaM
- Key Concepts
- Big Picture
- Variables in General
- ADSL
- BDS
- Miscellaneous ADaM Notes
- Implementation
- Validation
- Metadata changes
- Value Level metadata
- Controlled Terminology
- Tools/Resources

Why ADaM

- Number of CDISC submissions is increasing
- Preferred and requested by many reviewers and clients
- FDA has invested in CDISC – time/training/tools
- CDISC standards part of PDUFA plan
- Provides an industry accepted standard
- More efficient in the long term
- Streamline/standardize process of creating analysis datasets and TFLs

ADaM: Key Concepts

- ADaM 2.1 is the current production version
- ADaM is more of a set of guidelines than a data model (like SDTM)
- The ADaM guidance describes the general structure, metadata, and content typically found in Analysis Datasets.
  - Models for four types of data structures
  - Analysis dataset metadata
  - Analysis variable metadata
  - Analysis results-level metadata

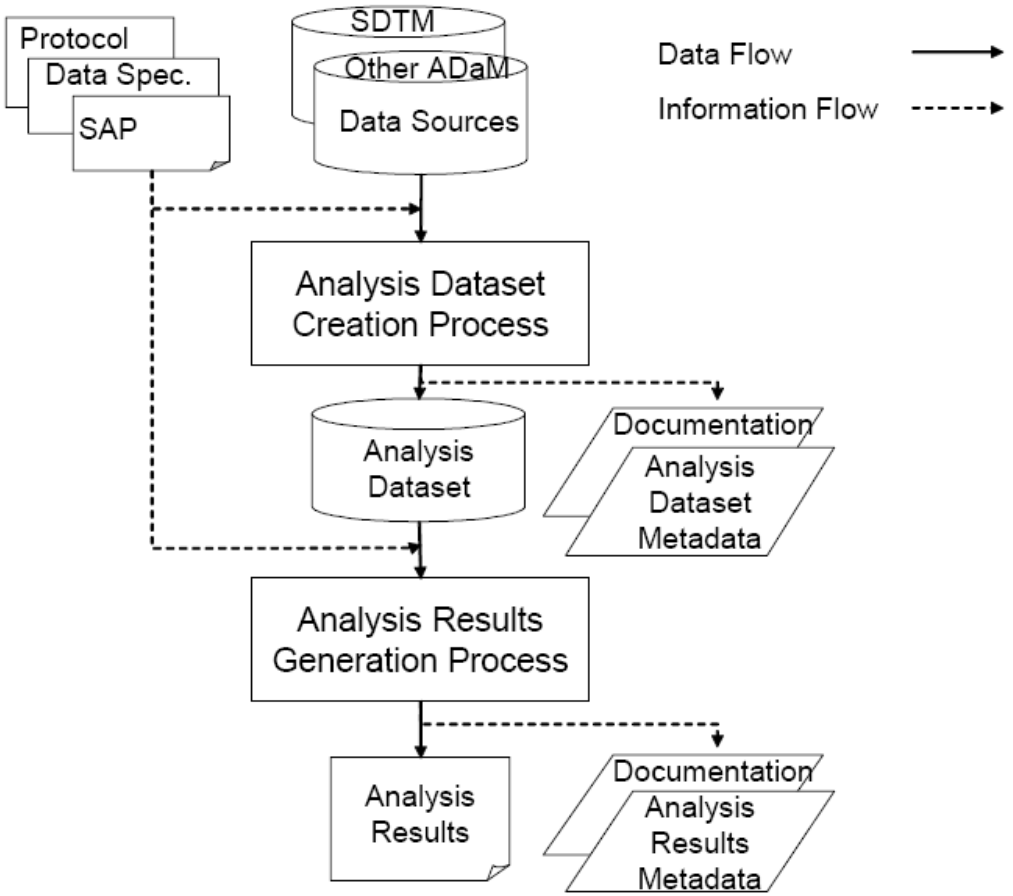
ADaM: Key Concepts

- Include all variables necessary to support statistical analyses
- Create optimal number to allow review/analysis with minimal data manipulation
- Minimum requirement is a subject level dataset
- Redundancy with SDTM is okay
- Maintain the values and attributes of SDTM variables if copied into analysis datasets without renaming (i.e., adhere to the “same name, same meaning, same values” principle of harmonization)
- Documented by machine readable metadata

ADaM: Key Concepts

- Analysis datasets should be "analysis-ready"
  - Analyzed with little or no programming
  - Analyzed with no complex data manipulation
  - "One Procedure Away"
- **Goal: reduce programming for statistical reviewers at FDA**
- Documented by machine readable metadata
- **Traceability**
  - there should be a clear understanding of the path from results to ADaM (analysis) datasets to SDTM (clinical) datasets

ADaM: Key Concepts Traceability



Glossary

- **ADaM** - Analysis Dataset Model
- **ADSL** – Subject Level Analysis Dataset
- **BDS** – Basic Data Structure
- **Controlled Terminology** - finite set of values that represent the only allowed values for a data item
- **Analysis parameter** – row identifier used to uniquely characterize a group of values that share a common definition
  Example: Systolic Blood Pressure
- **Analysis timepoint** – row identifier used to classify values within an analysis parameter into temporal or conceptual groups used for analyses
  Example: Week 2

Glossary

- **Analysis value**
  - The character (AVALC) or numeric (AVAL) value described by the analysis parameter
  - Values of certain functions are considered to be analysis values. Examples: baseline value (BASE), change from baseline (CHG).
- **Parameter-invariant** – A derived column is parameter-invariant if, whenever it is populated within an analysis dataset, it is always calculated the same way within the analysis dataset
  Example: whenever CHG is populated, it is always calculated as AVAL - BASE, regardless of the parameter

Big Picture

- Dataset types
  - ADSL: one record per subject and is required
  - BDS: basic datasets structure (e.g., vitals, ecg, or labs)
    - One record per subject per test per analysis time point
  - ADAE and ADTTE (Time to Event) now available in draft
  - If one of above structures don't work, create non-ADaM structure using ADaM concepts, defined variables and metadata model as much as possible.
  - All analysis datasets use prefix AD

Variables in General

- Character/numeric variable pairs
  - Suffixes of N and C are used. The most commonly used variable does not have a suffix.
    - Character treatment is considered most commonly used, so TRTP and TRTPN.
    - Numeric test result is considered most commonly used, so AVAL and AVALC.
- Flags/indicators
  - Character flag variables must end in FL; numeric versions must end in FN.
    - When a flag is used, the character version is required.
- SDTM datasets
  - If SDTM datasets are used as input, SDTM variables may not be changed.

Variables in General

- Length constraints
  - Variable names (8), label names (40), character variables (200)
- Date, time, and datetime variables
  - Date variables end in DT; time variables end in TM; datetime variables end in DTM.
  - *DT, *TM, and *DTM must be internally consistent.
  - Imputation flags end in DTF and TMF (as necessary).
  - Numeric versions of date/time variables are required
- Days
  - Relative day variables end in DY. There is no day zero.

Variables in General

- Timing
  - Timing start variables end in SDT and STM. Timing end variables end in EDT, ETM.
- Population flags/indicators
  - Subject-level flags must be Y/N (null values are not allowed).
  - Record-level flags end in RFL and RFN.

Variables in General #1

- Suppose you have a dataset ADQS with date variable:
  - QSSDT (Start Date of Questionnaire)
- Name the corresponding variables for:
  - Time
  - Datetime
  - Start Day
  - Date Imputation Flag
  - Time Imputation Flag
- [IG pg 25-26]

Answer

- QSSTM (Start Time of Questionnaire)
- QSSDTM (Start Date/Time of Questionnaire)
- QSSDY (Start Day Questionnaire)
- QSSDTF (Start Date Imputation Flag of Questionnaire)
- QSSTMF (Start Time Imputation Flag of Questionnaire)

Variables in General #2

- Specify the controlled terminology for
  - QSSDTF (Start Date Imputation Flag of Questionnaire)
  - QSSTMF (Start Time Imputation Flag of Questionnaire)
- [IG pg 11-12]

Answer

- D, M, Y
  - Aka, (DATEFL)
- S, M, H
  - Aka, (TIMEFL)

Key Concepts: Analysis Datasets
Data Structures: ADSL

- Subject level analysis dataset - **REQUIRED**
- Structure is 1 record/subject
- Variables
  - Some variables are required
  - Subject identifiers
  - Subject level population flags (ITT, Safety…)
  - Treatment variables
  - Timing variables (key dates)
  - Demographics (age, sex, race…)
  - Grouping variables
  - Describe subjects or events prior to treatment
  - Baseline values for critical variables
  - Factors affecting response to treatment
  - Other study-specific relevant variables

Data Structures: ADSL

Variable Name	Variable Label	Type	Codelist / Controlled Terms	Core
SEX	Sex	Char	(SEX)	Req
RACE	Race	Char	(RACE)	Req
RACEGRy	Pooled Race Group y	Char		Perm
RACEGRyN	Pooled Race Group y (N)	Num		Perm
Population Indicator(s)				
FASFL	Full Analysis Set Population Flag	Char	Y, N	Cond
SAFFL	Safety Population Flag	Char	Y, N	Cond
ITTFL	Intent-To-Treat Population Flag	Char	Y, N	Cond
PPROTFL	Per-Protocol Population Flag	Char	Y, N	Cond
COMPLFL	Completers Population Flag	Char	Y, N	Cond
RANDFL	Randomized Population Flag	Char	Y, N	Cond

Required Variable

Official CDISC Controlled Terminology

ADSL

- Study identifiers (required):
  - STUDYID, USUBJID, SUBJID, SITEID
- Demographics (required):
  - AGE, AGEU, SEX, RACE
- Treatment variables (required):
  - ARM, TRTxxP
    - xx = period number, must be 2-digit

ADSL

- Population flags (as necessary):
  - FASFL, SAFFL, ITTFL, RANDFL
- Trial dates (as necessary):
  - RANDDT,
  - TRTSDT, TRTEDT (overall treatment start and end dates),
  - TRxxSDT, TRxxEDT (period treatment start and end dates),
  - APxxSDT, APxxEDT (period start and end dates, irrespective of treatment)
    - xx = period number, must be 2-digit

ADSL #1

- There is to be a sensitivity analysis based on all subjects with at least one treatment-emergent adverse event. Provide the following information for a corresponding analysis flag variable:
  - Variable Name
  - Label
  - Type
  - Controlled Terminology
- [IG pg 15]

Answer

- SENSFL
- Sensitivity Analysis Population Flag
- Character
- Y, N

ADSL #2

- A clinical trial has one treatment period. Subjects are randomized to either:
  - 'active'
  - 'placebo'
- Name the treatment variable(s) that should be created in ADSL. [IG pg 15-16, 40]

Answer

- TRT01P (Planned Treatment for Period 01)
  - Required!
- TRT01PN (Planned Treatment for Period 01 (N))
  - Helpful for sorting.
- TRT01A (Actual Treatment for Period 01)
  - Strongly recommended.
- TRT01AN (Actual Treatment for Period 01 (N))
  - Helpful for sorting.
- ARM (Description of Planned Arm)
  - Required!

ADSL #3

- A clinical trial is a two-period crossover design with treatment arms:
  - 'active'
  - 'placebo'
- Assuming a particular subject is assigned to receive the active drug first, describe the treatment variables and values you would expect to find for this subject in ADSL. [IG pg 15-16, 40]

Answer

- ARM, TRT01P, TRT02P, TRTSEQP (and corresponding N variables)
  - ARM = 'Treatment Sequence 1: active, placebo'
  - TRTSEQP = 'active – placebo'
  - TRT01P = 'active'
  - TRT02P = 'placebo'
- TRT01A, TRT02A, TRTSEQA (and corresponding N variables) are also strongly recommended.

Key Concepts: Analysis Datasets
Data Structures: BDS Variables

- Some variables are required
- Copied from ADSL
- At least one population flag is required
- Subject identifiers (required):
  - STUDYID, USUBJID
- Treatment variables:
  - TRTP (required), TRTA (as necessary)
- Timing variables (as necessary):
  - ADT, ATM, ADTM, ADY, AVISIT(N), ATPT(N), APERIOD(C), APERSDT, APEREDT
    - At least 1 visit variable is required
- Parameter variables
  - PARAM, PARAMCD, AVAL(C) (required)
  - PARAMTYP, BASE, CHG, PCHG, R2BASE, BASECATy, CHGCATy, SHIFTy, CRITy (as necessary)

BDS Variables

- Analysis descriptors (as necessary):
  - DTYPE
- Analysis visit windowing (as necessary):
  - AWTARGET, AWDIFF, AWLO, AWHI, AWU
- Range variables (as necessary):
  - ANRLO, ANRHI, ANRIND, BNRIND
- Flag variables (as necessary):
  - ABLFL, ONTRTFL, LVOTFL, PPROTRFL

BDS #1

- A clinical trial is a two-period crossover design. Name the treatment variables that should be created for a BDS dataset. [IG pg 21]

Answer

- TRTP (Planned Treatment)
  - Required!
- TRTPN (Planned Treatment (N))
  - Helpful for sorting.
- TRTA (Actual Treatment)
  - Strongly recommended.
- TRTAN (Actual Treatment(N))
  - Helpful for sorting.

BDS Example #1

Row	USUBJID	TRTP	TRTPN	TRTA	TRTAN	APERIOD	APERIODC	PARAMCD	AVISIT	AVISITN	AVAL	ABLFL	BASE	CHG	TRT01P	TRT02P	TRTSEQP
1	1001							ALT	Screening	-1	16		16		active	placebo	active - placebo
2	1001							ALT	Baseline	0	16	Y	16		active	placebo	active - placebo
3	1001	active	1	active	1	1	Period 1	ALT	Day 1	1	18		16	2	active	placebo	active - placebo
4	1001	active	1	active	1	1	Period 1	ALT	Day 2	2	17		16	1	active	placebo	active - placebo
5	1001	placebo	2	placebo	2	2	Period 2	ALT	Day 4	3	14		16	-2	active	placebo	active - placebo
6	1001	placebo	2	placebo	2	2	Period 2	ALT	Day 5	4	10		16	-6	active	placebo	active - placebo
7	1002							ALT	Screening	-1	12		11		placebo	active	placebo - active
8	1002							ALT	Baseline	0	11	Y	11		placebo	active	placebo - active
9	1002	placebo	2	placebo	2	1	Period 1	ALT	Day 1	1	14		11	3	placebo	active	placebo - active
10	1002	placebo	2	active	1	1	Period 1	ALT	Day 2	2	15		11	4	placebo	active	placebo - active
11	1002	active	1	active	1	2	Period 2	ALT	Day 4	3	14		11	3	placebo	active	placebo - active
12	1002	active	1	active	1	2	Period 2	ALT	Day 5	4	15		11	4	placebo	active	placebo - active

BDS #2

- A clinical trial is a two-period crossover design. Dataset ADSL has corresponding treatment and timing variables:
  - TRT01P, TRT02P (period treatments)
  - AP01SDTM, AP01EDTM (period 1 start and end)
  - AP02SDTM, AP02EDTM (period 2 start and end)
- What BDS variables would one use in conjunction with the above ADSL variables in order to derive TRTP? [IG pg 21 (kinda sorta)]

Answer

- ADTM (Analysis Date/Time)
  - ADSL.TRT01P if ADLB.ADTM between ADSL.AP01SDTM and ADSL.AP01EDTM
  - ADSL.TRT02P if ADLB.ADTM between ADSL.AP02SDTM and ADSL.AP02EDTM

BDS Example #2

Row	USUBJID	TRTP	TRTPN	TRTA	TRTAN	PARAMCD	AVISIT	ADTM	AVAL	BASE	CHG	AP01SDTM	AP01EDTM	AP02SDTM	AD02EDTM
1	1001					ALT	Screening	09NOV11:09:00	16	16		10NOV11:10:00	13NOV11:12:00	13NOV11:12:01	15NOV11:12:00
2	1001					ALT	Baseline	10NOV11:09:10	16	16		10NOV11:10:00	13NOV11:12:00	13NOV11:12:01	15NOV11:12:00
3	1001	active	1	active	1	ALT	Day 1	11NOV11:09:30	18	16	2	10NOV11:10:00	13NOV11:12:00	13NOV11:12:01	15NOV11:12:00
4	1001	active	1	active	1	ALT	Day 2	12NOV11:09:00	17	16	1	10NOV11:10:00	13NOV11:12:00	13NOV11:12:01	15NOV11:12:00
5	1001	placebo	2	placebo	2	ALT	Day 4	14NOV11:09:45	14	16	-2	10NOV11:10:00	13NOV11:12:00	13NOV11:12:01	15NOV11:12:00
6	1001	placebo	2	placebo	2	ALT	Day 5	15NOV11:09:30	10	16	-6	10NOV11:10:00	13NOV11:12:00	13NOV11:12:01	15NOV11:12:00
7	1002					ALT	Screening	09NOV11:11:00	12	11		10NOV11:12:00	13NOV11:12:15	13NOV11:12:16	15NOV11:12:30
8	1002					ALT	Baseline	10NOV11:11:10	11	11		10NOV11:12:00	13NOV11:12:15	13NOV11:12:16	15NOV11:12:30
9	1002	placebo	2	placebo	2	ALT	Day 1	11NOV11:12:30	14	11	3	10NOV11:12:00	13NOV11:12:15	13NOV11:12:16	15NOV11:12:30
10	1002	placebo	2	active	1	ALT	Day 2	12NOV11:11:00	15	11	4	10NOV11:12:00	13NOV11:12:15	13NOV11:12:16	15NOV11:12:30
11	1002	active	1	active	1	ALT	Day 4	14NOV11:09:45	14	11	3	10NOV11:12:00	13NOV11:12:15	13NOV11:12:16	15NOV11:12:30
12	1002	active	1	active	1	ALT	Day 5	15NOV11:10:30	15	11	4	10NOV11:12:00	13NOV11:12:15	13NOV11:12:16	15NOV11:12:30

BDS #3

- Name the BDS variable meant to contain the units of the data. [IG pg 27]

Answer

- PARAM (Parameter)
  - E.g., PARAM = ‘LDL Cholesterol (mg/dL)’

Miscellaneous Notes

- Keeping raw/SDTM variables is not necessary, but repurposing them is prohibited.
- BDS records for analysis are not always identifiable by AVISIT alone. It is often the case that some combination of DTYPE, PARAMTYP, ANLzzFL, and *RFL will be required (varies by complexity).

Miscellaneous Notes

- In BDS, the strong default is to add rows, not columns. To quote the implementation guide, "Avoid undue horizontalization." E.g.,:
  - Log transformations
  - SI and US units in same study
  - Outcomes based on multiple records (e.g., average-based baselines, cumulative outcomes, etc.)
  - Repeated analysis on multiple populations (e.g., ITTRFL and PPROTRFL)
  - Imputation-based analysis (e.g., LOCF)
  - Crossover-necessitated record reutilization (e.g., period 2 baseline is also period 1 endpoint)

Miscellaneous #1

- Several parameters from dataset ADLB are to be analyzed on both linear and log scales. When creating additional records to contain the log results, describe how you would differentiate them from the original linear scale results. [IG pg 45]

Answer

- PARAM (Parameter)
  - PARAM = 'Log10(LDL Cholesterol (mg/dL))'
- PARAMCD (Parameter Code)
  - PARAMCD = 'L10LDL'
- PARAMTYP (Parameter Type)
  - PARAMTYP = 'DERIVED'

Miscellaneous Example #1

Row	USUBJID	PARAMCD	PARAM	PARAMTYP	AVISIT	AVISITN	AVAL	ABLFL	BASE	CHG	TRT01P
1	1001	LDL	LDL Cholesterol (mg/dL)		Screening	-1	147		148		active
2	1001	LDL	LDL Cholesterol (mg/dL)		Baseline	0	148	Y	148		active
3	1001	LDL	LDL Cholesterol (mg/dL)		Week 1	1	155		148	7	active
4	1001	LDL	LDL Cholesterol (mg/dL)		Week 2	2	160		148	12	active
5	1001	LDL	LDL Cholesterol (mg/dL)		Week 3	3	162		148	14	active
6	1001	LDL	LDL Cholesterol (mg/dL)		Week 4	4	152		148	4	active
7	1001	L10LDL	Log10(LDL Cholesterol (mg/dL))	DERIVED	Screening	-1	2.167		2.17		active
8	1001	L10LDL	Log10(LDL Cholesterol (mg/dL))	DERIVED	Baseline	0	2.17	Y	2.17		active
9	1001	L10LDL	Log10(LDL Cholesterol (mg/dL))	DERIVED	Week 1	1	2.19		2.17	0.02	active
10	1001	L10LDL	Log10(LDL Cholesterol (mg/dL))	DERIVED	Week 2	2	2.204		2.17	0.03	active
11	1001	L10LDL	Log10(LDL Cholesterol (mg/dL))	DERIVED	Week 3	3	2.21		2.17	0.04	active
12	1001	L10LDL	Log10(LDL Cholesterol (mg/dL))	DERIVED	Week 4	4	2.182		2.17	0.01	active

Miscellaneous #2

- In this study the primary efficacy analysis is based on the average of the last two observations for a subject . When creating an additional record to contain this result, describe how you would differentiate this additional record from the original results records. [IG pg 47]

Answer

- AVISIT (Analysis Visit)
  - AVISIT = 'Endpoint'
- AVISITN (Analysis Visit (N))
  - AVISITN = 9999
- DTYPE (Derivation Type)
  - DTYPE = 'AVERAGE'

Miscellaneous Example #2

Row	USUBJID	PARAMCD	PARAM	DTYPE	AVISIT	AVISITN	AVAL	ABLFL	BASE	CHG	TRT01P
1	1001	LDL	LDL Cholesterol (mg/dL)		Screening	-1	147		148		active
2	1001	LDL	LDL Cholesterol (mg/dL)		Baseline	0	148	Y	148		active
3	1001	LDL	LDL Cholesterol (mg/dL)		Week 1	1	155		148	7	active
4	1001	LDL	LDL Cholesterol (mg/dL)		Week 2	2	160		148	12	active
5	1001	LDL	LDL Cholesterol (mg/dL)		Week 3	3	162		148	14	active
6	1001	LDL	LDL Cholesterol (mg/dL)		Week 4	4	152		148	4	active
7	1001	LDL	LDL Cholesterol (mg/dL)	AVERAGE	Endpoint	9999	157		148	9	active

Miscellaneous #3

- A study has 5 scheduled post-baseline visits. The SAP requires by-visit summaries at visits 3-5 based on each of:
  - Observed
  - LOCF
  - WOCF
- Describe the BDS records you would create to support such analyses. [IG pg 49]

Miscellaneous Example #3

Row	USUBJID	PARAMCD	PARAM	DTYPE	VISITNUM	AVISIT	AVISITN	AVAL	ABLFL	BASE	CHG	TRT01P
1	1001	SIXM	Six Minute Walk (m)		-1	Screening	-1	405		410		active
2	1001	SIXM	Six Minute Walk (m)		0	Baseline	0	410	Y	410		active
3	1001	SIXM	Six Minute Walk (m)		1	Week 1	1	392		410	-18	active
4	1001	SIXM	Six Minute Walk (m)		2	Week 2	2	415		410	5	active
5	1001	SIXM	Six Minute Walk (m)		3	Week 3	3	408		410	-2	active
6	1002	SIXM	Six Minute Walk (m)		-1	Screening	-1	385		385		placebo
7	1002	SIXM	Six Minute Walk (m)		0	Baseline	0	385	Y	385		placebo
8	1002	SIXM	Six Minute Walk (m)		1	Week 1	1	370		385	-15	placebo
9	1002	SIXM	Six Minute Walk (m)		2	Week 2	2	365		385	-20	placebo
10	1002	SIXM	Six Minute Walk (m)		3	Week 3	3	380		385	-5	placebo
11	1002	SIXM	Six Minute Walk (m)		4	Week 4	4	395		385	10	placebo
12	1002	SIXM	Six Minute Walk (m)		5	Week 5	5	415		385	30	placebo

Miscellaneous #3 Mock

		Result				
Visit	N	Mean	Std	Min	Med	Max
Baseline(1)	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 3	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 4	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 5	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 3 LOCF	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 4 LOCF	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 5 LOCF	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 3 WOCF	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 4 WOCF	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X
Week 5 WOCF	XX	XXX.XX	XX.XXX	XXX.X	XXX.XX	XXX.X

Answer

- AVISIT, AVISITN, DTYPE
  - For the observed analysis, use the data as is.
  - For the LOCF analysis, create additional records with DTYPE = ‘LOCF’ corresponding to each unrepresented visit.
  - For the WOCF analysis, create additional records with DTYPE = ‘WOCF’ corresponding to each unrepresented visit.

Miscellaneous Example #3

Row	USUBJID	PARAMCD	PARAM	DTYPE	VISITNUM	AVISIT	AVISITN	AVAL	ABLFL	BASE	CHG	TRT01P
1	1001	SIXM	Six Minute Walk (m)		-1	Screening	-1	405		410		active
2	1001	SIXM	Six Minute Walk (m)		0	Baseline	0	410	Y	410		active
3	1001	SIXM	Six Minute Walk (m)		1	Week 1	1	392		410	-18	active
4	1001	SIXM	Six Minute Walk (m)		2	Week 2	2	415		410	5	active
5	1001	SIXM	Six Minute Walk (m)		3	Week 3	3	408		410	-2	active
6	1001	SIXM	Six Minute Walk (m)	LOCF	3	Week 4	4	408		410	-2	active
7	1001	SIXM	Six Minute Walk (m)	WOCF	1	Week 4	4	392		410	-18	active
8	1001	SIXM	Six Minute Walk (m)	LOCF	3	Week 5	5	408		410	-2	active
9	1001	SIXM	Six Minute Walk (m)	WOCF	1	Week 5	5	392		410	-18	active
10	1002	SIXM	Six Minute Walk (m)		-1	Screening	-1	385		385		placebo
11	1002	SIXM	Six Minute Walk (m)		0	Baseline	0	385	Y	385		placebo
12	1002	SIXM	Six Minute Walk (m)		1	Week 1	1	370		385	-15	placebo
13	1002	SIXM	Six Minute Walk (m)		2	Week 2	2	365		385	-20	placebo
14	1002	SIXM	Six Minute Walk (m)		3	Week 3	3	380		385	-5	placebo
15	1002	SIXM	Six Minute Walk (m)		4	Week 4	4	395		385	10	placebo
16	1002	SIXM	Six Minute Walk (m)		5	Week 5	5	415		385	30	placebo

Additional Thoughts

- Note that on the created LOCF records, the SDTM variables such as VISITNUM are retained from the record that begets the created one. This is done for traceability. It is only the ADaM variables such as AVISITN that are finessed in support of the subsequent analysis.
- Note that one does not create LOCF or WOCF records corresponding to visits that are represented in the data. This gives rise to analysis metadata (aka, annotations) something like:
  - Observed Analyses:
    - where AVISITN in (3 4 5) and DTYPE = '';
  - LOCF Analyses:
    - where AVISITN in (3 4 5) and DTYPE in ('' 'LOCF');
  - WOCF Analyses :
    - where AVISITN in (3 4 5) and DTYPE in ('' 'WOCF');

ADaM Implementation

- All analysis databases must contain ADSL
- Create analysis datasets using the BDS structure where appropriate
  - Lab
  - ECG
  - Vitals
  - PE
  - Other finding like datasets
- Use draft ADAE model for adverse events
- Use draft ADTTE for time to event

ADaM Implementation

- Where ADaM models don't apply create your own structure
  - ADaM identifier variables are required
  - Use ADaM variables when possible
  - Use ADaM principles whenever possible
- Always use ADaM metadata model
- Use controlled terminology when available

ADaM: Validation

- Additional validation necessary
- Datasets/metadata validate to the ADaM datasets and metadata standards
- Published set of ADaM Checks
- OPENCDISC

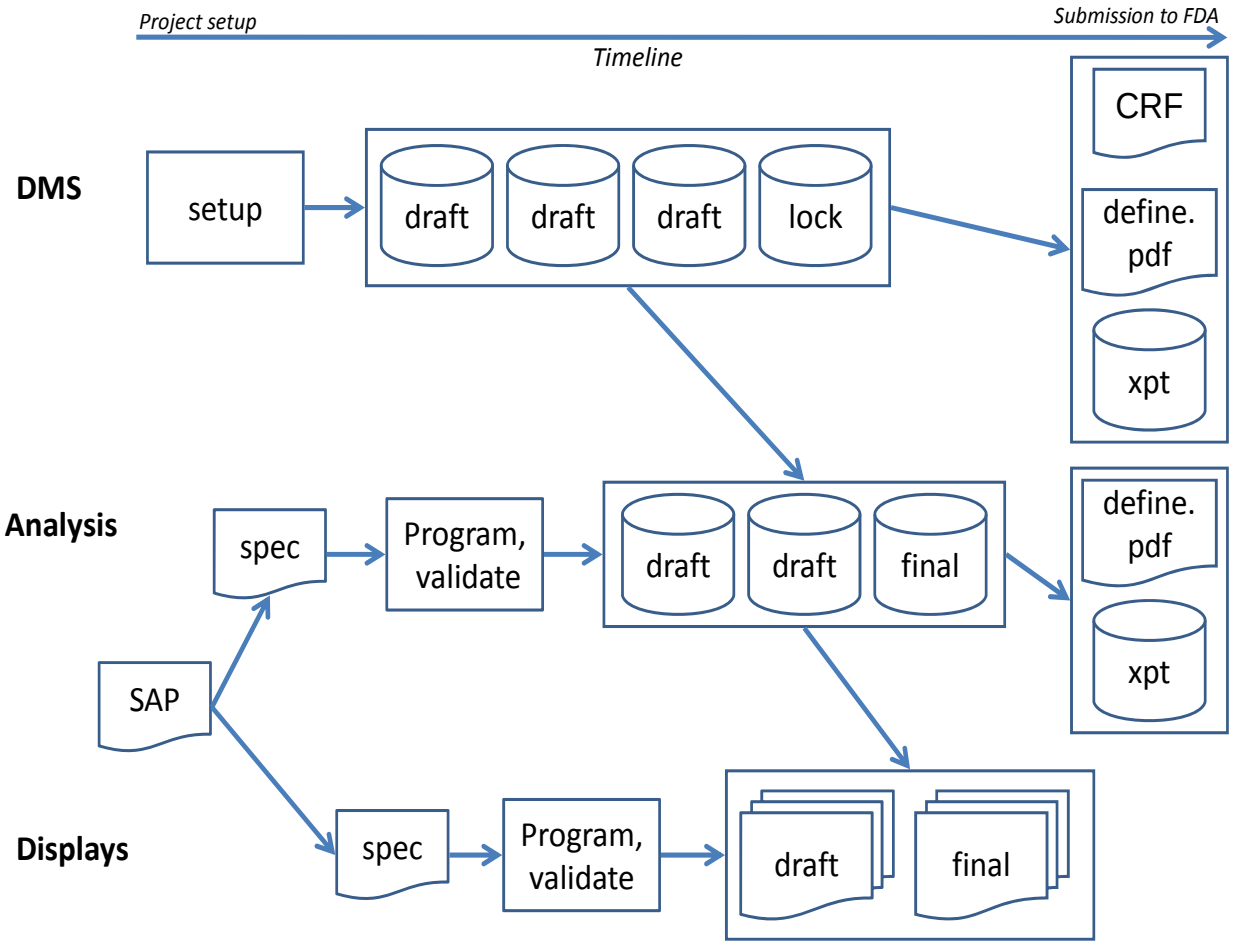
Part 4

Implementation of CDISC Standards within Your Organization

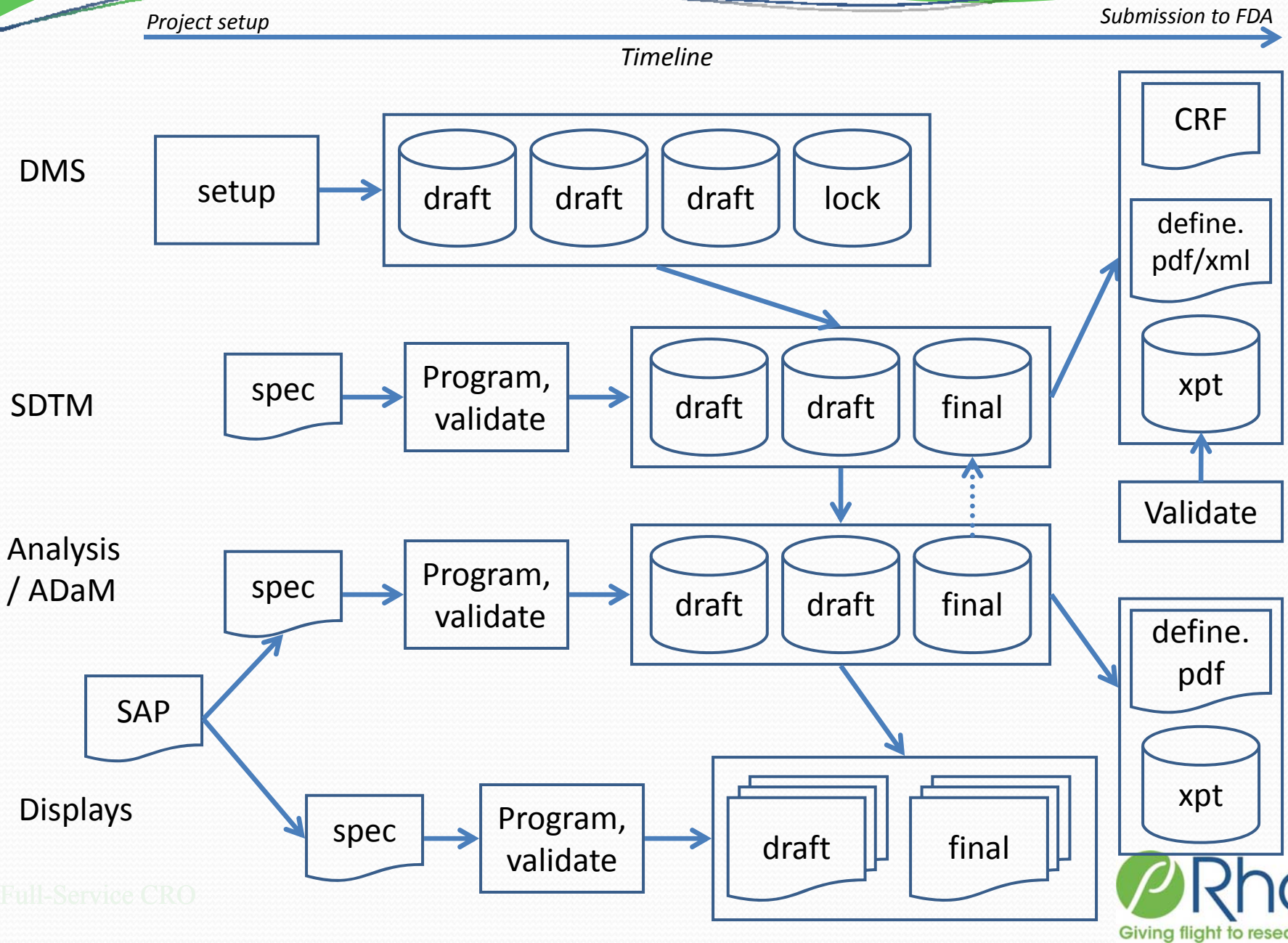
Overview

- General Workflow Discussion
- Resourcing
- Project Planning
- Training
- Technology/Tools
  - Demonstrations
- Future Considerations

Clinical Trial Workflow before CDISC



Clinical Trial Workflow with CDISC

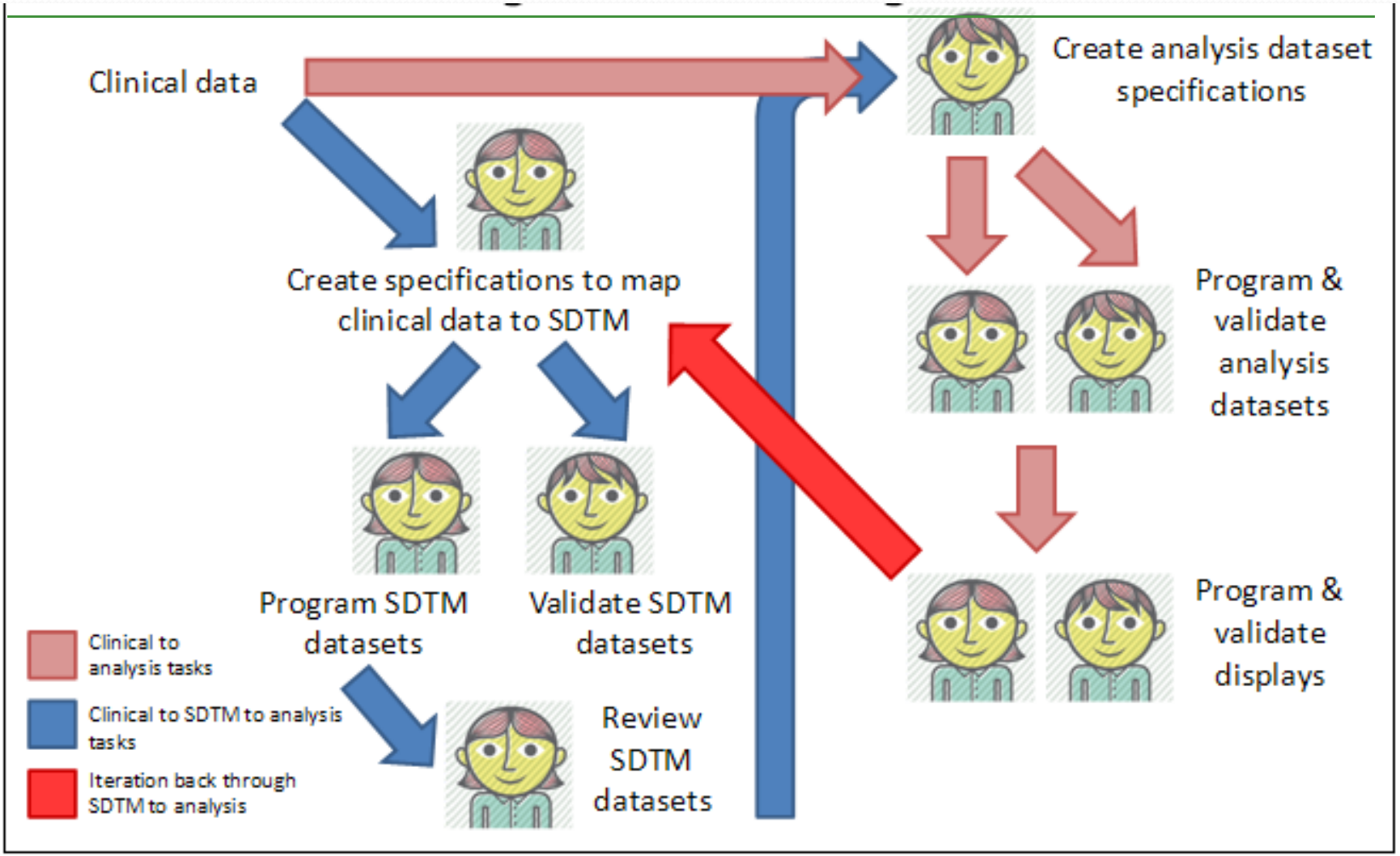


Rho
Giving flight to research

CDISC Workflow: Optimal Scenario

- Clinical database uses CDISC SDTM standard
- Analysis database uses ADaM standard
- The SDTM database used as input for ADaM
- **Traceability** is a key principle
- To maintain traceability, the cleanest path is from clinical data to SDTM to ADaM to displays/analysis.

Resourcing



Resourcing

- More work + new skill sets
- Traditional skills
- Familiarity with SDTM, ADAM, ISO standards
- XML + XSL + related XML technologies
- Internet browsers
- HTML
- JavaScript
- Database (Oracle, ACCESS, etc.)
- Version control software

Project Planning

- More components to projects that incorporate the CDISC SDTM and ADaM models
- Expect more of everything except time
  - More moving parts
  - More handoffs
  - More resources
  - More deliverables
- More up front planning a must

Project Planning

- Goals
  - # days from DBL to Top line results unchanged
  - High quality deliverables
  - Control costs
- Extensive up front planning between the SDTM team, data management, statistical programmers, and statisticians

Project Planning

- Continued planning and communication through lifecycle of the project
  - Study setup - DM and SDTM team plan mapping from DM to SDTM
  - Data collection
    - Program SDTM with interim data
    - SDTM and analysis teams coordinate spec'ing and programming analysis and displays
  - Changes to clinical data
    - Changes to SDTM
    - Communicate to SDTM team and analysis team
  - SDTM Validation - Changes to SDTM and analysis
  - Analysis issues – Changes to SDTM and maybe DM

Training

- Non-Technical Training
  - Understanding the ‘what’, not the ‘how’
- Technical Training
  - SDTM
  - SDTM Lite
  - ADaM
  - ODM/XML
  - CDASH

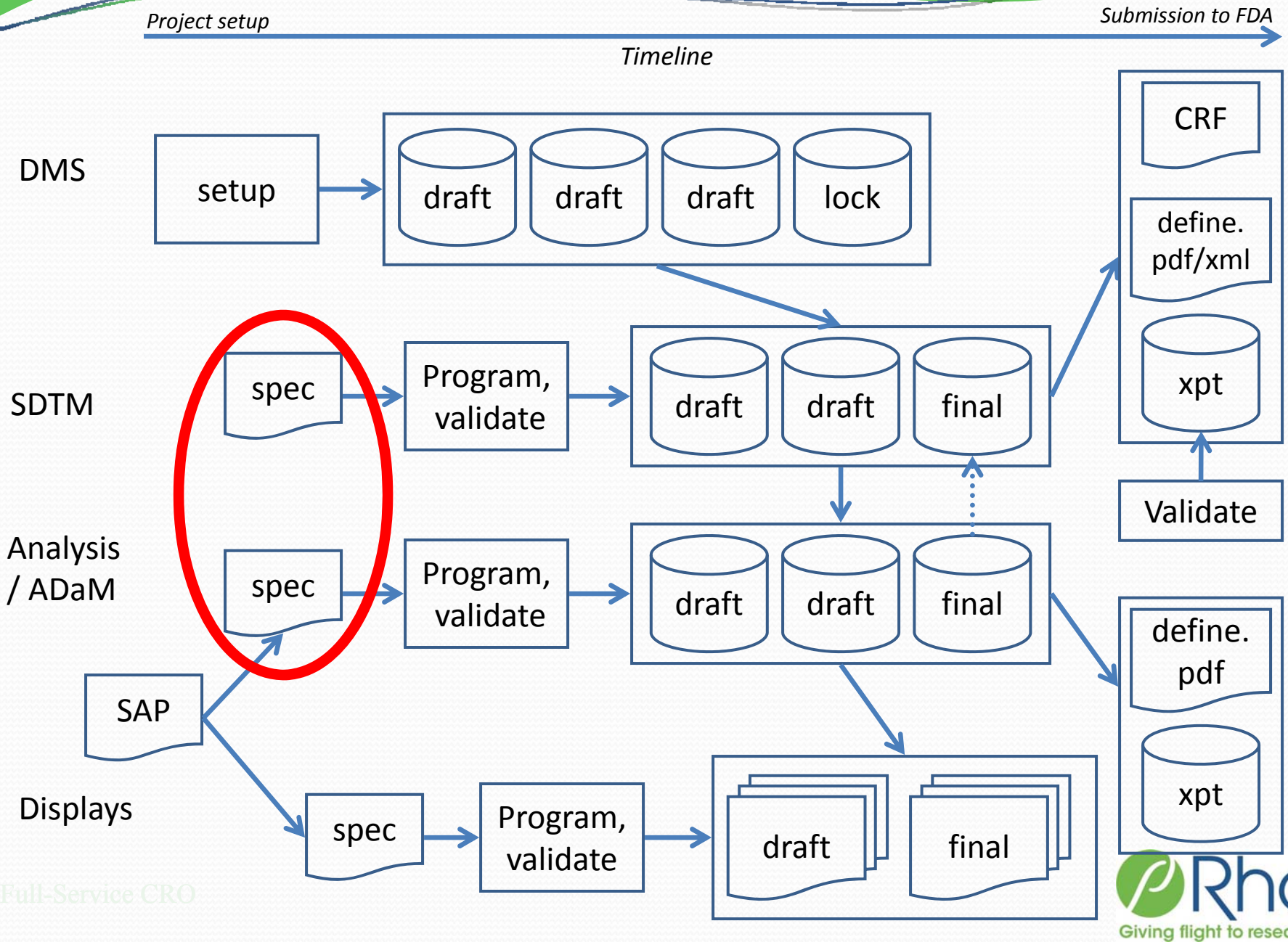
Technology/Tools

- Tools essential to successful implementation
- It's all about the metadata
- Multi-purpose metadata
- XML-related technologies
- Validation tools

Technology/Tools: Metadata

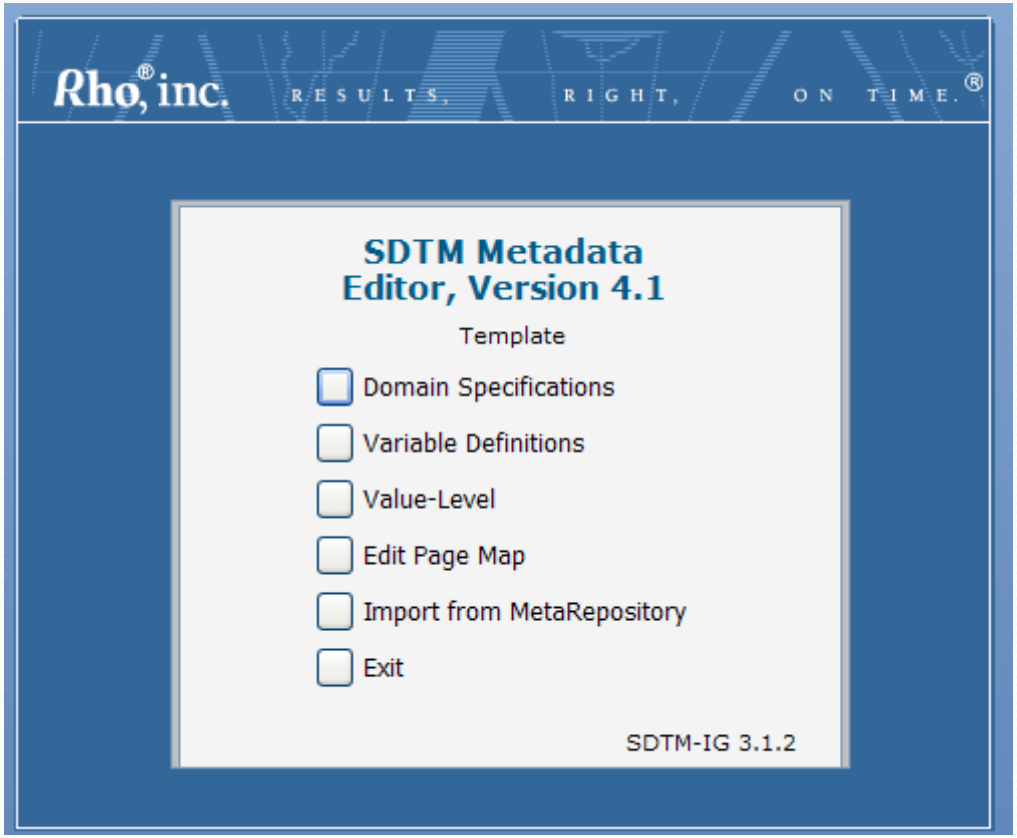
- Machine readable metadata now a must have
- Standards are metadata
- Most of the metadata is pre-determined
- Standards should utilize standard metadata libraries
  - Facilitates compliance to standards
  - Standard attributes
  - Standard controlled terminology
  - Repository
- Repurposing a must

Technology/Tools Demonstrations: Specifications



Rho
Giving flight to research

SDTM Specifications



SDTM Domain Metadata

Rho® SDTM Metadata Editor Rev. 4.1, 2010-08-24

Domain Specifications for Template Study

S:\Submissions\Training\Sct_workshop\Sdtm_excersises\Sdtm.mdb

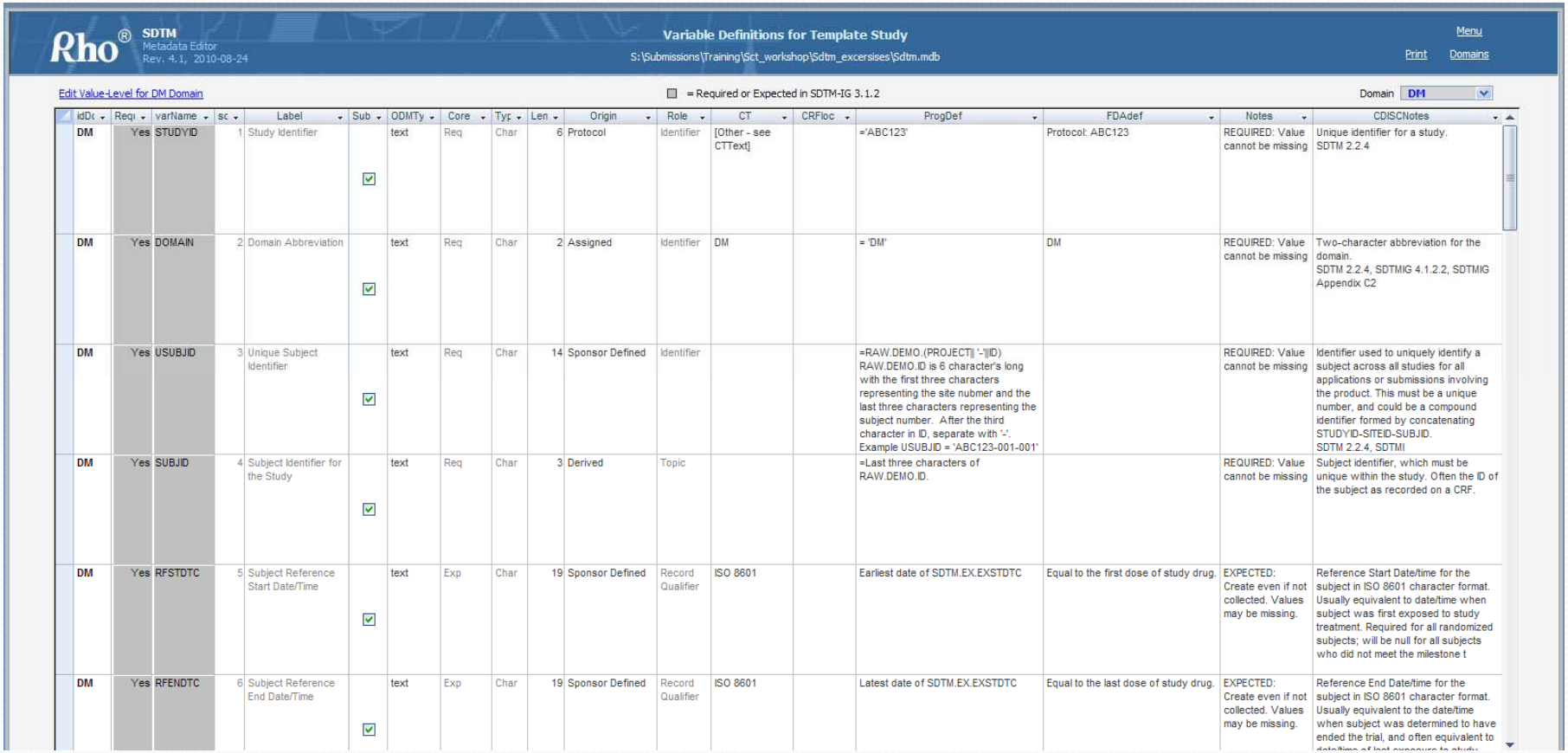
Menu | Print | Variables

Copy Domain

☐ = Standard SDTM-IG 3.1.2 Domain

Domain	Description	SubmitDB	Purpose	Keys	Class	Input_notes	Output_notes	Notes	Structure	LastEditDateTim	LastEditedBy
AE	Adverse Events	☐	Tabulation	STUDYID, USUBJID, AETERM, AESTDTC, AEDTC	Events				One record per adverse event per subject		
CE	Clinical Events	☐	Tabulation	STUDYID, USUBJID, CETERM, CESTDTC	Events				One record per event per subject		
CM	Concomitant Meds	☑	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC	Interventions	Input dataset is RAW.CMED.			One record per recorded medication occurrence per subject	5/15/2012 9:22:13 AM	Carol Baker
CO	Comments	☐	Tabulation	STUDYID, USUBJID, CODTC, COVAL, RDOMAIN	Special Purpose				One record per comment per subject		
DA	Drug Accountability	☐	Tabulation	STUDYID, USUBJID, DATESTCD, DADTC	Findings				One record per drug accountability finding per subject		
DM	Demographics	☑	Tabulation	STUDYID, USUBJID	Special Purpose	Input datasets are RAW.DEMO, RAW.ENRL and SDTM.EX.			One record per subject	5/14/2012 1:57:43 PM	Carol Baker
DS	Disposition		Tabulation	STUDYID,	Events				One record per		

SDTM Variable Metadata



Rho® SDTM Metadata Editor Rev. 4.1, 2010-08-24

Variable Definitions for Template Study

S:\Submissions\Training\Sct_workshop\Sdtm_excersises\Sdtm.mdb

Menu | Print | Domains

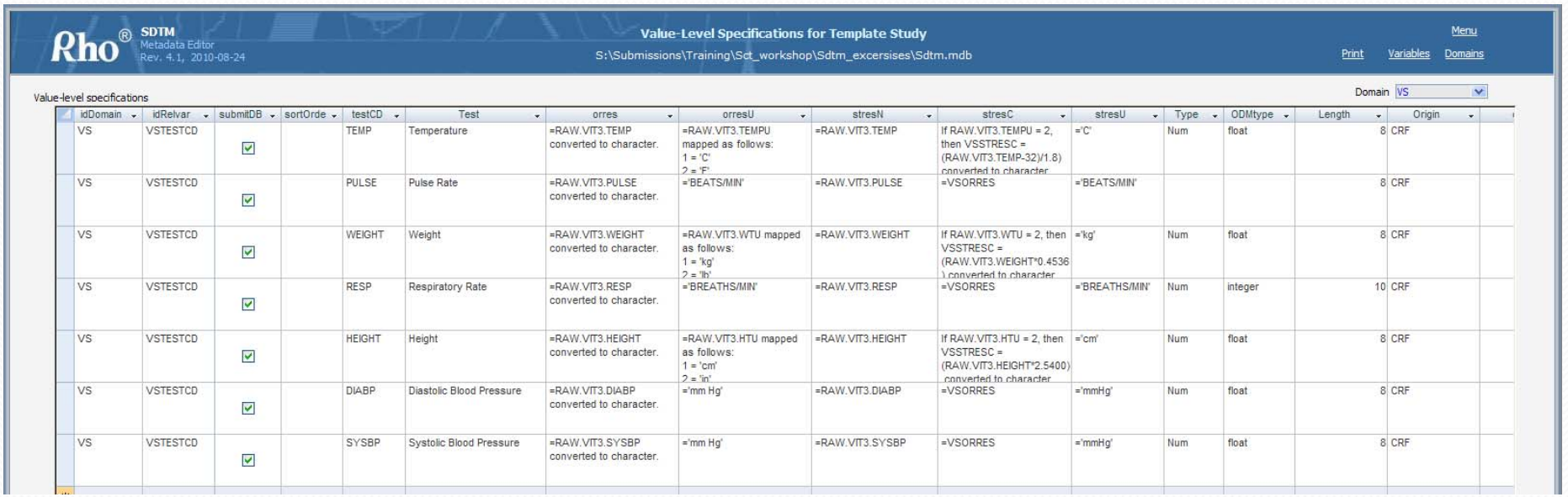
Edit Value-Level for DM Domain

= Required or Expected in SDTM-IG 3.1.2

Domain: DM

idDe	Requ	varName	sc	Label	Sub	ODMTy	Core	Typ	Len	Origin	Role	CT	CRFloc	ProgDef	FDAdef	Notes	CDISCNotes
DM	Yes	STUDYID	1	Study Identifier	☑	text	Req	Char	6	Protocol	Identifier	[Other - see CTText]		='ABC123'	Protocol: ABC123	REQUIRED: Value cannot be missing	Unique identifier for a study. SDTM 2.2.4
DM	Yes	DOMAIN	2	Domain Abbreviation	☑	text	Req	Char	2	Assigned	Identifier	DM		= 'DM'	DM	REQUIRED: Value cannot be missing	Two-character abbreviation for the domain. SDTM 2.2.4, SDTMIG 4.1.2.2, SDTMIG Appendix C2
DM	Yes	USUBJID	3	Unique Subject Identifier	☑	text	Req	Char	14	Sponsor Defined	Identifier			=RAW.DEMO.(PROJECT‖ '-'‖ID) RAW.DEMO.ID is 6 character's long with the first three characters representing the site nubmer and the last three characters representing the subject number. After the third character in ID, separate with '-'. Example USUBJID = 'ABC123-001-001'		REQUIRED: Value cannot be missing	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. This must be a unique number, and could be a compound identifier formed by concatenating STUDYID-SITEID-SUBJID. SDTM 2.2.4, SDTMI
DM	Yes	SUBJID	4	Subject Identifier for the Study	☑	text	Req	Char	3	Derived	Topic			=Last three characters of RAW.DEMO.ID.		REQUIRED: Value cannot be missing	Subject identifier, which must be unique within the study. Often the ID of the subject as recorded on a CRF.
DM	Yes	RFSTDTC	5	Subject Reference Start Date/Time	☑	text	Exp	Char	19	Sponsor Defined	Record Qualifier	ISO 8601		Earliest date of SDTM.EX.EXSTDTC	Equal to the first dose of study drug.	EXPECTED: Create even if not collected. Values may be missing.	Reference Start Date/time for the subject in ISO 8601 character format. Usually equivalent to date/time when subject was first exposed to study treatment. Required for all randomized subjects; will be null for all subjects who did not meet the milestone t
DM	Yes	RFENDTC	6	Subject Reference End Date/Time	☑	text	Exp	Char	19	Sponsor Defined	Record Qualifier	ISO 8601		Latest date of SDTM.EX.EXSTDTC	Equal to the last dose of study drug.	EXPECTED: Create even if not collected. Values may be missing.	Reference End Date/time for the subject in ISO 8601 character format. Usually equivalent to the date/time when subject was determined to have ended the trial, and often equivalent to

SDTM Value Level Metadata



Rho® SDTM Metadata Editor Rev. 4.1, 2010-08-24

Value-Level Specifications for Template Study

S:\Submissions\Training\Sct_workshop\Sdtm_excersises\Sdtm.mdb

Menu | Print | Variables | Domains

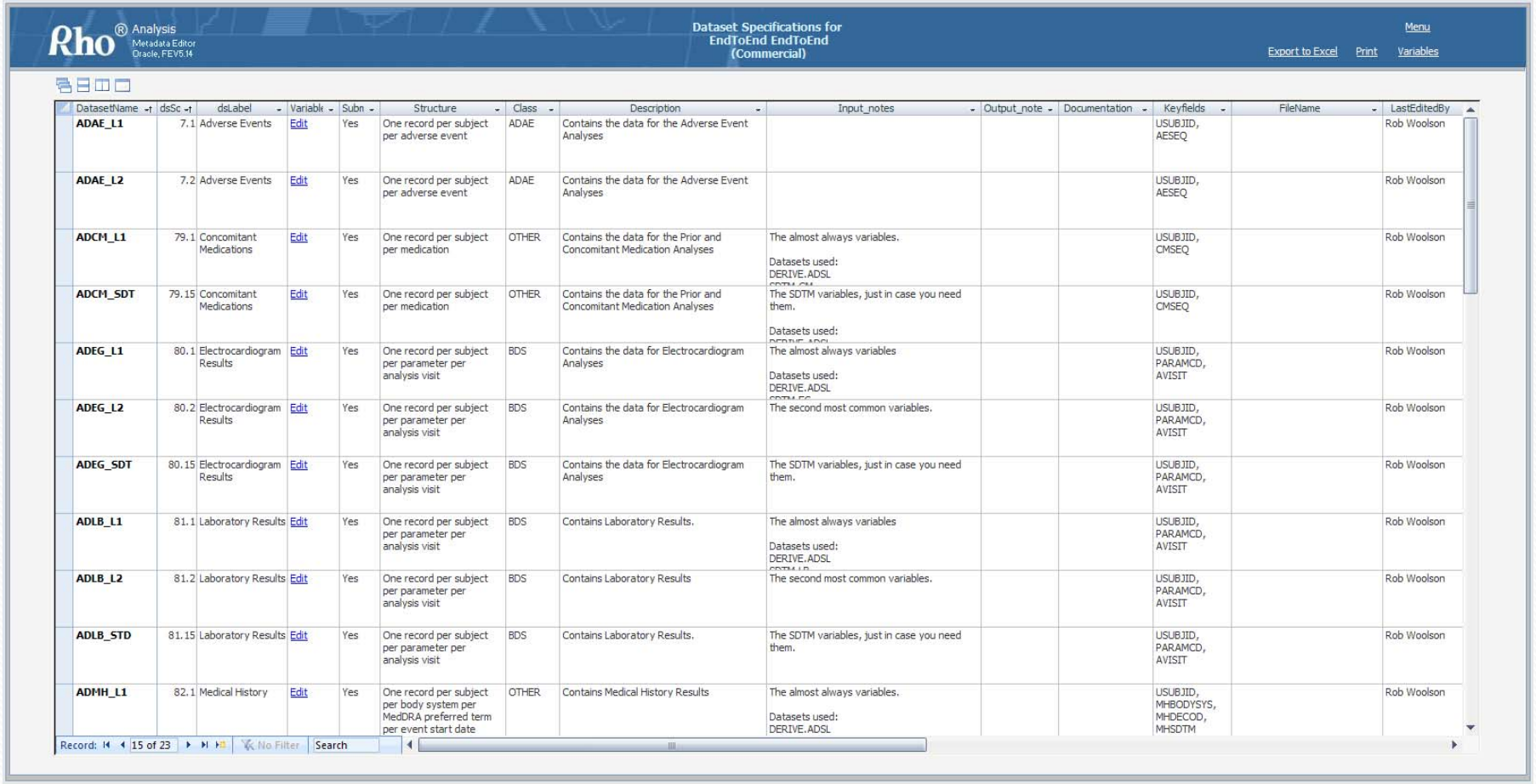
Value-level specifications

Domain VS

idDomain	idRelvar	submitDB	sortOrde	testCD	Test	orres	orresU	stresN	stresC	stresU	Type	ODMtype	Length	Origin
VS	VSTESTCD	☑		TEMP	Temperature	=RAW.VIT3.TEMP converted to character.	=RAW.VIT3.TEMPU mapped as follows: 1 = 'C' 2 = 'F'	=RAW.VIT3.TEMP	If RAW.VIT3.TEMPU = 2, then VSSTRESC = (RAW.VIT3.TEMP-32)/1.8) converted to character	='C'	Num	float	8	CRF
VS	VSTESTCD	☑		PULSE	Pulse Rate	=RAW.VIT3.PULSE converted to character.	='BEATS/MIN'	=RAW.VIT3.PULSE	=VSORRES	='BEATS/MIN'			8	CRF
VS	VSTESTCD	☑		WEIGHT	Weight	=RAW.VIT3.WEIGHT converted to character.	=RAW.VIT3.WTU mapped as follows: 1 = 'kg' 2 = 'lb'	=RAW.VIT3.WEIGHT	If RAW.VIT3.WTU = 2, then VSSTRESC = (RAW.VIT3.WEIGHT*0.4536) converted to character	='kg'	Num	float	8	CRF
VS	VSTESTCD	☑		RESP	Respiratory Rate	=RAW.VIT3.RESP converted to character.	='BREATHS/MIN'	=RAW.VIT3.RESP	=VSORRES	='BREATHS/MIN'	Num	integer	10	CRF
VS	VSTESTCD	☑		HEIGHT	Height	=RAW.VIT3.HEIGHT converted to character.	=RAW.VIT3.HTU mapped as follows: 1 = 'cm' 2 = 'in'	=RAW.VIT3.HEIGHT	If RAW.VIT3.HTU = 2, then VSSTRESC = (RAW.VIT3.HEIGHT*2.5400) converted to character	='cm'	Num	float	8	CRF
VS	VSTESTCD	☑		DIABP	Diastolic Blood Pressure	=RAW.VIT3.DIABP converted to character.	='mm Hg'	=RAW.VIT3.DIABP	=VSORRES	='mmHg'	Num	float	8	CRF
VS	VSTESTCD	☑		SYSBP	Systolic Blood Pressure	=RAW.VIT3.SYSBP converted to character.	='mm Hg'	=RAW.VIT3.SYSBP	=VSORRES	='mmHg'	Num	float	8	CRF

ADaM Specifications

ADaM Dataset Metadata



Rho® Analysis
Metadata Editor
Oracle, FEV5.14

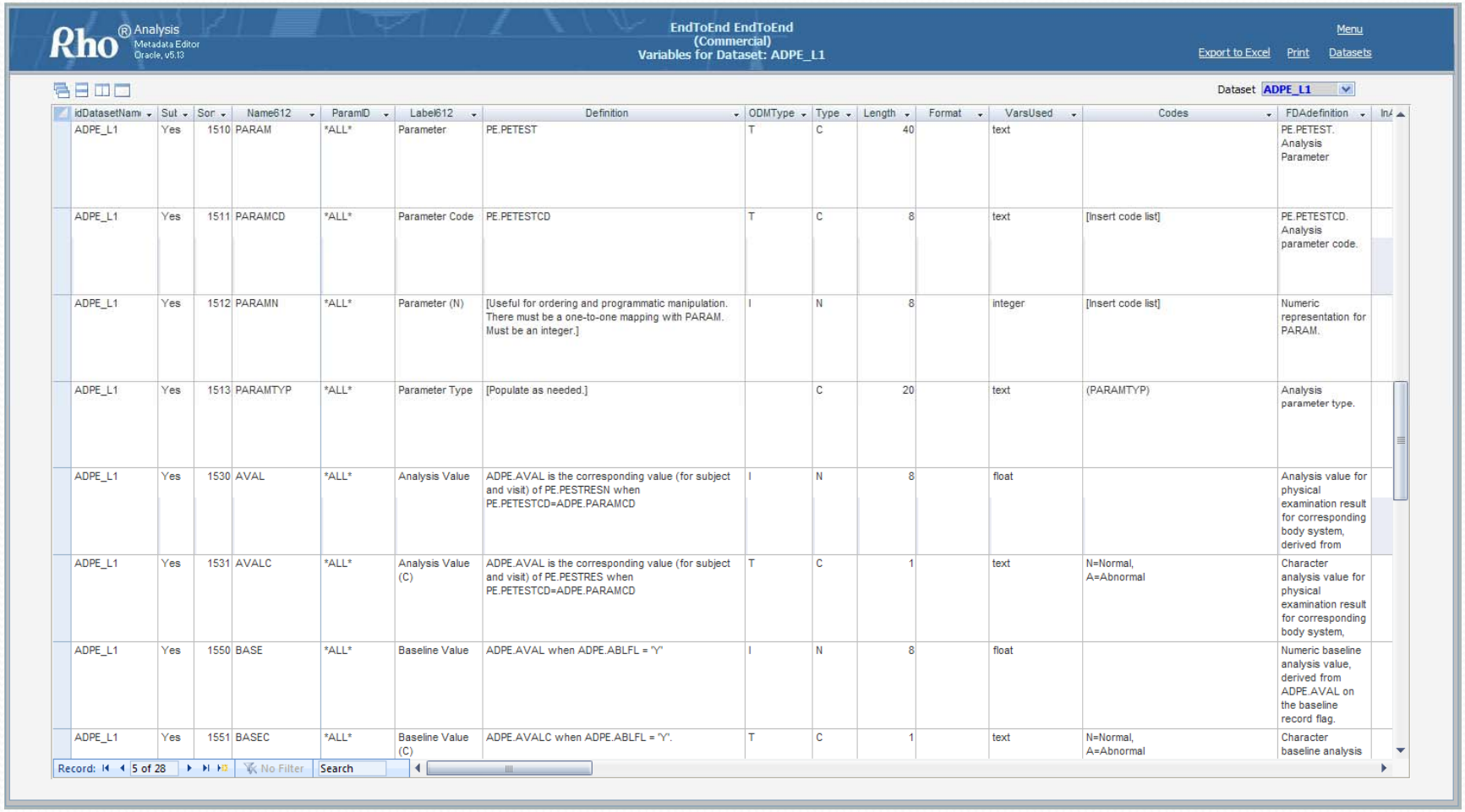
Dataset Specifications for
EndToEnd EndToEnd
(Commercial)

Menu
Export to Excel | Print | Variables

DatasetName	dsSc	dsLabel	Variable	Subm	Structure	Class	Description	Input_notes	Output_note	Documentation	Keyfields	FileName	LastEditedBy
ADAE_L1	7.1	Adverse Events	Edit	Yes	One record per subject per adverse event	ADAE	Contains the data for the Adverse Event Analyses				USUBJID, AESEQ		Rob Woolson
ADAE_L2	7.2	Adverse Events	Edit	Yes	One record per subject per adverse event	ADAE	Contains the data for the Adverse Event Analyses				USUBJID, AESEQ		Rob Woolson
ADCM_L1	79.1	Concomitant Medications	Edit	Yes	One record per subject per medication	OTHER	Contains the data for the Prior and Concomitant Medication Analyses	The almost always variables. Datasets used: DERIVE.ADSL			USUBJID, CMSEQ		Rob Woolson
ADCM_SDT	79.15	Concomitant Medications	Edit	Yes	One record per subject per medication	OTHER	Contains the data for the Prior and Concomitant Medication Analyses	The SDTM variables, just in case you need them. Datasets used:			USUBJID, CMSEQ		Rob Woolson
ADEG_L1	80.1	Electrocardiogram Results	Edit	Yes	One record per subject per parameter per analysis visit	BDS	Contains the data for Electrocardiogram Analyses	The almost always variables Datasets used: DERIVE.ADSL			USUBJID, PARAMCD, AVISIT		Rob Woolson
ADEG_L2	80.2	Electrocardiogram Results	Edit	Yes	One record per subject per parameter per analysis visit	BDS	Contains the data for Electrocardiogram Analyses	The second most common variables.			USUBJID, PARAMCD, AVISIT		Rob Woolson
ADEG_SDT	80.15	Electrocardiogram Results	Edit	Yes	One record per subject per parameter per analysis visit	BDS	Contains the data for Electrocardiogram Analyses	The SDTM variables, just in case you need them.			USUBJID, PARAMCD, AVISIT		Rob Woolson
ADLB_L1	81.1	Laboratory Results	Edit	Yes	One record per subject per parameter per analysis visit	BDS	Contains Laboratory Results.	The almost always variables Datasets used: DERIVE.ADSL			USUBJID, PARAMCD, AVISIT		Rob Woolson
ADLB_L2	81.2	Laboratory Results	Edit	Yes	One record per subject per parameter per analysis visit	BDS	Contains Laboratory Results	The second most common variables.			USUBJID, PARAMCD, AVISIT		Rob Woolson
ADLB_STD	81.15	Laboratory Results	Edit	Yes	One record per subject per parameter per analysis visit	BDS	Contains Laboratory Results.	The SDTM variables, just in case you need them.			USUBJID, PARAMCD, AVISIT		Rob Woolson
ADMH_L1	82.1	Medical History	Edit	Yes	One record per subject per body system per MedDRA preferred term per event start date	OTHER	Contains Medical History Results	The almost always variables. Datasets used: DERIVE.ADSL			USUBJID, MHBODYSYS, MHDECOD, MHSDTM		Rob Woolson

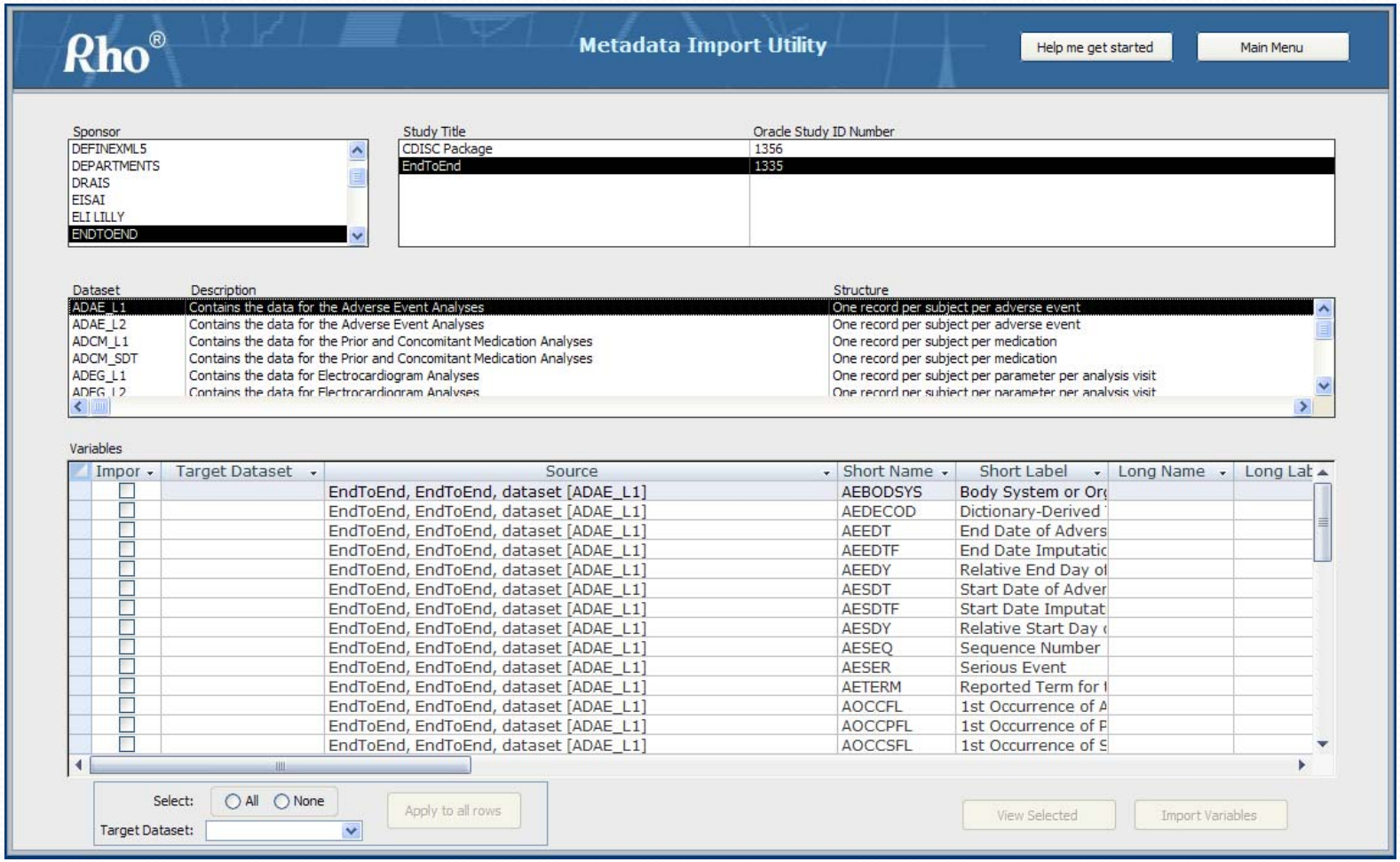
Record: 15 of 23 | No Filter | Search

ADaM Variable Metadata

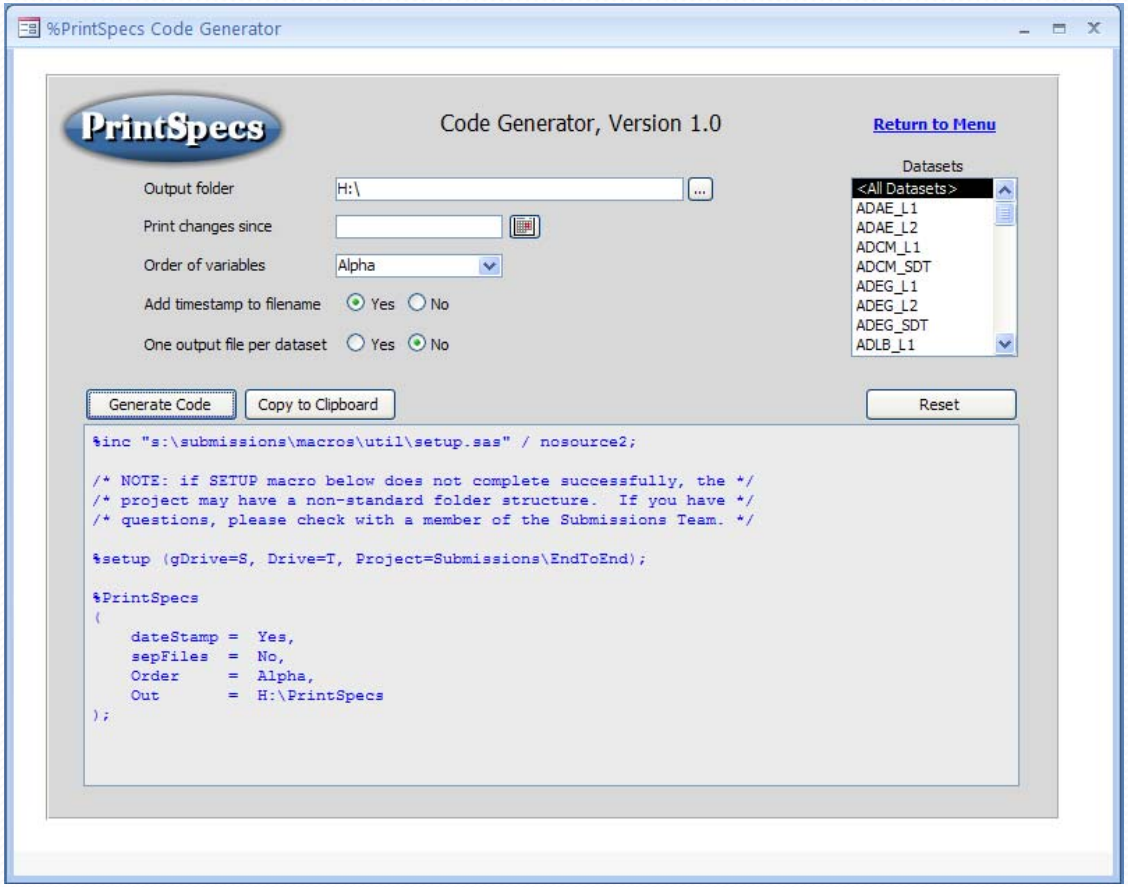


idDatasetNam	Sut	Sor	Name612	ParamID	Label612	Definition	ODMType	Type	Length	Format	VarsUsed	Codes	FDAdefinition	In
ADPE_L1	Yes	1510	PARAM	*ALL*	Parameter	PE.PETEST	T	C	40		text		PE.PETEST. Analysis Parameter	
ADPE_L1	Yes	1511	PARAMCD	*ALL*	Parameter Code	PE.PETESTCD	T	C	8		text	[Insert code list]	PE.PETESTCD. Analysis parameter code.	
ADPE_L1	Yes	1512	PARAMN	*ALL*	Parameter (N)	[Useful for ordering and programmatic manipulation. There must be a one-to-one mapping with PARAM. Must be an integer.]	I	N	8		integer	[Insert code list]	Numeric representation for PARAM.	
ADPE_L1	Yes	1513	PARAMTYP	*ALL*	Parameter Type	[Populate as needed.]		C	20		text	(PARAMTYP)	Analysis parameter type.	
ADPE_L1	Yes	1530	AVAL	*ALL*	Analysis Value	ADPE.AVAL is the corresponding value (for subject and visit) of PE.PESTRESN when PE.PETESTCD=ADPE.PARAMCD	I	N	8		float		Analysis value for physical examination result for corresponding body system, derived from	
ADPE_L1	Yes	1531	AVALC	*ALL*	Analysis Value (C)	ADPE.AVAL is the corresponding value (for subject and visit) of PE.PESTRES when PE.PETESTCD=ADPE.PARAMCD	T	C	1		text	N=Normal, A=Abnormal	Character analysis value for physical examination result for corresponding body system,	
ADPE_L1	Yes	1550	BASE	*ALL*	Baseline Value	ADPE.AVAL when ADPE.ABLFL = 'Y'	I	N	8		float		Numeric baseline analysis value, derived from ADPE.AVAL on the baseline record flag.	
ADPE_L1	Yes	1551	BASEC	*ALL*	Baseline Value (C)	ADPE.AVALC when ADPE.ABLFL = 'Y'.	T	C	1		text	N=Normal, A=Abnormal	Character baseline analysis	

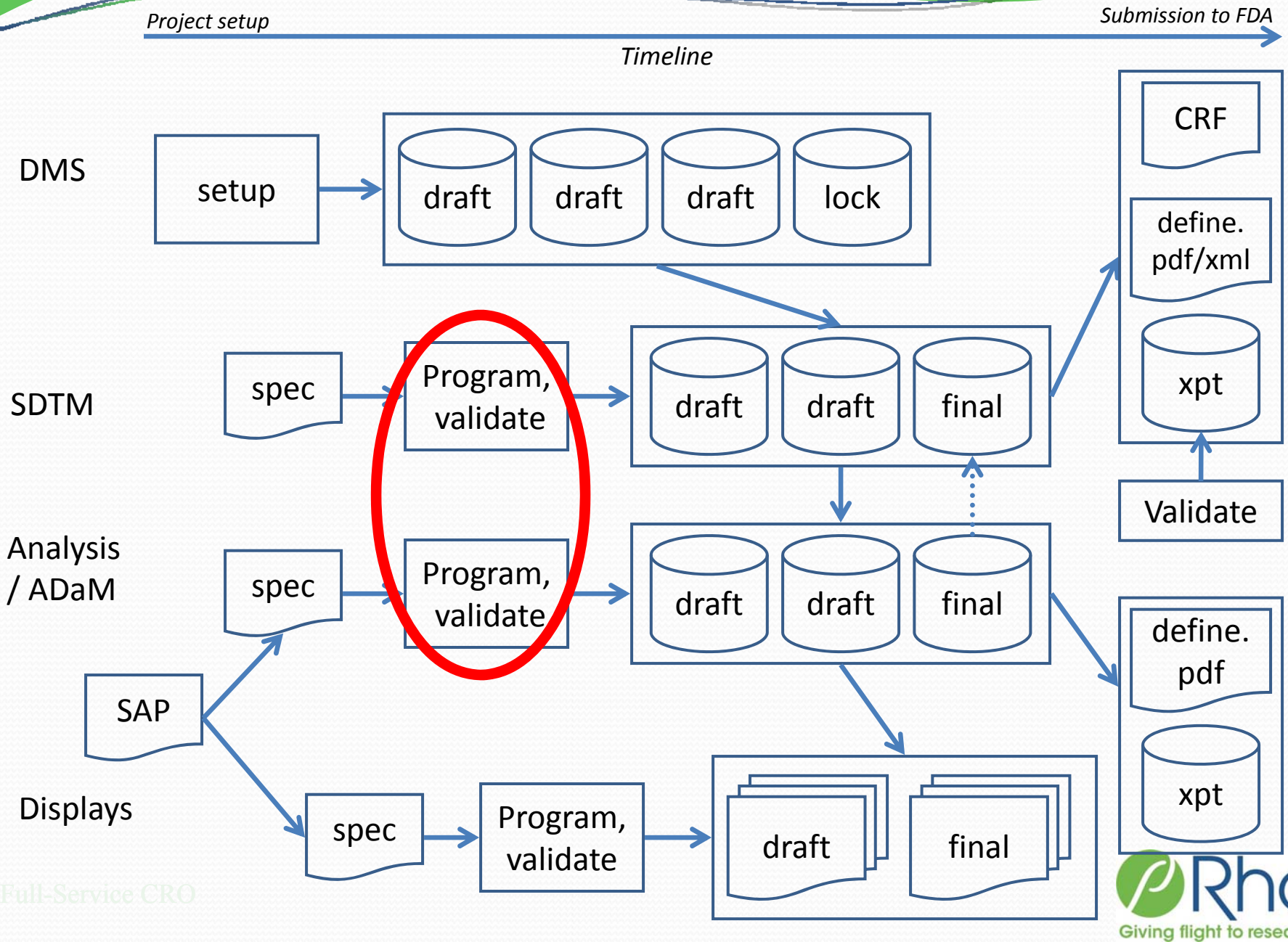
ADaM Import Tool



ADaM PrintSpecs Code Generator



Technology/Tools Demonstrations: Programming/Validation

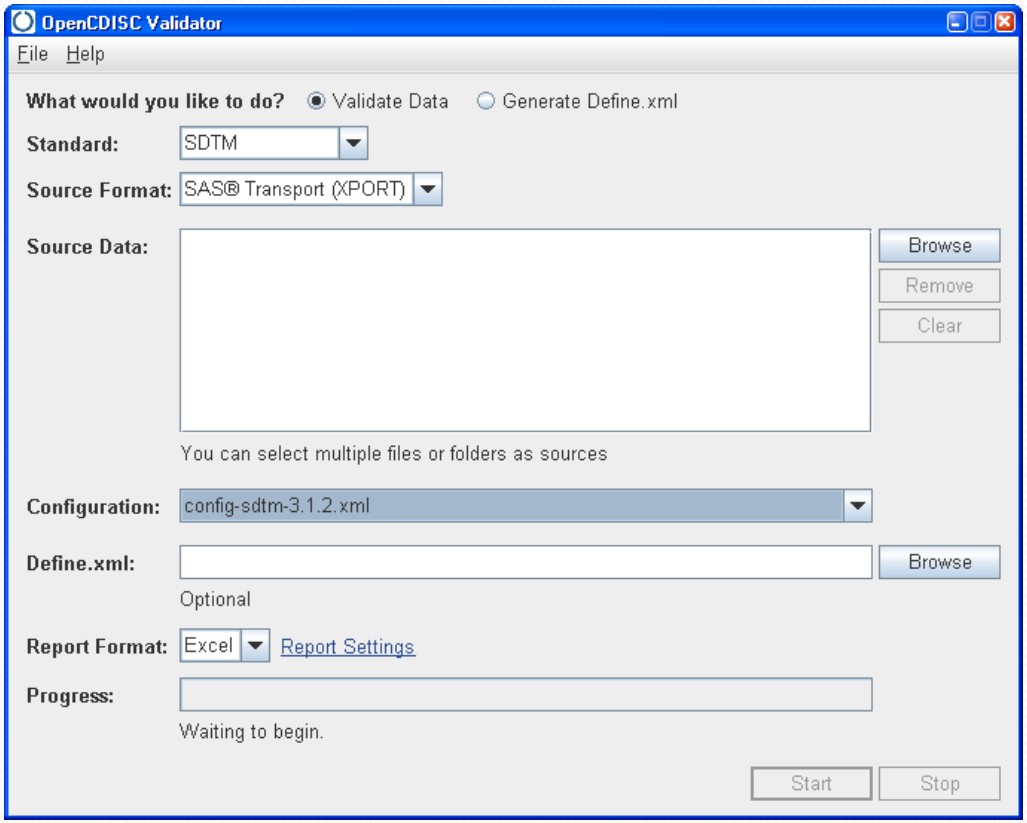


Rho
Giving flight to research

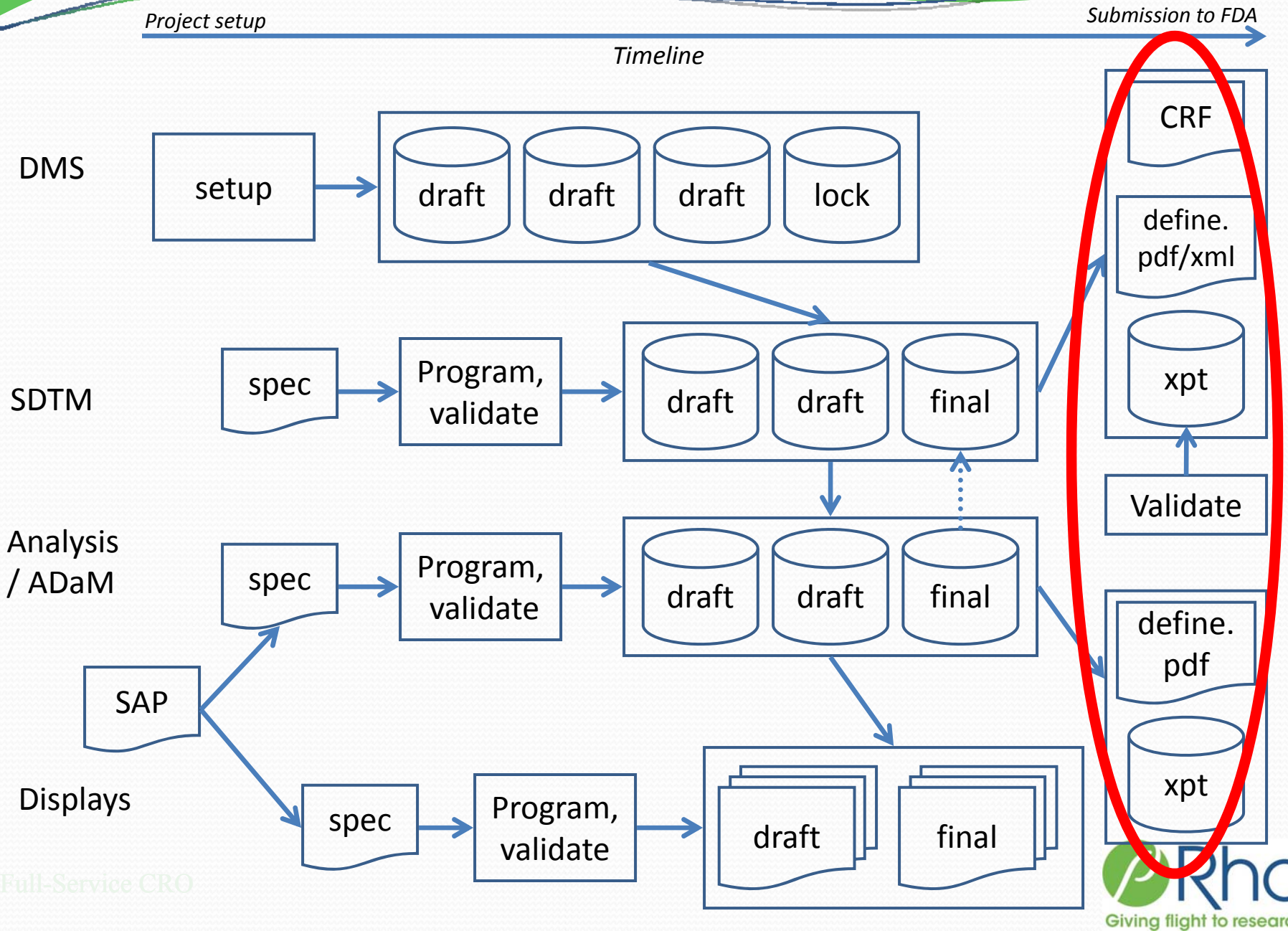
Variable Attribute Assignment Automation

```
645         %ATTRIB (in=metadata.dsvars,
646                  vname=name612,
647                  labname=label612,
648                  lenname=length,
649                  typename=type,
650                  forname=format,
651                  dsetname=DatasetName,
652                  dsn=&prog.,
653                  order=SortOrder,
654                  where=%nrstr(submitdb='Yes')
655                 );
** MACRO ATTRIB VERSION 4 IS EXECUTING **

************************A*T*T*R*I*B************************************************
Attrib-> Input Metadata Dataset is:                    [ metadata.dsvars ]
Attrib-> Metadata location is:                         [ T:\Biostat\EndtoEnd\Metadata\Analysis ]
Attrib-> Variable holding variable name is :           [ name612 ]
Attrib-> Variable holding variable label is :          [ label612 ]
Attrib-> Variable holding variable type is:            [ type ]
Attrib-> Variable holding variable length is :         [ length ]
Attrib-> Variable holding variable format is :         [ format  ]
Attrib-> Variable holding dataset name is :            [ DATASETNAME ]
Attrib-> Data set name as specified in metadata is:    [ ADSL ]
Attrib-> Variable holding order is:                    [ SORTORDER ]
Attrib-> Where statement specified as:                 [ submitdb='Yes' ]
Attrib-> Macro variable storing attribute statements   [ _attrib   ]
Attrib-> Macro variable for storing list of variables  [ _keeplist ]
Attrib-> Attrib statements were stored for   41 variables for dataset ADSL
*******************************************************************************

656
657         data adsl;
658            %unquote(&_ATTRIB);
659            set clinical.adsl;
660            fasfl=fullset;
661            saffl=safety;
662            ittfl=itt;
663
664            keep &_keeplist.;
665         run;

667         %put %unquote(&_ATTRIB);
attrib   STUDYID length=$100 format=$100. label="Study Identifier" USUBJID length=$100 format=$100. label="Unique Subject Identifier"
SUBJID length=$100 format=$100. label="Subject Identifier for the Study" SITEID length=$100 format=$100. label="Study Site [...]

668         %put &_keeplist.;
STUDYID USUBJID SUBJID SITEID ARM TRT01P TRT01PN TRT01A TRT01AN FASFL SAFFL ITTFL PPROTFL COMPTRFL COMPSTFL COMPLFL RANDFL ENRLFL
TRTFL SEX AGE AGEU RACE ETHNIC WEIGHTB HEIGHTB BMIB PPROTREA DISCTRS DISCSRS PROTDEV ANYPV BIRTHDT BIRTHDTM RANDDT TRTSDT TRTSTM
TRTSDTM TRTEDT TRTETM TRTEDTM
```

SDTM/ADaM Open CDISC Validation Tool



Technology/Tools Demonstrations: Define File Creation/Validation and Other Submission Deliverables



Rho
Giving flight to research

Define.xml Domains

Annotated Case Report Form (blankcrf.pdf) Supplemental documentation (readme.pdf) Printable version of define.xml (define.pdf)

Variables Value Metadata Code Lists

41 Datasets

Print All Print Next

Dataset	Description	Class	Structure	Purpose	Keys	Location
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID USUBJID AECAT AETERM AESTDTC AEENDTC	ae.xpt
CE	Clinical Events	Events	One record per event per subject	Tabulation	STUDYID USUBJID CETERM CECAT CESTDTC	ce.xpt
CM	Concomitant Medications	Interventions	One record per recorded medication occurrence per subject	Tabulation	STUDYID USUBJID CMCAT CMSCAT CMTRT CMSTDTC	cm.xpt
CY	Cytology	Findings	One record per cytology result per subject	Tabulation	STUDYID USUBJID CYTESTCD CYDTC	cy.xpt
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID USUBJID	dm.xpt

Define.xml Variables

Annotated Case Report Form (blankcrf.pdf) Supplemental documentation (readme.pdf) Printable version of define.xml (define.pdf)

Top Variables Value Metadata Code Lists

Dataset **DM** (22 variables): Demographics

SUPPDM Print All Print dm.xpt Previous Top Next

Var Seq	Variable	Label	Type	CT or Format	Origin	Role	DM Comment
1	STUDYID	Study Identifier	text		CRF	Identifier	Cover pg. 1
2	DOMAIN	Domain Abbreviation	text	DM	Assigned	Identifier	DM
3	USUBJID	Unique Subject Identifier	text		Derived	Identifier	Unique subject identifier created by concatenating Study Identifier, Site Number, and Subject Number.
4	SUBJID	Subject Identifier for the Study	text		CRF	Topic	Cover pg. 1
5	RFSTDTC	Subject Reference Start Date/Time	text	ISO 8601	Derived	Record Qualifier	Date of first dose of study medication.
6	RFENDTC	Subject Reference End Date/Time	text	ISO 8601	Derived	Record Qualifier	Date of last dose of study medication.
7	RFPENDTC	Date/Time of End of Participation	text	ISO 8601	Derived	Record Qualifier	Last known alive date. Maximum across all recorded data of dates at which subject is known to have been alive.
8	DTHDTC	Date of Death	text	ISO 8601	Derived	Record Qualifier	Date of Death
9	DTHFL	Subject Death Flag	text	(NY)	Derived	Record Qualifier	Indicator for subjects who have died.
10	SITEID	Study Site Identifier	text		CRF	Record Qualifier	Cover pg. 1
11	BRTHDTC	Date/Time of Birth	text	ISO 8601	CRF	Record Qualifier	Demographics pg. 2

Project Tracker: Results Level Metadata

Rho — Giving flight to research | A Contract Research Organization

Welcome, Rob Woolson! RhoNET™ | Change Password | My Profile | Logout

Need help? Got a question? Email Project Tracker Support

Project Tracker : Project Summary

PROJECT SUMMARY | ITEMS | TASKS | DELIVERIES | FOOTNOTES | NEW WINDOW | REFRESH | FAQ | HELP

PROJECTS

TLF Library

Filter Settings

TLF Library (Code: 004005-RHO000-000, ID: 4011)

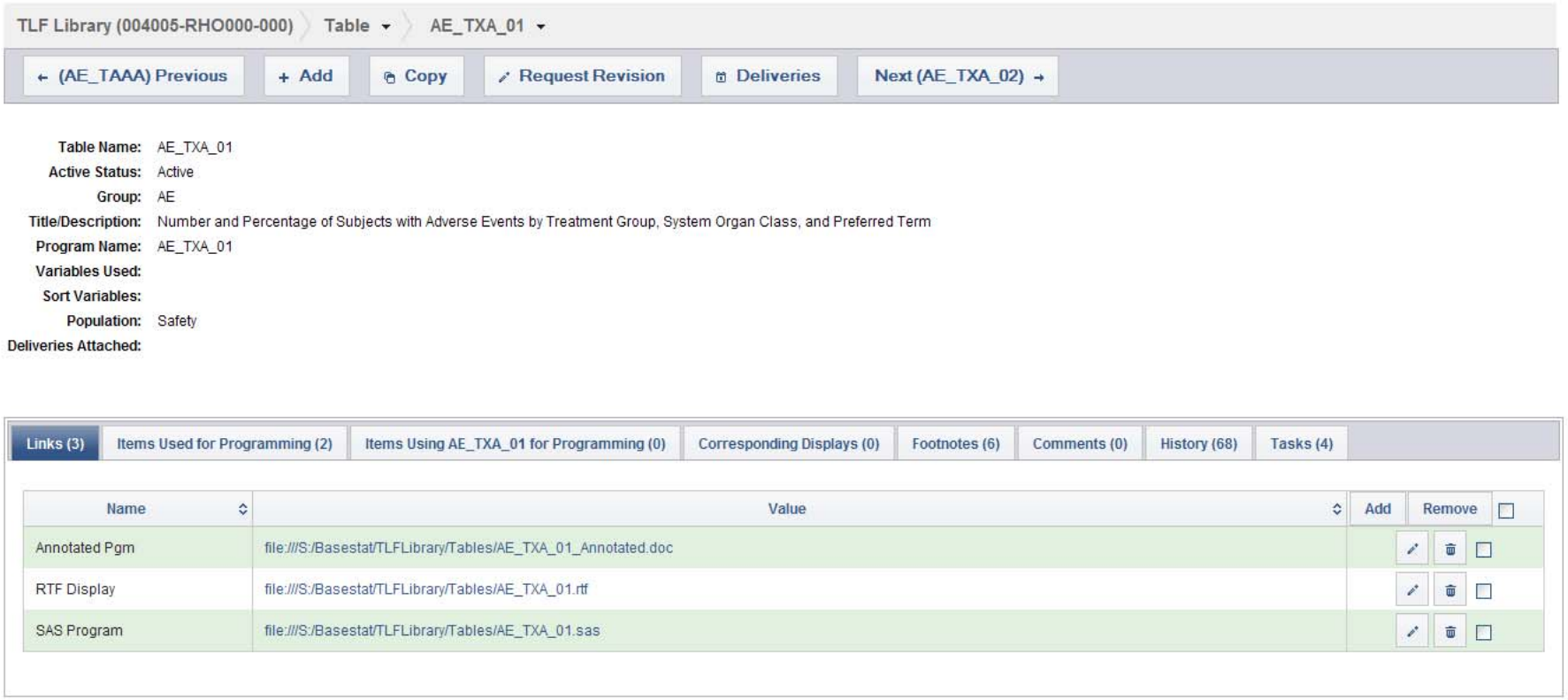
	Total Items	Items Not Started	Create Program Items Complete	Create Validation Program Items Complete	Items Ready for QC	Items In Progress	Items Complete	Percent Items Complete
Total	143	0	0	0	0	0	143	100%
- Analysis Datasets	11	0	0	0	0	0	11	100%

Items	Name	Project	Status	Tasks Not Started	Tasks In Progress	Tasks Complete	Percent Tasks Complete
	+ ADAE [details]	TLF Library	Completed	0	0	1	100%
	+ ADEG [details]	TLF Library	Completed	0	0	1	100%
	+ ADLB [details]	TLF Library	Completed	0	0	1	100%
	+ ADMD [details]	TLF Library	Completed	0	0	1	100%
	+ ADMH [details]	TLF Library	Completed	0	0	1	100%
	+ ADPE [details]	TLF Library	Completed	0	0	1	100%
	+ ADSL [details]	TLF Library	Completed	0	0	1	100%
	+ ADSTART [details]	TLF Library	Completed	0	0	1	100%
	+ ADVS [details]	TLF Library	Completed	0	0	1	100%

+ Figures	0	0	0	0	0	0	0	100%
+ Listings	0	0	0	0	0	0	0	100%
- Tables	130	0	0	0	0	0	130	100%

Items	Name	Project	Status	Tasks Not Started	Tasks In Progress	Tasks Complete	Percent Tasks Complete
	+ AE_GXA_01 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXA_01 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXA_02 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXB_01 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXB_02 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXC_01 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXC_02 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXD_01 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXD_02 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXE_01 [details]	TLF Library	Completed	0	0	3	100%
	+ AE_TXE_02 [details]	TLF Library	Completed	0	0	3	100%

Project Tracker: Results Level Metadata



The Future

- CDASH Development
- SDTM
  - Continued process improvement
  - Code generating specifications
- ADaM
  - Evolving standard
  - Results level metadata
- Integrate all of the above
  - Integrate all of the above in reverse (“Tables First”)

Conclusion

- CDISC standards essential part of business
- More than a collection of standards
- Impact on DM, programmers, and statisticians
- Continuous evolution
  - Project operations
  - Technology
  - Validation
  - Training/skills sets

Resources

FILE	DESCRIPTION
CDISC web site	CDISC web site:
http://www.fda.gov/ForIndustry/DataStandards/default.htm	FDA Resources for Data Standards
http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm249979.htm	CDER Data Standards Program
http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/ucm209137.htm	CBER CDISC Resources

Thank you for attending this workshop!

Send questions or comments to:

Jeff Abolafia (jeff_abolafia@rhoworld.com)
Carol Baker (carol_baker@rhoworld.com)
Rob Woolson (rob_woolson@rhoworld.com)

Special Thanks to:

Frank DiIorio, CodeCrafters, Inc., Philadelphia PA
Frank@CodeCraftersInc.com